OECD publishing

BBVA BIG DATA ON ONLINE CREDIT CARD TRANSACTIONS

THE PATTERNS OF DOMESTIC AND CROSS-BORDER E-COMMERCE

OECD DIGITAL ECONOMY PAPERS

March 2019 **No. 278**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

This report was prepared by the Working Party on Measurement and Analysis of the Digital Economy (MADE). It makes use of credit card payment data of customers of the Spanish multinational bank BBVA to shed new light on the determinants of cross-border e-commerce.

The report was discussed approved and declassified by written procedure by the Committee on Digital Economy Policy on 18 May 2018 and was prepared for publication by the OECD Secretariat.

The report was drafted by Jan Tscheke (OECD), Hicham Ganga (BBVA), Javier Alonso Meseguer (BBVA) and Vincenzo Spiezia (OECD).

Note to Delegations:
This document is also available on O.N.E. under the reference code:

DSTI/CDEP/MADE(2017)6/FINAL

Table of contents

Tables

Figures

Executive summary

This report uses a standard gravity setup to analyse the determinants of e-commerce, using data on online payment flows within and across regions and countries to proxy for commercial online transactions. Data on payments is provided by the multinational bank BBVA and comprises online credit card payments by private Spanish customers of the bank. So-called *card not present* transactions, credit card transactions where no payment card was physically involved are used to proxy for the value of e-commerce transactions.

Gravity equations have been used for decades by trade economists to explain trade flows between countries or regions and belong to the standard toolkit of empirical trade analysis not least due to their astonishing empirical performance (see Feenstra, 2002). The gravity model used in this report follows the literature and relates domestic and cross-border transaction values to the economic size of the trading partners as well as fundamental trade cost determinants such as distance and borders. Additional control variables are added to account for the specificities of e-commerce transactions.

The analysis proceeds at different levels of data aggregation. In all cases, the data provided by BBVA aggregates daily *real time* transaction level data into yearly data. For the purpose of this study, the individual level data is then further aggregated to provide aggregate *macro* flows between geographic entities. First, domestic payment transactions are aggregated at the regional level according to the product category assigned by BBVA to the transaction. This delivers bidirectional trade flows between any two of the 19 Spanish regions in the data, distinguished by the type of products associated with each merchant. Second, purchases of clients in each of the 19 regions are aggregated according to whether the purchase is domestic (Spain) or cross-border, involving any particular of the potential partner countries in e-commerce. As product category information is only available for domestic merchants, analysis with an international dimension does not consider the product dimension. Finally, payment flows are also used in a more granular aggregation at the individual client level, providing two additional data sets referred to as the *micro level*. Here observations represent either domestic transactions or cross-border transactions for a given individual vis-à-vis merchants in a given Spanish region or foreign country. This aggregation at the individual client level will be referred to as the *micro level*.

Each of the different levels of aggregation provides a distinct perspective on e-commerce transactions, by allowing augmentation of the gravity model with possible trade determinants at the regional, product category, country or even individual level. Classical trade determinants, such as distance, border or trade agreements, are used in all cases. Explanatory variables more specifically related to e-commerce, such as credit card usage, the number of secure internet servers in a given country or the cost of business start-up procedures are added at the cross-border level, whereas product category specific effects are considered for domestic e-commerce. Finally, the micro level data allows to further account for the age and gender of the individual engaging in e-commerce.

Across specifications, the results show that the gravity model applies well to credit card payments, explaining up to 95% of the variation in the data. Overall, classical trade determinants like distance and state or country borders are confirmed to have a statistically and economically significant effect on e-commerce trade flows within and across countries. The economic size of either trade partner is also found to be an important determinant of trade flows between any two regions or countries. In general, the analysis therefore confirms earlier findings in the literature for the special case of e-commerce. With respect to the economic size of the effects, the analysis finds potentially large border effects for trade between any two regions or countries, implying that individuals tend to purchase more from their home region or domestically than from other places. The estimates also suggest that the effect of distance might be slightly less important for e-commerce transactions overall when compared to results for offline trade, in line with earlier findings from Hortaşçu *et al.* (2009). While the *death of distance* (Cairncross, 1997) as a trade determinant is clearly rejected by the data, the findings suggest important heterogeneity in both the border and the distance effect that future work might find fruitful to analyse in more detail.

