LUND UNIVERSITY
School of Economics and Management

The Internet and Trade

An assessment of the effects of internet usage on trade volumes

Master Thesis in Economics
Lund University - Department of Economics
NEKN01 - Autumn 2015
Author: Fredrik Svensson Rosell
Supervisor: Joakim Gullstrand

Abstract

This paper aims to assess the effects of internet usage on bilateral trade volumes. Increasingly companies are conducting business online. If the effect of an increased internet usage on trade is positive then it would be yet another reason for policy makers to invest in the development and expansion of IT infrastructure.

Three main models are specified using the fixed effects estimator, including the least squares estimator and a Poisson maximum likelihood estimator. This paper uses an extensive panel data set of 180 countries and the most recent internet usage data available (2000-2014) to examine the effects of internet usage on bilateral trade. By using a gravity model specification the results are indicating a significant positive relationship between internet usage and bilateral trade performance. The effect is larger when more weight is given to countries with smaller internet usage rates.

The spatial relationship of this effect is also examined via a Hausman-Taylor estimation and a random effects model. The results from these models indicates that there is a proximity effect, the larger the distance between trading countries, the smaller the effect of internet usage on trade.

Key words: Gravity model, fixed effects, internet usage, trade, panel data.

Table of contents

1 Introduction

The development and expansion of internet access around the world has contributed in a lot of ways to economic development. An increased internet access has shown to have significantly improved economic growth, productivity, employment, increasing the consumer surplus as well as it improves firm efficiency (Katz, 2012). These are but a few of the direct or indirect spillover effects from an increased use of the internet by businesses and citizens.

In recent years the online business sector has expanded significantly as consumers increasingly buy and sell products online. As an example, in 2014 the online business company Alibaba was listed in the New York stock exchange in the biggest IPO in history. The internet has made it easier for businesses and consumers to buy and sell products across national borders. How the internet usage affects trade might have important implications when it comes to national policies regarding investing in IT infrastructure. The positive effects of trade on economic growth and welfare are well known. So if an increased internet usage has a positive effect on bilateral trade performance it would be another reason for policy makers to develop the internet infrastructure to reach as many citizens as possible. This paper will, through a gravity model framework, try to assess the impact of this increase of internet usage on countries' trade performances. Thus, the main research question can be formulated as following: *What is the effect of internet usage on bilateral trade performance?*

Even though many studies have been made on the relationship between internet usage on international trade, most of the studies have used either smaller samples or older data. As the IT sector has developed a lot in recent years, it would be of importance to study the effects on more recent data than what previously have been used, especially considering the fast development of online businesses and e-commerce. It would also be beneficial to use a larger data set than previous studies. This paper is using the most recent data available together with a more extensive panel data set in order to determine the global impact of internet usage on bilateral trade performance.

Another question that will be examined in this paper will build on the result of the first question and examine the geographical dimension of this potential effect. In other words; *what is the effect of internet usage on bilateral trade when considering the geographical distance between the two trading countries?*

There are reasons to believe that this effect could be either increasing or decreasing as the geographical distance increases. Even though previous research has indicated a proximity effect, it could be that the internet in recent years has increased the trade with economies further away.

This paper will be structured in five main chapters. Firstly, in chapter 2 there will be a review of previous studies in the field. Chapter 3 will include the theoretical framework of the gravity model of trade and how the internet usage is applied in the model. Chapter 4 will cover the data used in the paper as well as descriptive statistics of the data set. In chapter 5 the empirical framework will be established as well as the robustness specifications. Chapter 6 will include the results from the models as well as the results from the robustness specifications and this chapter will be followed by a conclusion.

2 Communication and trade: previous studies

A lot of studies have been made on how the usage of different means of telecommunications is affecting the international trade. In one of the first thorough panel data analyses Hardy (1980) found that the usage of telephones per capita had a significant effect on the GDP per capita one year later. As the telecommunications sector has developed significantly since then, later studies have been conducted on other means of communication such as mobile subscriptions or internet usage, just to mention a few. The results are in general showing that the use of different means of telecommunications such as telephones, fixed and mobile internet and mobile subscriptions is increasing trade volumes. The older studies mostly used cross-section or time-series data, while the more recent studies are increasingly using panel data. Most of the previous studies have been using the gravity model of trade to investigate the effect of communication on trade volumes. This paper will also use the gravity framework in order to determine the effects of internet usage on trade. Some of the previous studies in the field are presented in Table 1 below:

Table 1: Summary of previous studies on the topic, sorted after year of publication.

AUTHOR(S)	*YEAR*	*METHOD*	*ANGLE*	*SAMPLE*	*RESULTS*	*PERIOD*
Freund and Weinhold	2002	Cross section and panel data.	Looking on trade in services by using data from the US.	Trade in services from 31 countries to the US.	Positive. Internet development abroad increased the trade of service imports to the US.	1995-1999
Freund and Weinhold	2004	Cross section and ts.	Internet development and trade.	60 mixed countries.	Positive. Proximity biased growth.	1997-1999
Clarke and Wallsten	2006	Cross section, and ts.	Using countries' IT regulations as instrument variable.	26 developed 80 developing	Positive, different results for developed and developing when exporting.	1991-2001
Vemuri and Siddiqi	2009	Panel data.	Telephone lines, PCS and Internet users as Internet variables.	65 countries	Positive.	1985-2005
Thiemann and Fleming	2012	Panel data.	Vegetable and fruit trade only. Internet, fixed phones and mobile penetration as variables for ICT.	30 of the greatest trading countries of each fruit/vegetabl es.	Positive on imports. Negative on exports.	1995-2009

In chronological order, Freund and Weinhold (2002, 2004) conducted two studies on the impact of internet development and its effects on trade by using the number of web hosts as a measure of internet development. The first of the two studies found that a 10% increase of the number of web hosts in the US led to a 1,7% increase in the US import of services and a 1,1% increase in the US export of services (Freund and Weinhold, 2002). In the second study Freund and Weinhold found that a 10% increase in the number of webhosts were equivalent with a 0,2% increase in exports of goods (2004). Another result from this study was that it was a bigger effect from internet hosts on trade when countries were closer located geographically, i.e. there was a proximity bias (Freund and Weinhold, 2004).

Interestingly, Blum and Goldfarb (2006) find that this proximity effect was present even for products which have almost no trade costs at all, such as purely online products and services. The effect was also larger for taste based products such as music and their conclusion is that this is due to cultural differences. Even though online products and services have almost no trade costs at all, consumers tend to buy these products and services from countries that are geographically located closer than further away.

Clarke and Wallsten (2006) found that a higher internet penetration rate in developing economies tends to improve export performance from developing economies to developed economies. In their study they find no such relationship from neither developing to other developing nor from developed to either developed or developing countries. They conclude that this might be because almost all enterprises in developed countries already had access to internet in the time period of the study. Another conclusion they make is that being connected to the internet in developing countries is probably a great advantage if the firm is aiming to export to developed economies (where most firms probably already are connected).

Vemuri and Siddiqi (2009) made a thorough panel data analysis on the ICT infrastructure and the availability of internet for commercial transactions and found that both were significantly positive on international trade (Vemuri and Siddiqi, 2009).

Thiemann and Fleming (2012) studied the impact of ICT on the trade of fruits and vegetables. By using panel data on only the biggest fruit and vegetable trading economies they found that an increased internet penetration rate in the importing country is positively affecting the imports of fruit and vegetables. However, the same relationship is negative when it comes to exports of fruits and vegetables (Thiemann and Flemming, 2012). Thiemann and Flemming (2012) conclude that neither of their ICT variables had a very strong positive effect on the trade of fruits and vegetables but adds that their data set was not of best quality.

3 Gravity and the internet

This chapter will include an introduction to the theory of the gravity model and how the internet usage variable is normally treated in gravity models.

3.1 The gravity model of trade

The gravity model was first used to explain patterns of trade by Jan Tinbergen (1962). In its simplest form the gravity model of trade explains bilateral trade flows between two economies by the "economic mass" and the geographical distance between the two economies. The model is similar to Isaac Newton's gravity theory in which Newton explained how two bodies in the universe attracts each other with a force that is proportional to the product of the bodies' masses and disproportional to the distance between them. In the gravity model the same logic is applied to the trade flows between two economies. Empirically it was shown that a large amount of the trade flows between two economies can be explained by the "economic mass" (or GDP) and the geographical distance between the two economies. The intuition behind the model is easy to understand. Firstly, a country is trading more with larger markets than with smaller due to the fact that there are more potential producers (for imports) and more consumers (for export) than a smaller country. Secondly, the trade flows might be smaller the larger the geographic distance between any two countries due to higher trade costs such as higher transportation costs, higher communication costs or due to other trade costs that are increasing with distance.

