Worldwide Gravity in Online Commerce

Bo Cowgill and Cosmina Dorobantu*

Abstract

This paper analyzes geographical patterns of cross-country Internet transactions using proprietary data from Google. Covering 146 countries from 2008-2011 and over 10 billion online transactions, the data allow us to examine a fast growing area of trade. The paper finds the effect of distance on trade to be around -0.53, indicating that distance has a negative effect on trade even in the virtual marketplace. The study also finds that cultural characteristics, such as shared languages or religions, have a large impact on e-commerce, while economic ties, such as a common currency, have an insignificant effect. The paper underlines the importance of accounting for selection into trade in worldwide gravity estimations and identifies two exclusion restrictions that can be used when examining online trade flows.

1 Introduction

Few empirical results in economics have been more widely confirmed than the negative relationship between distance and trade. Since the 1960s, economists have estimated gravity equations to measure the relationship of distance and trade between countries. Head and Mayer (2014) collected over 2,500 of these estimates and calculated the mean of the estimated coefficient on distance. They find that the average of the estimated distance elasticity is -0.93, indicating that a 10% increase in the distance between two countries leads to a 9.3% decrease in trade between them, all else being equal.[1]

*The authors thank Google for sharing the data used in this paper. We also thank Avinash Dixit, Keith Head, Beata Javorcik, Bertin Martens, Marc Melitz, Dennis Novy, Ferdinand Rauch, Andrés Rodríguez-Clare, Tony Venables, Pierre-Louis Vézina, Noam Yuchtman, and seminar participants at Oxford, the European Trade Study Group, and the European Commission for helpful comments. Dorobantu acknowledges financial support from the Economic and Social Research Council.

[1] Head and Mayer (2014) expands Disdier and Head (2008)'s meta-analysis of 1,467 estimates of the distance coefficient in gravity settings. The authors generate the -0.93 average distance coefficient by

Our paper's main contribution is to advance the nascent literature on the role of distance in the virtual world. Researchers are interested in understanding how distance affects online trade for several reasons. First, information costs are lower online. A company that has an online presence can sell to a foreign country without sending its analysts there to do extensive market research. Similarly, consumers wishing to buy a product from a foreign company face close to zero search costs, as all they need to do in order to connect with a foreign seller is conduct an Internet search. Allen (2014) shows that information costs account for a significant portion of the distance effect in the offline world. In the online environment, where information costs are lower, how much does distance matter for trade?

Second, online transactions are no longer a negligible part of the global economy. Online commerce recently passed the $1 trillion mark and its growth shows no signs of abating. An obvious question is whether the rules governing trade in the offline world also apply online: does distance still matter in a world that is more connected every day? The business press, represented by books such as Frances Cairncross' *The Death of Distance: How the Communication Revolution is Changing our Lives* or Thomas Friedman's *The World is Flat*, has an answer: distance is no longer relevant. Are these authors right? Is distance irrelevant for this new – and fast growing – form of commerce?

Despite the academic and business interest in this topic, only four other papers estimate the elasticity of distance for online trade flows in a worldwide gravity setting: Lendle et al. (2016), Lendle and Vézina (2015), Hortaçsu et al. (2009), and Blum and Goldfarb (2006). Their findings vary significantly. The lowest estimate the previous researchers obtain is -0.31, while the highest is -1.17.

In our paper, we improve the precision of these estimates by using a comprehensive data set and a methodology that corrects known econometric biases. In terms of the

averaging over all estimates included in their meta analysis, regardless of whether these estimates were obtained using a naive or a structural gravity approach.

data, our sample reflects over 10 billion transactions conducted online over a four year timeframe, 2008-2011. Notably, our data set covers 146 countries in the world, which is a significant improvement over the other four studies. Lendle et al. (2016), the most extensive online worldwide gravity analysis to date, covers only 61 countries. We show that such limited coverage biases the distance estimate downwards.

In terms of methodology, we improve on previous online gravity work by using the methodology proposed by Helpman et al. (2008). At a worldwide level, in online trade flows, we observe many zeros. In our data, only 25% of all country pairs trade in both directions, about 30% trade in one direction only, and 45% do not trade at all. If the high number of zero trade flows between countries is the result of an economically meaningful selection process, estimators that do not account for this selection process suffer from selection and omitted variable biases.

Prior to us, no other researchers that examined the effect of distance on online trade accounted for selection into trade. We show that a simple OLS estimation that does not account for selection into trade biases the coefficient of distance upwards by as much as 17%.

One of the reasons why researchers do not account for selection into trade is that the approach requires finding an exclusion restriction. Our paper's second contribution is identifying two exclusion restrictions that can be used in the context of online trade. The first one measures the difficulty of building warehouses, which is the primary physical asset that online retails need when expanding abroad. The second one relates to countries using a common currency. We show that for online trade flows – unlike for traditional trade flows – a common currency affects the fixed, but not the variable, cost of exporting.

Our analysis reveals that conditional on exporting, a 10% increase in the distance between two countries reduces the volume of trade between them by approximately

5.3%. We conclude that at the worldwide level, distance is certainly still alive in the online world.

Section 2 presents an overview of the relevant literature. It describes the studies done to date examining the effects of distance both in the online and the offline world and provides an overview of Helpman et al. (2008). Section 3 introduces the data we use for our analysis. Section 4 describes the empirical approach. Section 5 presents the results and Section 6 concludes.

2 Relevant Literature

2.1 Gravity Equations

In order to identify the role of distance, trade economists typically resort to gravity equations. These equations have been estimated since the 1960s, based on a simple concept borrowed from physics: trade between two countries depends positively on their sizes and negatively on the distance between them.[2] Although gravity equations perform very well empirically, they had no theoretical foundations until a little over a decade ago.

Anderson and van Wincoop (2003) used a demand-side structure and relied on the simplifying Armington assumption, namely that goods are differentiated by place of origin, to derive the gravity equation. Around the same time, Eaton and Kortum (2002) used a supply-side structure to arrive at the gravity equation, in a model employing Ricardian technology with heterogeneous productivity for each country and good. A few years later, Chaney (2008), Helpman et al. (2008), and Melitz and Ottaviano (2008) integrated gravity within Melitz (2003) style heterogeneous firm models. More recently,

[2] Rauch (2015) provides a comprehensive overview of the similarity between gravity in physics and gravity in trade.

Arkolakis et al. (2012) showed that a wide class of models, with or without heterogeneous firms, can serve as the theoretical foundation from which the gravity equation derives.

Head and Mayer (2014) provide an excellent summary of the theoretical, as well as the empirical results, of the gravity literature.[3] The variables included in empirical gravity analyses typically measure the effect of the geographic, cultural or economic closeness between trading partners. The two authors examined over 2,500 estimates reported in 159 academic publications to evaluate the main empirical findings. Although our primary interest relates to the role of distance, we also examine the effect on online trade of other geographical, cultural, and economic factors. We summarize below the results Head and Mayer (2014) obtain for each of these factors.

2.1.1 Distance

The physical distance between two countries is a ubiquitous variable in any gravity equation. Head and Mayer (2014) analyze over 2,500 estimates of the distance coefficient in gravity settings and find that the average of the estimated distance elasticity is -0.93. Distance, in other words, significantly reduces trade and this empirical result holds across a wide range of samples and methodologies.

While researchers widely accept the fact that distance has a negative effect on trade, there is no consensus in the literature as to why distance reduces trade. Over the years, several theories have been proposed. We summarize the most relevant ones below.

Transport Costs

One of the most obvious explanations as to why distance has a negative effect on trade

[3] Similar summaries are provided in Anderson (2011) and Baltagi et al. (2014).

is transport costs: the further apart two countries are, the more costly it is to ship goods between them. However, although shipping costs increase with distance, researchers like Grossman (1998), Hummels (1999), and Allen (2014) convincingly argue that freight shipping costs, by themselves, cannot fully explain why trade flows decline with distance; the magnitude of the distance effect is simply too large to be explained by shipping costs alone.

