January 12, 2021

Is the Internet Bringing Down Language-based Barriers to International Trade?*

by

Erick Kitenge § and Sajal Lahiri ‡

Abstract

This paper investigates the interaction between internet access and language similarities in international trade. We apply recent developments in gravity analysis to bilateral aggregate export data from about 205 countries over the period 1954-2014. We find a positive impact of both internet and language similarities, but a negative interaction between them in international trade. Our results reveal that the language elasticity of trade is lower when more people have access to the internet and that the impact of the internet is lower when the trading partners use similar languages. That is, we find that the reliance on language similarities for international trade has been reduced because of the internet. Our results also pass a number of robustness checks. One of the policy conclusions is that the international institutions such as the World Trade Organization (WTO) should put more emphasis on the expansion of Information Technologies (IT) access among the population in their member countries.

JEL Classification: F10, F40
Key words: Languages, Internet Access, Gravity Model, International trade

§ College of Business, Prairie View A&M University, Prairie View, TX. 77446, U.S.A.; email: emkitenge@pvamu.edu
‡ School of Analytics, Finance, and Economics, Southern Illinois University Carbondale, Carbondale, Illinois 62901, U.S.A.; E-mail: lahiri@siu.edu

* We are thankful to Jeffrey Bergstrand and Yoto V. Yotov for their suggestions and help. We appreciate Thomas Zylkin's help in sharing with us the newly developed ppml-panel-sg-new Stata command and his comments on an earlier draft. We are also grateful for comments received at the 93rd Annual Conference of the Western Economic Association International (held in Vancouver, Canada during 26-30 June, 2018) and at the 20th Annual Conference of the European Trade Study Group (held in Warsaw, Poland during 13-15 September, 2018). Usual caveats apply.

1 Introduction

Does the internet reduce the impact of language-based barriers to international trade? This is the main question that this paper addresses. An answer to this question has straightforward implications for policy makers.

Different types of transaction costs are known to affect international trade, and these costs include inadequate and ineffective communication among people that are trading with each other. In particular, when trading partners are linguistically distant and their language differences are very pronounced, they are likely to trade little with each other (Baier and Bergstrand, 2004; Isphording and Otten, 2013; Fidrmuc and Fidrmuc, 2016). From the meta-analysis results provided by Head and Mayer (2014a), based on estimations from 159 papers, trade is likely to be, on an average, 72% (the median being 63%) higher when trading partners use similar languages.

Internet access, on the other hand, could facilitate trade both *directly* by helping the formation of trade networks and *indirectly* by reducing language-based barriers mentioned above. Internet can help trade directly by providing, to potential exporters and/or importers, information needed to make sound plans. Through the Internet, exporters and/or importers could identify potential competitors, find best commercial deals, know about trade policies in foreign countries and legal aspects of international trade, explore financing options, and target specific countries to export to. Moreover, economic agents can either promote export or search import opportunities by owning websites, by conducting web marketing campaigns, and by using e-marketplaces. With internet access, the time for an information system – which links the source of information to the receiver (source → transmitter → channel → receiver → destination) – to be completed is considerably shortened (Harris, 1995; Miller, 1951). The internet can also boost trade by maximizing the conversion of one-time clients into regular clients by sending personalized letters and available promotions (UNICTAD, 2004). There are many other channels through which internet access could impact trade.

For example, "The internet has become the world's largest marketplace, opening for business 24 hours a day, seven days a week, providing endless opportunities to buyers and sellers. With the internet, suppliers can more easily find information about new markets and are able to advertise to numerous buyers at once, thus reducing the fixed costs of entry into foreign markets." (Lin, 2015, pp. 409).

Lin (2015) found that a 10 percentage point increase in the Internet users increases international trade by 0.2-0.4% due to a reduction of information cost for traders, while Freund and Weinhold (2004) found that a 10 percentage point increase in the growth of web hosts in a country leads to about a 0.2 percentage point increase in export growth by reducing market-specific fixed costs of trade.

Turning now to the indirect effect of the Internet on trade via its effect on language-based transaction costs, it should first of all be noted that the ability to communicate has been improving with technological progress in the form of Internet access across the globe. A language is a tool within the information system, that contribute to the transmission of information from its source to the receiver. However, information can be noisy because of errors or mismatches that occur during the decoding by the receiver (Miller, 1951). The level of noisiness in the information depends upon the magnitude of the errors. Internet access in this context is seen as a communication tool apt to boost trade by reducing information noisiness (Harris, 1995). With internet access, the ability to decode the information is improved regardless of differences between the languages used at the source and at the destination. The resulting saving in time and the reduction in the level of noise in the information reduce information costs and undermine the need for language similarities in international trade.

It is possible to perceive the impact of internet on languages in many other different ways. By bringing economics agents on a virtual platform, regardless of their physical locations, the internet network increases the predisposition and the ability to communicate with each other. The increased predisposition offsets the necessity of other communication tools such as languages. With more internet access, the power for language similarities to connect people and drive trade is dampened. Consequently, with the internet, which has high capabilities of translation, individuals do not need to speak similar languages in order to trade. The offsetting power of internet access on language-based barriers to trade could be more significant in the future during the ongoing digital revolution characterized by the development of blockchain technologies, virtual currencies, and smart contracts in international trade (de Caria, 2016).

Therefore, the internet access will not only impact production because of higher input market accessibility, lower information cost, faster and easier dissemination of knowledge, and the resulting increased productivity, but will also reduce information costs between all the countries due to the increased decoding ability regardless of the language at the source either for speeches or for written documents (e.g., transactions documents, products labels, etc.). This offsetting power of internet access over languages is what this paper will try to measure. In other words, one would not only expect the coefficient of Internet to be positive in explaining bilateral trade, but at the same time the coefficient of the interaction term between Internet and Language similarities between countries should be expected to be negative.

In order to explain bilateral trade, it is now quite standard to use the gravity model which was first introduced by Tinbergen (1962) in order to predict market integration or bilateral trade in terms of economic sizes and bilateral distances. Since then, many other important factors have been added to the gravity model. For example, Anderson and van Wincoop (2003) found that national borders reduce trade between industrial countries, while Eichengreen and Irwin (1998) and Campbell (2010) established that history and culture matter for international trade because of habit persistence in consumer behavior. According to Rose and van Wincoop (2001) and Campbell (2013), bilateral trade is higher when trading countries belong to the same currency union. Many other factors have been used to explain the volume of bilateral trade such as, but not limited to, colonial relationships, trade agreements, WTO membership, landlockedness, and, as mentioned before, language differences. Being intuitive, flexible, grounded on solid theoretical foundations, reflecting a more realistic general equilibrium environment, and endowed with a higher predictive power, the gravity model has become a workhorse in the analysis of international trade (Yotov et al., 2016).

Apart from introducing new explanatory variables, in recent years there have been many methodological innovations in the gravity analysis of bilateral trade. First, when the number of countries is large, there are likely to be many zero observations for bilateral trade and therefore estimating a log-linearized gravity model with the Ordinary Least Squares (OLS) method may become problematic. Santos Silva and Tenreyro (2006) proposed the Poisson Pseudo-Maximum Likelihood (PPML) method which accounts for zero trade values and also deals with the problem heteroscedasticity arising due to the log-transformation of the gravity equation.

Second, in recent years, bilateral trade costs are captured by introducing pairwise fixed effects (Agnosteva et al., 2014; Egger and Nigai, 2015). Baier and Bergstrand (2007) suggested the use of pairwise fixed effects to deal with the endogeneity of explanatory variables which can depend on other omitted cross-sectional variables. However, these pairwise fixed effects will absorb all the time-invariant bilateral covariates.[1] This can be problematic if one of those variables is the focus of analysis. In such situations, using a two-step procedure, one can recover the estimates for time-invariant bilateral variables in the presence of pairwise fixed effects, as it has been done by Anderson and Yotov (2016).

Third, recent papers using panel data have incorporated exporter-time and importer-time fixed effects which are called multilateral resistances in trade (Anderson and van Wincoop, 2003). The omission of the multilateral resistances has been branded as the 'gold medal' mistake by Baldwin and Taglioni (2006). The theory behind the gravity equation suggests that trade between two locations is a function of bilateral barriers and/or drivers, and of

[1] Some of the time-invariant bilateral covariates which have been used in gravity analysis are contiguity and former colonial link (Hummels, 2007; Shiue, 2002).

inward and outward multilateral resistances. The information cost is a component of trade costs and multilateral resistances reflect the influence of other countries when, say, country i (the exporter) is trading with country j (the importer). Accordingly, higher trade costs between country i and countries other than i and j will enhance export to country j, while higher trade costs between country j and the other countries will boost import from country i (Anderson and van Wincoop, 2003). However, it can be a challenge to incorporate as covariates time-varying country-specific variables and even sometimes time-varying bilateral variables, in the presence of multilateral resistances as the former cannot be identified (Head and Mayer, 2014a).

Finally, given that the adjustment of trade in response to changes in other covariates takes time, Cheng and Wall (2005) suggested using intervals data instead of pooled data or continuous panel data. Trefler (2004) used 3-year intervals data, Baier and Bergstrand (2007) used 5-year intervals data, Olivero and Yotov (2012) found similar results using 3-year and 5-year intervals data, and Anderson and Yotov (2016) used 4-year intervals data.

