# Sheffield Economic Research Paper Series

# SERP Number: 2013010

**ISSN 1749-8368**

David Cuberes

## Are Internet and Face-to-Face Contacts Substitutes or Complements? Evidence from Internet Traffic between Cities

**July 2013**

Department of Economics
University of Sheffield
9 Mappin Street
Sheffield
S1 4DT
United Kingdom
www.shef.ac.uk/economics

# *Are Internet and Face-to-Face Contacts Substitutes or Complements? Evidence from Internet Traffic between Cities*[*]

David Cuberes
*University of Sheffield*

This paper uses a new dataset on Internet flows between cities around the world to study whether electronic communication and face-to-face contacts are substitutes or complements. In order to test these competing hypotheses I estimate a regression of bilateral Internet traffic on physical distance between pairs of cities and several city and country-specific variables that include a control for cities' population, countries' population and per capita GDP, the number of Internet users, the intensity of trade between countries, and several dummies that aim to capture city specific effects and the degree of familiarity between residents of different countries. The estimates reveal a strong and robust negative effect of distance on the intensity of electronic communications, suggesting that Internet and face-to-face contacts are more likely to be complements than substitutes.

JEL classification: R12

Keywords: cities; Internet; face-to-face contacts; death of distance

*"Together, [Internet's] apparent ubiquity and invisibility give its users a sense of placelessness, of freedom from the traditional constraints of physical distance. But this placelessness is an illusion. The Internet is where its users are." (Kolko, 1999)*

## 1. Introduction

In her famous 2001 book *The Death of Distance*, Frances Cairncross discussed how the growing ease and speed of communication was creating a world where distance mattered less and less. Her book comments on several major changes likely to result from the so called "IT revolution", including the fact that workers do not need to be physically attached to a city to perform their duties. Indeed, the great improvements in information technology over the past thirty years have led some to suggest that the informational functions of physical proximity will eventually become obsolete (Gilder 1995).[1] Gaspar and Glaeser (1998) challenge this view and argue that, to address this question, one needs to find out whether face-to-face communication and electronic

[*] I thank Rafael González-Val, Jeffrey Lin, Jenny Roberts, Jesse Shapiro, Jens Suedekum, Aki Tsuchiya, Yves Zenou and seminar participants at the 2010 NARSC meetings in Denver for useful comments. I acknowledge the financial support of the Ministerio de Ciencia e Innovación and FEDER funds (proyecto SEJ2007-62656). Please send comments to d.cuberes@sheffield.ac.uk. Department of Economics, University of Sheffield, 9 Mappin St, S1 4DT, Sheffield, United Kingdom. +44 7530371845.

[1] As claimed in Kolko (1999) similar predictions have been made by a few commentators, including Peter Drucker (1989), Bill Gates, Nathan Myhrvold, and Peter Rinearson (1995), and the National Research Council (1998).

communication are indeed substitutes or complements. While the two forms of communication can be substitutes if electronic communication is a faster and more efficient way of transmitting information than face-to-face communication, theoretically, the two types of communication could also be complements if people may expect to use both types of connection when forming relationships.[2]

This paper uses of a new dataset on bilateral Internet flows between cities to shed light on the relation between physical distance between cities and the intensity of Internet communication between them. I use physical distance as a proxy for the cost of establishing face-to-face contacts (see footnote 8 for more details on this).[3] A finding of a positive correlation between these two variables would provide evidence in favor of the hypothesis that direct face-to-face communications and electronic communications are indeed substitutes since cities that are further away would need to *substitute* that the lack of face-to-face exchanges with more frequent (or longer) Internet interactions. If, on the other hand, the relationship is negative, this would seem to indicate that residents of cities that are geographically close to each other happen to engage in more intense electronic communication, i.e. Internet and face-to-face contacts are complements.

The paper is organized as follows. The next section summarizes the literature on the link between Internet and communication technologies (ICTs) and several aspects of geography and the spatial distribution of population. In Section 3 I briefly discuss a theoretical framework that helps interpreting the relationship between Internet – or electronic communication in general - and face-to-face contacts. Section 4 presents the empirical strategy. The data used in the analysis is presented in Section 5, where I also discuss two measurement error problems. Section 6 shows the main results and, finally Section 7 concludes the paper.

## 2. Literature

Several theoretical models have explored how changes in ICT can affect a country's geographical structure. Ioannides et al. (2008) develop a model based on Rossi-Hansberg and Wright (2007) to show that ICT improvements lead to a decentralization of economic activity.[4] On the other hand, in the theory presented in Gaspar and Glaeser (1998), it is possible that a better ICT technology induces an increase in population concentration.[5]

[2] When comparing the two alternative ways of communicating, it is important to point out that the "good" exchanged in each of them may not be of the same quality, i.e. the information transmitted in face-to-face contacts may well be considered of "higher quality". Unfortunately, the nature of the data does not allow me to explore this point further.

[3] This seems a reasonable assumption even after the secular decline in transportation costs, since what matters is that these costs are an increasing function of distance, independently of the strength of this relationship

[4] In Henderson and Mitra (1996) the internal structure of cities becomes more decentralized as a result of improvements in ICT.

[5] There is a quite large literature addressing the question of how improvements in communication technologies have diminished the role of distance between countries on their bilateral volume of trade. For instance, Jacks (2009) uses nineteenth century trade data to argue that distance between countries became a less important determinant of trade flows during that century than it was in the past. Other recent work in international trade finds that, in spite of its decline as a fundamental factor to explain bilateral country trade, distance still matters (Berthelon and Freund 2008, Carrère and Schiff 2005, and

My analysis relates more directly to the empirical evidence presented in Gaspar and Glaeser (1998) who present a model in which face-to-face contacts may increase the demand for electronic contacts (telephone calls in their case), and so people who are physically closer to one another speak over the phone more often than those who live far apart from each other. To provide some suggestive evidence for this, they refer to a study by Moyer (1977) that shows that, in the mid-1970s, more than 40% of phone calls in the U.S. were made to places within a two mile radius, and more than 75% were made to places within a six mile radius. Second, they report the results of a regression estimated by Imagawa (1996), where the dependent variable is the number of telephone calls (and their duration) between pairs of Japanese prefectures, the main administrative subdivision in the country. The set of controls included in his regression are, among others, the sum across the two prefectures logarithm of population and the logarithm of income. Imagawa finds a robust negative correlation between distance and both the number and duration of telephone calls, suggesting that face-to-face communication – proxied by the inverse of physical proximity - and electronic communication are complements.[6]

A related exercise is the empirical analysis carried over by Ioannides et al. (2008). They use data on the number of telephone lines and on Internet usage in different countries and conclude that, increases in the former have significantly led to a more concentrated distribution of city sizes. One interpretation of their findings is that these technological changes can be seen as a complement to face-to-face contacts. Kolko (1999) uses data on Internet domain names to test different hypotheses about the relationship between city size and Internet use, also concluding that cities – i.e. places with intense face-to-face interactions - and electronic communication tend to be complements. Consistent with this, Zook (2000, 2001) finds that most of the dotcom Internet addresses and most of the top Internet companies in terms of volume of electronic commerce are located in major cities like San Francisco, New York, and Los Angeles. My paper clearly differs from Ioannides et al (2008) in that I have access to *bilateral traffic* data between cities as opposed to aggregate measures of usage per city. On the other hand, while my regression analysis is similar to Kolko (1999) and Imagawa (1996), an important difference is that I use more recent data on Internet flows rather than on telephone calls, and that my sample includes a large number of countries. Moreover, using bilateral Internet trade data instead of data on Internet domains has the advantage of being able to estimate the effect of distance between cities on the intensity of Internet usage.