Although the model showed to empirically estimate bilateral trade flows very well, the theory behind the model was incomplete for a long time. The first attempt of filling the theoretical gap in the gravity model was made when Leamer and Stern (1970) made the "potluck assumption". All produced goods are thrown into a pot and then all nations consume a share of the pot proportional to their income. The trade between two countries then equals the share of the first country times the share of the second country. Anderson (1979) later developed the theory with the assumption that each nation produced goods that were imperfectly substitutable with other nation's goods. Bergstrand (1989, 1990) and Deardorff (1995) later linked the theories of the Heckscher-Ohlin model and the theory of monopolistic competition to the gravity equation. By putting the theoretical pieces together, Anderson and Wincoop (2003) finally developed the gravity model to the following structure:

$$x_{ij} = \frac{y_i y_j}{y^W} \left[\frac{\tau_{ij}}{P_i P_j} \right]^{1-\sigma} \quad (1)$$

where x_{ij} is the bilateral trade flow between country i and j, y_i is the economic level (*nominal GDP*) of country i, y_j is the economic level of country j, y^W is the economic level of the world, P_i is the price level of country i, P_j is the price level of country j and τ_{ij} is the trade cost function between country i and j and σ is a constant measuring the elasticity. The product of the economic levels, $y_i y_j$, is the economic mass.

The trade cost function τ_{ij} is a function of factors that are creating some kind of trade cost. In the literature the most common trade cost is geographical distance. Other than the geographical distance, normally the trade cost function also includes several dummy variables such as common language, colonial links, contiguity and whether or not there are any trade agreements between the countries (Anderson and Wincoop, 2004). In this paper, the trade cost of interest is internet usage, which may be seen as mainly decreasing the information cost in trade.

3.2 Internet intensity in the gravity model

The trade cost function, τ_{ij} in the basic gravity equation is of importance when trying to determine the effect of some dependent variable in relation to trade. Anderson and Wincoop (2004:691f.) broadly defines trade costs as "*all costs incurred in getting a good to a final user other than the marginal cost of producing the good itself: transportation costs (both freight costs and time costs), policy barriers (tariffs and nontariff barriers), information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs (wholesale and retail)*". Internet usage, or lack thereof, can in this sense be treated as an information cost, but indirectly internet usage may affect the other types of trade costs as well. In this paper the trade cost function will include internet usage and indirectly, via the fixed effects estimator, also the country-pair fixed effects. Since the fixed effects model absorbs the bilateral country-pair time invariant fixed effects, the only trade cost in the fixed effects estimation output will be internet usage.

An important note on internet usage is that the internet usage needs to interact between the countries in every country pair, to reflect the internet interaction or internet communication between the countries. Thus, internet usage will be transformed into what will be called *internet intensity*. Basically the internet intensity variable is defined as *the product of the internet usage rate between two countries at time t*. In this way the internet intensity will be similar to the GDP mass in equation (1).

Since the FE model does not directly account for time invariant effects such as distance, language or colonial connection, a Hausman-Taylor (HT) estimation and a random effects (RE) model will be used in estimating the geographical effect on internet intensity. These models allows for time invariant variables to be included directly in the estimation. This is important as the geographical model needs to also control for geographical distance.

4 Data

This chapter will discuss most of the decisions and issues regarding the data selection. Firstly, there is a presentation of the sources of the data used in the paper. Then, as the sample of countries is quite important, a discussion on the sample selection is presented. Finally there is a section about the issue regarding zero trade data.

4.1 Data sources

The panel data set used in this paper contains 483300 pairwise observations. The bilateral trade data is collected from UN's Comtrade database.[1] The trade data used is "Total of all HS commodities" and the data used in this paper is import data. The import values are chosen because there is simply more data reported on bilateral imports than on bilateral exports. A limitation from the UN trade data is that the different rules of origin differ from each country to the next, meaning that the term partner country does not necessarily indicate any direct trade relationship. However, this issue is probably of minor importance due to the very large data set in this study.

The "internet usage" variable is gathered from the International Telecommunications Union (ITU).[2] The internet usage variable is "Percentage of individuals using the internet" and the data is gathered from national ministries, national statistics agencies or estimated by the ITU. The data on GDP[3] in current USD is collected from World Development Indicator (WDI). Finally, the data on

[1] http://comtrade.un.org/data/
[2] http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx
[3] http://data.worldbank.org/indicator/NY.GDP.MKTP.CD

geographical distance[4] and other time invariant gravity variables are collected from the *Centre d'Etudes Prospectives et d'Informations Internationales* (CEPII).

Table 2: Descriptive statistics for the main variables, bilateral trade, GDP and internet usage.

	Existing	*Missing/Zeroes*	*Perc. missing*	*Mean*	*Std. Dev.*
Bilateral trade	284494	198806	41,14	5,88*10^8	5,56*10^9
GDP	479183	4117	0,85	3,03*10^11	1,24*10^12
Internet usage rate	477214	6086	1,26	26,2	27,0

The internet usage variable and the GDP variable are interpolated and extrapolated since there are very few missing values for the two variables. The raw data values are later used for robustness. When variables are expressed as logarithmic functions they are first transformed by adding one unit. If the variables are not transformed in this way, the values of the logarithmic functions will approach negative infinity as the values approach zero, and consequently the estimates will be biased.

In order to check for multicollinearity in the data set a correlation matrix was performed. It does not seem to be a problem with multicollinearity in the data set. The correlation matrix (*See Table 3*) does not indicate any large correlations between my three main variables trade, GDP and internet intensity.

Table 3: Correlation matrix between the dependent variable, ln(trade) and the two independent variables, ln(GDP mass) and ln(Internet intensity).

	Ln(trade)	*Ln(GDP mass)*	*Ln(Internet intensity)*
Ln(trade)	1		
Ln(GDP mass)	0,7448	1	
Ln(Internet intensity)	0,3374	0,4749	1

An important note is that since the internet usage variable from ITU is in some cases based on estimates from ITU instead of estimates from national statistical

[4] http://www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp?id=6

agencies and ministries it might not be completely correct. However, the ITU data is used since it is a very comprehensive data available on national internet usage and internet penetration around the world. The ITU data has been used in studies before, but mostly when investigating trade of certain goods such as fruits and vegetables (Thiemann and Fleming, 2012) or wine (Thiemann and Fleming, 2011). It has also been used in smaller studies such as Clarke and Wallsten (2006) or in region specific studies such as Kurihara (2013). It has not been used in such an extensive way as in this study.

4.2 The sample

The sample used in this paper is chosen to include the largest amount of countries with existing internet usage data from ITU for the time period of 2000-2014. The motivation of using the largest possible data set available is to add variance in the data set, especially in the distance variable. After adding the GDP data and the distance variable from CEPII the data set finally consists of 180 countries over 15 years (*for the full country sample, see Appendix: Table 15*). The data set consists of 483300 paired observations (15*180*180) of which there is existing trade data for 284494 observations. Thus, the existing trade data makes up 58,9% of the sample. The rest is missing data, either because there is reported a zero trade for these years or because there was no trade between the countries in these years.

When using bilateral trade data there is often missing values and sometimes the data can be manipulated or biased in some way. For validity testing, the two most important variables in the data set, the bilateral trade value and the GDP, was tested using Benford's law (s*ee Appendix: Table 1 and 2*). Benford's law states that if values are distributed over multiple magnitudes, the first digit should follow a certain distribution where the digit 1 is most frequent and digit 9 is the least frequent. Benford (1938) showed that the frequency of the first digit, in these types of distributions, closely follows the logarithmic function:

$$F_a = log\left(\frac{a+1}{a}\right) \quad (2)$$

where F_a is the frequency of the digit a in the first place of the used numbers. By looking at the distributions the values seem to follow the Benford distribution fairly well. Two χ^2-tests were performed on the bilateral trade values and the GDP values to test whether they follow the Benford distribution or not. However, both p-values were less than 0,05, indicating that they do not follow the Benford distribution.