Harrigan and Venables (2006) and Evans and Harrigan (2005) show that besides shipping costs, goods also incur time costs as they are transported between countries. The larger the distance, the more time it takes for goods to be delivered. Harrigan and Venables (2006) show that delivery time for intermediate goods is costly for producers. As it takes time for components to be delivered, producers must place orders in advance, while they still face uncertainties regarding final demand. Furthermore, producers might see their production processes disrupted by unsynchronized deliveries. Harrigan and Venables (2006) argue that producers' desire to minimize the uncertainties they face as a result of delivery times for intermediate goods alters the efficient spatial organization of production.

In related work, Evans and Harrigan (2005) highlight the implications of time costs for retailers. In particular, they show that timely deliveries – which can only be achieved from nearby locations – reduce retailers' costs by allowing them to respond to fluctuations in demand without having to hold costly inventories.

Although the studies highlighted above are convincing, if distance proxies only for transport costs – be they freight shipping costs or time costs – then we would expect proximity to play no role in gravity models where the goods traded do not incur transport costs. Blum and Goldfarb (2006), however, examine website visits made by American Internet users to 46 foreign countries and find a distance elasticity of -1.17. Since website visits incur no transport costs, Blum and Goldfarb (2006)'s result indicates that distance must capture other trade impediments, besides shipping or time costs. We explore some of these frictions below.

Search Costs

Several authors argue that distance also captures search costs. In a seminal work, Rauch (1999) notes that it is costly for buyers and sellers to find each other, and shows that geographical proximity has a bigger impact on matching buyers and sellers of differentiated products (which are subject to greater informational impediments) than on matching buyers and sellers of homogenous products.

In related work, Chaney (2013) proposes a theoretical model of input-output linkages between firms to highlight the importance of the costs involved in finding suppliers and consumers. In Chaney (2013)'s model, firms have two options to engage in trade. The first one is to pay a direct cost to make a foreign contact, while the second one is to interact directly with one's existing contacts and get introduced to their contacts. This need for direct contact in order for trade to take place allows Chaney (2013) to explain why distance is such a large deterrent to trade.

More recently, Allen (2014) argues that distance also captures the costs of obtaining information about market conditions elsewhere. He shows that in a perfect competition model where producers incur costs in order to obtain information about market conditions in other locations, trade flows decline with distance not only because of transport costs, but also due to the fact that the probability of searching a destination decreases.

Trust

Besides transport and search costs, some researchers suggest that trust also contributes to the large distance effects researchers find in the data. Guiso et al. (2009) measure bilateral trust between countries and, using a gravity setting, show that the distance between two countries has a negative and significant effect on bilateral trust. In a related study, Huang (2007) show that distance affects countries that have high uncertainty-aversion more, and that these countries export disproportionately less to remote partners than a gravity equation would predict. Finally, in studies using data from eBay, Lendle et al. (2016) and Hortaçsu et al. (2009) show that distance has a bigger effect on transactions involving untrustworthy sellers, that have poor ratings on the site.

Localized Tastes

Ferreira and Waldfogel (2013) and Blum and Goldfarb (2006) show that in gravity settings, distance might also account for the existence of localized tastes. Ferreira and Waldfogel (2013) examine trade in recorded music, which faces low trade costs, and find that a one percent increase in distance reduces music trade by 0.42 percent. Blum and Goldfarb (2006) show that distance reduces the number of website visits for taste-dependent website categories, such as music, games, pornography or gambling. At the same time, the authors find that distance has no effect on the number of website visits for non-taste dependent website categories, such as software and general, technological or financial information sites.

Other geographical factors

Another common geographical factor considered in gravity equations is whether two countries share a land border or not. Most researchers find contiguity to have a positive and significant effect on trade. Head and Mayer (2014) report the average of the estimated coefficient on a dummy variable equal to 1 if two countries share a border to be around 0.5. Other geographical characteristics considered in gravity equations are whether the trading countries are landlocked or islands. Due to the limitations that these geographical characteristics place on transport options, they are generally found to negatively impact trade (see Limao and Venables (2001), Silva and Tenreyro (2006), Helpman et al. (2008)).

Cultural factors

Cultural closeness between two countries has a positive effect on the amount of trade between them. Head and Mayer (2014) calculate the average estimated coefficient of a dummy variable equalling 1 if two countries speak the same language to be around 0.4. The average estimated coefficient of a dummy variable equalling 1 if two countries share a colonial link is even higher, taking a value of around 0.8. Other cultural characteristics considered by researchers in gravity settings are whether the trading partners have similar religious beliefs or whether their legal systems share the same origin. These variables are generally found to positively and significantly influence trade (see Helpman et al. (2008)).

Economic factors

Free trade agreements, as well as a common currency, facilitate trade between two countries. In a gravity setting, the effect of a free trade agreement is measured by including a dummy variable that equals 1 if two countries are part of a free trade area. Head and Mayer (2014) report the average estimated coefficient of this dummy variable to be around 0.4. The effect of sharing a currency on trade is usually measured by including a dummy variable equal to 1 if two countries use the same currency. Head and Mayer (2014) calculate the mean of the estimated coefficient on this dummy variable to be around 0.2.

2.2 Internet Trade and Gravity Equations

A small literature at the intersection of computer science, economics and quantitative marketing has studied the relationship between economic geography and the Internet. This literature has primarily focused on consumers' substitution between online and offline purchasing channels (for example, see Forman et al. (2009), Brynjolfsson et al. (2009) or Goolsbee (2001)).

Very few researchers have used online transaction data to address trade-related questions. As noted in the introduction, Lendle et al. (2016), Lendle and Vézina (2015), Hortaçsu et al. (2009), and Blum and Goldfarb (2006) are the only other studies to estimate worldwide gravity equations on data relating to online activities. However, each of these researchers have faced data limitations, some of which we are able to improve upon.

Lendle et al. (2016) work with eBay data and they have access to the most comprehensive data set out of the four papers mentioned above. The authors perform a gravity analysis focusing on 61 countries and they justify limiting their analysis to 61 countries by noting that 92% of all offline trade flows happen between these 61 states.

By omitting the countries that trade little with each other, however, Lendle et al. (2016) risk underestimating the effects of distance. If distance is part of the reason why some countries trade with each other very little, or do not trade at all, excluding the locations that do not trade much with each other will bias the coefficient on distance downward.

Lendle and Vézina (2015) replicate Lendle et al. (2016)'s analysis on an eBay data set covering sellers located in 21 emerging economies. The sellers export to 204 destinations. The authors find a distance elasticity of -0.31 for eBay trade flows, smaller than the -0.44 elasticity reported in Lendle et al. (2016). Lendle and Vézina (2015)'s study faces limitations due to the fact that the data set covers only 21 exporting countries, and also due to the authors' decision to use a linear estimator, which is known to produce biased results in a gravity setting with zero trade flows.[4]

Hortaçsu et al. (2009) look at geographical trade patterns using data from eBay and MercadoLibre, a South American website dedicated to e-commerce and online auctions. The eBay data cover only US buyers and sellers over a four-month time frame and the location of the buyer is known for 27% of the transactions. This limits the authors to examine trade patterns only within the United States and only on a quarter of the transactions registered in their data set. The MercadoLibre data is more complete, but it only allows for an analysis of trade flows between nine South American countries.

Blum and Goldfarb (2006) use data from a defunct ISP company to look at the Internet activities of 2,654 American Internet users, who visited websites from 46 countries over the span of three months. Their data are also limiting. They only cover page views, the country of origin of the website visited is difficult to determine and there are questions about the representativeness of the sample, given that there were over 100 million Internet users in the United States in the year examined by the authors. They find that distance matters more for taste-dependent differentiated products and less for homogenous products.

Table 1 summarizes the distance elasticities obtained by the studies summarized above. As noted in the introduction, we are able to improve upon the existing estimates, by using a comprehensive data set which covers 146 countries, and by employing a methodology that corrects known econometric biases.

[4]Zero trade flows are not a problem in Lendle et al. (2016), as all 61 countries trade with each other. In Lendle and Vézina (2015), however, they are, as not all 21 countries export to all 204 destinations.