Veenhof et al. (2008) use Canadian data to show that heavy Internet users spend less face-to-face time with people who live near them (household members, friends, and relatives), pointing to a significant substitution effect. Pons-Novell and Viladecans-Marsal (2006) analyze the possible effects of Internet use and e-commerce on residential locations using survey data from survey data from 1,500 homes in the city of Barcelona, Spain. By analyzing whether the geographical distribution of new technologies differs across different city sizes, they conclude that, in terms of the decision to connect to the Internet, the relationship with the cities' off-line commercial offerings is one of

---

Disdier and Head 2008). As I show below, my results are in line with these findings since they suggest that distance is also a crucial determinant of the volume of electronic communications between cities.

[6] The finding that the more people see each other, the more they speak over the phone has also been reported using U.S. data in Wellman (1979), Fischer (1982), and Wellman and Tindall (1993).

complementarity. Their results concerning the decision to buy goods or services online is inconclusive, though. Finally, also using Canadian data, Mok et al. (2009) analyze, among other things, the role of distance in affecting the frequency of e-mail contact, in addition to face-to-face and phone contact. Their results show that e-mail contact is generally not affected by geographical distance. However, email communication tends to increase for transoceanic relationships greater than 3000 miles apart, again suggesting a substitution effect. They also report that face-to-face contact remains strongly related to short distances (within five miles), while distance has little impact on how often people phone each other at the regional level (within 100 miles). Once again, a clear advantage of my paper is that I use a large number of countries to approach a similar question, which allows me to interpret my results as applying more globally.

## 3. Theoretical framework

The empirical exercise carried out in this paper can be motivated with the theory developed in Gaspar and Glaeser (1998). They present a model in which, in principle, electronic communication and face-to-face contacts can be complements or substitutes. In their model, people can use electronic communication – Internet in my case – or face-to-face contacts as a technology to communicate with each other. In this sense, it is clear that they are substitutes. However, if some relationships involve both electronic and face-to-face communication, declines in the cost of electronic communication – for instance due to the widespread use of the Internet- raise the overall number of relationships. Since a fraction of these new relationships may occur face-to-face, the Internet and face-to-face contacts may indeed be complements. As argued in Mok et al. (2009), a clear example of this situation is when emails are used to sustain contact in-between meetings and/or to arrange future meetings.[7]

In their model agents have a choice of carrying out a private project, which gives a random return *R* or a joint one, which return is $\alpha f(i)$, where $\alpha$ is a random variable, *f* is an increasing and concave function, and *i* denotes the intensity of the interaction with another individual. Once an individual engages in a joint project she learns the quality of her match and then chooses whether to pursue the relationship or not. If she does, this can be done through face-to-face communication or through electronic communication. Finally, the intensity of the relationship must also be decided.

The electronic communication produces a relationship $i = \beta_E t$ where *t* is time spent in the interaction and $\beta_E > 0$. Face-to-face communication, on the other hand, generates an interaction $i = Max[0, \beta_F (t - t_F)]$ where $\beta_F > 0$ and $t_F$ represents the fix time component needed to set up face to face interactions.[8] Gaspar and Glaeser make the crucial assumption that $\beta_F > \beta_E$, reflecting the fact that face-to-face interactions produce more "high quality" outcomes than electronic ones (see my footnote 2). Both

[7] According to Hampton and Wellman (2002, 2003) and Shklovski et al. (2008) people prefer face-to-face and phone contact rather than electronic communication for the initial development of relationships. Arguably, this may have changed in recent years with the popularity of social networks like Facebook.
[8] Note that if $t_F > t$ face-to-face communication is too costly to set up and so agents choose to engage in electronic communications only. Moreover, it is reasonable to assume that $t_F$ should be an increasing function of distance between agents, although this model abstracts form this.

types of communication require the use of time as their only input, with an opportunity cost $c > 0$.

The optimal match intensity is chosen by setting the marginal benefit $\alpha f'(i)$ of an interaction equal to its marginal cost of intensity $c$. It is easy to show that in equilibrium $i_F^*(\alpha) > i_E^*(\alpha)$ i.e. the intensity of face-to-face interactions is higher than that of electronic ones. Gaspar and Glaeser further derive a cutoff level of $\underline{\alpha}$ above which any further contact – after the initial one – is feasible, and a second one $\alpha^*$ above which face-to-face communication is strictly preferred to the electronic one. It follows from their analysis that $\frac{\partial \alpha^*}{\partial \beta_E} > 0$. This shows that improvements in the efficiency of electronic communication – for instance an increase in the Internet speed – reduce the number of face-to-face interactions, indicating that the two forms of communication are substitutes. On the other hand, the total number of interactions rises as telecommunications technology improves i.e. $\frac{\partial j^*}{\partial \beta_E} > 0$ where $j^*$ is the individual who is indifferent between engaging in one type of interaction or the other.[9] The reason why this happens in the model is that the expected returns from an electronic communication increase as a result of the better technology.

Finally, Gaspar and Glaeser derive an expression that shows that an increase in the efficiency of electronic communication $\beta_E$ has an ambiguous effect on the total amount spent in face-to-face contacts. There are two different effects in place. First, holding the cutoff between face-to-face and electronic communications constant, the total time spent interacting (which is partially spent on face-to-face interactions) increases, making the two types of communication complements. Second, holding the number of interactions constant, the cutoff can increase or decrease depending on several factors. First, the elasticity of the number of contacts with respect to the benefits of making a new contact. Second, the relationship between an improvement in the efficiency of electronic communication and total expected revenues from contacts. Third, the elasticity of the time spent in face-to-face contacts with respect to the cutoff $\alpha^*$. Finally, the sign of this term depends on the extent to which $\alpha^*$ changes when electronic communication improves.

## 4. Empirical Strategy

In this section I propose a simple test of whether face-to-face contacts and the Internet (as a proxy for different types of electronic communication) are complements or substitutes. If the two are complements, face-to-face contact – implicitly measured as the inverse of the distance between two cities – should be associated with a higher demand for electronic interactions, and therefore the correlation between bilateral Internet flows and physical distance should be negative. If, on the other hand, these two

[9] This individual is implicitly defined as the one whose benefits of engaging in either of the two modes of communications are the same.

forms of communication are substitutes, the relationship between Internet traffic and physical distance should be positive.

To implement this, I estimate a regression similar to the one run by Takuo Imagawa as cited in Gaspar and Glaeser (1998):

$$\log I_{i_k j_{k'}} = \alpha + \beta_1 \log dist_{i_k j_{k'}} + \beta_2 X_{i_k j_{k'}} + \gamma' Z_{kk'} + u_{i_k j_{k'}}$$

where $I_{i_k j_{k'}}$ is the Internet traffic between city *i* located in country *k* and city *j* located in country *k'*. The variable $dist_{i_k j_{k'}}$ is the shortest distance (in kilometers) between cities $i_k$ and $j_{k'}$. $X_{i_k j_{k'}}$ is a vector of city-specific controls that includes a control of the cities' population (in logs), a dummy that takes a value of one if a city is the capital of the country, and a dummy that takes a value of one if the two cities in a pair are located in the same country.[10] The vector $Z_{kk'}$ includes different country-level covariates: a control of the countries' population and GDP per capita (both in logs), a dummy that takes a value of one if the two countries where the cities are located share the same main language, a dummy that takes a value of one if the country where the city is located has English as its main language, a dummy that takes a value of one if the two countries where the cities are located belong in the same world region, and a dummy that takes a value of one if the two countries where the cities are located belong in the same income group. These controls are meant to capture the degree of "familiarity" between the two cities and/or countries that engage in some type of Internet exchange. This acknowledges the fact that a bias in this regression occurs if people who live close to one another have more things in common than those who live far away, as suggested in Gaspar and Glaeser (1998). I also include as a covariate a country's number of Internet users (in logs) to capture its degree of communication's infrastructure. Finally, $u_{i_k j_{k'}}$ denotes the error term. Following Spolaore and Wacziarg (2009), spatial correlation may result from the fact that my dependent variable measures a bilateral flow. I therefore check whether my results are robust to using two-way clustered standard errors.

