

Does the Internet Defy the Law of Gravity?*

Bernardo S. Blum Avi Goldfarb[†]
University of Toronto University of Toronto

August 2005

Abstract

We show that gravity holds in the case of digital goods that are consumed over the Internet and have no trading costs. Therefore, trade costs, although possibly important, cannot account for the effects of distance on trade. In particular, we show that Americans are more likely to visit websites from countries that are physically close than from countries that are far, even after controlling for country-level Internet expertise, language, income, immigrant stock, and many other factors. Furthermore, we show that this effect only holds for digital products that depend on taste, such as music, games, and pornography. For these, a 1% increase in physical distance reduces the probability an American will visit the website by 3.25%. For less taste-dependent products, such as software, distance has no statistical effect.

A more polished version is forthcoming in the *Journal of International Economics*.

*We are grateful to Shane Greenstein, Juan Carlos Hallak, Walid Hejazi, Keith Head, Ig Horstmann, Edward Leamer, Joanne Oxley, Dan Treffer, the editor, and two anonymous referees for helpful comments. Zeynab Ziaie Moayyed provided excellent research assistance. Research support was provided by SSHRC Grant # 538-02-1013. All opinions and errors are our own.

[†]Please address correspondence to bblum@rotman.utoronto.ca and agoldfarb@rotman.utoronto.ca.

1 Introduction

One of the clearest and most robust findings in economics is that international trade is subject to “gravity”: a country trades more with countries that are large and nearby than with countries that are small and far away. There is a broad consensus that the distance effect in gravity proxies for distance-related trade costs, although the sources of these trade costs remain an open question.

In this paper, we analyze data on Internet activities by US consumers on non-US websites and show that trade costs, although important, cannot fully account for the effects of distance on trade. In particular, we show that even though transportation, time, and distribution costs are near zero over the Internet and online search costs do not depend on distance, trade in purely digital goods is significantly reduced by physical distance. This effect is also not legal or financial, since it holds even when no money changes hands. Furthermore, distance matters even when controlling for immigrant stock (a “phoning home” effect), language, income levels, and country-level Internet sophistication. The typical explanations for gravity clearly do not apply in this context.

Our study employs data on 2,654 US Internet users. We distinguish in the data between two types of digital products. For both, consumption is realized entirely over the Internet (i.e., there is no product to be shipped). In the first, there is a financial transaction (e.g., software). The second is free (e.g., free online games).¹ We find that for both categories a country’s physical distance from the US reduces the number of visits by US households to websites from that country. A 1% increase in physical distance from the US reduces the number of visits by US consumers to a website that involves a financial transaction by 2.7%. For websites without a financial transaction, the equivalent effect is 1.1%. Given that consumption of these products/services is free of all distance-related trade costs, we conclude that trade costs cannot account for the entire distance effect found in previous gravity studies.

¹There is a third type of website where a product is purchased online and physically shipped to the consumer. We do not look at these websites since distance should matter due to shipping costs.

Interestingly, we find that not all types of digital products have their demands reduced by distance. In particular, we find that websites of taste-dependent products such as music, games, and pornography suffer the effects of distance, while the websites of products like software and financial information do not. We interpret this to mean that taste-dependent differentiated products are affected by distance while more homogeneous products are not. Even after we control for key determinants like language, number of immigrants, and Internet penetration, US Internet users access more taste-dependent websites of companies from countries that are physically closer to the US than from countries farther away. For such products and services, a 1% increase in physical distance reduces the number of visits an American will make to the website by 3.25%. For less differentiated products and services, distance has no statistical effect on demand.

We therefore argue that distance proxies for *taste*. Distance appears to capture taste similarities, and countries located closer to each other either tend to have more similar tastes or are more able to cater to each other's consumer tastes. North American music, for example, is very different from Indian music. Similar stories apply to games and pornography.

The literature on empirically identifying the sources of the distance effect in gravity is small.² The papers most closely related to ours are Rauch (1996, 1999). In these papers, the gravity effect is estimated for homogeneous and differentiated products separately in an attempt to link gravity with search costs. It is found that distance has a smaller effect on the trade of homogeneous goods, which the author argues have lower search costs. It is then concluded that search costs account for part of gravity. We show that distance also proxies for taste. Given that the demand for differentiated products should be more taste dependant than the demand for homogeneous goods, Rauch (1996, 1999) might be partly attributing to search costs the effects of differences in preferences. In a series of papers, Hummels (1999, 2001a, 2001b) measures and estimates a variety of trade costs and their relation

²See Harrigan (2003) and Anderson and Wincoop (2004) for surveys of trade costs and their relation to physical distance. There is a larger literature on the income effect of gravity. Harrigan (2003) also surveys this literature, and Feenstra et al (2001) study the implications of differences in the income effect.

to distance. However, none of these costs apply to the products/services we study in this paper. In a related paper, Leamer and Storper (2001) argue that, like other innovations that reduce communication and trade costs, the Internet pushes for agglomeration and deagglomeration at the same time, and the dominant impact is yet to be determined. Our results point to a new force working in favor of agglomeration: taste similarity. Finally, Disdier and Head (2004) run a Meta-Analysis on 1052 estimates of the distance effect on trade. They show that in other studies using methods, controls, units of observation, and sample periods similar to ours, the average elasticity of trade in physical goods with respect to distance varies from 1.08% to 1.24%. In our sample, we find that the comparable OLS estimate of the distance effect amounts to 1.12%. Therefore, digital products are as subject to gravity as the average physical product.³

Cairncross (1997), Bakos (1997), and Shapiro and Varian (1999) respectively argue that the Internet reduces communication, search, and transportation costs to near zero. Goolsbee (2000) and Stevenson (2003) provide domestic US evidence that the Internet reduces the importance of distance. Freund and Weinhold (2004) provide evidence that the Internet increases trade (presumably by lowering communications costs). All these papers suggest that distance should not matter for goods that are distributed digitally. We show that to be the case for digital services that are not taste dependent, but not for taste-dependent categories.

The remainder of the paper is organized as follows. The next section develops a discrete choice model for web surfing that will be used later as a guide to the empirical analysis. Section 3 describes the data, and section 4 discusses the empirical strategy and the results. Section 4.1 shows distance matters online, section 4.2 investigates the role of taste in explaining the distance effect, and section 4.3 extensively assesses the robustness of the

³For websites that do not involve a financial transaction, the distance effect is smaller than the average of the studies in Disdier and Head (2004), with the distance elasticity equal to 0.9%. For websites that involve a financial transaction, the distance elasticity is 1.8%, therefore larger than the average effect in the studies cited above. For taste-dependent websites, the distance elasticity is equal to 3.25%. This effect is larger than the averages found in the literature for trade in physical goods and larger than the distance effect found in studies where products were disaggregated, such as in Rauch (1999) where goods were classified as “Organized Exchange Goods,” “Reference Priced Goods,” and “Differentiated Commodities.”

results, using random censoring tobits and Heckman selection models. Section 5 concludes.