Table 1: Distance Elasticities Obtained by Previous Researchers in a Worldwide Setting

Study	Distance Effect	Country Coverage	Estimator Used	Dependent Variable
Lendle et al. (2016)	-0.440*** (0.032)	61 exporters 61 importers	OLS	Value of sales (logs)
Lendle and Vézina (2015)	-0.313*** (0.048)	21 exporters 204 importers	OLS	Value of sales (logs)
Hortaçsu et al. (2009)	-0.382*** (0.030)	9 exporters 9 importers	OLS	Value of sales (logs)
Blum and Goldfarb (2006)	-1.167*** (0.114)	1 exporter 46 importers	OLS	Number of site visits (logs)

2.3 Overview of Helpman et al. (2008)

Worldwide gravity exercises have been criticized on two main grounds: first, that the estimating equation is not carefully derived from theory and second, that OLS estimates suffer from econometric biases due to ignoring the zero trade flows. Researchers like Silva and Tenreyro (2006) and Eaton et al. (2012) propose econometric methods that accommodate zero trade flows to estimate gravity equations. However, Helpman et al. (2008) is the only paper to address both criticisms outlined above by proposing a theoretical model in which zero trade arises due to an economically meaningful selection process and an empirical methodology that accounts for selection into trade.[5]

Helpman et al. (2008) start from a Melitz (2003) model with asymmetric countries. Consumers face a CES utility function. Each country has a number of firms, and each firm produces a distinctive product. A firm produces a unit of output with a cost-minimizing combination of inputs, $c_j a$. The parameter a is firm specific and it reflects differences across firms in productivity levels. The cost c_j is country specific and it

[5] Besides the theoretical appeal of Helpman et al. (2008)'s proposed estimated method, there is some evidence that it performs better than the PPML estimator proposed by Silva and Tenreyro (2006) for data sets where zero trade values are frequent. See Martin and Pham (2008).

reflects differences across countries in factor prices. A firm from country j can sell its products on the domestic market, or it can export to country i. If the firm wants to export to country i, however, it faces two additional costs: a fixed cost of exporting and a variable cost that is specified in the model as a "melting iceberg" transport cost. Profit maximization makes it obvious that while all firms will sell on the domestic market, exporting will be feasible only for a fraction of firms that have a productivity level a above the threshold required for exports from j to i to be profitable.

Based on the theoretical model, the authors derive the following estimating equation:

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + w_{ij} + u_{ij}$$

where:

$$\begin{aligned}
m_{ij} &= \text{country } j\text{'s exports to country } i \\
\lambda_j &= \text{fixed effect of the exporting country, } j \\
\chi_i &= \text{fixed effect of the importing country, } i \\
d_{ij} &= \text{symmetric distance between } i \text{ and } j \\
w_{ij} &= \text{fraction of firms (possibly zero) that export from } j \text{ to } i \\
u_{ij} &= \text{i.i.d. unmeasured trade frictions}
\end{aligned}$$

Traditional estimates are biased due to two econometric problems. First, omitting the fraction of firms that export from j to i creates an omitted variable bias. This fraction is endogenous; it is a function of the productivity threshold required for exports from j to i to be profitable, which in turn is a function of the variables controlling for the trade barriers between the two countries. Omitting w_{ij} from the estimating equation induces an upward bias in the coefficient measuring the effect of the symmetric distance between i and j, as this coefficient picks up the effect of trade barriers on firm-level trade and on the proportion of exporting firms.

Second, omitting zero trade flows creates a selection bias. Helpman et al. (2008) underline the fact that there are many zero trade flows observed in the real world. In 1997, only about 40% of country pairs traded in both directions. An additional 12% traded in one direction only, and a surprising 48% of country pairs did not trade at all. We ran a similar analysis for our data from Google and found that between 2008-2011, 25% of all country pairs traded in both directions, about 30% traded in one direction only, and 45% did not trade at all. Given the high percentage of zero trade flows in the online world, it is just as important as it is in the traditional, offline world, to correct for the selection bias created by omitting the zero trade flows. Excluding the country pairs with zero trade flows places a downward bias on γ, the trade barrier coefficient.

To correct these biases, Helpman et al. (2008) start by defining a latent variable, Z_{ij}:

$$z_{ij} = \ln Z_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij}$$

where:

z_{ij} = ratio of variable export profits for the most productive firm to the fixed export costs for exports from j to i

ξ_j = fixed effect of the exporting country, j

ζ_i = fixed effect of the importing country, i

d_{ij} = symmetric distance between i and j

ϕ_{ij} = observed measure of any additional country-pair specific fixed trade costs

$\eta_{ij} = u_{ij} + v_{ij}$, where v_{ij} is IID $N(0, \sigma_{ij})$

The latent variable z_{ij} is not observed, but positive trade between j and i is observed when $z_{ij} > 0$ and zero trade is observed when $z_{ij} = 0$. Let T_{ij} be an indicator variable equal to 1 when country j exports to i and 0 otherwise, and ρ_{ij} be the probability that j exports to i. We can then specify the Probit equation:

$$\rho_{ij} = \Pr(T_{ij} = 1) = \Pr(z_{ij} > 0) = \Phi(\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^* \phi_{ij})$$

If $\hat{\rho}_{ij}$ is the predicted probability of exports from j to i, we can easily obtain $\hat{z}_{ij}^* = \Phi^{-1}(\hat{\rho}_{ij})$, the predicted value of the latent variable z_{ij}^*. Since W_{ij}, the fraction of firms that export from j to i, is a monotonic function of Z_{ij}, we can use $\hat{z}_{ij}^* = \Phi^{-1}(\hat{\rho}_{ij})$ to obtain a consistent estimate for $w_{ij} = \ln W_{ij}$, the omitted variable in the gravity equation specified above. The proposed methodology, thus far, corrects the first of the two biases identified above, namely the omitted variable bias.

To correct the second bias identified above, Helpman et al. (2008) propose to use the standard Heckman (1979) correction for sample selection. For this, they use the same Probit equation specified above to obtain the inverse Mills ratio: $\hat{\bar{\eta}}_{ij}^* = \phi(\hat{z}_{ij}^*)/\Phi(\hat{z}_{ij}^*)$. Accounting for this term in the gravity equation addresses the selection bias created by omitting zero trade flows from the estimating equation.

We use the two-step estimation procedure proposed by Helpman et al. (2008) and described above to estimate the effect of distance on online trade flows. Section 3 describes the data we use, while Section 4 outlines the precise empirical specification.

3 Data

3.1 Google Data

3.1.1 Data Source

The data we use come from Google's online advertising platforms and they were generated by the same conversion tracking software that enabled us to perform the US-Canada border effects analysis presented in Chapter 2. The software tracks conversions generated as a result of users clicking on an ad placed on Google or on one of Google's partner sites.

As mentioned in Section 4.1.1 of the previous chapter, the data set has some minor limitations. In particular, the conversion tracking code reliably generates conversion counts (rather than values). Additionally, a count is recorded in the data set only if a user reaches an advertiser's thank you page within 30 days of the initial click on the ad. Finally, while we believe that we are able to attribute most buyers and sellers to the correct region, it is possible that we are attributing a small part of the buyers or sellers to the wrong location.[6]

3.1.2 Coverage

For our worldwide analysis, we use data aggregated across conversion types and sectors of economic activity. The data covers a large number of businesses advertising online. In 2007, Google announced that AdWords, its advertising program, passed the one million mark in terms of number of advertising accounts. This is the last publicly available figure. It is meaningful in that it indicates that a year prior to when our

[6] As a reminder for the reader, for buyers, we use their IP address to determine their location, while for sellers, we use their self-reported mailing address.

data coverage starts, there were already more than one million businesses opting for Google's advertising platform. In 2007, Google also announced publicly that no client accounts for more than 10% of the company's revenues, indicating that its advertising program is not dominated by one large business.

The data set we use covers all conversion counts recorded over a period of four years, from 2008 to 2011. Similar to Lendle et al. (2016), we average the observations over the four years of data that we have at our disposal.

The overall number of recorded conversion counts is in excess of 10 billion.[7] All but six countries are featured in our data set. At the time the data were collected, trade embargoes were in place between the United States and Burma, Cuba, Iran, North Korea, Sudan, and Syria. As a US-based company, Google could not generate revenues in these countries and therefore, it could not accept ads from businesses located in any of these six countries.