## 5. Data

The data on Internet traffic is measured as the traffic over Internet bandwidth between cities in the year 2008[11] in gigabits per second. The dataset reports traffic for the top connections i.e. the cities with the highest bilateral Internet flows.[12] This includes 15 routes for the Asian region, 30 for Europe, 5 for Latin America, and 20 for North America. Table 1 in the appendix lists all the routes ordered from highest to lowest

---

[10] I have also attempted to include as a control the cities' per capita real GDP, but the existing data, apart from being quite sparse, is not really comparable across countries. However, in results not reported here, I find that, in my sample of cities, the city's population and GDP per capita are quite strongly positively related, so that by including only the cities' population I should in part capture their economic relevance.

[11] The data source is *TeleGeography*. See http://www.telegeography.com/. A gigabit equals 1,000,000,000 bits, the smallest unit of data in a computer.

[12] This sample selection may bias our results if the complementarity effect that I find is stronger in routes with more bilateral traffic. While I acknowledge this possibility, there is not much that can be done to address this issue due to data availability.

traffic in 2008 along with the distance between the two cities involved. It is interesting to note that six of the ten more used routes involve traffic between U.S cities.

Bilateral Internet traffic data are more informative than the widely used percentage of Internet users, subscribers, or hosts.[13] Note that this measure of Internet use also offers some advantages with respect to the use of domain names, the dependent variable used in Kolko (1999). First, and most importantly, these data allow one to identify and quantify the degree of Internet usage between different locations. Second, it does not have any of the two shortcomings mentioned by Kolko: the fact that companies may have an e-mail but not a domain name, and that having a registered domain name does not necessarily mean that companies use Internet in their daily business process. Arguably, a shortcoming of this traffic data is that it does not allow one to distinguish between commercial and personal use.

**Measurement Error Issues**

The first problem I face with the Internet traffic data has to do with the fact that it is measured as an aggregate and may therefore contain electronic communications that can be considered as “aspatial”. For instance, a search on Google by an Internet user located in a Japanese city does not necessarily reflect a need to connect to the specific US city where the Google server is located. On the other hand, an e-mail sent between two people located in different cities it is a clear example of a desire for these people to establish a bilateral contact. While these are two extreme examples, the rest of the components of Internet traffic probably lie somewhere in between. One can however argue that this is not a major concern. Clearly, for the user, where the server is does not matter but, still, if one assumes that the location choice of the server is endogenous, it is then likely to depend on the local availability of high skilled people, for instance. This would induce co-agglomeration effects with other industries using high-skilled workers, which would translate in a negative impact of distance on Internet traffic.

On the other hand, even without considering the decision of where the server locates, what is most relevant for my exercise is what I will henceforth denote as the “spatial component” of Internet traffic i.e. any Internet traffic in which the location of the users involved in the communication matters for at least one of the users. Unfortunately, the bilateral traffic data is not disaggregated by applications, so the implicit assumption made in the paper is that most of the Internet traffic between cities has a significant degree of a spatial component. This certainly applies to email traffic, connections to read newspapers (people tend to read local newspapers that refer to a location near their residence than far away from it). Table 2 displays the aggregate Internet traffic by application in the year 2008.

TABLE 2 HERE

Of these components, one can safely assume that several of them have a clear spatial component, i.e. they involve bilateral (or multilateral) Internet traffic that seeks to

[13] See Andrés et al. (2009) for a discussion on the use of different aggregate indicators of Internet usage.

connect specific users located in different geographical locations. This seems obvious for Email and VolP/IM/Video Calling, as well as peer-to-peer (P2P) computer network.[14] One can also assume that a fraction of the traffic involved in "gaming" takes place between people located in different geographical areas and that the identity of your game partner may indeed be important. For example, it is likely that users may prefer to play with people they know in person or with people that speak their same language and/or live in the same country. The two largest shares in Table 1 are "Web browsing" and "Streamed & Buffered Audio / Video". Although I do not have information of the type of communications included in these categories, I make the conservative assumption that at least 50% of this traffic has a spatial component. With these assumptions, Table 2 suggests that around 54% of total Internet traffic involves communications in which the location of the parties involved matters.

The second data issue is that the bilateral Internet traffic data represent traffic between Internet servers, not Internet users directly. For example, Internet traffic going from New York to Dallas traverses several links (cities) before reaching their final destination. This implies that the data actually measure traffic between different cities, but do not necessarily measure traffic between the initial city and the final destination. In other words, a fraction of the traffic between New York and Dallas may indeed be due to a request in Los Angeles for some information in, for instance, San Francisco. I follow different strategies to attempt to control for the volume of this intermediate traffic. First, I include a dummy control for the population of any two cities involved in Internet communication. The rationale for this is that this variable should partially take into account the fact that larger cities are more likely to be central hubs that host a large fraction of intermediate traffic. Second, I include a dummy that takes a value of one if any of the two cities in a given route is the largest urban agglomeration of its country and zero otherwise.[15] Similarly, I construct a dummy that takes a value of one if any of the two cities in a given route is among the largest in the entire sample. In the regressions displayed below, the cutoff to define a large city is that its population is above the 75$^{th}$ percentile.[16] Finally, as Figure 1 shows, although there is some positive correlation (0.14) between city size and Internet traffic, the relationship is not very strong. Therefore, I construct a dummy that takes a value of one if any of the two cities in a given route is among the most transited ones, where the threshold is defined as the 90$^{th}$ percentile of the sample.[17] I also check whether an interaction term between distance and the dummy for the most transited cities is as important term in explaining bilateral Internet traffic.

FIGURE 1 HERE

[14] The latter are applications in which each computer in the network can act as a client or server for the other computers in the network, allowing shared access to various resources such as files, peripherals, and sensors without the need for a central server. P2P networks can be used for sharing content such as audio, video, data, or anything in digital format.

[15] These cities are Tokyo, New York, Seoul, Sao Paulo, Mexico City, Moscow, Beijing, Buenos Aires, Istanbul, Paris, London, Lima, Dusseldorf, Bogota, Taipei, Madrid, Milan, Singapore, Sidney, Warsaw, Brussels, Vienna, Copenhagen, Stockholm, Prague, Amsterdam, and Helsinki.

[16] Under this criterion, the group of large cities is composed of Tokyo, New York, Seoul, Sao Paulo, Mexico City, and Moscow.

[17] With this threshold the list of most transited cities is: New York, Washington, London, San Francisco, and Paris.

Distance (in kilometers) between cities is calculated by the author using the flight distance calculator from *Travelmath.com*.[18] Data on cities' populations is obtained from Demographia (2009), which collects the population of all large urban agglomerations around the world in different years between 2001 and 2007.[19] The countries' population and real per capita GDP in 2008 is from the Penn World Tables 6.3. (Heston, Summers, and Aten, 2009). The regional dummies used in the regression are defined in the World Bank Classification and include the following regions:[20] East Asia and Pacific (EAP), Europe and Central Asia, Latin American and the Caribbean, Middle East and North Africa, South Asia, and Sub-Saharan Africa. In order to include most of the countries in my sample, I add a dummy variable for Europe and another one for North America. The dummy for countries that belong in the same income group is constructed using the World Bank Country Classification - low-income, lower-middle-income, upper-middle-income, and high-income. The dummy for countries that share the same language is calculated by the author using information on each country's official language(s) from Ethnologue (2000). Finally, the information on the number of Internet users at the country level is from the International Telecommunication Union (2006). Table 3 presents summary statistics for the non-dummy variables. In our sample, around 39% of the country pairs share the same main language and 66% are located in the same world region, whereas 77% of them belong in the same income group. Finally, a remarkable 48% of the routes are between cities in Anglo-Saxon countries.