Using aggregate export data from 205 countries – possibly the largest coverage in a gravity analysis – over the period 1954-2014, this paper estimates a gravity model, taking into account the recent methodological developments discussed above. The three key explanatory variables are: the common language, the internet use, and the interaction of the two. Because of the reasons mentioned above, we use the PPML method to estimate the model, and our main results are obtained with 3-year intervals data. However, and we also carry out a series of robustness checks, which include the estimation of a log-linearized OLS regression.

Even though we explain our methodology in detail later on, here we state some of the salient features. First of all, common language is a time-invariant bilateral variable and it is absorbed by pairwise fixed effects.[2] To identify the coefficient of this variable, we use a two-step method proposed by Anderson and Yotov (2016). Second, as mentioned above, in the presence of multilateral resistance in the form of exporter-time and importer-time fixed

[2]The coefficient of its interaction with Internet access can however still be identified.

effects, using time-varying country-specific variables as covariates can be a challenge, and the internet variable from available sources is one such variable. In any case, the country-specific internet variable cannot fully capture the effect of internet on *bilateral* trade or how it mitigates the effect of common language on bilateral trade, as in order to examine its effect on time-variant bilateral trade, we need to consider a dyadic internet variable. For example, if one of the trading partners has no internet coverage, the level of coverage in the other partner may not have any effect on the level of trade between these two countries. The literature has in fact suggested using dyadic variables constructed from country-specific time-varying variables. However, most dyadic constructed variable pose a different challenge and that is in the interpretation of the estimates as shown by (Heid et al., 2017; Beverelli et al., 2018). One also has to be careful in the construction of the dyadic variable as some of them can be absorbed by the multilateral resistance terms. This paper uses the network theory from Jackson and Watts (2002) and constructs a dyadic internet variable that is meaningful and can be identified at the same time. It represents the value of the internet network between a pair of countries at a given time.

Our main findings are as follows. Internet access boosts international trade, but its interaction with language similarities is negative. That is, the language elasticity of trade is lower when more people in the trading partners have access to the internet and the impact of the internet is lower when the trading partners use similar languages. In other words, we show that the dependence on language similarities in international trade has been reduced because of the growth in internet access. We also find that non-high-income exporting countries benefit relatively more from internet access than high-income exporting countries.

The remainder of this paper is organized as follows. We describe the estimation methodology and the data in Section 2, we report and explain our empirical results in Section 3, and we present our concluding remarks in Section 4.

2 Estimation Methodology and Data

2.1 Econometric Specification

We estimate a gravity equation which is similar to the ones estimated by Anderson and van Wincoop (2003) and Bergstrand et al. (2015), but extend it by including common languages and Internet access in concert. The gravity equation has been derived theoretically by Anderson and van Wincoop (2003):

$$X_{ij} = \frac{E_j * Y_i}{Y_w} * \left(\frac{t_{ij}}{P_j * \Pi_i}\right)^{1-\sigma}, \quad \sigma > 1, \tag{1}$$

where, X_{ij} denotes the value of export from country i to county j, E_j is the total expenditure in country j, Y_i represents sales of goods by country i at destination prices, and Y_w is the world output. P_j and Π_i represent respectively inward and outward multilateral resistances, t_{ij} denotes trade cost factors on exports from country i to country j, and σ is the elasticity of substitution between the goods.[3] Equation 1 can be empirically estimated through the following equation:

$$\begin{aligned} X_{ijt} &= exp\,[\beta_0 + \beta_1\text{FTA}_{ijt} + \beta_2\text{INT}_{ijt} + \beta_3\text{LANG}_{ij} + \beta_4\text{INT}_{ijt} * \text{LANG}_{ij} \\ &\quad + \beta_5 ln\_\text{DIST}_{ij} + \beta_6\text{CNTG}_{ij} + \beta_7\text{CLNY}_{ij} + \beta_8\text{LAND}_{ij} + \beta_9\text{Trend} \\ &\quad + \beta_{10}\text{TREND} * \text{LANG}_{ij} + \beta_{11}\text{LIND}_{ijt} + n_{it} + \theta_{jt} + \delta_{ij}] + \epsilon_{ijt}, \end{aligned} \tag{2}$$

where X_{ijt} stands for the exports from country i (the exporter) to country j (the importer) at period t, ϵ_{ijt} is the error term, and n_{it}, θ_{jt} and δ_{ij} are respectively exporter-time, importer-time and bilateral, fixed effects.

FTA_{ijt} represents the membership of a Free Trade Agreement (FTA). This variable takes the value 1 if both countries i and j are members of the same Free trade agreement in year t, and zero otherwise.

[3] There are many different interpretation of P_j and Π_i, see Head and Mayer (2014a)

LANG_{ij} represents a bilateral common language variable. There are many ways that this variables can be defined, and we shall discuss that later on. Positive coefficients of this language variable would imply that language similarities boost international trade. Language similarities can influence international trade since they impact upon communication between agents in trading countries and therefore on trading costs.

DIST_{ij} stands for bilateral distances between capital cities of trading partners. This variable captures the remoteness between countries and is expected to have a negative sign since more distant countries are expected to trade less because of higher trade costs. LAND_{ij} is a binary variable that takes the value of 1 if at least one of the trading partners is landlocked. Note that the landlockedness variable in the literature is typically a country-specific variable (Mayer and Zignago, 2011). Here we construct a dyadic landlockedness variable that intends to capture non-shipping transportation costs in international trade. Since the majority of goods in international trade are transported by ocean, it might be more expensive for landlocked countries than for coastal ones to carry out international trade because of difficult and complex transportation conditions from the port to final consumers. Therefore, the coefficient of this variable is expected to be negative.

The contiguity (CNTG_{ij}) variable takes 1 (0) when the trading partners are (are not) contiguous. Colonial links (CLNY_{ij}) is a dummy variable taking the value of 1 when the trading partners have ever been in a colonial relationship. The coefficients of the CONT_{ij} and CLNY_{ij} are expected to be positive since we should expect lower transportation costs for contiguous countries and a higher level of trade when countries have ever been in colonial relationships.

We also add the interaction between the Trend variable and the language variable ($\text{Trend} * \text{LANG}_{ij}$) to account for any potential impact of other observable or unobservable time-varying factors that could be impacting the language elasticity of trade, prior to the internet era.

Linder (1961) introduced a theory that is based on consumer demand rather than on the supply side as it is the case with the well-known Heckscher-Ohlin theory. More specifically, according to him, countries with high income per capita spend a larger fraction of their income on high-quality goods. high-income countries will trade in high-quality but differentiated products, and countries with similar income per capita should trade more intensely with one another (Juan Carlos, 2010). Since capital-intensive goods are usually produced by high- income countries compared to labor-intensive goods, this means that countries with dissimilar incomes should trade less with each other, contrary to the Heckscher-Ohlin predictions. There have been studies that included the Linder's effect in Gravity analysis (Baltagi et al., 2003; Jerry G. and Marie C., 1987; Tang, 2003). In fact, Baltagi et al. (2003) showed that the absence of a variable Linder's effect could render the estimated coefficients biased. In our estimation we include a Linder's effect variable LIND_{ijt}.[4] A negative value of the coefficient will confirm Linder's hypothesis. We shall see if the presence affects other coefficients by also estimating the equation without it. As a robustness check, we shall consider different definitions of the variable.

The Internet access (INT_{ijt}) variable captures the connectedness between economic agents as a result of technological progress. Freund and Weinhold (2004) used the number of web hosts in a country which is not a time-dependent bilateral variable but a time-dependent country-specific one. In the Translog functions literature, researchers such as Sharma (1991) and Kitenge (2016) used the trend variable as a proxy of technological progress. We construct the Internet access variable that is a time-dependent bilateral/dyadic variable and which is in conformity with network externalities, according to which the benefits to users of a network should depend upon the fact that other users are also on the network (Harris, 1995). Internet access is captured by the value of the internet network in line with Jackson and Watts (2002). These scholars considered a social and economic network model that has a finite set of players $N = 1, ..., n$. The network relations among the players are represented by graphs

[4] In empirical studies Linder effects are captured by the following indicators: $[ln\_Y_i - ln\_Y_j]^2$; $|Y_i - Y_j|$; and $|ln\_Y_i - ln\_Y_j|$, where Y_i and Y_j denote the income per capita in the exporting country and in the importing countries, respectively.

whose nodes or vertices represent the players and whose links (edges or arcs) capture the pairwise relations. The complete network, denoted gN, is the set of all subsets of N of size 2. The value of a network is represented by $v : \{g \mid g \subset gN\} \rightarrow R$, where $v(g)$ represents the total utility or production of the network. The set of all such functions is V. The value function allows for a wide variety of applications and quite general forms of externalities. In some applications the value will be an aggregate of individual utilities or production values, $v(g) = \sum_i u_i(g)$, where $u_i : \{g \mid g \subset gN\} \rightarrow R$ (Jackson and Watts (2002), pp. 270-271).

In our case, players are the individuals with access to the internet and the value of each vertex represents the satisfaction and all the net benefits an individual could get from being part of the internet network. With N individual in a network, the value of the complete network would be

$$R_{pqt} = N_t * (N_t - 1) * r_{pqt}, \tag{3}$$

where r_{pqt} is the average net benefit for each tie from individual p to individual q included in the network in period t.