In particular, at the domestic *macro* level, where payment flows between Spanish regions are analysed taking into account the product type associated with the merchant, the analysis suggests that on average, an increase in the distance between client and merchant by 10% decreases trade by about 4.3%, while a statistically significant difference between e-commerce within the client's home region and purchases from other Spanish regions (*i.e.* the home bias or domestic border effect) is not confirmed in all specifications.

A more nuanced analysis shows that these results vary substantially by product category, with large domestic border effects for purchases from *hyper* (large supermarkets), *bars and restaurants*, *sports and toys* or *transportation*, but no significant border effects for categories like *fashion*, *contents* or *health*.[1] With respect to distance, *fashion* is also among the products that appear to be more easily purchased from further afield, with the effect of distance being only marginally significant. This is interesting given that the purchase of clothing involves actual packages being shipped. On the other hand, fashion is likely to be purchased relatively more often via e-commerce if the product cannot be found in the close surroundings. The effect is substantially different for purchases for example from *Hyper*(markets), where distance is a highly significant deterrent to online purchases, in line with the results for the domestic border effects.

The domestic level analysis further shows that the *home market bias* (a different way to describe the domestic border effect) is less pronounced for regions with a high average level of education and a high dissemination and use of digital technologies. In particular, the percentage of households engaging in online banking, the percentage of households with an internet connection and the frequency of internet usage are all positively associated with purchases from other regions. Additionally, clients appear to be more likely to purchase from other regions if the consumer price level in their home region is relatively high.

At the international level, where purchases are separated into purchases from Spain or cross-border purchases with respect to particular foreign countries, the literature findings of a negative distance effect are confirmed. Nevertheless, the estimated effects, varying between a 3%-9% reduction for each 10% increase in distance, tend to lie below the average literature finding of roughly 10.7% (Disdier and Head, 2008). With respect to the *border effect*, the estimates suggest large heterogeneity with respect to cross-border purchases from other Euro-area member countries, European Union members not part of the Euro-area and non-EU member countries. The border effect is found to be always negative and largest for countries that are not part of the European Union. Purchases from non-Euro member countries tend to be more likely than purchases from other European countries, an effect that can be explained with Great Britain playing such an important role as source of Spanish cross-border purchases.

In terms of absolute size, the estimated border effects, in particular with respect to countries not part of the EU, are implausibly high. Several factors that could explain this result are discussed. In particular, the literature suggests that omitted variables bias might be driving part of the results. This is a well-known problem in the estimation of border effects and relates to the discussion of *multilateral resistance* terms, which capture trade costs for each of the observed trading partners with respect to the rest of the world. The problem is not easily controlled for given the structure of the data used but the heterogeneity found in the effects across all specifications (and data variations) seems to support the hypothesis of actual border effects being picked up by the model.

The analysis further reveals that the existence of large multinational corporations and payment intermediaries in data on credit card transactions can have a potentially significant impact on the explanatory power and the estimated coefficients of the gravity model. This is a reflection of financial headquarters or payment intermediaries, whose location often determines the merchant address as registered by the bank, sometimes being quite different from the actual product location that determines the commercial transaction the gravity model ultimately aims to capture. The resulting measurement error can then lead to spurious results and a weak model performance.

As the analysis shows, dropping potential confounding merchants from the model significantly alters the economic size of the measured effects but leaves the qualitative results with respect to the effect of distance and borders on e-commerce mostly intact. Furthermore, control variables can help to explain the patterns of cross-border online purchases generally tend to gain in terms of statistical significance and thus explanatory power.

The estimation introduces several explanatory variables that are of particular interest in the context of e-commerce, including fundamental enablers of e-commerce such as the availability of credit cards in a given merchant country or the number of secure servers, but also the quality of the logistical system and regulatory factors such as index of *regulatory quality*, or the existence of a legal framework for electronic transactions or cybercrime prevention.[2] As the estimations suggest, all of these factors are both economically and statistically significantly related to the cross-border patterns of e-commerce purchases.