TABLE 3 HERE

## 6. Results

Figure 2 plots the unconditional correlation between distance across cities and bilateral Internet traffic. The correlation is strongly negative (-0.20) and significant at the 10% level, and it does not seem to be driven by any outlier. New York and Washington, DC are located at a relatively short distance from each other and have the highest flow of traffic in our sample. This flow is similar between London and New York, in spite of being much further apart from each other. The south-east of the figure displays cities located in extremely distant areas that tend to have a very low bilateral Internet traffic (see for instance the pairs Beijing-Los Angeles and Los Angeles-Singapore.)

FIGURE 2 HERE

Table 4 shows the OLS estimates of my empirical specification. The most remarkable finding for the purposes of this paper is that, across all specifications except the last one, geographical distance between cities (in logs) enters with a negative and statistically significant sign, indicating that cities that are located far apart from each other tend to

[18] Flight distance is chosen because it represents the shortest available way of travelling between most cities.

[19] The years at which the data is collected varies across countries, but it is safe to assume that these figures do not change dramatically in a six year span. The advantage of using this dataset is that it is readily comparable across countries.

[20]See http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/0,,pagePK:180619~theSitePK:136917,00.html

have a lower degree of bilateral Internet communication. This can be interpreted as strong evidence in favor of the hypothesis that face-to-face contacts (by assumption negatively correlated with physical distance) and Internet communication are complements. The effect of distance on Internet flows is large. My estimates suggest that a 10% increase in the distance between two cities reduces Internet traffic by about 47%. Interestingly, the magnitude and significance of this effect is not very much affected by the inclusion of other important determinants of Internet traffic in specifications 2-7.

The interpretation of the rest of the coefficients is often less clear, although it is worth reminding that the purpose of this paper is not to estimate a structural model that attempts to identify all the relevant determinants of Internet traffic between cities, but rather explore the relationship between Internet traffic between two cities and their physical distance. Moreover, the lack of precision in most of the coefficients is largely due to the low number of observations in my sample. The only statistically significant coefficients in my regressions are the sum of the logs of per capita GDP at the country level, which, enters negatively in specifications 4, 7 and 8. This is somewhat puzzling because, in principle, one would expect cities in richer countries to engage in more intense Internet traffic than in poorer ones.[21] As expected, in two specifications the fact that the two cities in a given route share the same language or are both Anglo-Saxon has a positive impact on Internet traffic, whereas being in the same income group has a surprising negative effect. Finally, the number of Internet users enters positively, although only significantly so in the last specification.

TABLE 4 HERE

As mentioned above, the structure of my dependent variable indicates that it may be appropriate to use two-sided clustered errors to account for spatial effects in the estimation. One problem with this technique is that the number of clusters in my sample – the 56 different cities - is too low in relation to the number of regressors. In order to estimate this model I therefore have to “partial out” some exogenous covariates from the regression. Failing to do this, the estimated covariance matrix of moment conditions is not of full rank and cannot be robustly estimated. In all cases I choose the set of city dummies (27 for the first city in every pair of cities with Internet traffic, and 29 for the second one) as the partialled out regressors, but the results are robust to other choices.[22] Table 5 displays the estimates. After accounting for potential spatial effects, the standard errors associated with the distance between cities sharply decreases, making it significant at the 1% level in all specifications. The size of the standard errors associated with the rest of controls also decrease in most cases. Cities’ and countries’ population has often a negative effect, whereas a country’s per capita GDP has now a positive impact on Internet exchanges. The number of Internet users still enters with a positive sign. The fact that one of the two cities in the route is a country’s capital is negatively related to Internet traffic, while in most cases the effect is positive if the two countries belong in the same income group or are both Anglo-Saxon. Being the largest city in the country is negatively associated with Internet traffic, although this is clearly due to the fact that in most countries the largest city is also the capital (note that in

[21] The log of the sum of the countries’ population is also significantly negative, although only in the last specification.
[22] Stata also includes the constant term in the partialled out set of regressors.

specification 5, the dummy for the capital turns insignificant). Interestingly, in the last column we see that the dummy for the most transited cities in the sample has a negative impact, although its interaction with distance is positive. This can be interpreted as suggesting that, although the relationship between Internet interaction and physical distance is negative, being one of the most transited cities partially reduces this effect.

TABLE 5 HERE

The main conclusion from Tables 4 and 5 is that, following Gaspar and Glaeser (1998) logic, the robust negative coefficient on distance can be interpreted in favor of the idea that Internet and face-to-face contacts are complements: after controlling for several relevant variables, Internet users tend to exchange more information through the web when they are nearby than when they are far away from each another.

One may argue that the previous regression would be less noisy using only cities that are located in the same country, because cultural and institutional differences among cities would then be wiped out. The obvious candidate, given the identity of the top routes listed in Table 2, would be the U.S. Table 6 presents the results of estimating the regressions using such data. The first two columns display the estimation using robust standard errors[23], while specifications 3-7 correct them to take into account spatial effects (i.e. using two-way clustered standard errors). The main finding of using this subsample is that the coefficient of distance is still negative and of similar size and significance as the one obtained using the full sample. This is quite remarkable given the fact that there are very few observations available to estimate these regressions. The rest of the variables have qualitatively similar results as in the regression with the full sample.

TABLE 6 HERE

A final robustness check I run is to test whether my results significantly change once I control for the intensity of bilateral trade between the two countries where the cities in a given route belong to.[24] The rationale for doing this is that Internet transit may just – or in large part- be reflecting the fact that two countries are very much related in some other dimension (trade, in this case). One problem with this strategy is that, due to my reduced sample size, the number of observations that can be used in the estimation drops to 48. Nevertheless, the negative coefficient on distance is maintained in all the standard regressions, although due to the reduced sample size, it is never significant (Table 7). However, as Table 8 shows, once I use the two-way clustered errors, the negative coefficient on distance is significant in all cases except in the specification that uses the interaction term between the most transited city and distance. In all cases, bilateral trade in goods enters positively but it is never significant at conventional levels. This robustness check confirms the validity of my finding and suggests that Internet flows between cities are not simply capturing trade in goods across countries.

[23] Adding more regressions results in estimates for which Stata cannot calculate the associated standard errors due to lack of degrees of freedom.
[24] I thank Scott Baier for sharing these data.

TABLE 7 HERE

TABLE 8 HERE

## 7. Conclusions

In a world where currently a huge fraction of interactions between humans takes place electronically, it is natural to ask the question of whether being located physically far away from other people matters at all for the intensity with which one communicates with them. From a theoretical point of view, Gaspar and Glaeser (1998) present a model in which a higher level of face-to-face interactions may either decrease or increase the demand for electronic communication, indicating that face-to-face interactions and electronic communications could in principle be substitutes or complements.

This paper uses a novel dataset on bilateral Internet traffic between cities to test these competing hypotheses and finds that, after controlling for several socioeconomic indicators, users located in cities that are closer to each other tend to have much larger flows of Internet traffic than those that are far apart. This result is very robust in a sample of international cities and it is maintained using only data from U.S cities, although the relationship is less precisely estimated. While it would be desirable to have access to more disaggregated data, my finding confirms some of the evidence provided in the literature, but, by using very different data and covering a large number of countries, it offers much more compelling evidence that Internet and face-to-face contacts are complements rather than substitutes.

## Appendix

TABLE 1 HERE

## References

Andrés, L., Cuberes, D. Diouf, D. Serebrisky, T., 2010. The diffusion of Internet: A cross-country analysis. Telecommunications Policy 34, 323-340.