In other words, if k players are in a connected component of a pairwise stable network, then there must be exactly $k - 1$ links. Thus in a complete network or a network without restrictions of k individuals, each individual can be linked to $k-1$ individuals. Since internet links can be both within and across countries, k in our case is the sum of the number of people with internet links in each of the two countries, *i.e.*, $k_{it} + k_{jt}$, and therefore the value of the complete network is:

$$R_{pqt} = (k_{it} + k_{jt}) * (k_{it} + k_{jt} - 1), \tag{4}$$

normalizing $r_{pq} = r = 1$

We shall use $\ln\_R_{ijt}$ in (4) as a proxy for Internet access (INT_{ijt}) for all periods after 1990 and zero before 1990, in our empirical investigations.

Our constructed internet network variable also allows us to circumvents the usual challenge to identify country-specific time-varying variables in the presence of multilateral resistances (i.e. exporter-time fixed effects, importer-time fixed effects). The importance of the functional form in the identification is well-recognized in the literature. Accordingly, Head and Mayer (2014b) stated that in the presence of importer-time and exporter-time fixed effects a variety of potentially interesting trade determinants can no longer be identified in a gravity equation. Notably, (1) anything that affects exporters propensity to export to all destinations (such has having hosted the Olympics or being an island), (2) variables that affect imports without regard to origin, such as country-level average applied tariff, (3) sums, averages, and differences of country-specific variables (pp.31). Baldwin and Taglioni (2006) demonstrated that the multilateral resistance terms should be accounted for by exporter-time and importer-time fixed effects in a dynamic gravity estimation framework with panel data. However, like Head and Mayer (2014b), Yotov et al. (2016) noted, "exporter-time and importer-time fixed effects will also absorb the size variables from the structural gravity model as well as all other observable and unobservable country-specific characteristics, which vary across these dimensions, including various national policies, institutions, and exchange rates" (p. 19). Therefore, the functional form of the internet variable is important. To address this type of challenges the literature has suggested using dyadic variables constructed from country-specific time-varying variables. However, many dyadic constructed variables may also pose a challenge with the interpretation of the estimates as shown by Heid et al. (2017) and Beverelli et al. (2018). Our constructed internet variable does not present such challenge given that it represents the value of the internet network.

Coefficients of the internet access variable (INT_{ijt}) and, more importantly, its interaction with the language variable LANG_{ij} are our main focus of interest. Normally, we expect Internet access to boost the volume of trade since technological improvement decreases information noisiness and provides access to markets regardless of physical locations. The coefficient of the interaction term between the Internet use and the language variable is expected to be negative in concurrence with the intuition provided in the introduction. Therefore, negative coefficients of the interaction terms between Internet access and the language variables would reflect the ability of the Internet to break down language-based barriers to trade. It will also reflect the higher ability for internet to boost trade when countries use different languages.

Taking advantage of our panel data set and following suggestions by Olivero and Yotov (2012), we use exporter-time (n_{it}) and importer-time (θ_{jt}) fixed effects. These variables capture, *inter alia*, the fact that trade between location i and location j is influenced by variables such as prices in other locations of the world, and these prices are influenced by bilateral distances between locations i and j on one hand with the other market locations on the other (Baier and Bergstrand, 2007). We also include pairwise fixed effects δ_{ij}. These fixed effects help us to deal with the type of endogeneity between the dependent variable and time-varying regressors like e.g. FTA_{ijt}, LIND_{ijt} and INT_{ijt} that arise because of presence of unobserved heterogeneity (Baier and Bergstrand, 2007; Yotov et al., 2016). These fixed effects are supposed to better capture bilateral trade costs. However, since pairwise fixed effects will absorb all the time invariant bilateral variables (one of them – LANG_{ij} – being our coefficient of interest), we drop such variables and consider the estimation of the following equation:

$$\begin{aligned} X_{ijt} &= exp\,[\beta_0 + \beta_1 \text{FTA}_{ijt} + \beta_2 \text{INT}_{ijt} + \beta_4 \text{INT}_{ijt} * \text{LANG}_{ij} \\ &\quad + \beta_9 \text{Trend} + \beta_{10} \text{TREND} * \text{LANG}_{ij} + \beta_{11} \text{LIND}_{ijt} + n_{it} + \theta_{jt} + \delta_{ij}] + \epsilon_{ijt}. \end{aligned} \tag{5}$$

In order to recover the coefficient of the bilateral time-invariant variables, building on the work of Agnosteva et al. (2014) and Egger and Nigai (2015), Anderson and Yotov (2016) proposed a two-step procedure to recover the estimates of all the time-invariant variables. This approach consists of estimating the gravity equation in the first step without time-invariant variables, then regressing the estimated pairwise fixed effects on time-invariant variables. Thus, after estimating the pairwise fixed effects from (5), we regress it on the variables that were stripped off from as shown below:

$$\begin{aligned}\hat{\delta_{ij}} &= exp\left[\beta_0 + \beta_3\text{LANG}_{ij} + \beta_5 ln\_\text{DIST}_{ij} + \beta_6\text{CONT}_{ij}\right. \\ &\quad \left. + \beta_7\text{CLNY}_{ij} + \beta_8\text{LAND}_{ij} + \bar{n}_i + \bar{\theta}_j\right] + \mu_{ij}, \end{aligned} \tag{6}$$

where a $\widehat{}$ over a variable denotes the estimated value of it, and $\bar{n}_i$ and $\bar{\theta}_j$ are exporter and importer fixed effects..

We estimate our gravity models in a multiplicative form by applying the Poisson Pseudo-Maximum Likelihood (PPML) method.[5] This method has the merit of incorporating zero export values that would be excluded if we used the OLS method to estimate a log-linearized gravity equation. It also addresses the issue of heteroscedastic error terms created by the log transformation of the gravity model (Santos Silva and Tenreyro, 2006).

The direct impact of internet access is equal β_2. Following Cameron and Trivedi (2009) and given that our main language variable is a dummy Common Official Language (COL)-, the total impact of internet between trading partners with similar languages is $\beta_2 + \beta_4$. The direct impact of COL is equal to β_3, while the indirect impact through internet access is β_4. The total impact of COL is $\beta_3 + \beta_4 * \text{IN}\bar{\text{T}}_{ijt}$.

2.2 Data

We use annual aggregate export data from 205 countries over the period 1954-2014 constructed by Fouquin and Hugot (2016).[6] We have complemented the data-set with data on intra-national (domestic) trade computed, following suggestions from Baier et al. (2016), by taking the difference between total domestic production and total export. Gravity variables data are retrieved from the *Centre d'tudes Prospectives et d'Informations Internationales* (CEPII).

[5]The ppml-panel-sg Stata command developed by Larch et al. (2017) enables us to handle higher fixed effects.

[6] These information is retrieved from the new dataset constructed by Fouquin and Hugot (2016) and available since November 2016 on the CEPII website. It contains *aggregate* trade data from 1827 until 2014 (http://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=9134).

FTA and Colonial ties data were constructed by Head et al. (2010), and language variables by Melitz and Toubal (2014); the other gravity variables are provided by Mayer and Zignago (2011).

From the country-specific landlockedness variable provided by Mayer and Zignago (2011), we construct a dyadic landlockedness variable, i.e., a bilateral one. To be more specific, we use the country-specific landlockedness provided by Mayer and Zignago (2011) to build a dyadic variable that takes the value of 1 when at least one of the trading partners is landlocked, and 0 otherwise. This bilateral dummy variable will enable us to incorporate in empirical studies transportation challenges and/or non-shipping transportation costs when at least one of the trading partners is landlocked.[7]

The value of k_{it}'s in the definition of the Internet access variable INT_{ijt} (see (4)) is the number of individuals with Internet access in country i in year t. To obtain this variable, we multiply the percentage of people with Internet access by the population, both taken from the from Word Development Indicators (WDI). This Internet is a world-wide public computer network. It provides access to a number of communication services including the World Wide Web and carries email, news, entertainment and data files, irrespective of the device used (not assumed to be only via a computer - it may also be by mobile phone, PDA, games machine, digital TV etc.). The International Telecommunication Union (ITU) collects the related primary information for numerous countries through questionnaires administered to various economic agents (e.g. households, businesses, etc.). Therefore, Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, and digital TV (WorldBank, 2017).

[7] More than 70% of international trade occurs between non-contiguous countries and a bulk of trade between these countries is transported by the ocean. Transportation costs can then be a function of various transport modes used to deliver products to final consumers such as, but not limited to, truck and railways in the exporting countries, ocean liner, and rail and truck in importing countries (Hummels, 2007). However, trade occurring within the same country are likely to occur through the main transport mode in the country and thus assimilated to trade between coastal countries. This variable is in line with Shiue (2002) who used the geographic location to capture trade costs due to the fact that coastal locations are exposed to lower transportation costs. Our constructed data on dyadic landlockedness is available upon request.

The internet data set is available only from 1990. Since there was no internet before 1990, we complete our internet variable with zero values for the years prior to 1990.

The lists of countries included in this paper are provided in the appendix (Table 13) and the summary statistics of the variables can be found in Table 1.

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
X_{ijt}	1,754,748	440	30,200	0	9,720,000
RTA_{ijt}	1,444,121	0.04	0.20	0	1.00
k_t	1,630,589	5.62	28.00	0	653.00
INT_{ijt}	1,576,275	12.07	14.24	0	41.98
COL_{ij}	1,444,121	0.16	0.37	0	1.00
ln_DIST_{ij}	1,444,121	8.70	1.08	0	9.89
$CNTG_{ij}$	1,444,121	0.02	0.13	0	1.00
$CLNY_{ij}$	1,444,121	0.01	0.12	0	1.00
$LAND_{ij}$	1,444,121	0.27	0.45	0	1.00
$LIND1_{ijt}$	1,475,021	4.51	5.65	0	54.18
$LIND2_{ijt}$	1,475,021	9.44	2.67	-9.47	18.47
$LIND3_{ijt}$	1,475,021	1.72	1.25	0	7.36

X_{ijt} and k_t are in millions.