Even though the tests showed that the trade and GDP distributions do not follow the Benford distribution on a 95% significance level, some information can be obtained. The p-value was lower for the GDP variable which could be because of the small variation in GDP within countries. The GDP variable is rather slowly moving, meaning that the first digit might be in the same decade for several years in a row. The bilateral trade variable has a larger variation within country pairs which might explain why it is significant on a 99% significance level but not on a 95% significance level. Considering the small amount of countries, the large amount of observations and the reasonable well fit, there is still reason to believe the data is not overly biased.

4.3 Zeroes

A common issue with bilateral trade data is regarding how to deal with zero trade values. As the data set in this paper is of panel data structure, and the trade values are bilateral trade values, many values are either missing or zero. A zero trade value might in this sense be either true or false. Thus a missing value could be a zero trade value because it is a zero value or it could be a zero because the trade in this period was not reported. According to UN, some countries do not necessarily report their trade statistics for every year[5]. Also, the UN comtrade database does not include estimates for missing data.

Because of the problem with zero trade values, two other models will be used except the main FE model. The first will be the same FE model, but treating all the zero trade values as zeroes, thus not omitting those observations in the estimation. The second will be a Poisson conditional maximum likelihood (ML) fixed effects estimator. The Poisson ML estimator has been recommended as appropriate to use in gravity models in papers by Santos Silva and Tenreyo (2006) and Westerlund (2011). One of the advantage of the Poisson ML estimator is that it has been proven to better deal with gravity estimations with a large proportion of zero trade values than the standard FE model.

[5] See UN comtrade disclaimer http://comtrade.un.org/db/help/uReadMeFirst.aspx .

5 Empirical specifications

This chapter includes a discussion on the choice of models for estimating the relationship between internet intensity and trade performance. Firstly, a discussion is presented between the choice of panel data model, the fixed effects or the random effect model. Secondly, the main model specifications are defined as well as robustness specifications and lastly there will be a discussion on causality.

5.1 The fixed effects model

The data set in this paper is of panel data structure and the normal procedure is to determine whether to use a fixed effects estimator or a random effects estimator. A Hausman test was conducted in order to test which of these models that should be applied (*see Appendix: Table 3-5*). The Hausman test tests whether or not the null hypothesis that the dependent variables and the individual intercept terms are uncorrelated (Verbeek, 2012:385 and Hausman, 1978). The results from the Hausman test indicate that the preferred method should be a fixed effects model (*See Appendix Table 5*). Due to this result, the main model specification will be a fixed effects model. Time dummies are added in the models to control for time fixed effects that vary over all country pairs such as global business cycles or global financial crises.

The fixed effect model is chosen in favor of the random effects model since the interesting aspect is to see how the bilateral trade flows are changing within the groups and across time. The country-pair specific effects that are time invariant like distance thus disappear from the estimation results. This characteristic of the fixed effects model safeguards the problem with omitted country-pair specific variables that otherwise should be included in the model. If we are interested in the individual intercepts, or believe that the individuals (in this case country pairs) cannot be seen as any random draw from some underlying population, then the fixed effects model is the preferred method. When the individuals are countries the fixed effects model is normally the most preferred (Verbeek, 2012:384).

The choice of a fixed effects model rather than a random effects model, when using the gravity model, is also strengthened in previous research. Egger (2000) argues that the preferred method should be to use a fixed effects model since most of the country specific effects in a gravity model are predetermined because of geographical, historical or political contexts.

Another fixed effects model, the Poisson ML estimator will also be used. The reason for using the Poisson ML estimator is because of how it treats zero trade values and because of the log-linear or linear-linear specification (Santos Silva and Tenreyo, 2006). Previous research suggests the log-log specifications in FE models in presence of heteroscedasticity will produce inconsistent estimates, and this heteroscedasticity is normally present in gravity models with very high values. There is normally also a sample selection bias that may occur when putting zero trade values as zeroes in the second FE model, or when excluding them in the first FE model. This bias is lowered in the Poisson ML estimation since the Poisson linking equation is in multiplicative form. This allows for keeping the trade variable as it is and not as a logarithmic function.

In order to account for the time invariant variables that are omitted in the fixed effects model, two models will be used which includes the time invariant variables. The first is the RE model and the second is a HT estimation, of the type proposed by Hausman and Taylor (1981). The HT estimator is an instrumental variable estimator that uses the individual means of strictly exogenous independent variables as instruments for the time invariant independent variables that are correlated with the individual effect (Baltagi, 2003). This requires a model with at least one strictly exogenous time variant independent variable and one time variant endogenous independent variable.

5.2 Model specifications

The main fixed effects panel data model for investigating the relationship between internet usage and international trade in this paper is a fixed effects model as following:

$$y_{ijt} = \alpha_{ij} + \delta_t + \beta x_{ijt} + u_{ijt} \tag{3}$$

where i and j are indices for the trading countries, *t* is the time index, y_{ijt} is the logarithm of the bilateral import flow between country *i* and *j* in time period *t,* α_{ij} is the country-pair intercept, δ_t is a year dummy that includes fixed effects in time that are affecting all observations equally, x_{ijt} are the logarithms of the explanatory variables that vary across trading partners and time and u_{ijt} is the error term. The explanatory variables x_{ijt} in this model consists of the logarithm of GDP mass $(GDP_{it} * GDP_{jt})$ and the logarithm of internet intensity which is specified as $(Int\_penetration_{it} * Int\_penetration_{jt})$. After making equation (1) log-linear and applying it to the FE estimator in equation (3), the main model will be specified as:

$$\ln(trade_{ijt}) = \alpha_{ij} + \delta_t + \beta_1 \ln(GDP\ mass_{ijt}) + \beta_2 \ln(Int.intensity_{ijt}) + u_{ijt} \tag{4}$$

where α_{ij} is unobserved country-pair specific effects, δ_t is the time fixed effects and u_{ijt} is the error term. Three models will be used using this type of specification (*see Table 4 for details about the specifications of the variables in the three main models*).

Table 4: Model specifications.

Model	*Estimator*	*Dep. Var.*	*Indep. Var. 1*	*Indep. Var. 2*
1	FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	The logarithm of the product of inter- and extrapolated internet usage at time t.
2	FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	The logarithm of the product of inter- and extrapolated internet usage with missing values as zeroes.
3	FE Poisson ML	Import trade value.	The product of inter- and extrapolated nominal GDP.	The product of inter- and extrapolated internet usage.

The first FE model is using the raw data on trade value in its logarithmic form as dependent variable. The second FE model is using a transformation of the trade value variable where the missing values are transformed into zeroes. Thus the second model will include values for every observation in the data set. The third FE model is a linear Poisson ML estimator. All three FE models are using robust standard errors.

In order to be able to account for time-invariant independent variables and to estimate the spatial effect, a fourth Hausman-Taylor (HT) model is used. Geographical distance is one of the most important variables to include in a gravity model as it explains a large part of the trade flows and this variable is omitted from the FE models. The HT estimator is constructed as following:

$$y_{ijt} = \beta x_{ijt} + \gamma z_{ij} + \epsilon_{ijt} + u_{ij} \quad (5)$$

where the time variant variables x_{ijt} are the same as in equation (2) plus the product of log(distance) and internet intensity, the time invariant variables z_{ij} in this model consists of; $\log d_{ij}$, which is the logarithm of the distance in km between country i and j, a common language dummy, which is 1 if at least 9% of the population of both countries speak the same language and 0 otherwise, a contiguity dummy, which is 1 if the two countries are adjacent to one another and

0 otherwise, a colonial connections dummy, which is 1 if the countries have a past colonial connection and 0 otherwise. The HT estimator requires that at least one of the time variant variables is defined as endogenous. In this case the GDP mass is defined as endogenous. However, the HT estimator then treats the exogenous variables as strictly exogenous, which is a rather strong assumption. Thus, the interpretation of the output should be treated with caution.

The HT estimation will be used to estimate how the internet intensity variable is affecting the trade performance when the geographical distance between two trading countries is increasing. A random effects (RE) model will also be included in order to be compared with the HT model. The HT and the RE model will be used in order to analyze the spatial relationship, whether the effect on internet intensity is increasing or decreasing with geographical distance.