Berthelon, M., Freund, C., 2008. On the conservation of distance in international trade. Journal of International Economics 75(2), 310-310.

Cairncross, F., 1997. The Death of Distance: How the Communications Revolution Will Change Our Lives, Cambridge, Harvard Business School Press.

Carrère, C., Schiff, M., 2005. On the geography of trade: Distance is alive and well. Revue Economique 56(6), 1249-1274.

Demographia, 2009. World Urban Areas & Population Projections: 5th Comprehensive Edition.

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

Doms, M., Dunne, T., Troske, K., 1997. Workers, wages, and technology. Quarterly Journal of Economics 112, 253-290.

Drucker, P., 1989. Information and the future of the city. Wall Street Journal. April 4.

Ethnologue, 2000. Languages of the World. www.sil.org/ethnologue/countries

Fischer, C., 1992. America Calling. Berkeley, CA: University of California Press.

Gaspar, J., Glaeser, E.L. 1998. Information technology and the future of cities. Journal of Urban Economics 43(1), 136-156.

Gates, B., Myhrvold, N., Rinearson, P., 1995. The Road Ahead. New York: Viking.

Gilder, G., 1995. Forbes ASAP. February 27. Quoted in Mitchell Moss, “Technology and Cities,” *Cityscape 3:3*.

Glaeser, E.L., 1998. Are Cities Dying? Journal of Economic Perspectives 12, 139-160.

Glaeser, E.L., 2000. The future of urban economics: Non-market interactions. *Brookings-Wharton Papers on Urban Affairs* 1, 101-150.

Glaeser, E.L., Mare, D., 1994. Cities and skills. *NBER Working Paper* 4728.

Graham, S., Marvin, S., 1996. Telecommunications and the City: Electronic Spaces, Urban Places. London: Routledge.

Greenstein, S., Prince, J. 2006. The diffusion of the Internet and the geography of the digital divide in the United States. NBER working paper 12182.

Greenstein, S., Lizardo, M., Spiller, P. 1997. The evolution of advanced large scale information infrastructure in the United States. NBER Working Paper 5929.

Hampton, K., Wellman, B. 2002. The not so global village of netville, in: B. Wellman and C. Haythornthwaite (Eds) The Internet in Everyday Life, 345-371. Oxford: Blackwell.

Hampton, K., Wellman, B. 2003. Neighboring on and offline in netville, the wired suburb, City and Community 1, 277-311.

Henderson, V.H, Mitra, A. 1996. The new urban landscape developers and edge cities. Regional Science and Urban Economics 26, 613-643.

Heston, A., Summers, R., Aten, B. 2009. Penn World Table Version 6.3, Center for International Comparison at the University of Pennsylvania (CICUP)

Imagawa, T., 1996. Telecommunciations and cities: Information from the prefectures. mimeo.

International Telecommunication Union, 2006. World telecommunicationindicators/ICTindicatorsdatabase. Geneva, Switzerland.

Ioannides, Y.M., Overman, H.G., Rossi-Hansberg, E., Schmidheiny, K., 2008. The effect of information and communication technologies on urban structure. Economic Policy 23(54), 201-242.

Jacks, D.S., 2009. On the death of distance and borders: Evidence from the nineteenth century. NBER working paper 15250.

Kolko, J., 2000a. Essays on information technology, cities, and location choice. *unpublished*.

Kolko, J., 1999. The death of cities? The death of distance? Evidence from the geography of commercial internet usage. Selected Papers from the Telecommunications Policy Research Conference.

Leamer, E.E., Storper, M. 2001. The economic geography of the Internet age. *working paper*.

Mok, D., Wellman, B., Carrasco, J. 2009. Does distance matter in the age of the Internet? Urban Studies 47(13), 2747-2783, November.

Moss, M., 1998. Technology and cities. Cityscape 3:3, 107-127.

Moyer, J.A., 1977. The telephone and the uses of time. in *The Social Impact of the Telephone* (I. de Sola Pool, Ed.), MIT, Cambridge.

National Research Council, 1998. Fostering Research on the Economic and Social Impacts of Information Technology. Draft. On-line at http://www2.nas.edu/cstbweb/5492.html

Pons-Novell, J. and Viladecans-Marsal, E., 2006. Cities and the Internet: The End of Distance? Journal of Urban Technology, 13(1), 109-131.

Rhee, H-J., 2008. Home-based telecommuting and commuting behavior. Journal of Urban Economics 63, 198-216.

Rossi-Hansberg, E., Wright, M., 2007. Urban structure and growth. *Review of Economic Studies* 74. April, 597-624.

Shklovski, I., Kraut, R., Cummings, J., 2008. Keeping in touch by technology, in: Proceedings of the Human Factors in Computer Science Conference. New York. ACM Press.

Spolaore, E., Wacziarg, R., 2009. The diffusion of development. Quarterly Journal of Economics 124(2), 469-529.

Telegeography, http://www.telegeography.com/

Travelmath, http://www.travelmath.com/flight-distance

Veenhof, B., Wellman, B., Quell, C., Hogan, B., 2008. Isolation, cohesion or transformation? How Canadians' use of the Internet is shaping society. Connectedness Report Series F0004M-16, Statistics Canada, December (http://www.statcan.gc.ca/pub/56f0004m/56f0004m2008016-eng.pdf).

Wellman, B., 1979. The community question. *American Journal of Sociology* 84, 1201-1231.

Wellman, B., Tindall, D., 1993. Reach out and touch some bodies. *Progress in Communication Science* 12, 63-94.

Zook, M.A., 2000. The web of production: the economic geography of commercial Internet content production in the United States. Environment and Planning A 32(3), 411-426.

Zook, M.A., 2001. Old hierarchies or new networks of centrality? The global geography of the Internet content market. American Behavioral Scientist 44(10), 1679-1696.