2 A discrete choice model for web surfing

In this section, we develop a discrete choice demand model of website choice based on Berry (1994) and Anderson et al (1992). The demand for websites by a US household is derived from utility maximization. Household-level decisions are then aggregated to the market level, enabling analysis of market-level data. In this framework, the physical distance between the country where the website is located and the US will matter only if it affects a household’s utility of browsing the website.

Suppose the utility that a household i gets from visiting a website from country c in service category s can be represented as:

$$U_{ic}^s = X_c^s \beta + \epsilon_c^s + v_i^s$$

where X_c^s is a vector of observed country and website characteristics, including the distance between country c and the US. Let ϵ_c^s represent unobserved country/website characteristics, assumed to be distributed $N(0, \sigma^2)$, and let v_i^s represent household unobserved characteristics. We assume that v_i^s is type two extreme value distributed, which implies that the market share of country c in website category s is given by:

$$S_c^s = n_c^s \frac{e^{X_c^s \beta + \epsilon_c^s}}{1 + \sum_w e^{X_w^s \beta + \epsilon_w^s}}$$

where n_c^s is the number of websites from country c in service category s .

The parameters of the model are only identified up to scale, and therefore it is necessary to define an “outside” choice for households. Since US users who visit foreign websites always have the option of visiting a US website, we let the choice of visiting an American

website offering the same service category s be the “outside” choice ($c = 0$). In this way, the *beta* parameters measure how country and website characteristics affect the marginal utility of visiting a category s website in country c relative to the marginal utility of visiting a category s website in the US.

Applying logarithms, the difference in market shares between country c ’s websites and US websites in service category s is given by a linear function of both countries’ and websites’ characteristics:

$$\log(S_c^s) - \log(S_0^s) = \log(n_c^s) - \log(n_0^s) + X_c^s \beta^s + \epsilon_c^s \quad (1)$$

With some algebraic manipulation, it can be shown that equation 1 is equivalent to the standard gravity equation, except for a rescaling of the constants. Notice first that the market share of country c in website category s can be written as the ratio of the number of visits to country c websites in category s and the total number of visits to category s websites in the world (i.e., $S_c^s = \frac{V_c^s}{V^s}$ and $S_0^s = \frac{V_0^s}{V^s}$). Second, because V_0^s and n_0^s do not vary by country within a service category s , they can be collapsed into a fixed effect variable (k^s). Equation 1 above can then be re-written as:

$$\log(V_c^s) = k^s + \log(n_c^s) + X_c^s \beta^s + \epsilon_c^s \quad (2)$$

where V_c^s is the number of visits the American consumers in our sample make to websites in category s in country c , and the vector X_c^s includes country- and website-specific variables that might affect a household’s utility when visiting a website. Some of the characteristics included in X_c^s are the language spoken in the country and the quality of the Internet connectivity. The usual reasons given for gravity suggest that distance should not matter; however, we include it to see if the usual reasons are complete explanations for gravity in

our context.

Equation 2 is equivalent to the traditional gravity equation in the sense that it relates the amount of an international transaction (in this case, visits to a foreign website) to countries' characteristics. Without loss of generality, in the empirical work we define X_c^s to be the logarithm of the countries' and websites' characteristics. In this case, the β parameters are not only utility function parameters but also measure the elasticity of the number of visits to a foreign website with respect to the countries' and websites' characteristics.⁴ Of particular interest will be the elasticity with respect to the physical distance between the country where the website is hosted and the US.

3 Data

3.1 Internet data

The main data, provided by Plurimus Corporation (which no longer operates), consist of every website visited by 2,654 American households from December 27, 1999 to March 31, 2000. We drop all websites where something needs be shipped because transportation costs are not zero in these cases. In all, the data consists of 2,927,213 different website visits. The raw data are in clickstream format, meaning that the web address and exact time of each website visit by each household can be identified. Furthermore, Plurimus categorized each website into one of 25 online service categories such as Email, Games, Music, and Software. Table 1 gives a complete list of the categories.

Portal/Search is the most commonly used category, although Pornography is by far the most common category of non-US websites. Since Vertical Portals are portals that are aimed to a particular group (e.g., gay.com, Chinese community site sina.com, etc.), they are commonly visited foreign websites largely by definition. The Internet category is an agglomeration of Internet-related services related to website design and hosting.

In the context of this paper, the data set has one main limitation. The geographic

⁴This makes our results readily comparable with the rest of the "Gravity Equation" literature.

distribution of the core data set is not representative. New York, Chicago, and Los Angeles are under-represented. Roughly half the sample comes from the Pittsburgh area. Another quarter is from North Carolina, another eighth from Tampa, and the remainder are from Texas, Ohio, and California.⁵

The geographic distribution of our sample prevents us from using US national immigration statistics as controls. Instead, we use US state-level statistics weighted by the states' representativeness in our sample. Given our controls, the geographic distribution of our sample has no obvious bias on the distance results. Furthermore, these potential biases will have no effect on the core finding of the paper: systematic differences occur across website category types.

An important feature of the data relates to download speed and distance. All users in the data are dial-up users. The bottleneck in downloading websites therefore occurs mostly in the last mile to a user's home. There also may be a bottleneck in the last mile from the backbone to the server location. Outside of the last mile in each direction, data travels quickly in the backbone wires, and the distance effect would be no more than fractions of a second. Given the speed of dial-up service, it is unlikely that any measured distance effect proxies for download speed.

While the raw data are at the household level, we aggregate to the country-service category level for three reasons. First, our explanatory variables of interest do not vary at the individual level. Second, a discrete choice household-level analysis of 2.9 million observations with 46 choices in 25 categories is computationally intractable, and simplifications would be difficult to interpret. Third, the data do not contain user characteristics other than the city where they live. This user anonymity is the cost of having such a rich dataset containing sensitive information, such as the number of visits to pornography websites. Unfortunately, the lack of user characteristics prevents us from exploring the relationship between user characteristics and browsing behavior.

⁵A geographically representative Plurimus data set of 421 households from March 1 to March 31, 2001 shows similar market shares for foreign firms. We do not use this data set because it only has 5% as many website visits. Therefore, there are not enough visits to foreign websites to be reliable.

Three characteristics of the dataset raise some concerns about its usefulness for studying the effects of distance on the demand for different websites. First, “Pornography” is by far the most commonly visited website category in foreign countries. Specifically, we worry that a few users might be behind most of the visits in this category and that the results we find are driven by this category. We find, however, that not to be the case. Of the 2,654 users in our sample, 1,970 accessed at least one pornography site in the three months covered in our sample. More impressively, 1,392 users (more 50% of the sample) accessed at least ten pornography sites, and 868 visited such sites at least 50 times in the same period. Visits to pornography sites are widespread in our sample. It is true that there are heavy users of pornography websites (as well as heavy users of every other type of website). The top 20 pornography visitors were responsible for 77,712 visits out of a total of 444,776 visits to pornography sites; however, the results in the paper hold when pornography websites are excluded from the sample. Pornography sites and pornography users are not driving the results of the paper.