5.3 Robustness specifications

A number of robustness specifications will be made in order to account for different ways of defining the variables (*See Table 5 for a summary of the robustness specifications)*. Firstly, a model is used which is using the raw values for GDP mass and internet intensity, without any interpolation or extrapolation. Secondly, a model is used which will not treat the intensity between the countries but the internet usage rates in both countries at time t. This model will estimate how the import trade value is affected by the internet usage rates in importing countries and exporting countries respectively.

In the third model the internet intensity variable is transformed into a minimum internet intensity variable. The intuition behind the minimum intensity variable is that the lowest amount of users determines the potential communication lines between the countries at time t, and this transformation will come closer to that value. In the literature this is referred to as a “communication bottleneck”. Intuitively, if one of the countries has a very high internet usage rate and the rate of the second country is very low, the actual number of possible internet communication lines between the countries at time t will depend on the country with the least number of users. This minimum internet intensity variable is the squared minimum internet usage rate of the two countries at time t. By transforming the internet intensity variable in this way, a higher weight will be given to the country with the lowest internet usage rate. Since the rates vary a lot across countries, with some countries with very low rates close to 0% and other that are close to 100% this transformation might be interesting to examine closer.

Lastly, a lagged model will be used in order to see the relationship between internet intensity and trade performance through time. This model will investigate the lag in the internet intensity in trade, assuming the causality from internet intensity to trade.

Table 5: Robustness specifications.

Estimator	*Dependent*	*Ind. Var. 1*	*Ind. Var. 2*	*Ind. Var. 3*	*Ind. Var. 4*	*Ind. Var. 5*
FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	The logarithm of the product of inter- and extrapolated internet usage.			
FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	The logarithm of the product of the minimum inter- and extrapolated internet usage.			
FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	Internet usage rate in importing country.	Internet usage rate in exporting country.		
FE	Logarithm of the import trade value.	The logarithm of the product of inter- and extrapolated nominal GDP.	The logarithm of the product of inter- and extrapolated internet usage at time t.	The logarithm of the product of inter- and extrapolated internet usage at time t-1.	The logarithm of the product of inter- and extrapolated internet usage at time t-2.	The logarithm of the product of inter- and extrapolated internet usage at time t-3.

5.4 Causality

A very important question to address before continuing is the causality problem which may cause endogeneity problems. The issue has been addressed in a lot of studies and the causality is quite clear from internet development (or development of other telecommunications sectors) to improved trade performance. Dutta (2001) found by using Granger causality that the causality between telecommunications infrastructure and economic activity in both industrialized and developing economies goes from developing the telecommunications infrastructure to an increased economic activity. Freund and Weinhold (2004) found that the growth of web hosts in a country had a significant positive effect on trade. Vemuri and Siddiqi (2009) found by using both FE and RE models that an increased internet usage increases trade volumes.

Clarke and Wallsten (2006) mention that there seems to be a correlation between internet penetration rates and exports, although that there also seems to be a correlation between internet penetration rates and globalization. Thus, Clarke and Wallsten (2006) concludes that the direction of the causality between internet penetration rates and exports is not completely certain.

Challenging the above mentioned studies, Meijers (2014) used a different method of establishing the causality between internet usage, trade and GDP per capita. The findings were in line with most previous findings that internet usage causes international trade and that international trade causes economic growth. Based on the previous studies the causality will be assumed to go from internet usage to bilateral trade performance even though this relationship is not fully determined.

6 Results

Since the data set is of panel data structure, a Hausman test was first performed in order to establish whether the preferred panel data estimator should be of a random effects or a fixed effects model. The result from the Hausman test indicates that a FE model is preferred. Thus, in order to analyze the effects on internet usage on trade performance three FE models were specified. The first two are standard FE models and the third is a Poisson estimator which is assuming the trade values are following a Poisson distribution. All three main models are controlling for time fixed effects, thus dummy variables for every year is included in the models. The estimates from the three main FE models are summarized in in Table 6:

Table 6: Estimates from the three main FE models of which the first two are specified as in equation (4). For full regression output, see Appendix: Table 6-8. [6]

Estimator	*FE*	*FE w. zeroes*	*Poisson (linear)*
GDP mass	0,5583 ***	0,3850 ***	2,91e-27 ***
	(0,0187)	(0,0298)	(9,18e-28)
Internet intensity	0,0446 ***	0,2821 ***	0,000173 ***
	(0,0088)	(0,0162)	(6,01e-06)

The results from the three FE models indicate that the effect on internet intensity is positive. When controlling for GDP mass and by controlling for year fixed effects the results are strictly positive. As the internet intensity increases between two countries, the trade performance between the same countries is improved. An increased internet intensity in countries, or as it is specified – an increased internet usage mass between countries – leads to an improved trade performance between

[6] *All estimations were made in Stata/IC 12.0 and the output is presented in the Stata format.*

the same countries. All coefficients are significant on a 0.01 significance level. While the relationship is strictly positive in the FE models the interpretation of the coefficients is a bit complicated. Since both the dependent and the independent variables in the LS estimates are logarithms, the interpretation of the coefficients will be that a one percent increase in the internet intensity variable will lead to an increase in percent of the trade value according to the coefficient value. In the first model, a one percent increase in internet intensity will lead to a 0,045 percent increase in the bilateral trade value. The effect of internet intensity on trade is thus rather small compared to the effect of the GDP mass. Thus, the economic significance of this effect is rather low.

The HT and the RE estimations including the interaction variable between distance and internet intensity show that the effect of internet intensity on trade is significantly negative when the distance is increasing (*See Table 7*).

Table 7: HT and RE estimations with an interaction variable between distance and internet intensity. The internet intensity and the interaction variable are defined as exogenous in the HT model. For full regression output, see Appendix: Table 9-10.

Estimator	*HT*	*RE*
ln(GDP mass)	0,8365751 ***	1,042716 ***
	(0,0070766)	(0,0049667)
ln(Internet intensity)	0,0292969 ***	0,1625059 ***
	(0,0097509)	(0,0164394)
*ln(Distance)*ln(Internet intensity)*	**-0,014046 *** **	**-0,0159914 *** **
	(0,0009846)	**(0,0016675)**
ln(Distance)	-1,334194 ***	-1,258492 ***
Contiguity	1,186945 ***	1,130844 ***
Common Language	0,7255503 ***	1,004339 ***
Colonial link	2,10947 ***	1,265253 ***

All the time invariant variables, distance, contiguity, common language and colonial connection are also significant with the predicted sign in the HT and RE models. The time variant effects also show the predicted signs, which is further strengthening the results of the interaction variable. When controlling for distance it seems to be a proximity bias, i.e. the effect of internet intensity on trade is larger when countries are located more close geographically. This result reinforces the result of previous findings by Freund and Weinhold (2004) and Blum and Goldfarb (2006).

6.1 Robustness discussion

For an increased robustness in the results several other models where specified to control for various complications with the data. These specifications include using the raw un-polated data, the minimum squared internet intensity, the internet usage separately for importers and exporters and a model including lagged internet intensity. A table summary of the robustness estimations is presented below:

Table 8: Output from the robustness estimations using fixed effects. The models include internet intensity using raw data, the minimum and squared internet intensity, the internet usage treated separately for exporters and importers and a model with lagged internet intensity. For full regression output, see Appendix: Table 11-14.

Estimator	*Coeff.*	*Std. Dev.*
FE (using un-polated data)		
ln(GDP mass)	0,5651685 ***	(0,019049)
ln(Internet intensity)	0,0462514 ***	(0,0088964)
FE (using minimum internet intensity)		
ln(GDP mass)	0,565302 ***	(0,0178217)
ln(internet minimum intensity)	0,0500751 ***	(0,0058628)
FE (internet usage separately)		
ln(GDP mass)	0,5453057 ***	(0,0187401)
ln(internet usage importer)	0,0885189 ***	(0,0121307)
ln(internet usage exporter)	0,0426086 ***	(0,0120095)
FE (lagged internet intensity)		
ln(GDP mass)	0,5731283 ***	(0,0221533)
ln(internet intensity)	0,0066846	(0,0184032)
Lag 1	0,0069139	(9,0205981)
Lag 2	-0,023799	(0,0169384)
Lag 3	0,0554598 ***	(0,0134577)

The result from the first three fixed effects estimations holds after several robustness models. The effect of internet intensity on bilateral trade is positive when using the un-polated GDP and trade data. It also holds when defining the internet intensity variable as the squared minimum internet usage rate in each country pair. Interestingly the effect is larger in this model compared to the first FE model. If a larger weight is given to the country with the lowest rate, the effect is larger. This is indicating that the marginal return of internet intensity on trade is larger in countries with a smaller rate of internet users.