3.1.3 Comparability to Estimates Obtained by Helpman et al. (2008)

At the worldwide level, we are unable to obtain online trade data broken down by sector of economic activity. While we do not know the precise composition of trade flows on Google, there are likely to be important differences between the bundle of goods and services reflected in our dataset and the bundle of products reflected in the data used by Helpman et al. (2008).

For their analysis, Helpman et al. (2008) use the bilateral trade flows reported in Feenstra's "World Trade Flows" database. This database records trade figures for 34 manufacturing industry codes and covers transactions in intermediate inputs such as primary metals and industrial chemicals. By contrast, we use data from Google, which is

[7] Due to confidentiality reasons, we cannot provide precise figures.

an all-purpose search engine that promotes a wide variety of online goods and services primarily targeted to end users. As such, our data are more likely to reflect transactions in consumer goods (such as books, antiques, or consumer electronics) than transactions in intermediate inputs. Furthermore, our data are likely to include transactions classified under the service sectors, such as legal advice, web-hosting, banking or web design, which are not covered by the data used by Helpman et al. (2008). Finally, our dataset is also likely to include sales of digital goods, such as software or music downloads, which are not reflected in the data used by Helpman et al. (2008).

We reproduce the results that Helpman et al. (2008) obtain in the Appendix. However, due to fact that the bundle of goods and services reflected in our dataset differs from the bundle of products reflected in the data used by Helpman et al. (2008), we cannot interpret the differences between the coefficients estimated by us and the coefficients estimated by Helpman et al. (2008) as providing a direct and precise comparison between trade barriers in the online world versus the offline world.

3.2 Explanatory Variables

We include in our regressions the same explanatory variables that Helpman et al. (2008) used in their paper. For each variable, we will discuss the data source we use, as well as our a priori hypotheses.

3.2.1 Distance

In our study, we are primarily interested in uncovering the empirical relationship between distance and online trade. We use the CEPII distances database as our data source for the distance measure, $distance_{ij}$.[8] To measure distance, we follow Head and

[8]This database is available online at http://www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp

Mayer (2009)'s suggestion to calculate the distance between two countries based on the bilateral distances between the biggest cities of those two countries and weighting those inter-city distances by the share of the city in the overall country's population.[9]

We expect distance to have a negative effect on both the extensive and the intensive margins of trade. In other words, we expect distance to have a negative effect both on the probability of two countries trading and on the volume of trade between them.

As far as the extensive margin is concerned, we expect to find a negative effect because a business wanting to export online will face some fixed costs. Evidently, a firm wanting to export online goes through a process that is different from that undertaken by a firm looking to export offline. On Google's platforms, a US company wanting to export to France can do so instantaneously, by ticking a box in its Google account and making its online ads visible within France. In theory, therefore, the US company wanting to export to France faces zero fixed costs. In practice, however, it is hard to imagine a profit maximizing firm making the decision to export without undertaking some initial investment to make its product appealing and accessible to foreign consumers. Before exporting to France, a US online retailer might want to translate its website and marketing materials into French. It might also want to ensure that it can process payments made with French credit cards, or that it can safely ship and store its products once they arrive in France. The costs outlined here are small compared to the investment a traditional firm needs to undertake to export offline, but they are unlikely to be zero.

As for the intensive margin, we expect to find that distance has a negative effect on the volume of online trade for several reasons. First, the physical goods sold through Google incur transport costs similar to those incurred by the manufacturing goods sold through traditional trade channels. This is because shipping physical goods to far away

[9] We also experiment with using simple (unweighted) distance measures and we find that they produce similar results.

places – regardless of whether the order was placed online or offline – is costly and can take a long time. Second, for some services exchanged over the Internet, the convenience of working with a provider that is close by will make transactions between online shoppers and service providers less likely over long distances. Finally, the existence of localized tastes is likely to affect online exchanges much like it affects the volume of transactions conducted through the traditional trade channels.

Although we expect distance to have a negative effect on the intensive margin of online trade, this effect is likely to be smaller online. There are several reasons for this. First, search costs are lower online. On Google, buyers and sellers find each other with a simple Internet search, the cost of which is low and uncorrelated with the distance between the transacting parties. Second, transport costs are zero for some of the transactions conducted online. As noted in section 3.1.3, Google is a generic, all-purpose search engine and the transactions reflected in our dataset cover a wide variety of digital goods and services, in addition to physical goods. For some of the digital goods and services reflected in our dataset, the transport costs are likely to be zero or close to zero. Finally, trust might have a smaller effect on online trade flows. Internet users and online businesses can usually rate and review each other online, which can act as a deterrent to dishonest trades, whether the transacting parties are nearby or far away from each other.

3.2.2 Other Geographic Characteristics

Similar to Helpman et al. (2008), we control for three other geographical characteristics besides distance: contiguity, as well as whether the trading partners are islands or landlocked countries. To do so, we build three dummy variables. The first one, $land\ border_{ij}$, takes the value of 1 if two countries share a land border and 0 otherwise. We use the CEPII distances database to construct this dummy. The second one, $island_{ij}$, takes the value of 1 if two countries are islands and 0 otherwise. We use the CIA's *World Factbook* land boundary measure to construct this variable.[10] In our study, we consider an island to be a country whose land boundary is reported by the CIA to measure 0 kilometers. Finally, the third dummy variable, $landlock_{ij}$, takes the value of 1 if two trading partners have no coastline and 0 otherwise. The CEPII distances database provides the data necessary to construct this third dummy.

In line with other studies using offline trade data, we expect contiguity to have a positive effect on online trade. A priori, it is difficult to predict whether the fact that two trading partners are landlocked or islands will have a positive or negative effect on online trade. In studies conducted using traditional trade figures, these geographical characteristics are generally found to negatively impact trade due to the limitations they place on transport options. With online trade, we might see a positive effect precisely because of these limitations: if the inhabitants of a remote island do not find in stores the foreign products that they want because it is not profitable for offline retailers to transport them, they might purchase these products online. Geographical isolation, in this case, would encourage, rather than discourage, online trade.

3.2.3 Cultural Characteristics

Four variables allow us to control for the cultural closeness between two countries. The first one is $legal_{ij}$, a dummy variable taking the value of 1 if two countries' legal system share the same origin and 0 otherwise. According to the CIA's *World Factbook*, the data source we use to construct this dummy variable, each country's legal system has, at its origins, one of five elements: civil law, common law, customary law, pluralistic law or religious law. If two countries' legal system share the same element, we assign a value

[10] The CIA's *World Factbook* is available online at https://www.cia.gov/library/publications/the-world-factbook/

of 1 to $legal_{ij}$. In line with studies conducting using traditional trade data, we expect to find a positive and significant effect of $legal_{ij}$ on online trade.

The second variable that allows us to control for cultural closeness is $language_{ij}$. This is a dummy variable taking the value of 1 if two countries share an official language and 0 otherwise. We use the CEPII distances database to construct this dummy variable. Our prior is that countries that speak the same language will trade more with each other in the online world. Translating the website and the marketing materials constitutes a large part of the fixed cost incurred by a company wanting to export online. After it starts exporting, the company will likely communicate with its customers via email or over the phone about the products offered or the details of the orders placed online. Speaking the same language will facilitate that communication. Our prior is that a common language, by lowering both the fixed and the variable cost of exporting, will have a positive influence on online trade.

The third variable is $colony_{ij}$, a dummy variable that takes the value of 1 if two countries were ever in a colonial relationship and 0 otherwise. Our data source for this variable is, once again, the CEPII distances database. Since countries that share a common colonial history are culturally closer, we expect to find a positive effect of $colony_{ij}$ on online trade.

Finally, the fourth variable we use to control for cultural similarity is $religion_{ij}$. Similar to Helpman et al. (2008), we use the following formula to generate this variable: $religion_{ij} = (\%\ Protestants_i \times \%\ Protestants_j) + (\%\ Catholics_i \times \%\ Catholics_j) + (\%\ Muslims_i \times \%\ Muslims_j)$. We obtain the percentage figures from CIA's *World Factbook*. A priori, we expect countries with similar religious beliefs to trade more with each other.