The second concern is that search engines might spuriously cause a distance effect if distance from searcher plays any role in ordering sites. We do not believe this to be the case for three reasons. First, industry studies (e.g., Sullivan, 1999) state that search engines at the time our data were collected used two main methods to rank websites in search results. One method was keyword matches. Websites that used the search terms more frequently, or used them in their titles, would be ranked higher. There is no reason to believe close countries would do this any more frequently than far away countries if we control for language, income, and Internet hosts. The other main method for ranking websites at the time was by examining previous users’ clickthrough behavior at the search engine. Since this is based on user behavior in the first place, it is not an explicit search engine bias toward close country websites.⁶ Second, the databases of the leading search engines contained hundreds of millions of websites, and websites from particular countries

⁶ A third method of ranking, used mainly by Google, ranked pages by the number of other websites linking to the website. Such a method should not bias our result either. At the time, Google had only a 0.6% share of the search engine market.

were not systematically missing from the databases (Lawrence and Giles, 1998). A third and final reason why we do not believe search engines pose any problem to our analysis is the cross-category variation we find in the data. To be consistent with our results, search engines would need to rank close websites higher than far away websites for taste-related categories but not for other categories. While we cannot completely reject this possibility, we believe it is unlikely.

A final concern regarding our dataset is that some websites classified as "no purchase" might direct users to sites where there are purchases involved. For example, one could go to a free game-playing site but then be offered through advertisements or links opportunities to purchase the game on CD. This could spuriously cause a distance effect in the data. Because a primary commercial reason for compiling this data was to distinguish between purchase and non-purchase sites, we are confident that the ones classified as "no purchase" indeed do not have purchases involved. However, it is possible that some of the no purchase sites contain advertisements (with links) of sites where the user might be able to physically buy products. In this case, the visit to this second link is recorded in our data as a new visit to a purchase website. It is possible that the first non-purchase site identifies the country from where the user is accessing the Internet (the US in this case) and selectively advertises sites from countries close to the US but not in the US. In this case, any distance effect we may find would be spuriously generated by such a selective algorithm. However, our data allow us to rule out this possibility because the data include the sequence of sites visited by each user. Therefore, we identify how many times a user visits a purchase-free website and immediately after visits a purchase website from a country located closer to the US than the first purchase-free site. This happens only nine times out of the 2,927,213 visits (85,226 foreign visits) in the data, making us confident that such a possibility is not driving our results.

3.2 Country data

The main data are supplemented with a number of complementary data sets. Our research assistant identified the country of origin of each of the roughly 9,000 websites in the click-stream data set. First, country suffixes (like .de and .uk) were taken to identify the country of origin. For generic suffixes like .com, .net, and .org, the country of origin was identified by the location of the company's head office. This is typically found through the "About Us" link on the webpage. The 46 countries, plus the United States, are listed in table 2. In our main specification, we do not look at countries with no visits at all in the data. The zero market share truncation therefore only applies to countries with at least one visit to a website by a member of the dataset. In unreported results, we find the missing countries are generally far away from the US and their inclusion would reinforce our results.⁷

A large number of country-specific characteristics were also compiled from a variety of sources, including the CIA World Factbook, the US Census, the International Telecommunications Union, the PRS Group, and the US Department of Transportation.

The distance between the various countries and the United States is defined as the distance in kilometers between capital cities. This data is standard in the trade literature and is available online from multiple sources.⁸ Data on GDP and GDP per capita come from the CIA World Factbook. Statistics from 2003 are used.

Data on the number of Internet hosts and the number of Internet users for 2000 come from the International Telecommunications Union (www.itu.int). Political stability and rule of law measures for 2000 come from the PRS Group's International Country Risk Guide.⁹ From the 2000 US census (www.census.gov), we use data on immigrant population in each US state by country of origin. Data on the number of American travelers abroad come from the US Department of Transportation's 1995 travel survey. Appendix table A1 provides descriptive statistics.

⁷Unreported results are available in our online appendix at <http://www.rotman.utoronto.ca/agoldfarb>.

⁸See, e.g., the webpage maintained by Jon Haveman at <http://www.eiit.org>.

⁹We thank Joanne Oxley and Walid Hejazi for providing us with these data sets. Oxley and Yeung (2001) show that rule of law is an important predictor of Internet usage across countries.

4 Empirical specifications and results

We estimate the empirical counterpart of equation 2 using the data discussed in the previous section. Following Freund and Weinhold (2004), we use hosts as a proxy for the number of websites in a country (n_c^s). To ensure robustness, we also include alternative proxies for n_c^s : the number of Internet users and the country's GDP.

There is one more estimation detail that needs explanation. Many countries' websites are never visited in particular categories. For example, there are no visits to Financial Information websites based in Belize or Chat websites in Australia. The typical treatment for these cases in the discrete choice literature is to assume that the product is unavailable in these countries. This, however, is not consistent with our data, as there exist websites in almost all country-service category combinations. Since our data is a sample of the US population, websites with very small market shares are unlikely to appear in the data. Therefore, we assert that the data are truncated at a very small number, and so we use censored regression models to analyze the data. We assume that the share data is truncated at the inverse of the number of observations in a service. For example, there are 61,348 observations in the "Chat" service. Australia, with zero visits in Chat, is assumed to have its number of visits truncated at $1/61,348$. To our knowledge, the censored regression model has not previously been used in the discrete choice context. All qualitative results are robust to different truncation values and to endogenous modeling of the censoring equation (section 4.3). If we ignore the truncation of the data and use OLS, we get the same qualitative significance results, although the coefficients are (as expected) closer to zero.

4.1 Distance matters online

Table 3 traces the development of our core specification. The regressions are censored regressions weighted by the number of international visits in the category. In column 1, we show the most basic gravity regression. As described above, we regress the log of the number of visits in each country-service category on category fixed effects, the log of distance, and

the log of GDP. There are 25 categories and 46 countries (excluding the US) for a total of 1150 observations. In column 1, GDP proxies for the number of websites in the country. Column 2 adds the log of the number of hosts as an explanatory variable. This is likely to be a better proxy than GDP for the number of websites in a country. In both of these columns, distance has a negative and significant impact on visits.

Two country characteristics that are likely to affect a consumer’s utility from visiting a website are the language in which it is written and the level of development of the country. Column 3 adds a dummy variable for whether English is spoken in the country, and column 4 adds the log of GDP per capita. This last variable can proxy for non-homotheticity in consumer preferences but also for the country’s infrastructure and the quality of the Internet connectivity. The change in log likelihood shows that each of these variables adds considerable explanatory power. For this reason, we label column 4 our “main model.” These controls reduce the magnitude of the estimated relationship between distance and visits; however, the relationship remains negative, statistically significant, and economically large. The coefficient in the main model (column 4) suggests that a 1% increase in distance is correlated with a 1.8% reduction in visits.¹⁰

Columns 5 and 6 compare categories where a purchase is likely to occur (e.g., Software) with categories where purchases do not occur (e.g., Music).¹¹ While distance matters in both cases, it matters more for purchased goods than for unpurchased goods. This is consistent with trust being correlated with distance.