Figure 1: Internet traffic and city size

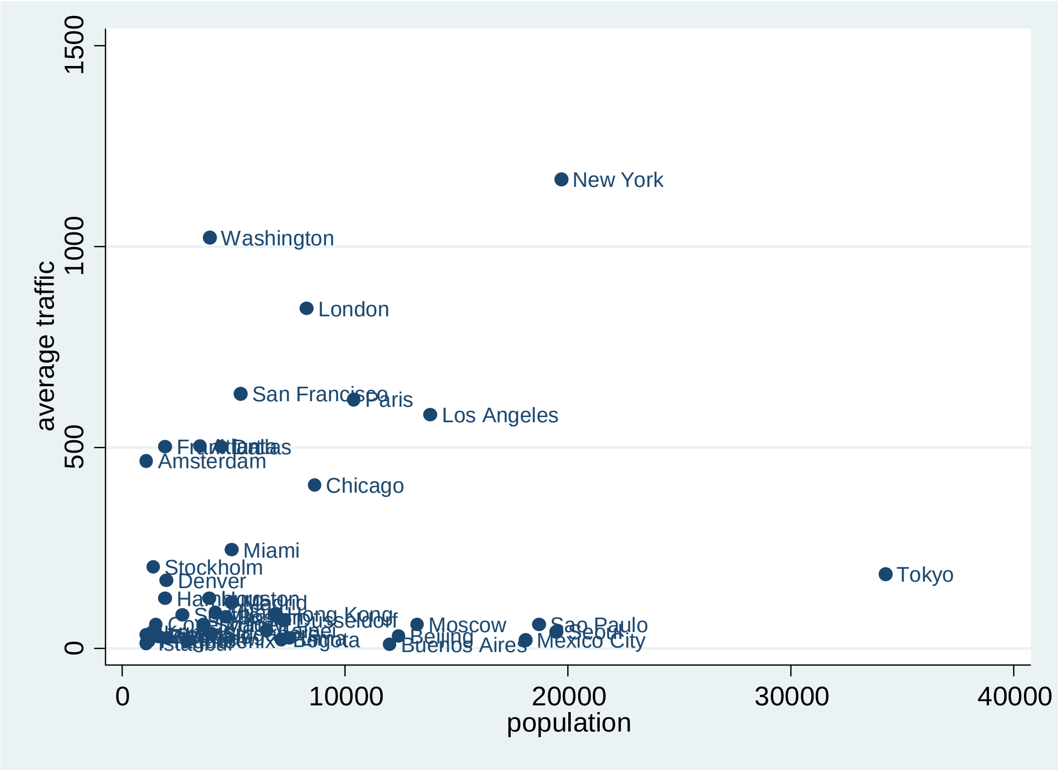

Figure 2: Bilateral Internet traffic and distance between cities

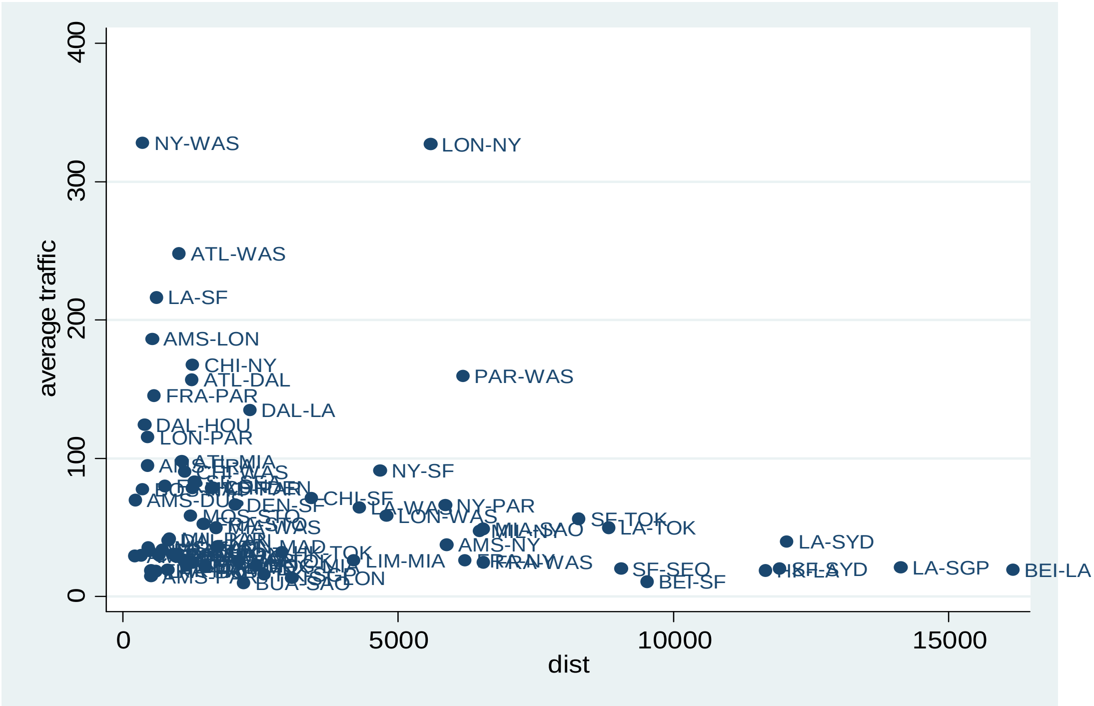

Table 1: Top Internet routes

| Route | Traffic | Distance | Route | Traffic | Distance |
|---|---|---|---|---|---|
| New York, U.S. - Washington, U.S. | 327.9 | 329 | Dallas, U.S. - Kansas City, U.S. | 40.2 | 729 |
| London, U.K. - New York, U.S. | 327.1 | 5586 | Los Angeles, U.S. - Sydney, Australia | 39.3 | 12063 |
| Atlanta, U.S. - Washington, U.S. | 248.2 | 872 | Amsterdam, Netherlands - New York, U.S. | 37.4 | 5881 |
| Los Angeles, U.S. - San Francisco, U.S. | 216.4 | 559 | London, U.K. - Madrid, Spain | 36.2 | 1264 |
| Amsterdam, Netherlands - London, U.K. | 186.2 | 359 | Amsterdam, Netherlands - Hamburg, Germany | 35.0 | 367 |
| Chicago, U.S. - New York, U.S. | 167.4 | 1149 | Frankfurt, Germany - Vienna, Austria | 33.7 | 599 |
| Paris, France - Washington, U.S. | 159.5 | 6180 | Helsinki, Finland - Stockholm, Sweden | 32.7 | 398 |
| Atlanta, U.S. - Dallas, U.S. | 156.7 | 1161 | Hong Kong, China - Tokyo, Japan | 31.7 | 2891 |
| Frankfurt, Germany - Paris, France | 145.5 | 480 | Hamburg, Germany - London, U.K. | 30.5 | 723 |
| Dallas, U.S. - Los Angeles, U.S. | 134.8 | 1996 | Copenhagen, Denmark - Stockholm, Sweden | 30.0 | 523 |
| Dallas, U.S. - Houston, U.S. | 124.2 | 362 | Copenhagen, Denmark - Hamburg, Germany | 29.6 | 289 |
| London, U.K. - Paris, France | 115.0 | 341 | Amsterdam, Netherlands - Brussels, Belgium | 29.3 | 174 |
| Atlanta, U.S. - Miami, U.S. | 97.7 | 972 | Hamburg, Germany - Stockholm, Sweden | 28.5 | 812 |
| Amsterdam, Netherlands - Frankfurt, Germany | 94.8 | 362 | Frankfurt, Germany - Warsaw, Poland | 26.1 | 892 |
| New York, U.S. - San Francisco, U.S. | 90.9 | 4139 | Frankfurt, Germany - New York, U.S. | 26.0 | 6219 |
| Chicago, U.S. - Washington, U.S. | 90.1 | 958 | Lima, Peru - Miami, U.S. | 25.8 | 4198 |
| San Francisco, U.S. - Seattle, U.S. | 82.7 | 1092 | Taipei, Taiwan - Tokyo, Japan | 25.4 | 2107 |
| Frankfurt, Germany - London, U.K. | 79.6 | 639 | Frankfurt, Germany - Washington, U.S. | 24.8 | 6548 |
| Chicago, U.S. - Denver, U.S. | 78.3 | 1479 | Dallas, U.S. - Denver, U.S. | 24.7 | 1065 |
| Madrid, Spain - Paris, France | 78.0 | 1054 | Bogotá, Colombia - Miami, U.S. | 22.6 | 2432 |
| Boston, U.S. - New York, U.S. | 77.2 | 306 | Seoul, Korea, Rep. - Tokyo, Japan | 22.1 | 1159 |
| Chicago, U.S. - San Francisco, U.S. | 70.7 | 2990 | Dallas, U.S. - Mexico City, Mexico | 21.3 | 1497 |
| Amsterdam, Netherlands - Düsseldorf, Germany | 69.6 | 179 | Los Angeles, U.S. - Singapore, Singapore | 20.9 | 14134 |
| Denver, U.S. - San Francisco, U.S. | 66.2 | 1529 | San Francisco, U.S. - Sydney, Australia | 20.1 | 11935 |
| New York, U.S. - Paris, France | 66.0 | 5851 | San Francisco, U.S. - Seoul, Korea, Rep. | 19.9 | 9048 |
| Los Angeles, U.S. - Washington, U.S. | 64.2 | 3700 | Hong Kong, China - Taipei, Taiwan | 19.3 | 812 |
| Moscow, Russia - Stockholm, Sweden | 58.3 | 1232 | Beijing, China - Los Angeles, U.S. | 19.3 | 10083 |
| London, U.K. - Washington, U.S. | 58.0 | 5915 | Frankfurt, Germany - Prague, Czech Republic | 18.6 | 414 |
| San Francisco, U.S. - Tokyo, Japan | 55.9 | 8286 | Hong Kong, China - Los Angeles, U.S. | 18.6 | 11674 |
| Frankfurt, Germany - Milan, Italy | 52.4 | 1188 | Los Angeles, U.S. - Phoenix, U.S. | 18.2 | 575 |
| Miami, U.S. - Washington, U.S. | 49.9 | 1485 | Hong Kong, China - Singapore, Singapore | 15.8 | 2569 |
| Los Angeles, U.S. - Tokyo, Japan | 49.6 | 8830 | Amsterdam, Netherlands - Paris, France | 14.3 | 428 |
| Miami, U.S. - São Paulo, Brazil | 48.6 | 6546 | Istanbul, Turkey - London, U.K. | 13.0 | 2504 |
| Milan, Italy - New York, U.S. | 47.0 | 6482 | Beijing, China - San Francisco, U.S. | 10.4 | 9525 |
| Milan, Italy - Paris, France | 41.6 | 643 | Buenos Aires, Argentina - São Paulo, Brazil | 9.8 | 1676 |