3.2.4 Economic Characteristics

Two variables measure the extent of the economic ties between two trading partners. The first one is $currency_{ij}$, a dummy variable taking the value of 1 it two countries use the same currency and 0 otherwise. We use the CIA's *World Factbook* to construct this variable. The second one is FTA_{ij}, a dummy variable taking the value of 1 if two countries belong to a free trade agreement and 0 otherwise. To construct this variable, we use the World Trade Organization's Regional Trade Agreements Information System,[11] which contains a list of all regional trade agreements in place, along with their date of entry into force. As common currencies and free trade agreements encourage trade between two countries, we expect these two variables to have a positive and significant effect on online trade.

3.2.5 Exclusion Restriction

An exclusion restriction is needed for the first stage of the two-step estimation procedure proposed by Helpman et al. (2008). The authors' theoretical model suggests that a valid exclusion restriction is one that affects the fixed cost of exporting, so it enters with a non-zero coefficient in the first stage Probit equation, but has no effect on the variable cost of exporting.

Helpman et al. (2008) propose two indicators variables as exclusion restrictions, the first of which takes the value of 1 when the cost of setting up a business in both countries is above the median, while the second takes the value of 1 when the number of days and procedures required to set up a business in both countries is above the median. Although these variables affect the fixed cost of exporting in the offline trade world, they do not necessarily affect the fixed cost of exporting online. Indeed, when we test

[11] Available online at http://rtais.wto.org/ui/PublicMaintainRTAHome.aspx

these two indicator variables on our data set, we find that they do not have enough explanatory power in the first stage estimation to act as valid exclusion restrictions. The reason for this is simple: in the virtual world, in order to start exporting, a business does not need to set up an entity abroad.

In e-commerce, warehouses, rather than business entities, are a key component of international expansion. Online retailers often store inventory in their destination markets,[12] so securing warehouse space is one of the first steps they make when expanding abroad. There are two reasons why online retailers keep inventory in warehouses overseas. First, it speeds up the delivery of the goods to the final consumers: once a user from the destination market places an order, the good is delivered to them directly from the warehouse. Second, it makes returns and exchanges easier for the buyers.

A reasonable exclusion restriction for our analysis, therefore, would capture the costs of setting up a warehouse, rather than the cost of setting up a business. The World Bank's *World Development Indicators* database contains data on the number of procedures required to build a warehouse.[13] We use this data to build an indicator variable taking the value of 1 when the number of procedures required to build a warehouse in both countries is above the median and 0 otherwise. We find that this variable does very well as an exclusion restriction for our data set and satisfies the requirement of influencing the fixed, rather than the variable, cost of exporting online.

[12] A recent eBay report, *Overseas Warehousing: New Trend in Cross-Border Commerce*, noted that over a quarter of Hong Kong's eBay sellers with sales over $10,000 store their goods in overseas warehouses.

[13] Data on the number of procedures to build a warehouse were not available for the following countries and territories: Andorra, Aruba, Barbados, Curacao, Holy Sea, Hong Kong, Kosovo, Libya, Liechtenstein, Macau, Malta, Monaco, Montenegro, Nauru, Netherlands Antilles, Palestine, San Marino, Sint Maarten, Somalia, Taiwan, Turkmenistan, and Tuvalu.

Table 2: Summary Statistics

Variable	Nr. Obs	Mean	Std. Dev	Min	Max
$distance_{ij}$	19,425	8,121	4,524	33	19,650
$land\ border_{ij}$	19,425	0.016	0.124	0	1
$island_{ij}$	19,425	0.043	0.203	0	1
$landlock_{ij}$	19,425	0.042	0.201	0	1
$legal_{ij}$	19,425	0.370	0.483	0	1
$language_{ij}$	19,425	0.162	0.368	0	1
$colony_{ij}$	19,425	0.007	0.081	0	1
$religion_{ij}$	19,425	0.169	0.223	0	1
$currency_{ij}$	19,425	0.013	0.112	0	1
FTA_{ij}	19,425	0.058	0.234	0	1
$exclusion\ restriction_{ij}$	19,425	0.244	0.429	0	1

Table 2 provides summary statistics for all the explanatory variables described above.

4 Empirical Approach

To estimate the effect of distance using online trade flows, we use the two-step estimation procedure proposed by Helpman et al. (2008). The first stage focuses on the extensive margin of online trade, and it is a Probit regression that yields a prediction for the probability that country j exports to country i. The second stage focuses on the intensive margin, and it is a gravity equation estimated on the subset of positive trade flows, which examines the volume of exports from j to i, conditional on exporting.

The Probit equation takes the following form:

$$\begin{aligned}\rho_{ij} = \beta_0 + \beta_1 \ln(distance)_{ij} + \beta_2\ land\ border_{ij} + \beta_3\ island_{ij} + \beta_4\ landlock_{ij} + \\ \beta_5\ legal_{ij} + \beta_6\ language_{ij} + \beta_7\ colony_{ij} + \beta_8\ religion_{ij} + \beta_9\ currency_{ij} \\ \beta_{10}\ FTA_{ij} + \beta_{11}\ exclusion\ restriction_{ij} + \lambda_i + \chi_j + \varepsilon_{ij}\end{aligned}$$

where:

> $\rho_{ij} = 1$ if the number of conversions that users in country i generate for businesses located in country j is positive, 0 otherwise.
>
> $distance_{ij}$ = population weighted distance, in km, between the largest cities of i and j.
>
> $land\ border_{ij} = 1$ if two countries share a land border, 0 otherwise.
>
> $island_{ij} = 1$ if both countries are islands, 0 otherwise.
>
> $landlock_{ij} = 1$ if both countries have no coastline, 0 otherwise.
>
> $legal_{ij} = 1$ if both countries' legal systems share the same origin, 0 otherwise.
>
> $language_{ij} = 1$ if both countries share an official language, 0 otherwise.
>
> $colony_{ij} = 1$ if the two countries were in a colonial relationship, 0 otherwise.
>
> $religion_{ij} = (\%Protestants_i \times \%Protestants_j) + (Catholics_i \times Catholics_j) + (\%Muslims_i \times \%Muslims_j)$
>
> $currency_{ij} = 1$ if both countries use the same currency, 0 otherwise.
>
> $FTA_{ij} = 1$ if both countries belong to a free trade agreement, 0 otherwise.
>
> $exclusion\ restriction_{ij} = 1$ if the number of procedures required to build a warehouse is above the median in both countries, 0 otherwise.

In the econometric model, λ_i and χ_j are importer and exporter fixed effects, respectively, and ε_{ij} is the error term.

We estimate the first stage Probit regression above in order to obtain a predicted probability of export, $\hat{\rho}_{ij}$, which in turn allows us to obtain consistent estimates for $\hat{\bar{w}}^*_{ij}$ (the fraction of exporting firms) and $\hat{\bar{\eta}}^*_{ij}$ (inverse Mills ratio). These estimates are needed as additional controls in the second stage estimation, as they correct the omitted variable and sample selection biases (see Section 2.3).

Helpman et al. (2008) propose three different ways to estimate the second stage regression. The dependent variable in the second stage estimation is, invariably, m_{ij}, which represents the exports from j to i. For us, m_{ij} is the number of conversion counts that users in country i generate for businesses located in country j. Since the coefficient associated with $\hat{\bar{w}}^*_{ij}$ does not enter the model linearly, an obvious first candidate for the second stage estimation is nonlinear least squares (NLS). A second candidate for the second stage estimation is an OLS regression, which includes among the independent variables the usual gravity controls, as well as $\hat{\bar{\eta}}^*_{ij}$ and a polynomial in $\hat{\bar{z}}^*_{ij}$ (predicted values of the latent variable; see Section 2.3). The authors obtain this estimating equation by relaxing the assumption governing the distribution of firm heterogeneity, which eliminates the nonlinearity induced by $\hat{\bar{w}}^*_{ij}$, and the need for a nonlinear estimator. Finally, the third candidate for the second stage estimation is a non-parametric functional form that emerges as a possibility after relaxing the joint normality assumption for the unobserved trade costs. Although these three estimation methods rely on different assumptions, similar to Helpman et al. (2008), we find that they yield very similar results.