When the internet variable is kept without any mass or intensity, i.e. the internet usage rates are treated separately for importing and exporting countries, the results still show a positive effect for internet usage in both importing countries and exporting countries. The effect is larger in importing countries than in exporting countries. This result is strengthening previous research by Thiemann and Fleming (2012) who also found that the effect of internet penetration rate in the importing country was positive on imports. Thiemann and Fleming found the effect of internet penetration rate in exporting countries was negative on imports. However, the result in this study shows that the internet penetration rate in exporting countries also is positive on imports.

Finally, the model including lagged internet intensity shows that the effect is significant on the three year lag, meaning that the effect on trade will be the largest three years after an increase in the internet intensity. Both the one year lag and the two year lag were insignificant.

7 Conclusion

The aim of this paper was to answer what the effect of internet usage has on bilateral trade performance. A data set in the form of a panel consisting of 180 countries over the time period 2000-2014 was used in order to examine this relationship. The results from the main fixed effects models show that the relationship is significantly positive. In this sense, the results are in line with previous studies on the topic, although that this study is using a larger data set and a more recent time period than previous studies. The internet usage has a significant positive effect on bilateral trade performance. This result holds for several robustness estimations with different variable transformations and variable definitions.

The findings in this paper also show that the effect seems to be smaller the larger the distance between the two countries. The longer the distance between two countries the smaller the effect of internet intensity on bilateral trade flows. This proximity bias was also found in a previous study by Freund and Weinhold (2004). The effect of internet intensity is also larger when more weight is given to countries with less developed IT infrastructure, indicating that the gains are larger in countries that still are developing their IT infrastructure. Another interesting finding is that the effect of internet usage in both importing and exporting countries are positive on bilateral import values.

The overall results suggest that if a larger proportion of a country's population achieves access to internet there are potential welfare gains to be made through an increased trade. This suggests that policies for developing the IT infrastructure in order to reach more users will have positive effects on the country's trade performance. The effect is also larger the less developed the IT infrastructure is, indicating that the benefit of expanding internet access is larger for countries with lower internet access rates.

8 References

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 106-116.

Anderson, J. E., & Van Wincoop, E. (2001). *Gravity with gravitas: a solution to the border puzzle* (No. w8079). National bureau of economic research.

Anderson, J. E., & Van Wincoop, E. (2004). *Trade costs* (No. w10480). National Bureau of Economic Research.

Baltagi, B. H., Bresson, G., & Pirotte, A. (2003). Fixed effects, random effects or Hausman–Taylor?: A pretest estimator. *Economics letters*, *79*(3), 361-369.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 551-572.

Bergstrand, J. H. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The review of economics and statistics*, 143-153.

Bergstrand, J. H. (1990). The Heckscher-Ohlin-Samuelson model, the Linder hypothesis and the determinants of bilateral intra-industry trade. *The Economic Journal*, 1216-1229.

Blum, B. S., & Goldfarb, A. (2006). Does the internet defy the law of gravity?. *Journal of international economics*, *70*(2), 384-405.

Clarke, G. R., & Wallsten, S. J. (2006). Has the internet increased trade? Developed and developing country evidence. *Economic Inquiry*, *44*(3), 465-484.

Deardorff, A. V. (1995). Determinants of bilateral trade: does gravity work in a neoclassical world?. *Nov*, *7*, 7-28.

Dutta, A. (2001). Telecommunications and economic activity: An analysis of Granger causality. *Journal of Management Information Systems*, *17*(4), 71-95.

Egger, P. (2000). A note on the proper econometric specification of the gravity equation. *Economics Letters*, *66*(1), 25-31.

Fleming, E., Thiemann, F., & Mueller, R. A. (2011). Information and communication technology and the global flow of wine: a gravity model of ICT in wine trade. In *2011 International Congress, August 30-September 2, 2011, Zurich, Switzerland* (No. 114353). European Association of Agricultural Economists.

Freund, C., & Weinhold, D. (2002). The Internet and international trade in services. *American Economic Review*, 236-240.

Freund, C. L., & Weinhold, D. (2004). The effect of the Internet on international trade. *Journal of international economics*, *62*(1), 171-189.

Hardy, A. P. (1980). The role of the telephone in economic development. *Telecommunications policy*, *4*(4), 278-286.

Hausman, J. A. (1978), Specification tests in Econometrics, *Econometrica*, 46, 1251-1271.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society*, 1377-1398.

Katz, R. (2012). The impact of broadband on the economy: Research to date and policy issues. *Geneva: International Telecommunication Union*.

Kurihara, Y., & Fukushima, A. (2013). Impact of the prevailing internet on international trade in Asia. *Journal of Sustainable Development Studies*, *3*(1).

Leamer, E. E., & Stern, R. M. (1970). *Quantitative international economics*. Transaction Publishers.

Meijers, H. (2014). Does the internet generate economic growth, international trade, or both?. *International Economics and Economic Policy*, *11*(1-2), 137-163.

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

Thiemann, F., Fleming, E. M., & Mueller, R. A. (2012, August). Impact of information and communication technology (ICT) on international trade in fruit and vegetables: A gravity model approach. In *2012 Conference, August 18-24, 2012, Foz do Iguacu, Brazil* (No. 123839). International Association of Agricultural Economists.

Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy. *Books (Jan Tinbergen)*.

Vemuri, V. K., & Siddiqi, S. (2009). Impact of commercialization of the internet on international trade: A panel study using the extended gravity model. *The International Trade Journal*, *23*(4), 458-484.

Verbeek, M. (2012). A guide to Modern Econometrics. 4th edition. Wiley.

Westerlund, J., & Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, *43*(6), 641-649.

APPENDIX

Table 1: Benford's test on the trade value variable. The count of observations, the distribution of the first digit in the variable and the Benford distribution.

	First digit	Count	Percent	Benford
	1	88393	**30,054**	**30,103**
	2	51954	**17,664**	**17,609**
	3	36957	**12,565**	**12,494**
	4	28923	**9,834**	**9,691**
	5	23264	**7,910**	**7,918**
	6	19607	**6,666**	**6,695**
	7	16915	**5,751**	**5,799**
	8	14768	**5,021**	**5,115**
	9	13336	**4,534**	**4,576**
TOTAL		294117	**100**	**100**

χ^2-test p-value = 0,044165762

Table 2: Benford's test on the GDP variable. The count of observations, the distribution of first digit in the variable and the Benford distribution.

	First digit	Count	Percent	Benford
	1	150897	**31,490**	**30,103**
	2	84667	**17,669**	**17,609**
	3	54595	**11,393**	**12,494**
	4	43855	**9,152**	**9,691**
	5	40812	**8,517**	**7,918**
	6	29714	**6,201**	**6,695**
	7	26313	**5,491**	**5,799**
	8	27566	**5,753**	**5,115**
	9	20764	**4,333**	**4,576**
TOTAL		479183	**100**	**100**

χ^2-test p-value = 0,000

Table 3: Output from the FE estimation for the Hausman test.

```
. xtreg ln_tradevalue ln_gdp_mass ln_ip_mass_transformation, fe

Fixed-effects (within) regression               Number of obs      =    284694
Group variable: idnew                           Number of groups   =     27307

R-sq:  within  = 0.1172                         Obs per group: min =         1
       between = 0.6230                                        avg =      10.4
       overall = 0.5571                                        max =        15

                                                F(2,257385)        =  17083.84
corr(u_i, Xb)  = 0.5061                         Prob > F           =    0.0000
```

ln_tradevalue	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_gdp_mass	.5681654	.0084446	67.28	0.000	.5516141	.5847166
ln_ip_mass_transformation	.0284197	.00478	5.95	0.000	.0190511	.0377883
_cons	-13.56741	.3906067	-34.73	0.000	-14.33299	-12.80183
sigma_u	3.1677404					
sigma_e	1.4763027					
rho	.82156042	(fraction of variance due to u_i)				

```
F test that all u_i=0:     F(27306, 257385) =    30.37       Prob > F = 0.0000
```