In order to assure the robustness of the distance effect on Internet visits, we run a number of checks to investigate the robustness of our results with respect to sample selection and different estimation methods. Column 7 in table 3 shows the results when our

¹⁰In our online appendix, we show distance matters even with controls for rule of law, political stability, number of immigrants from the country living in the US, number of US travelers to the country, whether the country uses the latin alphabet, and an interaction term between whether the country speaks English and the number of website hosts.

¹¹While today there is a great deal of within-category variation in whether services are purchased, in 2000 there were many categories where purchase was extremely rare. For example, purchased music websites do not exist in this data. In defining the service categories, Plurimus explicitly took into account whether money changed hands.

main specification is estimated without weighting the data by the number of visits in each category. In this way, we avoid the results being driven by the largest categories, such as Pornography. The distance effect continues to be large and significant. In order to check if the data truncation and modeling specification are driving the results, column 8 of table 3 shows the results using OLS rather than a censored model. The results are shown in column 8 of table 3. As expected, the estimated distance effect is smaller, but it is still large and statistically significant.¹²

The results so far indicate that, within our sample, distance has a large effect on the number of visits a country’s website receives. It is possible, however, that our sample itself is biased towards distance. That would be the case if, for example, countries closer to the US were selectively excluded from the sample. We address this possibility in two ways. First, we show that the countries included in our sample are not biased to be particularly close to or far away from the US. In particular, we estimate a logit model where the dependent variable is equal to one if the country’s websites were visited at all (i.e. the country is in our sample) and zero otherwise. The independent variables are the same as in our main specification and the sample now contains 106 countries, all countries for which we found gravity data. The results show that physical distance does not affect the likelihood that a country is in our sample.¹³ It is not the case that countries located closer to or farther away from the US are selectively absent of our sample. We also aggregate the number of visits at the country level rather than at the country-category level in column 9 of table 3. In this specification, all the countries in the sample have a positive number of visits (i.e., there are no zeros in this specification) and again distance has a statistically significant and economically large effect on visits. In our online appendix, we also show the table 3 results hold without Canada and without both Canada and Mexico. With this cumulative evidence, we are confident that sample bias is not the source of the estimated distance effect.

¹²We estimated our main specification using OLS only on observations with positive number of visits. Even without the zeros, distance still has a significant effect on the number of visits a website gets. The estimates are available online.

¹³The estimated parameters are available online.

The final robustness check (table 3 column 10) asks if the results hold when the data on visits to US websites are included in the analysis. Even though it is not possible to geographically locate a website within a country, we follow a methodology proposed by Leamer (1997) and used by Blum and Leamer (2004) to proxy for the internal distance between the user and the website visited. In this methodology, we imagine that countries are circular and ask what is the expected distance between randomly selected points within a circle of radius r . The answer is approximately the radius r . We then use the radius of a circle with an area equal to the US as a proxy for the average distance between US users and US websites.

This method also allows us to account for internal distances of all the countries in our sample when calculating the distance between them and the US. This takes care of the fact that the distance between a US user and a Chinese website might be very different than the distance between that user and Beijing, but the distance between a user and a Portuguese website is likely to be similar to the distance between that user and Lisbon. The results show that the distance effect is reinforced when the internal distance of the countries and the US data are taken into account. The positive coefficient on the USA dummy also suggests a border effect in addition to the distance effect.

4.2 The net is global but tastes are local

In this section, we explore the differences in the effect of distance on different types of websites to uncover the reasons why distance still matters online when trade costs should be zero.

Table 4 shows parameter estimates for different types of websites as defined by Plurimus.¹⁴ We classify websites into two groups, websites that are taste-dependent and websites that are not taste-dependent. The taste dependent categories are the ones where there is no reason, other taste differences, for distance to affect the utility a household gets when visiting this site. The categories classified as taste-dependent are Music, Pornography, Gambling,

¹⁴Small sample sizes prevent us from estimating the model for each of the website categories in table 1.

and Games. On the other extreme, the non taste-dependent categories are the ones where tastes (and distance) play no role whatsoever. These are Financial Information, General Information, Internet, Software, and Technology Information.

For the other categories, the effect of tastes gets mixed-up with other possible reasons why distance might matter. Take, for example, the Travel Information category. Clearly, traveling to distant countries is more expensive than to nearby countries. Therefore, US travelers will be more likely to visit travel information sites from countries that are close, simply because they are more likely to travel to these countries. There might also be a taste component to their choices on where to travel and to which travel information sites to visit, but it is impossible to isolate this from other distance related effects. Similar stories apply for all the other categories. A large amount of sensitivity analysis was conducted to check for the robustness of the results with respect to the website categories defined as taste-dependent and non-taste-dependent. Results are also robust to excluding the Pornography category,¹⁵ and to several other changes in the categories labeled as taste dependent and not taste dependent. The online appendix shows that none of the results of the paper changed in any of the specifications considered.

Distance matters for taste-dependent websites but not for the ones that do not depend on tastes. When looking for music, US households tend to browse websites from nearby countries. When looking for software, it does not matter where the website is located.

To further confirm the effect of taste-differences on the demand for digital products, we estimate the same equation for OECD and non-OECD countries separately. If, as we suspect, physical distance proxies for differences in tastes, one should expect that among developed countries the distance effect should be smaller than among less developed countries. This is indeed the case. Even though the distance effect is still negative and statistically significant for both sub-samples, the elasticity of the number of visits with respect to distance is -0.9% for OECD countries and -4.4% for non-OECD countries.

¹⁵Excluding Pornography, the distance effect on visits to taste-dependent websites is reduced to -1.66. It is, however, still significantly different than zero and economically large.

Tables 5, 6a, and 6b assess the robustness of our results to sample selection, changes in estimation methods, and changes in the set of explanatory variables included in the model. Table 5 shows that the result that distance matters only for taste-dependent products is robust to not weighting the data by category size, and to ignoring the selection of the data and using OLS. It is also robust to including the US visits and to excluding Canada and Mexico from the analysis. The large (though insignificant) coefficient on the USA dummy in column 8 of table 5 suggests that even not taste dependent categories may be subject to a border effect.

Columns 1-8 in table 6a and 6b separately add measures of rule of law, political stability, the log of the number of immigrants by country of origin living in the US,¹⁶ the log of the number of Americans who traveled to that country, the log of the number of Internet users in that country, whether they use the Latin alphabet, and an interaction between English as the main language and the number of hosts in the country. Column 9 is a “kitchen sink” regression with all these variables. For both, taste-dependent and non taste-dependent categories, the variables included do not change the significance of the relationship between distance and visits.