5 Results

5.1 Main Results

Table 3 reports our paper's main results. Columns (1) and (2) underline the importance of expanding the country coverage in worldwide gravity analyses done on online data. In particular, in column (1), we use a simple linear estimator and report the results we obtain when, similar to Lendle et al. (2016), we limit the analysis to the 61 countries they covered. In column (2), we present the results we obtain when we expand the data set to 146 countries. This more than doubles the number of countries considered in the Lendle et al. (2016) paper.

Table 3: Worldwide Gravity Results – Google Data

	OLS	OLS	OLS	Probit	NLS	Polynomial	Non-Param.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$distance_{ij}$	-0.555***	-0.610***	-0.610***	-0.139***	-0.509***	-0.528***	-0.533***
	(0.044)	(0.029)	(0.029)	(0.011)	(0.034)	(0.038)	(0.039)
$land\ border_{ij}$	0.923***	0.743***	0.743***	0.221***	0.737***	0.786***	0.776***
	(0.144)	(0.142)	(0.142)	(0.043)	(0.138)	(0.142)	(0.143)
$island_{ij}$	0.399***	-0.029	-0.029	-0.114***	-0.020	-0.033	0.031
	(0.130)	(0.104)	(0.104)	(0.035)	(0.101)	(0.103)	(0.104)
$landlock_{ij}$	0.007	0.284**	0.284**	-0.005	0.221*	0.208*	0.213*
	(0.152)	(0.115)	(0.115)	(0.041)	(0.114)	(0.116)	(0.117)
$legal_{ij}$	0.127***	0.138***	0.139***	0.018	0.110***	0.115***	0.116***
	(0.049)	(0.038)	(0.038)	(0.014)	(0.037)	(0.037)	(0.037)
$language_{ij}$	0.635***	1.192***	1.192***	0.316***	1.031***	1.072***	1.070***
	(0.108)	(0.068)	(0.068)	(0.017)	(0.086)	(0.096)	(0.097)
$colony_{ij}$	0.545***	0.450**	0.450**	0.017	0.451**	0.457**	0.450**
	(0.180)	(0.202)	(0.202)	(0.098)	(0.192)	(0.196)	(0.198)
$religion_{ij}$	0.525***	1.051***	1.051***	0.071**	1.036***	1.038***	1.034***
	(0.136)	(0.084)	(0.084)	(0.031)	(0.081)	(0.083)	(0.083)
$currency_{ij}$	0.104	-0.009	-0.007	0.121**	0.099	0.122	0.101
	(0.092)	(0.151)	(0.151)	(0.056)	(0.153)	(0.156)	(0.157)
FTA_{ij}	0.328***	0.779***	0.779***	0.206***	0.558***	0.622***	0.615***
	(0.076)	(0.079)	(0.079)	(0.031)	(0.079)	(0.084)	(0.085)
$excl\ restriction_{ij}$			-0.009	-0.052***			
			(0.061)	(0.023)			
$\hat{\omega}^*_{ij}$ (predicted fraction of exporters)					0.302***		
					(0.091)		
$\hat{\eta}^*_{ij}$ (inverse Mills ratio)					1.719***	1.720***	
					(0.081)	(0.135)	
$\hat{\bar{z}}^*_{ij}$ (predicted latent variable)						1.363***	
						(0.361)	
$\hat{\bar{z}}^{*2}_{ij}$						-0.182*	
						(0.111)	
$\hat{\bar{z}}^{*3}_{ij}$						0.008	
						(0.011)	
Obs.	3,593	10,120	10,120	19,425	10,120	10,120	10,120
R^2	0.907	0.756	0.756	0.660		0.780	0.782

Notes: Exporter and importer fixed effects. The dependent variable is the number of conversion counts in logs. Only 61 countries considered in column (1); 146 countries considered in all other columns. Marginal effects at sample means and pseudo R^2 reported for Probit. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
*** Significant at 1%, ** significant at 5%, * significant at 10%

We find that limiting the analysis to the 61 countries with high levels of trade between them biases the coefficients of the variables intended to capture the closeness between trading partners downwards. Remoteness, be it geographic, cultural or economic, discourages trade. By excluding from a data set the country pairs that trade little with each other, one excludes the observations for which remoteness might matter the most. A comparison of the results generated using the smaller sample size of 61 countries and those generated using the larger sample size of 146 countries confirms this hypothesis.

With the increase in sample size, the elasticity of distance increases from -0.555 to -0.610. The coefficients of the other variables that measure cultural or economic closeness increase, as well. In particular, the coefficient of $language_{ij}$, $religion_{ij}$ and FTA_{ij} go up. All increases are statistically significant and point to the importance of expanding the country coverage of worldwide analyses beyond the countries that trade the most. Column (3) in Table 3 confirms the validity of the exclusion restriction used in the first stage regression. As noted in Section 3.2.5, a valid exclusion restriction should have no impact on the amount of trade between two countries, but it should have a statistically significant effect on the probability of two countries trading. The coefficients on our exclusion restriction reported in columns (3) and (4) of Table 3 confirm that it is exactly the type of variable that we need.

Column (4) of Table 3 reports the results of the Probit regression. This is the first stage estimation and it gives insight into the extensive margin of trade. The estimated coefficients reflect the effect that the independent variables have on the probability of two countries trading. Columns (5)-(7) of Table 3 report the results obtained in the second stage regressions for each of the estimators discussed in Section 4: nonlinear least squares, OLS with additional controls for $\hat{\bar{\eta}}_{ij}^*$ and a polynomial in $\hat{\bar{z}}_{ij}^*$, and finally, a non-parametric estimator. In columns (5)-(7), the estimated coefficients reflect the effect of the independent variables on the volume of trade between two countries, conditional on exporting. The results in these columns give us insight into the intensive margin of trade. We discuss the estimated coefficients below.

Distance

Distance has a negative and significant influence on both the extensive and the intensive margins of trade. This confirms our a priori belief that distance is not dead, even online. The second stage estimation results indicate that conditional on exporting, a 10% increase in the distance between two countries reduces the volume of trade between them by approximately 5.3%. Based on these results, we can safely conclude that among Google's advertising clients, distance lives on.

Besides uncovering the role of distance for online trade, the results reported in Table 3 also underline the importance of accounting for selection into trade when estimating worldwide gravity equations using online trade data. Using OLS biases the coefficient on distance upwards. The benchmark results, reported in column (2) of Table 3, show the elasticity of distance to be around -0.61. The magnitude of this coefficient is between 13% and 17% higher than the magnitude of the coefficients reported in columns (5)-(7) of Table 3, which are estimated using the methodology suggested by Helpman et al. (2008).

Other Geographic Characteristics

Our results show that besides distance, sharing a land border also significantly affects the volume of trade between two countries. The magnitude of the coefficient of the contiguity dummy in columns (5)-(7) is large: it predicts that conditional on exporting, online trade between two countries that share a land border is close to 120% larger than trade between countries that do not share a border. Contiguity also affects the extensive margin on trade. Column (4) of Table 3 shows that the coefficient of the *land* $border_{ij}$ dummy variable has a positive and significant impact on the probability of two countries trading online.

Countries that are islands or landlocked, characteristics which several studies done on data from the offline world found to negatively impact trade, no longer decrease the volume of trade in the virtual world. In columns (5)-(7), the coefficient of the dummy variable equalling 1 if both countries are landlocked is positive and statistically significant at the 10% level, while the coefficient on the dummy variable equalling 1 if both countries are islands is insignificant. We find some indication that the probability of exporting is negatively affected by both trading partners being islands. The coefficient on the $island_{ij}$ dummy variable, reported in column (4) is negative and significant.