Table 4: Output from the RE estimation for the Hausman test.

```
. xtreg ln_tradevalue ln_gdp_mass ln_ip_mass_transformation, re

Random-effects GLS regression                   Number of obs      =    284694
Group variable: idnew                           Number of groups   =     27307

R-sq:  within  = 0.1131                         Obs per group: min =         1
       between = 0.6172                                        avg =      10.4
       overall = 0.5544                                        max =        15

                                                Wald chi2(2)       =  74272.30
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

ln_tradevalue	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_gdp_mass	.9493971	.0047748	198.84	0.000	.9400388	.9587555
ln_ip_mass_transformation	-.1565284	.0031296	-50.02	0.000	-.1626624	-.1503945
_cons	-31.75042	.2181601	-145.54	0.000	-32.178	-31.32283
sigma_u	2.621275					
sigma_e	1.4763027					
rho	.75918931	(fraction of variance due to u_i)				

Table 5: The Hausman test for the FE and the RE coefficients from Table 3 and Table 4.

. hausman fixed random

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
ln_gdp_mass	**.5681654**	**.9493971**	**-.3812317**	**.0069652**
ln_ip_mass~n	**.0284197**	**-.1565284**	**.1849481**	**.003613**

```
                       b = consistent under Ho and Ha; obtained from xtreg
        B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                 chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =     4283.11
               Prob>chi2 =      0.0000
```

Table 6: The main FE model with interpolated and extrapolated GDP and internet intensity data.

```
Fixed-effects (within) regression               Number of obs      =    294117
Group variable: idnew                           Number of groups   =     27454

R-sq:  within  = 0.1159                         Obs per group: min =         1
       between = 0.6232                                        avg =      10.7
       overall = 0.5539                                        max =        15

                                                F(16,27453)        =    706.24
corr(u_i, Xb)  = 0.5065                         Prob > F           =    0.0000

                              (Std. Err. adjusted for 27454 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_i_gdp_mass	.5583365	.0187333	29.80	0.000	.5216183	.5950546
ln_i_ip_mas~n	.0446044	.0087752	5.08	0.000	.0274045	.0618042
year						
2001	-.0008733	.013703	-0.06	0.949	-.0277318	.0259853
2002	-.0506388	.0171329	-2.96	0.003	-.08422	-.0170575
2003	-.0800605	.0198459	-4.03	0.000	-.1189594	-.0411615
2004	-.0652272	.0230893	-2.82	0.005	-.1104833	-.019971
2005	-.0695405	.0261834	-2.66	0.008	-.1208613	-.0182196
2006	-.0776605	.0297899	-2.61	0.009	-.1360503	-.0192707
2007	-.0550445	.0336493	-1.64	0.102	-.1209988	.0109098
2008	-.0213672	.0380302	-0.56	0.574	-.0959082	.0531739
2009	-.1245329	.0373211	-3.34	0.001	-.1976841	-.0513817
2010	-.1235857	.0404515	-3.06	0.002	-.2028727	-.0442988
2011	-.066681	.0437569	-1.52	0.128	-.1524467	.0190848
2012	-.0593673	.0450736	-1.32	0.188	-.1477139	.0289792
2013	-.0822977	.0469943	-1.75	0.080	-.1744089	.0098134
2014	-.0815411	.0482155	-1.69	0.091	-.176046	.0129638
_cons	-13.12872	.8856546	-14.82	0.000	-14.86465	-11.39279
sigma_u	3.168751					
sigma_e	1.4886144					
rho	.81920692	(fraction of variance due to u_i)				

Table 7: FE model with all missing trade values transformed into zeroes.

Fixed-effects (within) regression		Number of obs =	**483300**
Group variable: **idnew**		Number of groups =	**32220**
R-sq: within = **0.0209**		Obs per group: min =	**15**
between = **0.6337**		avg =	**15.0**
overall = **0.4856**		max =	**15**
		F(**16,32219**) =	**264.67**
corr(u_i, Xb) = **0.6518**		Prob > F =	**0.0000**

(Std. Err. adjusted for **32220** clusters in idnew)

ln_tradevalue_transform	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_i_gdp_mass	**.3849913**	**.0298239**	**12.91**	**0.000**	**.3265353**	**.4434472**
ln_i_ip_mass_transformat~n	**.282124**	**.0162198**	**17.39**	**0.000**	**.2503326**	**.3139154**
year						
2001	**-.0112581**	**.0220774**	**-0.51**	**0.610**	**-.0545305**	**.0320144**
2002	**-.082077**	**.0288778**	**-2.84**	**0.004**	**-.1386786**	**-.0254754**
2003	**-.0431456**	**.0340431**	**-1.27**	**0.205**	**-.1098713**	**.0235802**
2004	**-.0649776**	**.0412504**	**-1.58**	**0.115**	**-.1458301**	**.0158748**
2005	**-.1982648**	**.0489118**	**-4.05**	**0.000**	**-.2941338**	**-.1023958**
2006	**-.1266029**	**.0561397**	**-2.26**	**0.024**	**-.2366388**	**-.016567**
2007	**-.0706643**	**.0637058**	**-1.11**	**0.267**	**-.1955301**	**.0542015**
2008	**-.2351986**	**.0710702**	**-3.31**	**0.001**	**-.3744989**	**-.0958983**
2009	**-.5078392**	**.0732918**	**-6.93**	**0.000**	**-.651494**	**-.3641844**
2010	**-.6378531**	**.0812914**	**-7.85**	**0.000**	**-.7971873**	**-.4785189**
2011	**-.858286**	**.0889023**	**-9.65**	**0.000**	**-1.032538**	**-.6840341**
2012	**-.8578929**	**.0926041**	**-9.26**	**0.000**	**-1.0394**	**-.6763854**
2013	**-1.171151**	**.0975611**	**-12.00**	**0.000**	**-1.362375**	**-.9799278**
2014	**-1.81911**	**.0991059**	**-18.36**	**0.000**	**-2.013361**	**-1.624859**
_cons	**-10.45457**	**1.369751**	**-7.63**	**0.000**	**-13.13933**	**-7.769804**
sigma_u	**5.7812683**					
sigma_e	**3.599398**					
rho	**.720655**	(fraction of variance due to u_i)				

Table 8: Linear Poisson maximum likelihood estimation, fixed effects.

```
Iteration 0:   log pseudolikelihood = -2.015e+13
Iteration 1:   log pseudolikelihood = -1.309e+13
Iteration 2:   log pseudolikelihood = -1.307e+13
Iteration 3:   log pseudolikelihood = -1.307e+13
Iteration 4:   log pseudolikelihood = -1.307e+13

Conditional fixed-effects Poisson regression    Number of obs      =    411765
Group variable: idnew                           Number of groups   =     27451

                                                Obs per group: min =        15
                                                               avg =      15.0
                                                               max =        15

                                                Wald chi2(0)       =    829.16
Log pseudolikelihood  = -1.307e+13              Prob > chi2        =         .

                                 (Std. Err. adjusted for clustering on idnew)
```

trade_zeroes	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
i_gdp_mass	2.91e-27	9.18e-28	3.17	0.002	1.11e-27	4.71e-27
i_ip_mass	.000173	6.01e-06	28.80	0.000	.0001612	.0001848

Table 9: HT estimation with interaction variable between distance and internet intensity.

Hausman-Taylor estimation		Number of obs =	**287890**
Group variable: **idnew**		Number of groups =	**27336**
		Obs per group: min =	**1**
		avg =	**10.5**
		max =	**15**
Random effects u_i ~ **i.i.d.**		Wald chi2(**7**) =	**52645.61**
		Prob > chi2 =	**0.0000**

ln_tradeva~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TVexogenous						
ln_i_ip_ma~n	**.0292969**	**.0097509**	**3.00**	**0.003**	**.0101855**	**.0484083**
ln_dist_in~r	**-.014046**	**.0009846**	**-14.27**	**0.000**	**-.0159757**	**-.0121163**
TVendogenous						
ln_i_gdp_m~s	**.8365751**	**.0070766**	**118.22**	**0.000**	**.8227052**	**.8504451**
TIexogenous						
ln_distance	**-1.334194**	**.0334761**	**-39.86**	**0.000**	**-1.399806**	**-1.268582**
contig	**1.186945**	**.1786256**	**6.64**	**0.000**	**.8368456**	**1.537045**
comlang_et~o	**.7255503**	**.0624023**	**11.63**	**0.000**	**.6032441**	**.8478565**
colony	**2.10947**	**.1937061**	**10.89**	**0.000**	**1.729813**	**2.489127**
_cons	**-15.20968**	**.4454943**	**-34.14**	**0.000**	**-16.08284**	**-14.33653**
sigma_u	**3.5499799**					
sigma_e	**1.4818023**					
rho	**.85162023**	(fraction of variance due to u_i)				

Note: **TV** refers to time varying; **TI** refers to time invariant.