Many of the previous results are confirmed when estimating the effects with data that considers purchases at the individual client level. In absolute terms, both the border effects and particularly the effect of distance are substantially lower when estimated at the micro level. In particular, the estimated coefficients suggest that for individuals, domestic purchases are about 2.6 times more likely than purchases from other Euro member countries, about 2.8 times more likely than purchases from non-Euro EU member countries and about 3.5 times more likely than purchases from other foreign countries. While these estimates should not be taken at face value, given the problems mentioned earlier, the relative size of these effects is still in line with patterns that would arise from actual border effects driving the results. Interestingly, the micro-level estimates seem to imply a smaller effect of Great Britain, implicitly captured in the effect of the common currency.

With respect to the effect for distance, the estimates for both domestic and cross-border e-commerce seem to suggest reverse effects for purchases at very close distances. In particular, the effect of distance seems to be positive for purchases within a range of 50km surrounding the client. Within this radius, e-commerce purchases are found to be increasing with distance, a result that would be in line with consumers substituting online with offline purchases in the immediate neighbourhood. While this is a novel effect, it principally seems to reaffirm earlier results by Hortaçsu *et al.* (2009), who also find that "driving distance" is relevant for e-commerce. Unfortunately, product category data was not provided at the micro-level for this study, and a deeper analysis of this hypothesis will therefore be left for future research. Applying the different product categories that were used in the macro aggregate data for domestic purchases might proof to be a fruitful approach for identification.

Overall, the estimates at the micro-level suggest much smaller effects of distance even beyond the 50km perimeter. Thus, on average, purchases are estimated to be reduced by about 0.55% to 0.75% for each 10% increase in distance, significantly lower than the estimated effect of 3%-9% at the macro-aggregate level and also much lower than the estimates for traditional trade obtained from the standard gravity literature.[3]

The micro-level estimates further show that men are much more likely than women to engage in online purchases and that the average value of purchases is increasing in age, confirming the results that were observed from simply averaging the data (see Table 11 in the Appendix). Additionally, the micro-data confirms that a reduction in start-up costs and complexity, as well as a high availability of credit cards in the merchant country are positively associated with e-commerce from these countries.

BBVA Big data on online credit card transactions: the patterns of domestic and cross-border e-commerce

Introduction

The gravity model of trade applies Newton's law of gravity to spatial relationships between economic actors or geographic regions, postulating distance and (economic) mass as key determinants of trade flows between countries or regions. Gravity models have a long history in explaining the degree of commercial and other geographic interactions between regions and have previously been applied in areas such as migration, retail or traffic due to their excellent empirical performance. This paper uses the gravity model of trade and applies it to data on online payments, in order to provide novel insights into the determinants of domestic and cross-border e-commerce.

In particular, the analysis of online purchases using the gravity model can help to assess whether and how the role of trade determinants has changed in the era of e-commerce. As the digital transformation has dramatically reduced the cost of long-distance communication, customers can now compare product prices and access products from merchants all over the world, principally irrespective of their geographic location. This reduction in informational costs is thought to reduce the effect of distance on international trade relationships, a notion resonating earlier voices that have proclaimed the "death of distance" due to the communication revolution (Cairncross 1997).[4]

The analysis presented in this report will help to shed new light on these and other characteristics of online trade. Specifically, the analysis will use online payment flows of private costumers to proxy for B2C e-commerce transactions between clients and merchants at the individual or regional level. The analysis thereby augments the simple gravity framework by incorporating old and new explanatory variables at the individual, regional or country level, thought to be relevant determinants of online transactions. Factors considered range from classical trade determinants, such as distance, borders or trade agreements, to novel controls including the dissemination of credit cards, the number of secure internet servers or cost of start-up procedures in a given country. At the individual client level, the data also allows to add controls for age and gender.

The results presented in this paper should be seen as an experimental attempt to apply existent empirical methods to a new type of data. This also implies dealing with several complexities that cannot always be solved using traditional literature prescriptions. Accordingly, the presented results will have to be carefully re-evaluated once more data and related research becomes available. Against the backdrop of this caveat, the results presented below suggest that the gravity model performs surprisingly well when applied to data on online payments. On average, the model confirms significant negative border effects but also seems to support earlier findings suggesting that distance might be a slightly less important deterrent of domestic and cross-border e-commerce when compared to estimates for classical offline trade.