3 Empirical Results

In this section we shall present our empirical results and carry out a number of robustness checks. These will be done in sections 3.3. and 3.5 respectively. However, before that in section 3.1, we try to present a *prima facie* evidence in support of our main hypothesis, viz. that the internet has been bringing down the language-based barriers to international trade. Then in section 3.2, we shall present the estimation of our benchmark model, without the internet variable but with a number of alternative language variables. The purpose of section 3.2 is to show the robustness of the standard gravity model — see, for example, Baier and Bergstrand (2007) and Anderson and Yotov (2016) — to our extended data in terms of both country coverage — 205 of them, and the number of years — our data is for the period 1954-2014. In section 3.4, we shall divide the countries into high-income and non-high-income groups and see if our hypothesis holds for one group and not for the other.

3.1 Internet and trade between countries with similar languages

In Figure 1, every point is one of the 205 countries in our data. We plot changes in the share of exports to countries with similar languages between the years 1998 and 2014, in the vertical axis, and the growth rates in internet coverage between the same period of time in the horizontal axis.[8] We chose 1998 as the base period because there are many zero and missing observations for internet coverage before that, particularly among the Non-Higher-Income (NIH) countries. The scatter in the figure gives a sense of raw correlation between the two variables, i.e. without controlling for anything. As we can see from this figure, there is a small negative relationship between the two, the cross-country correlation coefficient being -0.22. Thus, Figure 1 provides a *prima facie* evidence in support of our main hypothesis, viz. that the Internet has been bringing down the language-based barriers to trade.

3.2 Language Elasticity of Trade

In this sub-section, we shall use different language variables in the estimation a Gravity model without the internet or the Linder's variable and see how robust the model is across the language variables and in our extended dataset. We estimate the gravity equation using the PPML method, with multilateral resistance and bilateral fixed effects.

There are many variables that have been used to capture the similarities in languages between a pair of countries. Each one possibly capture a different aspect of the similarities and therefore is expected to have different effect on international trade. LANG_{ij} can either be the Common Official Language (COL_{ij}), the Common Spoken Language (COL_{ij}), or

[8] The growth in internet coverage is $ln(k_{i,2014}) - ln(k_{i,1998})$ with k_i being the number of individuals with internet access in country i.

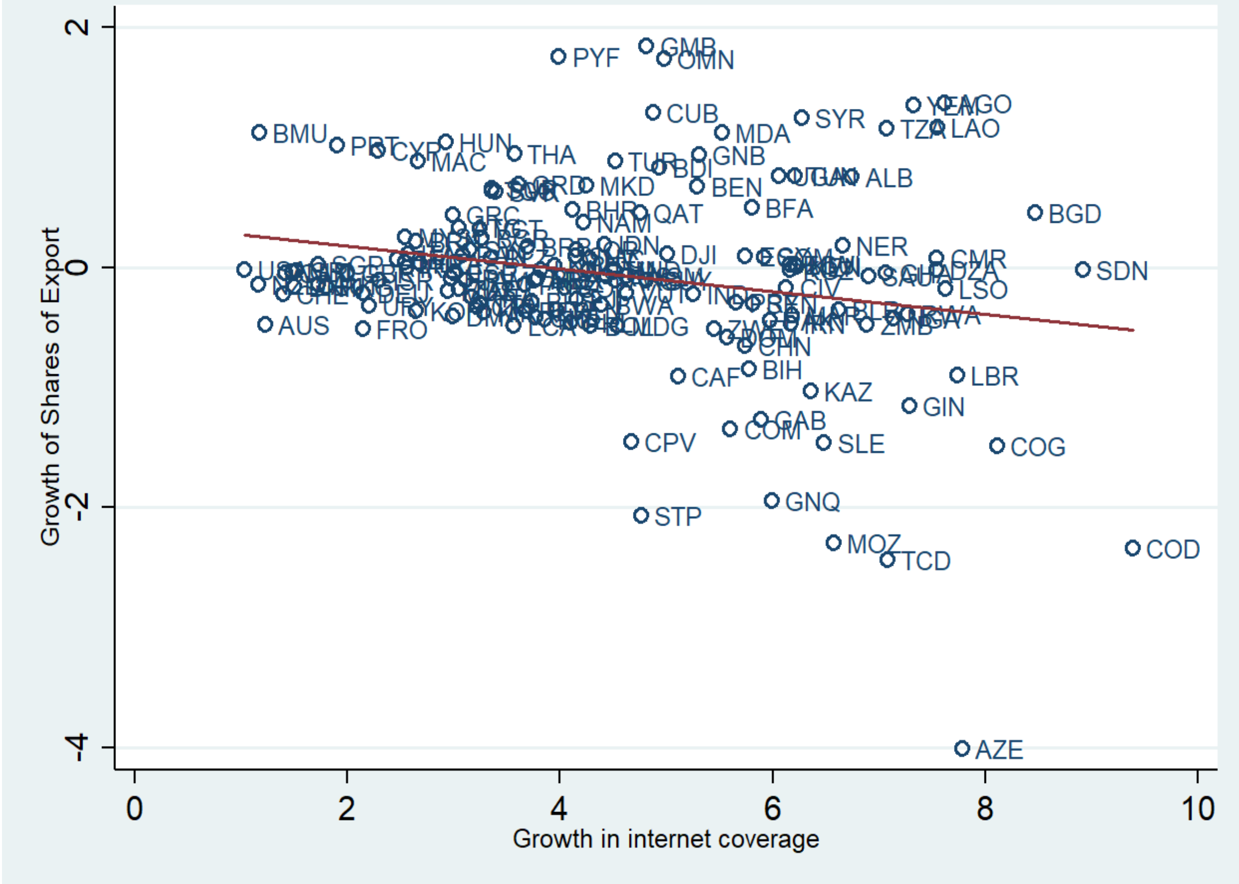

Fig 1.Growth in Internet coverage and Share of exports to countries with COL=1 from 1998 to 2014.

the Common Native Language (CNL_{ij}) (Crystal, 1985; Fielding et al., 2015). The common language could also be the Language Proximity ($LP1_{ij}$ & $LP2_{ij}$), or the aggregate index Common Language (CLE_{ij} & CL_{ij}). An official language is available to anyone in the country in a language the person understands. COL is a dummy variable that takes the value of 1 if two countries have similar official languages and 0 otherwise. Official languages reflect institutionalized support for translation from a chosen language into the others that are spoken at home. Assuming that every spoken language can be a native one, CSL and CNL are constructed together, and they show the percentage of individuals using similar spoken and native languages, respectively, across countries. An example to clarify the distinction between CNL and CSL is the following. In Kinshasa, the capital city of the Democratic Republic of the Congo,[9] Lingala is the spoken language while there are many native languages

[9] The official Language of the Democratic Republic of the Congo is French.

depending upon tribes of each inhabitant.

To account for the ease of communication between two countries if they have no common native language, Melitz and Toubal (2014) constructed language proximity variables (LP1 and LP2) to capture linguistic proximities between native languages. Then, they have proposed aggregate index of common language variables (CLE and CL) that summarize the evidence about the linguistic influences in an index resting strictly on exogenous linguistic factors. However, the aggregate language variables are not recommended for studies that specifically focus on linguistic proximities such our current study (Melitz and Toubal, 2014)

The table below gives the correlations between the different language variables. As can be seen, many of the correlations are quite high (more than 0.5) and therefore using them together many lead to multicollinearity. In this subsection we shall only use them one at a time. Later on in our regressions with the internet variable in it, we shall use some of the language variables together as one of our robustness checks and these will be given in Table 11.

Table 2: Language Correlation Matrix

	CSL	COL	CNL	LP1	LP2	CL	CLE
CSL	1.00						
COL	0.54	1.00					
CNL	0.66	0.50	1.00				
LP1	0.15	-0.15	-0.08	1.00			
LP2	0.13	-0.21	-0.14	0.83	1.00		
CL	0.72	0.52	0.80	0.39	0.43	1.00	
CLE	0.72	0.51	0.80	0.40	0.44	0.99	1.00

Table 3: Language elasticity of Trade (PPML method)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Trade agreements	0.405***	0.405***	0.405***	0.405***	0.405***	0.405***	0.405***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Common Official Language	0.352***						
	(0.000)						
Common Spoken Language		0.828***					
		(0.000)					
Common Native Language			0.036				
			(0.000)				
Language Proximity 1				0.068***			
				(0.000)			
Language Proximity 2					0.101***		
					(0.000)		
Common Language						0.437**	0.464**
						(0.000)	(0.000)
Geographic Distance	-0.767***	-0.721***	-0.797***	-0.787***	-0.786***	-0.758***	-0.744***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contiguity	0.858***	0.809***	0.852***	0.777***	0.786***	0.862***	0.801***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Colonial Links	1.171***	1.103***	1.303***	1.289***	1.280***	1.255***	1.206***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Landlockedness	-1.023***	-0.938***	-1.048***	-1.027***	-1.024***	-1.007***	-0.982***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
# Observations_1	426,072	426,072	426,072	426,072	426,072	426,072	426,072
R-square_1	0.999	0.999	0.999	0.999	0.999	0.999	0.999
# Observations_2	498,343	498,343	498,343	428,243	428,243	498,343	428,243
R-square_2	0.894	0.896	0.895	0.899	0.899	0.895	0.899

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses below each coefficient.. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity.