In summary, distance consistently reduces the number of visits the website of taste-dependent products receive in all model specifications and robustness checks. However, distance does not affect the number of visits websites of non taste-dependent (homogeneous) products.

4.3 Modeling the censoring of the data

In the previous sections, the standard censoring model (tobit) was used to deal with the fact that we only observe non-negative values for the number of visits websites of different countries receive. In our context, however, this model might be inappropriate because the

¹⁶We compute the number of immigrants using as weights the share of immigrants in the population by country of origin in each US state and the representativeness of each US state in our sample. In order for this data to be of a similar magnitude as the other coefficients in the analysis, we scale this measure up to the US population and take the logarithm.

censoring value may vary by country and website category. In this section, we estimate an econometric model that allows the censoring of the data to depend on a set of country and website characteristics. Equation 2 is then estimated together with a censoring condition, where the probability that a category s website in country c will be visited is a function of a set of country and website characteristics.

Suppose there is an unobserved fixed cost of visiting a website in category s in country c , for convenience measured in the same unit as used in the utility function. This might include, for example, costs associated with bad connectivity, acquiring information about the country, or learning the language. Therefore, we only observe a strictly positive number of visits to countries and website categories in which the benefits of visiting a website from that country and category outweigh the costs. Although unobserved, we can model the net benefit of visiting a website of a given category in a given country as the difference between the utility and the cost of visiting such a website. Assuming the cost function to be linear in country and website characteristics, the econometric model has the following equations:

$$\log(V_c^s)^* = k^s + \log(n_c^s) + \log(X_c^s)\beta^s + \epsilon_c^s \quad (3)$$

$$\log(V_c^s)^* = \log(V_c^s) \quad \text{if} \quad NB_c^s = \gamma^s + \log(X_c^s)\beta - \log(Z_c^s)\delta + v_c^s > 0 \quad (4)$$

$$\log(V_c^s)^* = 0 \quad \text{if} \quad NB_c^s = \gamma^s + \log(X_c^s)\beta - \log(Z_c^s)\delta + v_c^s < 0$$

where the set of variables Z_c^s affects the cost of visiting a website from country c in category s . Z_c^s and X_c^s determine the probability the websites of a country will be visited by the US users at all, and X_c^s determines the number of visits they get. The error term v_c^s is a linear combination of the error terms in the utility function and in the cost function.

Following Hallak (2005), we jointly estimate the system as a tobit with two equations imposing the cross-equation restrictions. In this specification, the effect of the independent variables on the utility of visiting a website does not vary between censored and uncensored observations. However, the independent variables may have a different impact on the cen-

soring *value* and on the utility of visiting a website because they also affect the “cost” part of the censoring equation. The parameter estimates of this model are shown in columns 1-4 in table 7.¹⁷ They confirm that distance is negatively and significantly correlated with the number of overall visits and of visits to taste-dependent websites specifically but not with the number of visits to non taste-dependent websites. They also indicate that distance affects the censoring of the data mostly to the extent that it affects the utility a user gets when visiting a website.

An alternative way to let the censoring of the data vary by country and website category is by estimating a Heckman selection model. In this case, the decision of how many visits to make to websites of a country in a given category is treated as independent of the decision to visit that country’s websites in the same category at all. Distance, for example, may be a determinant of how many visits users make to a country’s websites but not a determinant of whether the country’s websites are visited at all, or vice versa. The results of the Heckman model are shown in columns 5-8 of table 7. Again, distance is negatively and significantly correlated with the number of visits overall and for taste-dependent websites but not for non taste-dependent websites. As expected, distance also makes it less likely that a country’s websites will be visited at all.

In summary, even allowing for the censoring of the data to vary by country and website category distance still reduces the number of visits a website receives. The distance effect can now be split into two components: a) the effect on the probability the country’s website will be visited at all; and b) the effect on the number of visits a country’s websites get, given they are visited. Both of these effects are large and significant for taste-dependent categories but not for categories that are not taste dependent.

¹⁷Two exclusion restrictions are imposed and help in the identification of the parameters. The first is that the number of travelers to the country does not affect the cost function. The second is that, after controlling for language, the use of the Latin alphabet does not affect the utility of a user when visiting a website. Otherwise the parameters are identified off the functional forms. The same is true for the Heckman Selection model estimated below. In both cases, the results are robust to a number of different choices of exclusion restrictions.

5 Conclusions

We have shown that physical distance reduces trade even in online products and services that should be free of trade costs. In particular, we find that in an environment with near zero transportation, distribution, time, and other trade costs, physical distance still matters. We find evidence that countries that are close to each other either have similar tastes or are better able to cater to each other tastes. For products and services where taste matters, such as music, websites in countries near the US are more heavily visited than websites in far away countries. For sites of products where taste is less important, visits are uncorrelated with physical distance (aside from a possible border effect). This result holds even after we control for a number of effects. For the possibility of clustering in website availability, we control for the number of web hosts in each country. For the possibility of clustering in income, we control for per capita GDP. In order to control for the effect of immigrants “phoning home,” we control for the number of immigrants from each country in the US.

Our results provide a new explanation for the persistence of distance in the gravity model. It suggests that the distance effect in gravity will persist for a number of products even if transportation costs, search costs, and other trade barriers associated with distance are reduced to zero. For the distance effect to go away, there needs to be a homogenization of cultures. Tracking the distance effect for taste-based products and services in the future may provide an interesting way to measure worldwide cultural homogenization.

6 References

- Anderson, Simon P., Andre de Palma, and Jacque-Francois Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. The MIT Press: Cambridge MA.
- Anderson, James E., and Eric van Wincoop. 2004. Trade Costs. *Journal of Economic Literature*, 42#3 (Sep). 691-752.
- Bakos, Yannis. 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science*. 43#12 (Dec). 1676-1692.
- Berry, Steven T. 1994. Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics* 25#2 (Summer), 242-262.
- Blum, Bernardo, and Edward E. Leamer. 2004. Can FTAA Suspend the Law of Gravity and Give the Americas Higher Growth and Better Income Distributions? In *FTAA and Beyond: Prospects for Integration in the Americas*, ed. by Antoni Estevadeordal, Dani Rodrik, Alan Taylor, and Andres Velasco, Harvard University Press.
- Cairncross, Frances, 1997. *The Death of Distance*. Harvard University Press: Cambridge, MA.
- Disdier, Anne-Celia, and Keith Head. 2005. The Puzzling Persistence of the Distance Effect on Bilateral Trade. Manuscript, University of British Columbia.
- Feenstra, Robert, James A. Markusen, and Andrew K. Rose. 2001. Using the Gravity Equation to Differentiate Among Alternative Theories of Trade. *Canadian Journal of Economics*, 34(2), 430-447.
- Freund, Caroline L. and Weinhold, Diana, 2004. The Effect of the Internet on International Trade. *Journal of International Economics* 62(1), 171-189.
- Goolsbee, Austan. 2000. In a World without Borders: The Impact of Taxes on Internet Commerce. *Quarterly Journal of Economics* 115 (2 May), 561-576.
- Greene, William H. 1995. *Econometric Analysis*, 3rd edition. Prentice Hall: Upper Saddle River NJ.
- Hallak, Juan Carlos, 2005. Product Quality and the Direction of Trade. Forthcoming *Journal of International Economics*.
- Harrigan, James. 2003. Specialization and the Volume of Trade: Do the Data Obey the Laws? In *The Handbook of International Trade*, ed. Kwan Choi and James Harrigan, Basil Blackwell: London, UK.
- Hummels, David. 1999. Have International Transportation Costs Declined? Manuscript, Purdue University.
- Hummels, David. 2001a. Time as a Trade Barrier. Manuscript, Purdue University.