Language

Our prior regarding the importance of language to online trade was correct. A good proportion of the fixed cost that a company faces when exporting online comes from translating its website and marketing materials. In column (4) of Table 3, we can see that the coefficient on the dummy variable measuring the effect of two countries sharing a language has a large, positive, and statistically significant effect on the probability of two countries trading. Once it decides to export, a company needs to provide customer support services to its foreign clients. In columns (5)-(7), we can see that sharing an official language also has a large and positive effect on the intensive margin of trade. Conditional on exporting, the volume of trade between countries that share an official language is more than 180% higher than the volume of trade between countries that do not speak the same language. These results underline the importance of customers and businesses sharing an official language and interacting via a website written in a language that they can both understand.

Other Cultural Characteristics

Besides language, other cultural affinities, such as similarity in religious beliefs, colonial ties and common legal systems also have a positive influence on the volume of online trade. The index measuring similarity in religious beliefs, in particular, has a positive and statistically significant effect, both on the extensive and the intensive margins of online trade. The dummy variables measuring the effect of two countries having similar legal systems or colonial ties have a positive and statistically significant effect only on the intensive margin on trade. Conditional on exporting, the volume of trade between countries that have similar legal systems is 12% larger than the volume of trade between countries that do not. For countries that have colonial ties, the volume of trade conditional on exporting is 57% higher than for countries that do not.

5.1.1 Economic Characteristics

Belonging to a free trade agreement has a positive and statistically significant effect on the probability of countries trading and on the volume of online exports between them once the decision to export is made. All else being equal, the quantity of online trade, conditional on exporting, is around 80% higher for countries that belong to a free trade agreement than for countries that do not.

Using a common currency influences the extensive, but not the intensive margin of online trade. The coefficient of the dummy variable $currency_{ij}$, which takes the value of 1 if two countries have a common currency and 0 otherwise, is positive and significant in column (4) of Table 3, indicating that the indicator influences the probability of countries trading with each other.

This is a sensible result. Before exporting to a new country, an online company incurs costs in order to ensure that it can provide price information and accept electronic payments in the currency of the importing country. Assume a US software developer would like to start promoting his software programs to UK-based users. Currently, he only sells his programs within the United States. All prices on his website are expressed in US dollars and he can only accept credit card payments in US dollars.

Before he starts to advertise his products in the UK, he makes changes to his website in order to ensure that his UK-based users see price figures in pounds. For example, he might implement some code on his website that uses the information contained in his website visitors' IP addresses to determine their country.[14] If the code determines that the visitors are US-based, the prices on the site are expressed in US dollars. If the code determines that the visitors are UK-based, the prices on the site are expressed in pounds. These website changes are costly not only in terms of developing the code to determine users' country of residence, but also in terms of converting all prices from US dollars to pounds.[15]

Besides the changes he makes to his website, the American software developer also needs to ensure that he can accept credit card payments in British pounds. For example, he might need to set up a new merchant account or to switch from a single-currency to a multiple-currency merchant account.[16] Depending on the provider, opening a new merchant account or adding a currency to an existing merchant account might incur setup fees.

The example of the software developer illustrates the fixed costs that a business might incur in order to sell its products in a country that uses a different currency. The results reported in column (4) of Table 3 confirm the fact that $currency_{ij}$ influences the fixed

[14] Alternatively, he can add a drop-down menu to his website allowing users to choose their preferred country.

[15] The seller can program his website so that the US dollar to pound price conversion is done automatically, using up-to-date exchange rate information, or he can be perform the price conversion manually, by entering a 'fixed' pound price for each product.

[16] A merchant account is a bank account that enables companies, like our software developer, to accept and process debit or credit card payments.

cost of exporting and has a positive and statistically significant effect on the extensive margin of trade.

Once the initial setup is complete, $currency_{ij}$ is unlikely to influence the volume of online trade between two countries. In the example above, once the US-based software developer undertakes the necessary investment to make changes to his webpage, his site automatically – and costlessly – displays prices in British pounds to UK-based users. Furthermore, the new setup of his merchant account allows him to process credit card payments made in British pounds without him incurring additional costs. Due to the fact that following the initial setup, a company does not need to incur additional costs to sell to countries using different currencies, sharing a common currency should not impact the volume of trade between two locations. Our results in columns (5)-(7) of Table 3 confirm this hypothesis. The coefficient on the dummy variable $currency_{ij}$ is insignificant.

5.2 Alternative Exclusion Restriction

Given the importance of accounting for selection into trade when assessing the role of distance for online trade, we are interested in suggesting other potential exclusion restrictions that researchers can use in their analyses. The results presented in Section 5.1 reveal another solid candidate for an identifying restriction, namely $currency_{ij}$.

As we have seen, a common currency affects the fixed – and not the variable – cost associated with exporting online. Furthermore, this variable has the added advantage of being readily available at the worldwide level and easy to generate: it simply takes the value of 1 if two countries share a common currency.

Table 4: Results Using an Alternative Exclusion Restriction – Google Data

	OLS	Probit	NLS	Polynomial	Non-Param.
	(1)	(2)	(3)	(4)	(5)
$distance_{ij}$	-0.610***	-0.139***	-0.510***	-0.531***	-0.529***
	(0.029)	(0.011)	(0.034)	(0.038)	(0.039)
$land\ border_{ij}$	0.743***	0.222***	0.738***	0.788***	0.762***
	(0.142)	(0.043)	(0.138)	(0.142)	(0.143)
$island_{ij}$	-0.029	-0.114***	-0.015	-0.027	-0.026
	(0.104)	(0.035)	(0.101)	(0.103)	(0.103)
$landlock_{ij}$	0.284**	-0.004	0.219*	0.206*	0.202*
	(0.115)	(0.041)	(0.114)	(0.116)	(0.116)
$legal_{ij}$	0.138***	0.018	0.110***	0.115***	0.116***
	(0.038)	(0.014)	(0.037)	(0.037)	(0.037)
$language_{ij}$	1.192***	0.316***	1.030***	1.073***	1.061***
	(0.068)	(0.017)	(0.087)	(0.097)	(0.098)
$colony_{ij}$	0.450**	0.016	0.452**	0.457**	0.460**
	(0.202)	(0.098)	(0.192)	(0.196)	(0.197)
$religion_{ij}$	1.051***	0.072**	1.035***	1.038***	1.033***
	(0.084)	(0.031)	(0.081)	(0.083)	(0.083)
FTA_{ij}	0.779***	0.205***	0.567***	0.635***	0.611***
	(0.079)	(0.031)	(0.078)	(0.084)	(0.084)
$currency_{ij}$	-0.009	0.115**			
	(0.151)	(0.056)			
$\hat{\omega}^*_{ij}$ (predicted fraction of exporters			0.303***		
			(0.092)		
$\hat{\bar{\eta}}^*_{ij}$ (inverse Mills ratio			1.719***	1.709***	
			(0.082)	(0.136)	
$\hat{\bar{z}}^*_{ij}$ (predicted latent variable)				1.321***	
				(0.363)	
$\hat{\bar{z}}^{*2}_{ij}$				-0.169	
				(0.112)	
$\hat{\bar{z}}^{*3}_{ij}$				0.006	
				(0.011)	
Obs.	10,120	19,425	10,120	10,120	10,120
R^2	0.756	0.660		0.780	0.782

Notes: Exporter and importer fixed effects. The dependent variable is the number of conversion counts in logs. Common currency is the excluded variable. Marginal effects at sample means and pseudo R^2 reported for Probit. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
*** Significant at 1%, ** significant at 5%, * significant at 10%

In Table 4, we replicate the analysis we used to generate the results for Table 3, but we drop the indicator variable signaling whether the number of procedures required to build a warehouse is above the mean in both countries and we instead use $currency_{ij}$ as our exclusion restriction.

The results we obtain in Table 4 are very similar to those we report in Table 3, thus confirming the appropriateness of the common currency dummy as an alternative exclusion restriction.

6 Conclusion

In this study, we challenge the popular view that distance is dead in the virtual world. To do so, we use a proprietary data set from Google to perform the most extensive online worldwide gravity analysis to date. We find that conditional on exporting, a 10% increase in the distance between two countries lowers the volume of online trade between them by 5.3%. We conclude that distance continues to negatively affect trade, even for transactions done over the Internet.