The analysis also provides novel insights related to the specificities of e-commerce. In particular, the paper finds that both the effect of distance and the effect of (interregional) borders appear to be highly contingent on the type of product sold, reflecting that some purchases, for example involving supermarkets or bars and restaurants, represent product types that are substantially less tradeable across distance and borders than fashion goods for instance. Surprisingly, and most likely related to this finding, the data also reveals that e-commerce today appears to comprise activities that are governed by fundamentally different regularities. Thus, within a close geographic radius of 50km around the individual client, the usually negative effect of distance on purchase volume is reversed, potentially indicating particular substitution effects between online and offline purchases for e-commerce activities that are happening close by.

The paper further identifies several country characteristics that help to explain the geographic distribution of online purchases, including such factors as the regulatory framework, the dissemination of e-commerce enablers like credit cards or secure servers as well as the quality of the postal system. Finally, the analysis also illustrates that large multinationals have become important determinants of international online payment flows and that the geographic separation of financial headquarters from the physical distribution centers has important implications for the explanatory power of the gravity model.

1. Related literature

This paper uses a standard gravity setup to analyse the payment flows observed in the data. Gravity equations have been used for decades by trade economists to explain trade flows between countries or regions and belong to the standard toolkit of empirical trade analysis not least due to their very good empirical performance (see Feenstra, 2002). A basic gravity equation usually relates trade between two countries or regions to factors like economic size, measured by GDP, as well as factors that affect trade costs, including distance, borders, free trade agreements, a common official language or a common currency. Optimally, gravity specifications also account for price differences between countries or regions that are arising with distance, transport costs and other trade barriers, including with respect to other trading partners. Due to its simplicity, the gravity equation has also been applied beyond traditional trade analysis, for example to explain the distribution of migrant or tourism flows across countries (*e.g.* Morley, Rosselló and Santana-Gallego, 2014 or Ramos, 2016).

Freund and Weinhold (2004) were among the first to apply the gravity model in a context related to the digital transformation. Specifically, they added the number of top-level domain names that were attributed to a country to the list of trade determinants in order to assess whether Internet penetration fosters international trade or reduces the importance of distance in international transactions. Their findings suggest that the Internet indeed stimulates trade but does not directly affect the relationship between distance and trade.

Blum and Goldfarb (2006) applied the gravity equation to digital goods consumed over the Internet, using click stream patterns to analyse whether distance remains an important determinant of trade for goods that can be traded without costs. They show that Americans are more likely to visit websites from nearby countries after controlling for language, income, the immigrant stock and other likely determinants of website choice. They also find that the effect only holds for taste-dependent digital products, such as music or games, whereas distance has no statistically significant effect for less taste-depend products like software. Their research directly relates to the study of online purchases as they find the effect of distance to matter more for website categories where purchases are likely to occur.

Hortaçsu, Martínez-Jerez and Douglas (2009) apply the gravity approach to online purchases of individuals, using intra-national transactions data from eBay and international transactions data from MercadoLibre. They find that distance remains an important deterrent to e-commerce between geographically separated buyers and sellers, but the estimated adverse effect of distance on trade is smaller than for non-Internet trade flow and highly nonlinear. Specifically, they find evidence for a strong *home bias*, implying that individuals tend to trade much more with counterparties located in the same city, while the effect of distance significantly diminishes after leaving the metropolitan area. They argue that location-specific goods, such as opera tickets, cultural factors, and the possibility of direct contract enforcement in case of breach may explain this result.

Lendle *et al.* (2016) extend the analysis of eBay transactions to trade between 61 countries and compare the distance effect between eBay transactions and total international trade flows. They find the effect of distance to be on average 65% smaller on eBay. The authors argue that this reduction is due to lower trade costs, specifically where the use of internet platforms reduces informational frictions. The distance effect increases with product differentiation and when trade partners speak different languages, as well as with the amount of corruption in the exporting country. The authors also control for shipping costs and find no significant change in the distance coefficient. They argue that remaining informational frictions and differences in taste might explain the remaining effect of distance.