- Hummels, David. 2001b. Toward a Geography of Trade Costs. Manuscript, Purdue University.
- Lawrence, Steve, and C. Lee Giles. 1998. Search the Worldwide Web. *Science*, 280 (April), pp. 98-100.
- Leamer, Edward E. 1997. Access to Western Markets and Eastern Effort. In *Lessons from the Economic Transition, Central and Eastern Europe in the 1990s*, ed. Salvatore Zecchini. Kluwer Academic Publishers: Dordrecht, pp. 503-526.
- Leamer, Edward E. and Michale Storper. 2001. The Economic Geography of the Internet Age. *Journal of International Business Studies*, 32(4), 641-655.
- Oxley, Joanne, and Bernard Yeung. 2001. E-commerce Readiness: Institutional Environment and International Competitiveness. *Journal of International Business Studies*, 32(4 December), pp. 705-724
- Rauch, James E. 1996. Trade and Search: Social Capital, Soga Sosha and Spillovers. NBER working paper # 5618, June.
- Rauch, James E. 1999. Network Versus Markets in International Trade. *Journal of International Economics* 48, 7-35.
- Shapiro, Carl, and Hal R. Varian, 1999, *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press: Boston.
- Stevenson, Betsy. 2003. The Internet, Job Search, and Worker Mobility. Manuscript, Stanford University.
- Sullivan, Danny. 1999. The Search Engine Report #4. October. Available at www.searchenginewatch.com.

Table 1: Service Category Descriptive Statistics

	category	Total Visits	Non-US visits	Non-US, Non-Canada visits	Example
1	Astrology	2,104	38	10	Astronet
2	Automotive information	14,116	1,188	930	Cartrackers
3	Brochureware	16,652	648	424	Nike.com
4	Chat	61,348	162	162	YahooChat
5	Community	144,209	1,064	964	Geocities
6	Comparison Shopping	8,064	364	364	MySimon
7	Ecards	22,715	668	45	Bluemountain
8	Email	193,369	402	373	Hotmail
9	Financial Information	130,121	226	158	YahooFinance
10	Financial Transactions	36,542	1,791	592	TDWaterhouse
11	Gambling	26,429	239	183	Lotto.net
12	Games	68,993	2,187	1,986	Boxerjam
13	General Information	65,244	1,533	1,109	Britannica
14	Internet	182,199	7,613	5,185	Desktoppublishing.com
15	Music	18,665	1,252	926	MP3.com
16	News/Sports	186,856	4,561	4,293	CNN
17	Non-profit/Education	42,901	387	295	Utoronto.ca
18	Pornography	442,967	31,302	26,803	Porncity
19	Portal/Search	874,224	7,393	5,322	Yahoo
20	Software	128,861	6,138	4,316	Tucows
21	Technology Information	39,530	210	182	CNET
22	Travel (Information)	17,185	298	298	CitySearch
23	Travel (Purchase)	27,940	81	61	Expedia
24	TV/Movies	22,174	136	136	TVGuide
25	Vertical Portals	153,805	15,345	13,698	Canada.com

Table 2: Country Descriptive Statistics

	country	Total country visits	Most common category	Most common category frequency	Second most common category	Second most common category frequency
1	Australia	1587	Email	370	Pornography	264
2	Austria	387	Pornography	294	Games	64
3	Belize	583	Pornography	491	General Information	85
4	Brazil	351	Pornography	281	News/Sports	69
5	Canada	16411	Pornography	4499	Internet	2428
6	China	297	Vertical Portals	178	Internet	119
7	Cyprus	128	Pornography	114	Vertical Portals	14
8	Czech Republic	1458	Vertical Portals	1334	Pornography	44
9	Denmark	1780	Vertical Portals	778	Internet	741
10	Estonia	185	Vertical Portals	185	---	---
11	Finland	466	News/Sports	260	Internet	159
12	France	2025	Pornography	907	Vertical Portals	784
13	Germany	3491	Vertical Portals	1532	Pornography	999
14	Greece	2	Non-profit/Education	2	---	---
15	Holland	13992	Pornography	8555	Software	3752
16	Hungary	64	Pornography	64	---	---
17	India	1093	Vertical Portals	875	News/Sports	218
18	Indonesia	41	Pornography	41	---	---
19	Iran	191	News/Sports	191	---	---
20	Ireland	767	Portal/Search	422	News/Sports	215
21	Israel	1033	Software	395	Comparison Shopping	340
22	Italy	559	Portal/Search	171	Vertical Portals	166
23	Japan	6801	Games	1606	Vertical Portals	1580
24	Luxembourg	621	Pornography	621	---	---
25	Malaysia	33	Internet	21	Portal/Search	12
26	Mexico	4148	Pornography	4093	Portal/Search	45
27	New Zealand	1217	News/Sports	1060	Internet	156
28	Norway	487	Pornography	280	Internet	123
29	Pakistan	4	Vertical Portals	4	---	---
30	Panama	3319	Pornography	3114	Internet	102
31	Poland	102	Technology Information	52	Pornography	33
32	Portugal	132	Vertical Portals	123	Internet	9
33	Russia	4796	Portal/Search	2028	Pornography	1665
34	South Africa	188	Internet	77	Gambling	64
35	South Korea	298	Pornography	172	Internet	92
36	Singapore	51	Internet	30	Pornography	21
37	Slovakia	646	Vertical Portals	404	Pornography	144
38	Slovenia	77	Music	77	---	---
39	Spain	2638	Vertical Portals	1301	Pornography	798
40	Sweden	2572	Vertical Portals	1690	Community	339
41	Switzerland	1064	Financial Transactions	580	Internet	254
42	Taiwan	199	Community	199	---	---
43	Thailand	48	Pornography	34	Internet	14
44	Turkey	379	Pornography	376	Non-profit/Education	3
45	Ukraine	62	Pornography	62	---	---
46	United Kingdom	8453	Pornography	1982	Portal/Search	1712
47	United States	2841987	Portal/Search	866831	Pornography	411665