PPML estimates of the Gravity equation are given in Table 3. In the first 5 columns, we show what happens when we introduce Common Official Language (COL_{ij}), Common Spoken Language (CSL_{ij}), Common Native Language (CNL_{ij}), Language Proximity (LP1_{ij} & LP2_{ij}), one at a time. In column (6) we use CLE -the Common Language Index based on a level specification-, and in column (7) we use CL -the Common Language Index based on log specification. All but the CNL_{ij} perform well.

Our coefficients using the PPML estimation method, with multilateral resistance variables and pairwise fixed effects, are smaller than those from Melitz and Toubal (2014). Our sample cover the period from 1954 to 2014, thus lower estimates than those from studies that do not go that far back could reflect an overtime increase in the language elasticity of trade. Borchert and Yotov (2017) found that the language elasticity of trade increased from 0.294 in 1986 to 0.315 in 1996. Coefficients of the other gravity variable seem to not be affected by the type of language variable considered.

3.3 Internet access and language similarities in international trade

We shall now present our main results by introducing internet (INT). As for the the variable capturing language similarities, we choose here Common Official Language (COL). Later on in Tables 12 and 13, we shall look at some of the other language variables to see how robust our results are. It is to be noted that COL_{ij} indicates the contribution of institutionalized support for translation from a chosen language into the others that are spoken at home, and institutional supports can be implemented through numerous tools such as, but not limited to, school curricula, cultural centers, and nowadays the internet.

The results are presented in Table 4.[10] In terms of the specification, the first column of Table 4 is the same as the first column in Table 3, except that it is augmented with Internet access, its interaction with COL_{ij}, the trend, and the interaction between the trend variable and COL_{ij}. Trend and the interaction between Trend and COL have been introduced to

[10]The coefficient of TREND cannot be identified in the presence of the multilateral resistances.

control for the possibility that the effect of COL_{ij} on trade has been declining even before internet came into the picture.

We use the two-step method described in 2.1 and with the exception of bilateral time invariant variables (e.g. common official language, geographic distance, contiguity, colonial links, and landlockedness), the other variables are included in the first step.

From the results reported in Table 4, international trade is higher when trading partners are involved in trading agreements, when they use similar official languages, when they are contiguous, and when if they have ever been in a colonial relationship. Geographic distance dampens international trade, and the volume of international trade is lower when at least one of the trading partners is landlocked. In addition, the volume of international trade is 0.01% higher when the value of internet network is 1% higher, and most importantly, the interaction between COL_{ij} and INT is negative. This means, the total impact of internet, $\beta_2 + \beta_4 * \text{COL}_{ij}$, is lower when countries use similar official languages, and the total impact of COL_{ij}, $\beta_3 + \beta_4 * \text{IN}\bar{\text{T}}_{ijt} + \beta_9 * \text{TREND}$, is lower when the value of INT is higher.

The higher impact of internet when trading partners are linguistically distant would reflect the greater ability for individuals or businesses to carry out business transactions, through the internet, with economic agents located in countries using different languages. The lower impact of COL_{ij}, when more individuals have access to the internet, means that the need for language similarities in international trade is lower when more people have access to the internet. Our results also reveal a positive interaction between the trend variable and COL_{ij}. This indicates that the language elasticity of trade could have been increasing (as also shown in Fig.2), with the total impact of COL_{ij} being 0.274 in 1954, 0.436 in 1972, 0.530 in 1990, but this total impact shifted down to 0.27 in 1993, and then increased to 0.303 in 2014. Even though the total impact of COL is also increasing after 2005 (about 3.19% every 3 years), the rate of increase is much lower than during the period before 1990 (about 6.9% every 3 years). Changes in the language elasticity of trade observed in this study possibly provide an answer to why Borchert and Yotov (2017) found the impact of language

Table 4: Internet Access and Official Languages in international Trade (PPML method)

	Without Linder Effects		With Linder Effects	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.271***	(0.000)	0.279***	(0.000)
Internet Access	0.010*	(0.05)	0.020***	(0.001)
Common Official Language	0.309***	(0.002)	0.265***	(0.002)
Internet Access*Common Official Language	-0.021***	(0.000)	-0.016***	(0.000)
Trend*Common Official Language	0.012***	(0.000)	0.009***	(0.001)
Geographic Distance	-0.751***	(0.000)	-0.697***	(0.000)
Contiguity	0.873***	(0.000)	0.601***	(0.000)
Colonial Links	0.888***	(0.000)	1.030***	(0.000)
Landlockedness	-1.003***	(0.000)	-0.892***	(0.000)
Linder Effects			-0.111***	(0.000)
# Observations_1	359136		359136	
R-square_1	0.999		0.999	
# Observations_2	393454		393454	
R-square_2	0.883		0.809	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=$[\ln(Y_i)-\ln(Y_j)]^2$.

to increase from 0.294 in 1986 to 0.315 in 1996, but decreased to 0.120 in 2006. Thus, the impact of language has been shifted down during the internet era. Fig.2 also shows that the total impact of COL_{ij} is a decreasing function of internet access. This result reflects the fact that, because of internet access, it is increasingly becoming less necessary to use the same language as the one used by a foreign vendor in order to carry out a transaction.

In the second column of Table 4, we include the Linder's effect and it is found to be negative, reflecting lower trade between countries with higher differences in terms of quality requirements. This finding is in line with the predictions by Linder (1961). As we noted before, according to Baltagi et al. (2003), the absence of a variable Linder's effect could render the estimated coefficients biased. Comparing the two columns of table 4, we note that the two are qualitatively similar. In Table 5, we examine if the differences in the

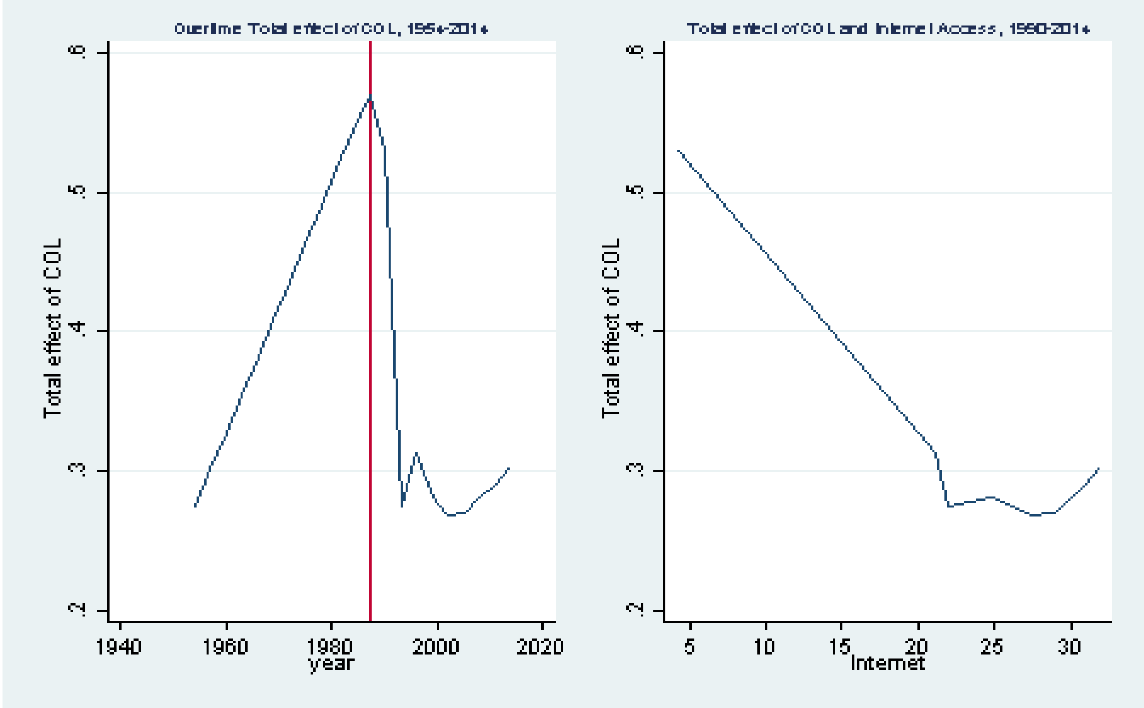

Fig 2. Total Effect of COL from the PPML estimation method

magnitudes of the coefficients are statistically significant. As it can be seen, only one of the coefficient, that of DIST, is statistically different.

3.4 Internet and language: differential effects by country groups

Recognizing that the impact of internet on language elasticity of trade could potentially depend upon the income levels of trading partners, we modify the equation estimated in Table 4. We introduce a dummy variable which has a somewhat different definition in the two columns of Table 6. In the first column, D_1 is equal to 1 if, corresponding to the trade variable X_{ijt}, the exporter country i is a Non-High Income (NHI) country, and 0 otherwise,[11] while in the second column, D_1 takes the value of 1 if the importer country j is an NHI, and 0 otherwise. Our second regression from Table 4 is augmented with D_1 and its interaction with our main variables of interest (internet access, COL, and the interaction between COL and internet access).

[11] We follow the World Development Indicator's definition of NHI and High Income (HI) countries (World-Bank, 2017)).