Besides the negative effect of distance on online trade, our study also uncovers a few other interesting facts relating to online trade. We find that certain cultural affinities matter greatly in the virtual world. A common official language, in particular, has a large effect on the volume of online trade between two countries. We also find that economic ties, such as a common currency, which facilitate trade in the offline world, have an insignificant effect on the intensive margin of trade in the virtual world. Our results indicate that the economic policies that encourage trade in the offline world might not be as successful at increasing the number of cross-border transactions online.

Our study also points to the importance of accounting for selection into trade when performing gravity analyses at the worldwide level. To aid future research, we identify two variables that can be used as exclusion restrictions in a selection model designed to uncover trade patters in the online world. The first variable measures the difficulty of building a warehouse, while the second one captures the benefits of sharing a common currency in a virtual environment.

In this study, we answer the important question of how much distance matters online. However, our estimates do not provide a precise and exact point of comparison between the role of distance online versus offline. Our conversion count measure does not have a direct equivalent in the offline world. Also, the composition of online and offline trade flows differs. In our future work, as more detailed data regarding the composition of the products traded online become available, we hope to be able to bring more preciseness to the online-offline comparison.

References

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Anderson, James E, "The gravity model," *Annual Review of Economics*, 2011, *3* (1), 133–160.

_ **and Eric van Wincoop**, "Gravity with gravitas: A solution to the border puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Baltagi, Badi H, Peter Egger, and Michael Pfaffermayr, "Panel data gravity models of international trade," 2014.

Blum, Bernardo S and Avi Goldfarb, "Does the internet defy the law of gravity?," *Journal of International Economics*, 2006, *70* (2), 384–405.

Brynjolfsson, Erik, Yu Hu, and Mohammad S Rahman, "Battle of the retail channels: How product selection and geography drive cross-channel competition," *Management Science*, 2009, *55* (11), 1755–1765.

Chaney, Thomas, "Distorted gravity: The intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

_ , "The gravity equation in international trade: An explanation," Technical Report, National Bureau of Economic Research 2013.

Disdier, Anne-Célia and Keith Head, "The puzzling persistence of the distance effect on bilateral trade," *Review of Economics and Statistics*, 2008, *90* (1), 37–48.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (2), 1741–1779.

— **, Samuel S Kortum, and Sebastian Sotelo**, "International trade: Linking micro and macro," Technical Report, National Bureau of Economic Research 2012.

Evans, Carolyn L and James Harrigan, "Distance, time, and specialization: Lean retailing in general equilibrium," *American Economic Review*, 2005, *95* (1), 292–313.

Ferreira, Fernando and Joel Waldfogel, "Pop internationalism: Has half a century of world music trade displaced local culture?," *Economic Journal*, 2013, *123* (569), 634–664.

Forman, Chris, Anindya Ghose, and Avi Goldfarb, "Competition between local and electronic markets: How the benefit of buying online depends on where you live," *Management Science*, 2009, *55* (1), 47–57.

Goolsbee, Austan, "Competition in the computer industry: Online versus retail," *Journal of Industrial Economics*, 2001, *49* (4), 487–499.

Grossman, Gene, "Comment," in "The Regionalization of the World Economy," University of Chicago Press, 1998.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Cultural biases in economic exchange?," *Quarterly Journal of Economics*, 2009, *124* (3), 1095–1131.

Harrigan, James and Anthony J Venables, "Timeliness and agglomeration," *Journal of Urban Economics*, 2006, *59* (2), 300–316.

Head, Keith and Thierry Mayer, "Illusory border effects: distance mismeasurement inflates estimates of home bias in trade," in "The Gravity Model in International Trade: Advances and Applications. Editors: Bergeijk and Brakman" Citeseer 2009.

— **and** — , "Gravity equations: Workhorse, toolkit, and cookbook," *Handbook of International Economics*, 2014, *4*, 131–195.

Heckman, James J, "Sample selection bias as a specification error," *Econometrica*, 1979, *47* (1), 153–161.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

Hortaçsu, Ali, F Asís Martínez-Jerez, and Jason Douglas, "The geography of trade in online transactions: Evidence from eBay and MercadoLibre," *American Economic Journal: Microeconomics*, 2009, *1* (1), 53–74.

Huang, Rocco R, "Distance and trade: Disentangling unfamiliarity effects and transport cost effects," *European Economic Review*, 2007, *51* (1), 161–181.

Hummels, David L, "Toward a geography of trade costs," *Available at SSRN 160533*, 1999.

Lendle, Andreas and Pierre-Louis Vézina, "Internet technology and the extensive margin of trade: Evidence from eBay in emerging economies," *Review of Development Economics*, 2015, *19* (2), 375–386.

_ , Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina, "There goes gravity: eBay and the death of distance," *Economic Journal*, 2016, *126* (592), 406–441.

Limao, Nuno and Anthony J Venables, "Infrastructure, geographical disadvantage, transport costs, and trade," *World Bank Economic Review*, 2001, *15* (3), 451–479.

Martin, Will and Cong S Pham, "Estimating the gravity equation when zero trade flows are frequent," *MPRA Paper*, 2008.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

— **and Gianmarco IP Ottaviano**, "Market size, trade, and productivity," *Review of Economic Studies*, 2008, *75* (1), 295–316.

Rauch, Ferdinand, "The geometry of the distance coefficient in gravity equations in international trade," 2015.

Rauch, James E, "Networks versus markets in international trade," *Journal of International Economics*, 1999, *48* (1), 7–35.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Appendix

Helpman et al. (2008)'s Results (Reproduced)

	OLS (1)	Probit (2)	NLS (3)	Polynomial (4)	Non Param. (5)
$distance_{ij}$	-1.167*** (0.040)	-0.213*** (0.016)	-0.813*** (0.049)	-0.847*** (0.052)	-0.789*** (0.088)
$land\ border_{ij}$	0.627*** (0.165)	-0.087 (0.072)	0.871*** (0.170)	0.845*** (0.166)	0.863*** (0.170)
$island_{ij}$	-0.553** (0.269)	-0.173** (0.078)	-0.203 (0.290)	-0.218 (0.258)	-0.197 (0.258)
$landlock_{ij}$	-0.432** (0.189)	-0.053 (0.050)	-0.347** (0.175)	-0.362* (0.187)	-0.353* (0.187)
$legal_{ij}$	0.535*** (0.064)	0.049*** (0.019)	0.431*** (0.065)	0.434*** (0.064)	0.418*** (0.065)
$language_{ij}$	0.147* (0.075)	0.101*** (0.021)	-0.030 (0.087)	-0.017 (0.077)	-0.036 (0.083)
$colony_{ij}$	0.909*** (0.158)	-0.009 (0.130)	0.847*** (0.257)	0.848*** (0.148)	0.838*** (0.153)
$religion_{ij}$	0.281** (0.120)	0.141*** (0.034)	0.120 (0.136)	0.139 (0.120)	0.100 (0.128)
$currency_{ij}$	1.534*** (0.334)	0.216*** (0.038)	1.077*** (0.360)	1.150*** (0.333)	1.107*** (0.346)
FTA_{ij}	0.976*** (0.247)	0.343*** (0.009)	0.124 (0.227)	0.241 (0.197)	0.065 (0.348)
$reg.\ costs_{ij}$	-0.146 (0.100)	-0.108*** (0.036)			
$days\ \&\ proc._{ij}$	-0.216* (0.124)	-0.061** (0.031)			
$\hat{\bar{\omega}}^*_{ij}$			0.840*** (0.043)		
$\hat{\bar{\eta}}^*_{ij}$			0.240** (0.099)	0.882*** (0.209)	
$\hat{\bar{z}}^*_{ij}$				3.261*** (0.540)	
$\hat{\bar{z}}^{*2}_{ij}$				-0.712*** (0.170)	
$\hat{\bar{z}}^{*3}_{ij}$				0.060*** (0.017)	
Obs.	6,602	12,198	6,602	6,602	6,602
R^2	0.693	0.573		0.701	0.706

Notes: Exporter and importer fixed effects. Marginal effects at sample means and pseudo R^2 reported for Probit. Regulation costs are excluded variables in all second stage specifications. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
*** Significant at 1%, ** significant at 5%, * significant at 10%