Table 3: Americans Visit Websites from Nearby Countries More than Faraway Countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Basic	Add hosts	Add English	Add GDP per Capita (MAIN MODEL)	Purchase Involved	No Purchase Involved	Unweighted	OLS	Country level (OLS)	Internal Distance
ln(distance)	-2.899	-2.214	-1.953	-1.785	-2.679	-1.123	-1.544	-1.167	-1.521	-2.324
	(0.353)**	(0.343)**	(0.335)**	(0.331)**	(0.641)**	(0.363)**	(0.311)**	(0.114)**	(0.430)**	(0.422)**
ln(gdp)	1.226	0.321	0.339	0.902	0.329	1.351	0.664	0.427	0.245	0.941
	(0.140)**	(0.181)+	(0.175)+	(0.213)**	(0.381)	(0.251)**	(0.224)**	(0.061)**	(0.230)	(0.203)**
ln(# hosts)		1.072	1.014	0.338	0.451	0.291	0.592	0.060	0.165	0.355
		(0.155)**	(0.150)**	(0.201)+	(0.360)	(0.234)	(0.216)**	(0.057)	(0.215)	(0.190)+
English			2.095	2.242	0.130	3.101	2.400	1.057	1.069	2.438
			(0.474)**	(0.469)**	(0.916)	(0.508)**	(0.462)**	(0.154)**	(0.584)+	(0.436)**
ln(gdp per capita)				1.798	1.108	2.350	1.801	0.644	0.612	1.629
				(0.406)**	(0.708)	(0.489)**	(0.434)**	(0.114)**	(0.433)	(0.386)**
USA Dummy										3.027
										(1.337)*
Observations	1150	1150	1150	1150	322	828	1150	1150	46	1175
LL	-1012.52	-984.97	-975.38	-965.62	-290.35	-656.80	-901.30	N/A	N/A	-1051.36
R ²	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.46	0.49	N/A

Standard errors in parentheses

All regressions (except the country level regression in column 9) include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

Table 4: Distance only matters for categories that depend on tastes

	(1)	(2)	(3)	(4)
	Taste Dependent [#]	Not Taste Dependent ^{##}	Non-OECD Countries Only	OECD Countries Only
ln(distance)	-3.248 (0.757)**	-0.979 (0.628)	-4.433 (1.495)**	-0.900 (0.292)**
ln(gdp)	0.131 (0.439)	0.375 (0.440)	1.028 (0.541)+	1.080 (0.253)**
ln(# hosts)	-0.313 (1.066)	2.738 (0.916)**	-0.084 (0.449)	0.806 (0.250)**
English	0.402 (0.411)	0.804 (0.432)+	0.388 (1.597)	2.248 (0.447)**
ln(gdp per capita)	0.518 (0.795)	2.177 (0.870)*	2.225 (1.292)+	2.017 (0.498)**
Observations	184	230	475	675
LL	-218.71	-214.65	-216.17	-689.59

Standard errors in parentheses

All regressions include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

[#]Taste-dependent categories: Gambling, Games, Music, and Pornography

^{##}Not Taste-dependent categories: Financial Information, General Information, Internet, Software, and Technology Information.

Table 5: Sensitivity Analysis—Robustness of Significance

	Taste Dependent					Not Taste Dependent				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Unweighted	OLS	With Internal Distances (Includes US)	No Canada	No Canada or Mexico	Unweighted	OLS	With Internal Distances (Includes US)	No Canada	No Canada or Mexico
ln(distance)	-2.419 (0.707)**	-2.348 (0.394)**	-4.213 (0.929)**	-4.231 (1.084)**	-4.411 (1.215)**	-1.198 (0.619)+	-0.637 (0.252)*	-1.226 (0.817)	-0.582 (1.074)	-1.519 (1.166)
ln(gdp)	-0.166 (0.483)	0.182 (0.211)	0.302 (0.427)	0.249 (0.466)	0.253 (0.468)	0.156 (0.455)	0.195 (0.135)	0.386 (0.421)	0.371 (0.467)	0.303 (0.458)
ln(# hosts)	0.698 (0.464)	-0.618 (0.535)	0.409 (0.392)	0.409 (0.427)	0.435 (0.436)	0.966 (0.448)*	1.186 (0.342)**	0.811 (0.413)+	0.813 (0.459)+	1.010 (0.471)*
English	2.053 (1.008)*	0.192 (0.197)	0.081 (0.998)	0.174 (1.157)	0.141 (1.163)	2.686 (0.909)**	0.196 (0.126)	2.816 (0.860)**	2.640 (1.054)*	2.598 (1.025)*
ln(gdp per capita)	0.566 (0.868)	0.286 (0.396)	0.391 (0.760)	0.555 (0.824)	0.481 (0.854)	1.910 (0.884)*	0.601 (0.253)*	2.046 (0.833)*	2.296 (0.928)*	1.703 (0.946)+
USA Dummy			1.175 (3.133)					2.868 (2.643)		
Observations	184	184	188	180	176	230	230	235	225	220
LL	-208.92	N/A	-231.35	-204.82	-201.85	-213.58	N/A	-227.91	-201.87	-199.47
R ²	N/A	0.23	N/A	N/A	N/A	N/A	0.25	N/A	N/A	N/A

Standard errors in parentheses

All regressions include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

Table 6a: Distance Matters In Taste Dependent Categories Even with Many Country-Level Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rule of Law	Political Stability	Immigrants	# Travelers	# Users rather than # Hosts	Use Latin alphabet rather than English	Interact English and # Hosts	All variables together
ln(distance)	-3.211 (0.767)**	-3.455 (0.742)**	-3.185 (0.940)**	-2.348 (0.909)*	-3.296 (0.761)**	-2.800 (0.805)**	-3.227 (0.758)**	-2.451 (1.057)*
ln(gdp)	0.101 (0.452)	-0.147 (0.454)	0.097 (0.529)	-0.146 (0.471)	0.701 (0.615)	0.155 (0.437)	0.165 (0.449)	0.199 (0.792)
ln(# hosts)	0.419 (0.415)	0.533 (0.408)	0.395 (0.416)	0.179 (0.430)		0.357 (0.409)	0.425 (0.416)	0.545 (0.494)
English	-0.181 (1.168)	-0.097 (1.041)	-0.325 (1.071)	-1.033 (1.130)	-0.149 (1.082)		1.538 (5.180)	1.314 (5.631)
ln(gdp per capita)	0.807 (1.321)	1.973 (1.041)+	0.556 (0.863)	0.621 (0.787)	1.280 (0.755)+	0.424 (0.792)	0.570 (0.808)	1.532 (1.278)
Rule of law	-0.271 (0.986)							0.785 (1.248)
Political stability		-1.895 (0.897)*						-2.301 (1.180)+
ln(# immigrants)			0.055 (0.487)					-0.174 (0.577)
ln(# travelers)				0.651 (0.394)+				-0.219 (0.427)
ln(# users)					-0.240 (0.644)			-0.641 (0.741)
Latin alphabet						1.100 (0.926)		0.532 (0.478)
English*ln(# hosts)							-0.149 (0.407)	1.545 (0.961)
Observations	184	184	184	184	184	184	184	184
LL	-218.68	-216.50	-218.71	-217.35	-219.13	-218.05	-218.65	-213.18