From the results reported in Table 6, the internet access has a higher direct impact on exports from NHI countries (-0.003+0.033=0.03). However, the effect of internet on imports is positive for both groups of countries and the difference in effects on the groups is statistically insignificant.

The total impact of 1% increase in internet access on exports from NHI countries to countries with common official languages (COL=1) and different official languages (COL=0) are -0.003+0.033-0.013-0.008=0.009% and -0.003+0.033-0.013=0.017% respectively. The total impact of 1% increase in internet access on exports from HI countries to countries with common official languages and different official languages are -0.003-0.013=-0.016% and -0.003% (which is statistically insignificant) respectively. The total impact of 1% increase in internet access on imports in NHI countries from countries with common official languages and different official languages are 0.036-0.017-0.013-0.006=0% and 0.036-0.017-0.013=0.006% respectively. Finally, the total impact of 1% increase in internet access on imports in HI countries from countries with common official languages and different official languages are 0.036-0.013=0.023% and 0.036% respectively.

Table 5: Tests of differences between coefficients without and with the Linder Effects

Variables	Without Linder Effects		With Linder Effects		a-b	Se (a-b)	t-value	Sig
	a) Coef.	Se	b) Coef	Se				
Trade Agreements	0.271***	0.047	0.279***	0.059	-0.008	0.075	-0.106	
Internet Access	0.010*	0.005	0.020***	0.006	-0.010	0.008	-1.280	
Common Official Language	0.309***	0.101	0.265***	0.086	0.044	0.133	0.332	
Internet Access*Common Official Language	-0.021***	0.003	-0.016***	0.002	-0.005	0.004	-1.387	
Trend*Common Official Language	0.012***	0.003	0.009***	0.003	0.003	0.004	0.707	
Geographic Distance	-0.751***	0.013	-0.697***	0.015	-0.054	0.020	-2.720	***
Contiguity	0.873***	0.166	0.601***	0.135	0.272	0.214	1.271	
Colonial Links	0.888***	0.102	1.030***	0.100	-0.142	0.143	-0.994	
Landlockedness	-1.003***	0.081	-0.892***	0.070	-0.111	0.107	-1.037	
Linder Effects			-0.111***	0.011				

*,**,and *** denote statistical significance at 10%, 5%, and 1%.

From the first column of Table 4, the negative interaction between Internet access and COL_{ij} is higher for exports from NHI countries (-0.013-0.008=-0.021) than for exports from

HI countries (-0.013). From the second column of Table 4, the negative interaction between internet access and COL_{ij} is higher for imports in NHI countries (-0.013-0.0055=-0.019) than for imports in HI countries (-0.013).

3.5 Robustness Checks

To check the robustness of our findings, we carry out a number of exercises. First, we consider two alternative variables to capture the Linder effects variables, and these are $ln|Y_i - Y_j|$ and $|ln\ Y_i - ln\ Y_j|$. The results reported in Table 7 show positive impacts of internet access and COL_{ij}, but a negative interaction between them, as it was the case before. However, in one of the two alternatives, the coefficient of the interaction term is higher than the internet coefficient, and in the other alternative it is the other way round. Coefficients of the other gravity variables are similar to those in Table 4.

Table 6: Internet Access and Official Languages in international Trade with Income-group dummies

	With Exporter dummies		With Importer dummies	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.282***	(0.000)	0.273***	(0.000)
Internet Access	-0.003	(0.716)	0.036***	(0.001)
D1*Internet Access	0.033***	(0.000)	-0.017	(0.102)
Common Official Language	-0.091	(0.257)	-0.298**	(0.026)
D1*Common Official Language	0.920***	(0.000)	0.997***	(0.000)
Internet Access*Common Official Language	-0.013***	(0.000)	-0.013***	(0.000)
D1*Internet Access*Common Official Language	-0.008***	(0.001)	-0.006**	(0.022)
Trend*Common Official Language	0.009***	(0.003)	0.008***	(0.006)
Geographic Distance	-0.689***	(0.000)	-0.709***	(0.000)
Contiguity	0.586***	(0.000)	0.669***	(0.000)
Colonial Links	1.175***	(0.000)	0.941***	(0.000)
Landlockedness	-0.856***	(0.000)	-0.896***	(0.000)
Linder Effects	-0.114***	(0.000)	-0.103***	(0.000)
# Observations_1	359136		359136	
R-square_1	0.999		0.999	
# Observations_2	393454		393454	
R-square_2	0.816		0.835	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. D1 denote Exports from or Imports in NHI countries. Linder Effects Variable=[ln(Yi)-ln(Yj)]2 .

Second, we replicate the results in the second column of Table 4, but with 2-year and 3-year intervals data. The results are reported in Table 8 and are consistent with those in Table 4. More specifically, we find positive coefficients for trade agreements, internet access, common official language, the interaction between trend and COL_{ij}, CNTG_{ij}, and CLNY_{ij}, and we find negative coefficients for the interaction between COL_{ij} and the internet access, ln DIST_{ij}, LAND_{ij}, and the Linder effects.

Table 7: Internet Access and Official Languages in international Trade with different Linder Effect variables

	Lind=ln|Yi-Yj|		Lind=|ln(Yi)-ln(Yj)|	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.280***	(0.000)	0.266***	(0.000)
Internet Access	0.017***	(0.001)	0.019***	(0.001)
Common Official Language	0.422***	(0.000)	0.316***	(0.000)
Internet Access*Common Official Language	-0.019***	(0.000)	-0.017***	(0.000)
Trend*Common Official Language	0.007**	(0.011)	0.008***	(0.006)
Geographic Distance	-0.689***	(0.000)	-0.672***	(0.000)
Contiguity	0.926***	(0.000)	0.703***	(0.000)
Colonial Links	0.991***	(0.000)	1.071***	(0.000)
Landlockedness	-0.945***	(0.000)	-0.921***	(0.000)
Linder Effects (Lind)	-0.068***	(0.000)	-0.454***	(0.000)
# Observations_1	359136		359136	
R-square_1	0.999		0.999	
# Observations_2	393454		393454	
R-square_2	0.835		0.778	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=[ln(Yi)-ln(Yj)]2 .

Third, instead of restricting the influence of internet access on COL only, we include interaction terms between the internet access variable and the other bilateral time-invariant variables (e.g. geographic distance ln DIST_{ij}, contiguity CNTG_{ij}, colonial links CLNY_{ij}, and landlockedness LAND_{ij}). The results reported in Table 9 confirm the negative interaction between COL and the internet access, and show a positive interaction with CNTG_{ij} and a negative interaction with CLNY$_{ij}$. The coefficient of COL$_{ij}$ on its own is not significant, but the coefficient of its interaction term with TREND is much higher than that in Table 4. We have computed the total effects of COL and have plotted it against trend and the internet access variable, the graph we obtained mirror exactly Fig.2.

Table 8: Internet Access and Official Languages in international Trade (2 and 4-year intervals data)

	2-year		4-year	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.255***	(0.000)	0.247***	(0.000)
Internet Access	0.018***	(0.005)	0.018***	(0.002)
Common Official Language	0.241	(0.112)	0.273*	(0.079)
Internet Access*Common Official Language	-0.016***	(0.000)	-0.017***	(0.000)
Trend*Common Official Language	0.011***	(0.000)	0.011***	(0.000)
Geographic Distance	-0.694***	(0.000)	-0.698***	(0.000)
Contiguity	0.570***	(0.000)	0.608***	(0.000)
Colonial Links	1.019***	(0.000)	0.993***	(0.000)
Landlockedness	-0.871***	(0.000)	-0.876***	(0.000)
Linder Effects (Lind)	-0.111***	(0.000)	-0.107***	(0.000)
# Observations_1	525672		265818	
R-square_1	0.999		0.999	
# Observations_2	569090		293298	
R-square_2	0.82		0.833	

We apply the PPML method to 2-year and 4-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=[ln(Yi)-ln(Yj)]2 .

Fourth, we use the following alternative definitions of the internet access variable INT_{ijt}: $ln(k_{it}) + ln(k_{jt})$ and $k_{it}/(max\{k\}) + k_{it}/(max\{k\})$. While our main internet access variable represent the value of the internet network, it is difficult to assign specific meaning to these alternative definitions. However,the purpose here is to see if the sign of the coefficient of the interaction term between COL$_{ij}$ and these alternate definitions of internet access continue to be negative. The results reported in Table 10. The coefficients of both alternative internet variables are dropped because of perfect multicolinearity with the multilateral resistance fixed effect variables, but their interactions with COL_{ij} can be identified and are negative.

Table 9: Interaction between internet Access and bilateral time-invariant variables

	Without Linder Effects		With Linder Effects	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.248***	(0.000)	0.255***	(0.000)
Internet Access	0.026***	(0.000)	0.037***	(0.000)
Common Official Language	0.011	(0.894)	0.058	(0.646)
Internet Access*Common Official Language	-0.026***	(0.000)	-0.023***	(0.000)
Internet Access*Geograhic Distance	0.0002	(0.542)	-0.0004	(0.101)
Internet Access*Contiguity	0.005*	(0.089)	0.007**	(0.019)
Internet Access*Colonial Links	-0.008***	(0.001)	-0.004**	(0.037)
Internet Access*Landlockedness	0.001	(0.627)	0.004	(0.178)
Trend*Common Official Language	0.021***	(0.000)	0.017***	(0.000)
Geographic Distance	-0.751***	(0.000)	-0.677***	(0.000)
Contiguity	0.766***	(0.000)	0.307**	(0.011)
Colonial Links	1.091***	(0.000)	1.222***	(0.000)
Landlockedness	-1.049***	(0.000)	-0.936***	(0.000)
Linder Effects (Lind)			-0.128***	(0.000)
# Observations_1	359136		359136	
R-square_1	0.999		0.999	
# Observations_2	393454		393454	
R-square_2	0.835		0.757	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=[ln(Yi)-ln(Yj)]2 .