*Source: TeleGeography in mid-2012*

Table 2: Internet traffic by application

| Application | Percentage of traffic |
|---|---|
| Web browsing | 25% |
| Streamed & Buffered Audio + Video | 27% |
| Peer-to-peer networking (P2P) | 14% |
| Online File Storage | 9% |
| Voice over Internet Protocol (VoIP) + Instant messaging (IM) + Video Calling | 5% |
| Gaming | 4% |
| Internet Protocol virtual private network (VPN) | 4% |
| Email | 3% |
| Other | 9% |

*Source: TeleGeography in mid-2012. Data based on the average shares by application*

Table 3: Summary statistics

| | Unit | Mean | St dev | Min | Max |
|---|---|---|---|---|---|
| Average traffic | Gigabits per second | 67.8 | 67.9 | 9.8 | 327.9 |
| Distance | Kilometres | 3049.1 | 3530.5 | 174 | 14134 |
| Sum of log of city's population | | 30.9 | 1.5 | 28.1 | 34.1 |
| Sum of log of country's population | | 23.2 | 2.1 | 17.7 | 26.7 |
| Sum of log of country's per capita GDP | | 21 | 0.7 | 18.6 | 21.7 |
| Sum of log of country's Internet users | | 35.1 | 2.2 | 30.2 | 37.7 |

Table 4: The determinants of bilateral Internet traffic. OLS regressions.

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] |
|---|---|---|---|---|---|---|---|---|
| Log distance | -0.5*** | -0.46*** | -0.46** | -0.46** | -0.52** | -0.46** | -0.41* | -1.02 |
| | (0.14) | (0.15) | (0.2) | (0.2) | (0.2) | (0.2) | (0.2) | (0.76) |
| Log population cities | | -0.11 | 0.15 | 0.25 | -0.66 | 0.08 | 0.21 | 0.12 |
| | | (0.25) | (0.4) | (0.21) | (0.51) | (0.63) | (0.2) | (0.53) |
| Capital dummy | | -0.79 | -0.82 | -0.82 | -0.49 | -0.82 | -1.02 | -0.49 |
| | | (0.54) | (0.57) | (0.57) | (0.59) | (0.57) | (0.59) | (0.73) |
| Log population country | | -0.01 | 0.17 | -0.3 | -1.36 | 0.91 | -0.5 | -2.05** |
| | | (0.19) | (0.16) | (0.71) | (1.04) | (0.73) | (0.73) | (0.79) |
| Log GDP per capita country | | -0.59 | 0.86 | -2.22** | -2.08 | 2.04 | -2.33** | -2.53* |
| | | (0.5) | (0.82) | (0.78) | (1.25) | (1.2) | (0.84) | (1.28) |
| Same language | | | -0.66 | 1.97* | 1.32 | -0.2 | 1.71** | 2.43 |
| | | | (1.68) | (0.93) | (0.97) | (0.78) | (0.66) | (1.44) |
| Same region | | | -0.13 | -0.13 | -0.15 | -0.13 | -0.05 | -0.46 |
| | | | (0.5) | (0.5) | (0.52) | (0.5) | (0.43) | (0.7) |
| Same income group | | | -0.05 | 0.59 | -3.47** | 0.24 | 0.11 | -2.95** |
| | | | (0.99) | (1.21) | (1.44) | (0.41) | (1.03) | (1.24) |
| Anglo-Saxon country | | | -1.2 | 1.43* | 0.51 | -0.74 | 1.67** | 2.17 |
| | | | (1.21) | (0.73) | (1.21) | (1.42) | (0.56) | (1.46) |
| Log Internet users country | | | | 0.16 | 1.4 | -0.33 | 0.32 | 1.91** |
| | | | | (0.64) | (0.89) | (0.7) | (0.7) | (0.73) |
| Largest city in the country | | | | | -0.93 | | | |
| | | | | | (0.84) | | | |
| Largest cities in the sample | | | | | | 0.04 | | |
| | | | | | | (1.82) | | |
| Most transited cities in the sample | | | | | | | -0.75 | -3.8 |
| | | | | | | | (1.59) | (4.6) |
| Most transited cities in the sample*Log distance | | | | | | | | 0.51 |
| | | | | | | | | (0.58) |
| Constant | 4.36*** | 22.00 | -18.22 | 45.21** | 56.93 | -45.42 | 47.47** | 41.03 |
| | (0.92) | (21.02) | (28.29) | (15.43) | (34.35) | (34.41) | (15.99) | (33.01) |
| Number of observations | 70 | 70 | 70 | 70 | 70 | 70 | 70 | 70 |
| $R^2$ | 0.91 | 0.91 | 0.92 | 0.92 | 0.93 | 0.92 | 0.92 | 0.92 |

*Note: Robust standard errors in parentheses.*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The dependent variable is the log of Internet traffic in 2008.*

Table 5: The determinants of bilateral Internet traffic. Two-sided clustered errors.