Table 10: RE estimation with interaction variable between distance and internet intensity.

```
Random-effects GLS regression                   Number of obs      =    287890
Group variable: idnew                           Number of groups   =     27336

R-sq:  within  = 0.1126                         Obs per group: min =         1
       between = 0.7153                                        avg =      10.5
       overall = 0.6517                                        max =        15

                                                Wald chi2(21)      =  76539.89
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                   (Std. Err. adjusted for 27336 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_i_gdp_mass	1.042716	.0049667	209.94	0.000	1.032982	1.052451
ln_i_ip_mass_transformation	.1625059	.0164394	9.89	0.000	.1302852	.1947265
ln_dist_inter	-.0159914	.0016675	-9.59	0.000	-.0192596	-.0127233
ln_distance	-1.258492	.0207504	-60.65	0.000	-1.299162	-1.217822
contig	1.130844	.1091446	10.36	0.000	.9169245	1.344764
comlang_ethno	1.004339	.0393795	25.50	0.000	.9271563	1.081521
colony	1.265253	.089411	14.15	0.000	1.09001	1.440495
year						
2001	.0169137	.0136559	1.24	0.216	-.0098515	.0436788
2002	-.0540273	.0164996	-3.27	0.001	-.086366	-.0216887
2003	-.2002838	.0183897	-10.89	0.000	-.236327	-.1642406
2004	-.3261188	.0200103	-16.30	0.000	-.3653382	-.2868995
2005	-.43773	.0212379	-20.61	0.000	-.4793556	-.3961045
2006	-.5527897	.0226644	-24.39	0.000	-.5972112	-.5083682
2007	-.6833665	.0235799	-28.98	0.000	-.7295823	-.6371508
2008	-.7824181	.0252431	-31.00	0.000	-.8318937	-.7329425
2009	-.8272354	.0262011	-31.57	0.000	-.8785887	-.7758821
2010	-.9167579	.0276314	-33.18	0.000	-.9709145	-.8626012
2011	-.9674233	.0289136	-33.46	0.000	-1.024093	-.9107538
2012	-.9782102	.0298891	-32.73	0.000	-1.036792	-.9196286
2013	-1.050675	.0310137	-33.88	0.000	-1.111461	-.9898889
2014	-1.066792	.0319877	-33.35	0.000	-1.129487	-1.004097
_cons	-25.80948	.3064453	-84.22	0.000	-26.4101	-25.20886
sigma_u	2.1963579					
sigma_e	1.4815782					
rho	.68726962	(fraction of variance due to u_i)				

Table 11: FE model with the raw GDP and internet intensity data.

```
Fixed-effects (within) regression               Number of obs      =    284694
Group variable: idnew                           Number of groups   =     27307

R-sq:  within  = 0.1176                         Obs per group: min =         1
       between = 0.6240                                        avg =      10.4
       overall = 0.5583                                        max =        15

                                                F(16,27306)        =    702.58
corr(u_i, Xb)  = 0.5061                         Prob > F           =    0.0000

                                   (Std. Err. adjusted for 27307 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_gdp_mass	.5651685	.019049	29.67	0.000	.5278316	.6025054
ln_ip_mass_transformation	.0462514	.0088964	5.20	0.000	.0288139	.0636889
year						
2001	.0042687	.0138547	0.31	0.758	-.0228872	.0314247
2002	-.0441929	.0172921	-2.56	0.011	-.0780862	-.0102995
2003	-.0726912	.0201328	-3.61	0.000	-.1121525	-.0332299
2004	-.0664358	.0234113	-2.84	0.005	-.1123231	-.0205485
2005	-.072048	.0263485	-2.73	0.006	-.1236925	-.0204036
2006	-.0754175	.0299118	-2.52	0.012	-.1340461	-.0167889
2007	-.0606336	.0337905	-1.79	0.073	-.1268648	.0055975
2008	-.0260568	.0381972	-0.68	0.495	-.1009253	.0488116
2009	-.1364163	.0374758	-3.64	0.000	-.2098706	-.0629619
2010	-.1332558	.0405955	-3.28	0.001	-.212825	-.0536866
2011	-.0778375	.0439523	-1.77	0.077	-.1639864	.0083113
2012	-.0680471	.0452127	-1.51	0.132	-.1566663	.0205721
2013	-.0910482	.047117	-1.93	0.053	-.1833999	.0013036
2014	-.0916378	.048556	-1.89	0.059	-.1868101	.0035345
_cons	-13.44798	.9009462	-14.93	0.000	-15.21388	-11.68208
sigma_u	3.1632631					
sigma_e	1.4760237					
rho	.82120085	(fraction of variance due to u_i)				

Table 12: FE model with the internet intensity variable defined as the squared minimum value of the internet usage rate of country i and j at time t.

```
Fixed-effects (within) regression               Number of obs      =    294117
Group variable: idnew                           Number of groups   =     27454

R-sq:  within  = 0.1163                         Obs per group: min =         1
       between = 0.6262                                        avg =      10.7
       overall = 0.5573                                        max =        15

                                                F(16,27453)        =    702.04
corr(u_i, Xb)  = 0.5038                         Prob > F           =    0.0000

                              (Std. Err. adjusted for 27454 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_i_gdp_mass	.565302	.0178217	31.72	0.000	.5303705	.6002335
ln_imios	.0500751	.0058638	8.54	0.000	.0385817	.0615685
year						
2001	.0029233	.0131444	0.22	0.824	-.0228404	.0286871
2002	-.0473592	.0153574	-3.08	0.002	-.0774604	-.0172579
2003	-.0787881	.0177949	-4.43	0.000	-.1136669	-.0439092
2004	-.068231	.0211444	-3.23	0.001	-.109675	-.026787
2005	-.0755115	.0243891	-3.10	0.002	-.1233153	-.0277077
2006	-.0883677	.0280001	-3.16	0.002	-.1432493	-.033486
2007	-.0722296	.0323258	-2.23	0.025	-.1355898	-.0088694
2008	-.0455917	.0369604	-1.23	0.217	-.1180359	.0268526
2009	-.1531634	.0358019	-4.28	0.000	-.223337	-.0829899
2010	-.1599908	.0391306	-4.09	0.000	-.2366887	-.083293
2011	-.1093118	.0427002	-2.56	0.010	-.1930063	-.0256172
2012	-.1070773	.0439436	-2.44	0.015	-.1932089	-.0209457
2013	-.135908	.0459937	-2.95	0.003	-.226058	-.045758
2014	-.1400112	.0472923	-2.96	0.003	-.2327065	-.0473159
_cons	-13.4102	.8491738	-15.79	0.000	-15.07463	-11.74578
sigma_u	3.148834					
sigma_e	1.4882857					
rho	.8173976	(fraction of variance due to u_i)				

Table 13: FE model with the internet usage rate for importing and exporting countries separately.