Fifth, We shall now try different alternatives to our main language variable COL. Around Table 3, we discussed different language variables that people have considered in the literature. Given the results in Table 3, the obvious options are common spoken language (CSL), language proximity 1 (LP1) and language proximity 2 (LP2). From Table 2, we see that the correlation between LP1 and LP2 is very high and that between CSL and the language proximity variables are very low. We therefore consider the three variables one at a time and also consider CSL and LP1, and CSL and LP2 together. The results are presented in table 11. The coefficient of CSL is not statistically significant when used on its own. However, in this case the coefficient of CSL*TREND is positive and significant as in the case of COL. The coefficient of CSL*INT is also negative and significant as in Table 4. When CSL is

Table 10: Internet Access and Official Languages in international Trade

	$INT_{ijt}=\ln(k_{it})+\ln(k_{jt})$		$INT_{ijt}=(k_{it}/\max k)+(k_{jt}/\max k)$	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.2580***	(0.000)	0.3927***	(0.000)
Internet Access	0.0172	(1,000)	n.i.	n.i.
Common Official Language	0.2775***	(0.0012)	0.1237*	(0.0965)
Internet Access*Common Official Language	-0.0153***	(0.000)	-0.4721***	(0.0004)
Trend*Common Official Language	0.0080***	(0.0068)	0.0038**	(0.0436)
Geographic Distance	-0.6980***	(0.000)	-0.7269***	(0.000)
Contiguity	0.6155***	(0.000)	0.7629***	(0.000)
Colonial Links	1.0011***	(0.000)	0.9142***	(0.000)
Landlockedness	-0.8732***	(0.000)	-0.9808***	(0.000)
Linder Effects (Lind)	-0.1086***	(0.000)	-0.0663***	(0.000)
# Observations_1	337969		359235	
R-square_1	0.9996		0.9995	
# Observations_2	370788		393543	
R-square_2	0.8191		0.8389	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=[ln(Yi)-ln(Yj)]2 .

used in conjunction with either LP1 or LP2, the coefficient of CSL becomes significant at 10% level, the interactions of the language variables with TREND are positive and significant, and the interactions of the language variables with internet access are negative and significant. Interestingly, the coefficients of LP1 and LP2 are negative, a finding which also intrigued Melitz and Toubal (2014).[12]

Finally, we regress the specification in Table 4 with Linder effect, using the OLS estimation method. Here we shall also run the regression with CSL instead of COL. The results are in Table 12. The coefficient of the interaction between language (COL or CSL) is negative

[12] Melitz and Toubal (2014) wrote, "We also find, rather uncomfortably, that linguistic proximity harms bilateral trade for this combination of languages, which is possibly simply a reflection of the earlier result that native English helps exceptionally since English figures prominently in the separate measure of LP2 in the same estimate (whose effect is now correspondingly higher). In other estimates for individual languages, we also find that LP2 helps to interpret foreign languages for Spanish and is harmful for French and Arabic. All these results about the significance of separate native languages in interpreting foreign languages based on linguistic proximity remain a mystery to us." (p. 361)

Table 11: Internet Access and Spoken Languages in international Trade (PPML estimations)

	Coef.	Prob.	Coef.	Prob.	Coef.	Prob.	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.303***	(0.00)	0.357***	(0.00)	0.363***	(0.00)	0.257***	(0.00)	0.267***	(0.00)
Internet Access	0.021***	(0.00)	-0.003	(0.53)	-0.003	(0.61)	0.021***	(0.00)	0.02***	(0.00)
Common Spoken Language	0.172	(0.23)	`		`		0.279*	(0.1)	0.274*	(0.09)
Language Proximity 1			-0.107***	(0.00)			-0.209***	(0.00)		
Language Proximity 2					-0.046	(0.21)			-0.235***	(0.00)
Internet Access*Common Spoken Language	-0.025***	(0.00)					-0.021***	(0.00)	-0.019***	(0.00)
Internet Access*Language Proximity 1			-0.001	(0.3)			-0.003**	(0.01)		
Internet Access*Language Proximity 2					-0.002	(0.37)			-0.005***	(0.00)
Trend*Common Spoken Language	0.02***	(0.00)					0.015***	(0.00)	0.014***	(0.00)
Trend*Language Proximity 1			0.002	(0.36)			0.005***	(0)		
Trend*Language Proximity 2					0.001	(0.78)			0.006**	(0.01)
Geographic Distance	-0.661***	(0.00)	-0.722***	(0.00)	-0.722***	(0.00)	-0.659***	(0.00)	-0.665***	(0.00)
Contiguity	0.466***	(0.00)	0.565***	(0.00)	0.58***	(0.00)	0.441***	(0.00)	0.445***	(0.00)
Colonial Links	1.06***	(0.00)	1.074***	(0.00)	1.074***	(0.00)	1.047***	(0.00)	1.068***	(0.00)
Landlockedness	-0.797***	(0.00)	-0.924***	(0.00)	-0.914***	(0.00)	-0.824***	(0.00)	-0.825***	(0.00)
Linder Effects	-0.118***	(0.00)	-0.097***	(0.00)	-0.097***	(0.00)	-0.12***	(0.00)	-0.119***	(0.00)
# Observations_1	333285		333285		333285		333285		333285	
R-square_1	0.999		0.999		0.999		0.999		0.999	
# Observations_2	347268		347268		347268		347268		347268	
R-square_2	0.774		0.810		0.811		0.765		0.777	

We apply the PPML method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=$[\ln(Y_i)-\ln(Y_j)]^2$.

and significant. However, the coefficient on internet is not significant, though positive. The coefficient of CSL is insignificant as in Table 11. However, the coefficient of the interaction term between language (COL or CSL) with TREND is positive and significant. All other coefficients are qualitative the same as they are in Table 4.

4 Conclusions

In this paper, we have tried to find a robust answer to our main research question: is the Internet bringing down language-based barriers to international trade? It is known that countries that do not have common languages, on average, tend to trade less between each other compared to countries that do have common languages. We examine if recent expansion of the Internet has reduced the dependence, as mentioned above, on a common

Table 12: Internet Access and Languages in international Trade (OLS estimations)

	Common Official Language		Common Spoken Language	
	Coef.	Prob.	Coef.	Prob.
Trade Agreements	0.387***	(0.000)	0.362***	(0.000)
Internet Access	0.004	(0.391)	0.004	(0.391)
Language	0.308***	(0.000)	0.008	(0.544)
Internet Access*Language	-0.007***	(0.000)	-0.008***	(0.000)
Trend*Language	0.015***	(0.000)	0.024***	(0.000)
Geographic Distance	-1.132***	(0.000)	-1.137***	(0.000)
Contiguity	0.791***	(0.000)	0.814***	(0.000)
Colonial Links	1.176***	(0.000)	1.363***	(0.000)
Landlockedness	-0.570***	(0.000)	-0.555***	(0.000)
Linder Effects	-0.046***	(0.000)	-0.044***	(0.000)
# Observations_1	224348		224348	
R-square_1	0.835		0.835	
# Observations_2	224348		224348	
R-square_2	0.87		0.866	

We apply the OLS method to 3-year intervals data over the period 1954-2014. *,**,and *** denote statistical significance at 10%, 5%, and 1%, respectively. P-values are reporter in the parentheses next to each coefficient. Fixed effect coefficients (multilateral resistances and pairwise in the first step, the country fixed effects in the second stage) and the constant terms are not reported for brevity. Linder Effects Variable=$[\ln(Y_i)-\ln(Y_j)]^2$.

language to trade. We have estimated gravity equations to do so. In particular, we made use of recent innovations in gravity analysis to study the impact of Internet access on the ability of common languages to drive international trade. From a dataset of aggregate export values from 205 countries over the period 1954-2014, our paper finds that the answer to our research question is very likely to be yes.

We are also able to provide a number of noteworthy observations. First, in line with the existing literature, distance between countries dampens trade while language similarities boost trade; trade tend to be higher for contiguous countries, for countries involved in trade agreements, and for countries that have been in a colonial relationship; and trade tend to be lower when at least one of the trading partners is landlocked. Second, Internet access boost international trade directly probably by helping the formation of trade networks. Third, and most importantly, Internet access reduces the impact of language-based barriers to trade and the internet boost relatively more international trade when the trading partners use different languages.

Our findings imply that with more accessibility to the Internet, the world has become 'flatter'– a term coined by Friedman (2007), and the effects of trade barriers such as language differences have been reduced. We also find that, controlling for internet and other country-specific technological innovations, captured by country-time fixed effects, language would increasingly be a barrier to international trade. Accordingly, policymakers should explicitly include IT in discussions/agreements on trade policies aiming at mitigating the negative effects of trade barriers such as language differences.

References

Agnosteva, D., Anderson, J. E., and Yotov, Y. V. (2014). Intra-national trade costs: Measures and aggregation. *National Bureau of Economic Research,* Working Paper, 19872.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 1:170.

Anderson, J. E. and Yotov, Y. V. (2016). Terms of trade and global efficiency effects of free trade agreements, 1990-2002. *Journal of International Economics*, 99[C]:279–298.