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] |
|---|---|---|---|---|---|---|---|---|
| Log distance | -0.5*** | -0.46*** | -0.46*** | -0.46*** | -0.52*** | -0.46*** | -0.41*** | -1.02*** |
| | (0.06) | (0.06) | (0.1) | (0.1) | (0.1) | (0.1) | (0.11) | (0.36) |
| Log population cities | | 0.009 | 2.97 | 1.48 | -4.99*** | 0.72 | 4.38** | -3.54*** |
| | | (0.34) | (5.59) | (1.47) | (0.93) | (0.57) | (2.13) | (1.31) |
| Capital dummy | | -0.79** | -0.82*** | -0.82*** | -0.49 | -0.82*** | -1.02*** | -0.49 |
| | | (0.32) | (0.31) | (0.31) | (0.33) | (0.31) | (0.34) | (0.45) |
| Log population country | | 0.17* | -2.34 | -4.21 | 6.61*** | 1.18 | -7.43** | -8.5*** |
| | | (0.09) | (5.28) | (4.93) | (1.99) | (0.74) | (3.39) | (2.91) |
| Log GDP per capita country | | 0.37 | 5.02 | 2.61** | -40.74*** | 2.46 | 4.88*** | 7.87*** |
| | | (0.31) | (6.54) | (1.12) | (7.39) | (1.94) | (1.87) | (2.26) |
| Same language | | | 0.54 | 1.86 | 14.17*** | 0.86 | 5.92** | -9.7*** |
| | | | (1.35) | (2.51) | (2.84) | (0.6) | (3.01) | (2.76) |
| Same region | | | -0.13 | -0.13 | -0.15 | -0.13 | -0.05 | -0.46 |
| | | | (0.27) | (0.27) | (0.26) | (0.27) | (0.27) | (0.35) |
| Same income group | | | -8.998 | -8.49 | 12.98*** | -0.68 | -20.33** | 4.15** |
| | | | (14.37) | (8.67) | (1.77) | (1.26) | (9.06) | (1.99) |
| Anglo-Saxon country | | | 0.007 | 1.32 | 13.36*** | 0.32 | 5.88* | -9.96*** |
| | | | (1.35) | (2.64) | (2.8) | (0.63) | (3.32) | (2.69) |
| Log Internet users country | | | | 4.22 | -9.01*** | -1.24 | 7.6** | 8.17*** |
| | | | | (4.78) | (2.18) | (0.86) | (3.39) | (2.8) |
| Largest city in the country | | | | | -0.94*** | | | |
| | | | | | (0.36) | | | |
| Largest cities in the sample | | | | | | 1.1 | | |
| | | | | | | (0.69) | | |
| Most transited cities in the sample | | | | | | | -0.75 | -3.8** |
| | | | | | | | (0.55) | (1.78) |
| Most transited cities in the sample*Log distance | | | | | | | | 0.51* |
| | | | | | | | | (0.28) |
| Number of observations | 70 | 70 | 70 | 70 | 70 | 70 | 70 | 70 |
| centered $R^2$ | 0.49 | 0.54 | 0.6 | 0.63 | 0.71 | 0.62 | 0.66 | 0.69 |

*Note: Two-way clustered errors in parentheses.*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The dependent variable is the log of Internet traffic in 2008.*

Table 6: The determinants of bilateral Internet traffic in U.S. routes only.

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
|---|---|---|---|---|---|---|---|
| Log distance | -0.55* | -0.55* | -0.55*** | -0.48*** | -0.55*** | -0.48*** | -2.01*** |
| | (0.08) | (0.08) | (0.01) | (0.000) | (0.01) | (0.000) | (0.000) |
| Log population cities | | 0.13** | | 4.07*** | 4.11*** | 4.07*** | 3.99*** |
| | | (0.004) | | (0.000) | (0.07) | (0.000) | (0.000) |
| Capital dummy | | | | -0.37*** | | | |
| | | | | (0.000) | | | |
| Largest city | | | | | -4.97*** | | |
| | | | | | (0.11) | | |
| Most transited cities | | | | | | -0.37*** | -30.47*** |
| | | | | | | (0.000) | (0.000) |
| Most transited cities*Log distance | | | | | | | 3.43*** |
| | | | | | | | (0.000) |
| constant | 7.9** | 4.64* | | | | | |
| | (0.4) | (0.56) | | | | | |
| Number of observations | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| Clustered errors | No | No | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.99 | 0.99 | | | | | |
| Centered-$R^2$ | | | 0.99 | 0.99 | 0.99 | 0.999 | 0.999 |

*Note: Columns (1) and (2) display standard robust standard errors in parenthesis and do not include city dummies. Columns (3) and (4) display two-way clustered errors in parentheses and include city dummies.*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The dependent variable in all regressions is the log of Internet traffic in 2008.*

Table 7: The determinants of bilateral Internet traffic. The effect of trade. OLS regressions.

| | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| Log distance | -2.26 | -2.28 | -2.19 | -0.37 |
| | (2.4) | (2.08) | (2.28) | (2.54) |
| Log population cities | -0.09 | -0.81 | -0.14 | -0.1 |
| | (0.61) | (0.63) | (0.35) | (0.28) |
| Log population country | 1.56 | 0.44 | 1.65 | 2.86 |
| | (2.28) | (0.8) | (2.19) | (2.56) |
| Log GDP per capita country | 1.36 | 0.35 | 2.38 | 5.37 |
| | (4.08) | (1.22) | (3.77) | (5.14) |
| Same language | 2.63 | -1.31 | -0.27 | -2.73 |
| | (1.94) | (3.36) | (2.00) | (3.41) |
| Same region | 0.27 | -1.99 | -0.15 | 0.8 |
| | (0.94) | (3.84) | (0.81) | (0.77) |
| Anglo-Saxon country | 1.1 | -2.63 | -1.21 | -2.78 |
| | (1.52) | (4.46) | (3.23) | (3.67) |
| Log Internet users country | -0.33 | 0.93 | -0.23 | -2.04 |
| | (1.52) | (1.18) | (0.93) | (2.23) |
| Largest city in the country | -0.85 | | | |
| | (1.44) | | | |
| Log of total bilateral trade | 0.04 | 0.06 | 0.03 | 0.33 |
| | (0.81) | (0.73) | (0.84) | (0.92) |
| Largest cities in the sample | | 0.66 | | |
| | | (1.24) | | |
| Most transited cities in the sample | | | -0.55 | 16.29 |
| | | | (1.66) | (16.44) |
| Most transited cities in the sample*Log distance | | | | -2.87 |
| | | | | (2.66) |
| Constant | -28.76 | -45.42 | -53.47 | -97.24 |
| | (85.39) | (34.41) | (83.86) | (102.65) |
| Number of observations | 48 | 48 | 48 | 48 |
| $R^2$ | 0.9 | 0.89 | 0.9 | 0.92 |

*Note: Two-way clustered errors in parentheses. The dependent variable in all regressions is the log of Internet traffic in 2008.*

Table 8: The determinants of bilateral Internet traffic. The effect of trade. Two-way-clustered errors.

| | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| Log distance | -2.26*** | -2.28*** | -2.19*** | -0.37 |
| | (0.48) | (0.5) | (0.5) | (0.6) |
| Log population cities | 7.33*** | 0.97*** | -52.44 | -9.79*** |
| | (1.67) | (0.38) | (33.94) | (2.39) |
| Capital | 3.77* | -7.13 | -22.27 | 11.52*** |
| | (1.96) | (6.61) | (13.81) | (2.26) |
| Log population country | -5.86*** | 3.11*** | 4.31** | -3.11*** |
| | (1.77) | (0.56) | (1.99) | (0.75) |
| Log GDP per capita country | 11.6*** | 1.13 | -96.33 | -22.1*** |
| | (2.98) | (2.84) | (62.27) | (6.46) |
| Same language | -4.85*** | 1.35*** | 12.21* | 7.29*** |
| | (1.68) | (0.41) | (7.07) | (2.22) |
| Same region | -12.03*** | 0.08 | 94.44 | 14.97*** |
| | (2.79) | (2.07) | (60.71) | (4.73) |
| Anglo-Saxon country | -6.37*** | 0.03 | 11.26* | 7.24*** |
| | (1.88) | (0.59) | (6.82) | (2.55) |
| Log Internet users country | 1.86** | -2.64*** | 37.46 | 13.27*** |
| | (0.76) | (0.51) | (24.29) | (2.99) |
| Largest city in the country | -0.85** | | | |
| | (0.42) | | | |
| Log of total bilateral trade | 0.04 | 0.06 | 0.03 | 0.33 |
| | (0.21) | (0.22) | (0.22) | (0.2) |
| Largest cities in the sample | | -0.61 | | |
| | | (1.85) | | |
| Most transited cities in the sample | | | -0.55 | 16.29*** |
| | | | (0.6) | (3.91) |
| Most transited cities in the sample*Log distance | | | | -2.87*** |
| | | | | (0.66) |
| Number of observations | 48 | 48 | 48 | 48 |
| Centered-$R^2$ | 0.75 | 0.79 | 0.73 | 0.71 |

*Note: Two-way clustered errors in parentheses.*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. The dependent variable in all regressions is the log of Internet traffic in 2008.*