```
Fixed-effects (within) regression               Number of obs     =    294117
Group variable: idnew                           Number of groups  =     27454

R-sq:  within  = 0.1163                         Obs per group: min =         1
       between = 0.6193                                        avg =      10.7
       overall = 0.5500                                        max =        15

                                                F(17,27453)       =    665.06
corr(u_i, Xb)  = 0.5068                         Prob > F          =    0.0000

                              (Std. Err. adjusted for 27454 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_i_gdp_mass	.5453057	.0187401	29.10	0.000	.5085742	.5820372
ln_iiro	.0885189	.0121307	7.30	0.000	.0647422	.1122956
ln_iipo	.0426086	.0120095	3.55	0.000	.0190694	.0661478
year						
2001	-.0065842	.0135274	-0.49	0.626	-.0330986	.0199301
2002	-.0642188	.0165892	-3.87	0.000	-.0967345	-.0317031
2003	-.0950234	.0190928	-4.98	0.000	-.1324462	-.0576007
2004	-.0827175	.0223187	-3.71	0.000	-.1264632	-.0389717
2005	-.0870971	.0254162	-3.43	0.001	-.1369142	-.03728
2006	-.0967601	.0289121	-3.35	0.001	-.1534293	-.0400909
2007	-.0757935	.0329262	-2.30	0.021	-.1403305	-.0112564
2008	-.0441348	.0373585	-1.18	0.237	-.1173593	.0290897
2009	-.1541674	.0366058	-4.21	0.000	-.2259166	-.0824182
2010	-.156814	.0397299	-3.95	0.000	-.2346866	-.0789415
2011	-.1011115	.0431187	-2.34	0.019	-.1856262	-.0165967
2012	-.097561	.0443711	-2.20	0.028	-.1845306	-.0105914
2013	-.1236056	.0463447	-2.67	0.008	-.2144435	-.0327677
2014	-.1236663	.0475488	-2.60	0.009	-.2168643	-.0304682
_cons	-12.58895	.8848498	-14.23	0.000	-14.3233	-10.8546
sigma_u	3.1874604					
sigma_e	1.4883527					
rho	.82099586	(fraction of variance due to u_i)				

Table 14: FE model with three lagged internet intensity variables.

```
Fixed-effects (within) regression               Number of obs      =    238725
Group variable: idnew                           Number of groups   =     27042

R-sq:  within  = 0.0783                         Obs per group: min =         1
       between = 0.6324                                        avg =       8.8
       overall = 0.5638                                        max =        12

                                                F(16,27041)        =    471.55
corr(u_i, Xb)  = 0.5129                         Prob > F           =    0.0000

                                         (Std. Err. adjusted for 27042 clusters in idnew)
```

ln_tradevalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_i_gdp_mass	.5731283	.0221533	25.87	0.000	.5297068	.6165499
ln_i_ip_mass_transformation	.0066846	.0184032	0.36	0.716	-.0293867	.0427558
lag1_liimt	.0069139	.0205981	0.34	0.737	-.0334595	.0472872
lag2_liimt	-.023779	.0169384	-1.40	0.160	-.0569791	.0094211
lag3_liimt	.0554598	.0134577	4.12	0.000	.029082	.0818376
year						
2004	.0035065	.0145663	0.24	0.810	-.0250441	.0320572
2005	-.0248387	.0191684	-1.30	0.195	-.0624097	.0127324
2006	-.039221	.0233324	-1.68	0.093	-.0849537	.0065117
2007	-.0260336	.0281339	-0.93	0.355	-.0811776	.0291103
2008	.0055298	.0329563	0.17	0.867	-.0590662	.0701258
2009	-.1000754	.032645	-3.07	0.002	-.1640614	-.0360895
2010	-.1033059	.0361613	-2.86	0.004	-.174184	-.0324279
2011	-.0528678	.0401236	-1.32	0.188	-.1315121	.0257765
2012	-.0469247	.0419613	-1.12	0.263	-.1291711	.0353216
2013	-.076264	.04454	-1.71	0.087	-.1635647	.0110367
2014	-.0790737	.0462654	-1.71	0.087	-.1697563	.0116089
_cons	-13.91389	1.045288	-13.31	0.000	-15.96271	-11.86507
sigma_u	3.2140632					
sigma_e	1.4527841					
rho	.83034979	(fraction of variance due to u_i)				

Table 15: List of all 180 countries included in the study and with the percentage of total bilateral trade values available in the period 2000-2014, from lowest to highest.

Country	Percent	Country	Percent	Country	Percent
FS Micronesia	11,9	Botswana	52,4	Mauritius	73,8
Guinea-Bissau	14,3	Benin	52,5	Luxembourg	74,5
Marshall Isds	14,4	Belize	53,5	Tunisia	75,1
Equatorial Guinea	16,2	Iran	54,5	Viet Nam	75,1
Kiribati	20,0	Mali	54,5	Sri Lanka	75,1
Chad	20,2	Guyana	55,0	Israel	76,0
Eritrea	20,7	Mauritania	55,3	Cyprus	76,2
Djibouti	20,8	China, Macao SAR	55,6	Chile	76,4
Tajikistan	21,8	Azerbaijan	55,7	Saudi Arabia	77,1
Bhutan	22,0	Armenia	56,1	Pakistan	77,7
Turkmenistan	22,7	Trinidad and Tobago	56,5	United Arab Emirates	77,8
Vanuatu	23,3	Paraguay	56,8	Croatia	78,0
Uzbekistan	23,9	Namibia	58,1	Peru	78,0
Lesotho	24,0	Cambodia	58,2	Morocco	78,2
Angola	24,6	Panama	58,6	Philippines	78,7
Liberia	25,2	Kuwait	58,7	Egypt	79,1
Lao People's Dem. Rep.	25,5	Georgia	59,8	Argentina	79,2
Tonga	26,0	Zambia	59,9	Hungary	80,3
Haiti	26,2	Zimbabwe	60,1	Slovakia	81,3
Sao Tome and Principe	28,0	Malawi	60,2	Bulgaria	81,3
Solomon Isds	28,5	Qatar	60,4	Ukraine	81,9
Comoros	29,0	Rep. of Moldova	60,5	Singapore	81,9
Samoa	29,4	Oman	60,9	Slovenia	81,9
Sierra Leone	30,6	Bangladesh	61,1	Colombia	82,0
Libya	31,3	El Salvador	61,8	Lebanon	82,0
Saint Lucia	34,1	Honduras	62,5	Greece	83,5
Afghanistan	34,4	Albania	62,5	Norway	84,4
Aruba	36,1	Bosnia Herzegovina	62,7	Portugal	84,7
Mongolia	36,3	Jamaica	62,9	Russian Federation	85,0
Papua New Guinea	36,8	Sudan	62,9	New Zealand	85,5
Brunei Darussalam	37,3	TFYR of Macedonia	63,3	Finland	86,0
Swaziland	38,0	Mozambique	63,4	Poland	86,7
Saint Kitts and Nevis	38,1	Nicaragua	63,5	Czech Rep.	87,0
Grenada	39,0	Bolivia	63,7	Turkey	87,0
Maldives	40,0	Latvia	63,8	Mexico	87,1
Central African Rep.	40,6	Uganda	64,6	China, Hong Kong SAR	87,4
Cabo Verde	41,4	Bahrain	65,8	Brazil	87,7
Faeroe Isds	41,7	Kazakhstan	65,9	Malaysia	87,9
Seychelles	42,0	Jordan	66,2	Australia	88,3
Congo	42,8	Barbados	66,3	Indonesia	88,4
Antigua and Barbuda	43,0	Guatemala	66,4	Sweden	88,4
Burundi	43,3	Kenya	66,9	Ireland	88,8
Andorra	44,3	Cameroon	67,0	Austria	89,1
St Vinc. and the Gdines	44,6	Belarus	67,0	India	89,1
Bermuda	44,9	Senegal	67,6	Denmark	89,7
Dominica	45,5	Venezuela	67,9	China	89,9
Guinea	46,0	Nigeria	68,1	Belgium	89,9
Nepal	46,1	Madagascar	68,5	Rep. of Korea	90,0
Gambia	46,1	Malta	69,4	Canada	90,1
Kyrgyzstan	46,5	Iceland	69,4	South Africa	90,1
Gabon	47,7	Ethiopia	69,6	Switzerland	90,2
Burkina Faso	48,9	Ghana	69,7	Thailand	90,2
Niger	48,9	Uruguay	70,7	Japan	90,3
Fiji	49,1	Dom. Rep.	70,9	Spain	90,4
Rwanda	49,5	Cote d'Ivoire	71,1	Italy	90,5
Yemen	50,0	Ecuador	71,3	Netherlands	90,6
Cuba	50,9	Algeria	71,4	United Kingdom	91,1
Togo	50,9	Costa Rica	72,4	USA	91,2
Bahamas	51,2	Estonia	73,3	Germany	91,2
Suriname	51,8	United Rep. of Tanzania	73,3	France	91,2