Baier, S. and Bergstrand, J. (2004). Economic determinants of free trade agreements. *Journal of International Economics*, 64(1):29–63.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, 71(1):72–95.

Baier, S. L., Yotov, Y. V., and Zylkin, T. (2016). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *CESifo Working Paper Series*, 6174.

Baldwin, R. and Taglioni, D. (2006). *Gravity for Dummies and Dummies for Gravity Equations*. National Bureau of Economic Research, NBER Working Paper, No. 12516, Cambridge, MA.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, 80(3):391.

Bergstrand, J. H., Larch, M., and Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78:307–327.

Beverelli, C., Keck, A., Larch, M., and Yotov, Y. (2018). Institutions, trade and development: A quantitative analysis. *Working paper*.

Borchert, I. and Yotov, Y. V. (2017). Distance, globalization, and international trade. *Economics Letters*, 153:32 – 38.

Cameron, A. C. and Trivedi, P. K. (2009). Microeconometrics using stata.

Campbell, D. (2010). History, culture, and trade: A dynamic gravity approach. *Journal Of Economic History*, 70(2):500.

Campbell, D. L. (2013). Estimating the impact of currency unions on trade: Solving the glick and rose puzzle. *World Economy*, 36(10):1278–1293.

Cheng, I. and Wall, H. J. (2005). Controlling for heterogeneity in gravity models of trade and integration. *Federal Reserve Bank Of St. Louis Review*, 87(1):49–63.

Crystal, D. (1985). Speaking of writing and writing of speaking. *Longman language Review*, 1:5–8.

de Caria, R. (2016). A digital revolution in international trade? the international legal framework for blockchain technologies, virtual currencies and smart contracts: Challenges and opportunities. *,* Working Paper, page Accessed on September/3/2018.

Egger, P. and Nigai, S. (2015). Structural gravity with dummies only: Constrained anova–type estimation of gravity models. *Journal of International Economics*, 97(1):81–99.

Eichengreen, B. and Irwin, D. A. (1998). The role of history in bilateral trade flows. in j. a. frankel (ed.). The regionalization of the world economy, NBER Project Report Series.

Fidrmuc, J. and Fidrmuc, J. (2016). Foreign languages and trade: evidence from a natural experiment. *Empirical Economics*, 50(1):31–49.

Fielding, D., Hajzler, C., and Macgee, J. (2015). Distance, language, religion, and the law of one price: Evidence from canada and nigeria. *Journal Of Money, Credit, And Banking*, 47(5):1007–1029.

Fouquin, M. and Hugot, J. (2016). Two centuries of bilateral trade and gravity data: 1827-2014. Working Papers 2016-14, CEPII.

Freund, C. and Weinhold, D. (2004). The effect of the internet on international trade. *Journal of International Economics*, 62:171–189.

Friedman, T. L. (2007). *The world is flat : a brief history of the twenty-first century*. New York : Picador/Farrar, Straus and Giroux, Holtzbrinck.

Harris, R. (1995). Trade and communication costs. *Canadian Journal of Economics*, 28:46–75, Special Issue.

Head, K. and Mayer, T. (2014a). Chapter 3 - gravity equations: Workhorse,toolkit, and cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 131 – 195. Elsevier.

Head, K. and Mayer, T. (2014b). Chapter 3 - gravity equations: Workhorse,toolkit, and cookbook. *Handbook of International Economics*, 4(Handbook of International Economics):131 – 195.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(4):1–14.

Heid, B., Larch, M., and Yotov, V. (2017). Estimating the effects of non-discriminatory trade policies within structural gravity models. *Cesifo Working paper-6735*.

Hummels, D. (2007). Transportation costs and international trade in the second era of globalization. *The Journal of Economic Perspectives*, 21(3):131–154.

Isphording, I. and Otten, S. (2013). The costs of babylon-linguistic distance in applied economics. *Review Of International Economics*, 21(2):354–369.

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265 – 295.

Jerry G., T. and Marie C., T. (1987). Bilateral trade flows, the linder hypothesis, and exchange risk. *The Review of Economics and Statistics*, 69(3):488.

Juan Carlos, H. (2010). A product-quality view of the linder hypothesis. *The Review of Economics and Statistics*, 92(3):453.

Kitenge, E. (2016). Effects of food and agricultural imports on domestic factors in the u.s. agricultural sector: a translog cost function framework. *Applied Economics Letters*, 23(2):132–137.

Larch, M., Wanner, J., Yotov, Y. V., and Zylkin, T. (2017). The currency union effect: A ppml re-assessment with high-dimensional fixed effects. *Drexel University School of Economics, Working Paper*.

Lin, F. (2015). Estimating the effect of the internet on international trade. *Journal Of International Trade & Economic Development*, 24(3):409–428.

Linder, S. B. (1961). An essay on trade and transformation.

Mayer, T. and Zignago, S. (2011). Notes on cepiis distances measures: the geodist database. CEPII, Working Paper, 25.

Melitz, J. and Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 92(2):351–363.

Miller, A. (1951). *Language and Communication*. McGraw-Hill Book Company, Inc.

Olivero, M. P. and Yotov, Y. V. (2012). Dynamic gravity: endogenous country size and asset accumulation. *Canadian Journal Of Economics*, 45(1):64–92.

Rose, A. K. and van Wincoop, E. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review*, 2:386.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88:641–658.

Sharma, S. (1991). Technological change and elasticities of substitution in korean agriculture. *Journal of Development Economics*, 35:147–172.

Shiue, C. H. (2002). Transport costs and the geography of arbitrage in eighteenth-century china. *American Economic Review*, 92(5):1406–1419.

Tang, D. (2003). Economic integration among the asia-pacific economic cooperation countries: Linder effect on developed and developing countries (1985-1999). *International Trade Journal*, 17(1):19.

Tinbergen, J. (1962). Shaping the world economy: Suggestions for an international economic policy. New York: Twentieth Century Fund. The first use of a gravity model to analyze international trade flows.

Trefler, D. (2004). The long and short of the canada-u.s. free trade agreement. *American Economic Review*, 94(4):870–895.

UNICTAD (2004). Use of the internet for efficient international trade. UNCTAD/SDTE/TIB/2003/3.

WorldBank (2017). World development indicators.

Yotov, Y. V., Larch, M., Monteiro, J., and Piermartini, R. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. World Trade Organization, Geneva.

Appendix: Tables

Table 13. List of countries

Aruba	Cabo Verde	Hungary	Mali	Sierra Leone
Afghanistan	Costa Rica	Indonesia	Malta	El Salvador
Angola	Cuba	Isle of Man	Myanmar	San Marino
Albania	Cayman Islands	India	Montenegro	Somalia
Andorra	Cyprus	Ireland	Mongolia	Serbia
United Arab Emirates	Czech Republic	Iran	Mozambique	South Sudan
Argentina	Germany	Iraq	Mauritania	São Tomé and Principe
Armenia	Djibouti	Iceland	Mauritius	Suriname
Antigua and Barbuda	Dominica	Israel	Malawi	Slovak Republic
Australia	Denmark	Italy	Malaysia	Slovenia
Austria	Dominican Republic	Jamaica	Namibia	Sweden
Azerbaijan	Algeria	Jordan	New Caledonia	Eswatini
Burundi	Ecuador	Japan	Niger	Seychelles
Belgium	Egypt	Kazakhstan	Nigeria	Syrian Arab Republic
Benin	Eritrea	Kenya	Nicaragua	Chad
Burkina Faso	Spain	Kyrgyz Republic	Netherlands	Togo
Bangladesh	Estonia	Cambodia	Norway	Thailand
Bulgaria	Ethiopia	Kiribati	Nepal	Tajikistan
Bahrain	Finland	St. Kitts and Nevis	Nauru	Turkmenistan
The Bahamas	Fiji	Korea	New Zealand	Timor-Leste
Bosnia and Herzegovina	France	Kuwait	Oman	Tonga
Belarus	Faroe Islands	Lao PDR	Pakistan	Trinidad and Tobago
Belize	Micronesia	Lebanon	Panama	Tunisia
Bermuda	Gabon	Liberia	Peru	Turkey
Bolivia	United Kingdom	Libya	Philippines	Tuvalu
Brazil	Georgia	St. Lucia	Palau	Tanzania
Barbados	Ghana	Liechtenstein	Papua New Guinea	Uganda
Brunei	Gibraltar	Sri Lanka	Poland	Ukraine
Bhutan	Guinea	Lesotho	Puerto Rico	Uruguay
Botswana	The Gambia	Lithuania	Dem. People's Rep. Korea	United States
Central African Republic	Guinea-Bissau	Luxembourg	Portugal	Uzbekistan
Canada	Equatorial Guinea	Latvia	Paraguay	St. Vincent and the Grenadines
Switzerland	Greece	Macao SAR, China	French Polynesia	Venezuela
Chile	Grenada	Morocco	Qatar	Virgin Islands
China	Greenland	Monaco	Russia	Vietnam
Côte d'Ivoire	Guatemala	Moldova	Rwanda	Vanuatu
Cameroon	Guyana	Madagascar	Saudi Arabia	Samoa
Dem. Rep. Congo	Hong Kong SAR, China	Maldives	Sudan	Yemen
Congo	Honduras	Mexico	Senegal	South Africa
Colombia	Croatia	Marshall Islands	Singapore	Zambia
Comoros	Haiti	North Macedonia	Solomon Islands	Zimbabwe