Standard errors in parentheses

All regressions include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

Table 6b: Distance Matters Much Less in Non Taste Dependence Categories Even with Many Country-Level Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rule of Law	Political Stability	Immigrants	# Travelers	# Users rather than # Hosts	Use Latin Alphabet rather than English	Interact English and # Hosts	All variables together
ln(distance)	-1.121 (0.626)+	-0.971 (0.631)	-0.670 (0.773)	0.051 (0.850)	-1.141 (0.629)+	-1.511 (0.702)*	-0.996 (0.630)	-0.552 (0.961)
ln(gdp)	0.541 (0.444)	0.389 (0.455)	0.159 (0.543)	-0.063 (0.513)	-0.448 (0.670)	0.259 (0.456)	0.388 (0.443)	-0.650 (0.863)
ln(# hosts)	0.669 (0.435)	0.795 (0.438)+	0.784 (0.437)+	0.633 (0.457)		1.011 (0.446)*	0.835 (0.448)+	0.271 (0.504)
English	2.084 (0.967)*	2.733 (0.917)**	2.711 (0.916)**	1.969 (0.988)*	2.808 (0.906)**		4.573 (6.314)	1.775 (6.114)
ln(gdp per capita)	0.428 (1.310)	2.095 (1.096)+	2.375 (0.929)*	2.153 (0.888)*	2.235 (0.767)**	2.168 (0.915)*	2.210 (0.881)*	0.168 (1.360)
Rule of law	1.578 (0.933)+							1.909 (1.235)
Political stability		0.111 (0.904)						-0.371 (1.155)
ln(# immigrants)			0.329 (0.484)					0.069 (0.583)
ln(# travelers)				0.818 (0.466)+				-0.005 (0.453)
ln(# users)					1.616 (0.690)*			0.984 (0.778)
Latin alphabet						-0.640 (0.895)		0.813 (0.610)
English*ln(# hosts)							-0.140 (0.475)	-1.573 (0.943)+
Observations	230	230	230	230	230	230	230	230
LL	-213.21	-214.64	-214.41	-212.97	-213.60	-218.80	-214.61	-208.42

Standard errors in parentheses

All regressions include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

Table 7: Modeling the Censoring of the Data

	Hallak's Random Censoring Model				Heckman Selection Model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All data [#]	No Purchase Only [#]	Taste Dependent Only [#]	Not Taste Dependent Only [#]	All data	No Purchase Only	Taste Dependent Only	Not Taste Dependent Only
	Visits Equation				Visits Equation			
Ln(distance)	-0.645 (0.197)**	-0.681 (0.265)*	-0.942 (0.360)**	-0.168 (0.466)	-0.885 (0.193)**	-0.921 (0.256)**	-0.901 (0.329)**	-0.119 (0.444)
Ln(gdp)	0.157 (0.151)	0.375 (0.194)+	0.099 (0.316)	-0.612 (0.413)	0.458 (0.130)**	0.745 (0.199)**	0.354 (0.230)	-0.607 (0.342)+
English	1.016 (0.299)**	1.314 (0.395)**	0.665 (0.547)	0.128 (0.860)	1.397 (0.233)**	1.834 (0.343)**	0.771 (0.459)+	0.025 (0.596)
Ln(# hosts)	0.200 (0.142)	-0.210 (0.207)	0.147 (0.297)	0.816 (0.330)*	-0.098 (0.105)	-0.125 (0.142)	-0.210 (0.190)	0.803 (0.323)*
Ln(gdp per capita)	0.447 (0.324)	0.615 (0.408)	0.176 (0.477)	-0.818 (1.052)	0.396 (0.224)+	0.448 (0.356)	0.216 (0.296)	-0.995 (0.803)
Ln(# travels)	0.014 (0.083)	-0.236 (0.142)+	1.92e-04 (0.0021)	0.574 (0.351)	-0.080 (0.100)	-0.394 (0.151)**	0.086 (0.137)	0.597 (0.327)+
	Fixed Cost (Censoring) equation				Selection equation			
Ln(distance)	0.042 (0.248)	0.481 (0.355)	-0.937 (0.359)**	1.014 (2.180)	-0.684 (0.135)**	-0.715 (0.184)**	-0.835 (0.296)**	-0.430 (0.286)
Ln(gdp)	-0.127 (0.156)	-0.256 (0.210)	0.100 (0.316)	-0.531 (0.759)	0.281 (0.052)**	0.421 (0.069)**	0.022 (0.113)	0.156 (0.127)
English	0.030 (0.371)	-0.546 (0.488)	0.662 (0.546)	-2.956 (4.041)	0.670 (0.146)**	1.180 (0.195)**	-0.559 (0.314)+	1.096 (0.330)**
Ln(# hosts)	-0.028 (0.151)	0.210 (0.178)	0.148 (0.297)	-0.182 (1.126)	0.108 (0.048)*	0.112 (0.062)+	0.127 (0.103)	0.212 (0.121)+
Ln(gdp per capita)	-0.309 (0.346)	-0.643 (0.461)	0.176 (0.476)	-3.185 (2.748)	0.573 (0.101)**	0.762 (0.137)**	0.115 (0.210)	0.767 (0.244)**
Latin Alphabet	0.182 (0.191)	0.117 (0.353)	-0.0015 (0.0156)	1.314 (2.019)	-0.267 (0.127)*	-0.473 (0.172)**	0.467 (0.266)+	-0.509 (0.298)+
Observations	1150	828	184	230	1150	828	184	230
LL	-781.42	-543.54	-173.46	-175.31	-1193.04	-701.09	-274.44	-204.61

Standard errors in parentheses

All regressions include category fixed effects

+ significant at 10% * significant at 5%; ** significant at 1%

[#]The results for Hallak's model are unweighted. The weighted likelihood would not converge.

APPENDIX

Table A1: Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum	# of Observations
# visits to country-category	74.110	408.863	0	8555	1150
Ln(distance)	8.932	0.535	6.598	9.703	1150
Ln(gdp)	5.597	1.630	0.247	8.698	1150
English	0.174	0.379	0	1	1150
Ln(# hosts)	12.170	2.040	5.808	15.778	1150
Ln(gdp per capita)	9.550	0.754	7.580	10.898	1150
Ln(# immigrants)	11.193	1.521	7.673	14.876	1150
Rule of law	0.936	0.946	-0.9	2.22	1150
Political stability	0.691	0.798	-1.85	1.73	1150
Ln(# travelers)	3.969	1.802	0	8.771	1150
Ln(# users)	14.383	1.671	9.616	17.667	1150
Latin alphabet	0.587	0.493	0	1	1150