Gomez-Herrera, Martens and Turlea (2013) go beyond e-commerce transactions enabled by eBay and use evidence from a consumer survey to approximate domestic and cross-border online trade in goods among the 27 EU Member States. They confirm that the standard gravity model performs well in explaining online cross-border trade flows and that distance-related trade costs appear to be lower for online than for offline trade, whereas the importance of other trade costs, such as language barriers, increases. They also control for online payment facilities and cost-efficiency of parcel delivery systems and find that these factors matter for online trade. They further confirm the strong home market bias that was found in earlier studies.

Finally, Cowgill and Dorobantu (2014) use data from Google's online advertising platform to analyse geographical patterns of cross-country Internet transactions among a large number of countries. Their data tracks conversion counts, *i.e.* the number of times users reach the sections of the sites where the advertisers placed the tracking code, and does not recover the actual value of the transaction. Their findings confirm that distance still matters in online trade and that additional measures of cultural and economic closeness are significant determinants of online trade.

In Cowgill and Dorobantu (2016), the authors apply their empirical strategy to transactions between Canadian and US states and re-examine the effect of a national border that has spurred a whole literature after McCallum's (1995) seminal work. They find intra-national trade to be 6.7 times higher than international trade and show that the effect is highest for sectors that feature services whose consumption is tied to a particular location, and goods that face large regulatory and bureaucratic hurdles at the border.

The approach of this paper is related to these studies as it also applies a gravity equation to online trade flows. But unlike any previous study, the present analysis looks at actual payment transactions in which the credit card was not physically involved. This is a very close proxy for goods and services ordered online (e-commerce) and preferable to more indirect measures based on consumer survey responses or conversion counts. In this respect, the only comparable studies are those based on eBay data. The advantage of the BBVA data is that it involves all transactions enabled by the credit card of a client, irrespective of the country, merchant or online platform involved.

2. BBVA credit card data – Overview

The data used in this report is the result of considerable investments by the BBVA group in data infrastructure and digital banking technology, resulting in the creation of *BBVA Data & Analytics*, an independent subsidiary specialised in Big Data applications. With a market share of 15% in Spain, several hundreds of million transactions are electronically processed by BBVA on a busy day (only 5.5% of these were related to online transactions in 2015). The information available for this study consists of all transactions that private BBVA customers in Spain have carried out with their credit card in 2015. For each transaction, the following data are recorded: randomised customer identifiers, customer's age, customer's gender, purchase amount, exact distance between customer and merchant, customer's region, merchant's region or merchant country when abroad.[5] The geographic location of client and merchant is identified by the address (in the case of domestic transactions) or the country that is related to the bank account.

The data is restricted to include only online transactions, proxied by *card not present* transactions, implying that the credit card was not physically involved for the transactions. This is usually the case when a customer realises a purchase via a home computer or mobile device, *i.e.* when a product was paid for online. The data that was available for the analysis is limited to transactions taking place in 2015. In principle however, the data is updated on a daily base and similar exercises could be repeated with several periods in the future.[6]

The total number of online transactions recorded was 45.8 million in 2015, with a total transaction value of several billion euros. Because business customers are excluded from the sample, close to 60% of the total transaction value are represented in the data analysed in this paper, which account for over 96% of all online transactions of *private* customers. About 50% of these transactions were outward bound, to a total of 115 countries. Yet, these cross-border payments are highly concentrated in only a few countries, with Great Britain, Ireland and the Netherlands alone explaining about 85% of transactions involving foreign merchants (see Figure 1).[7] This distribution is partly explained by the fact that the data refers to monetary transactions rather than trade flows. Thus, in many cases, monetary transactions will be linked to the geographic location of merchants' fiscal headquarter and not resemble the actual shipping route. It is, therefore, *a priori* far from obvious, whether traditional trade determinants will remain important for this kind of data, a question that this paper tries to address.

Figure 2 shows the origin of payment flows. Madrid (20%), Cataluña (19%) and Andalucía (15%) jointly account for more than half of the number of all transactions. The distribution of the total transaction value is very similar when looking at cross-border flows only, where the three regions mentioned above account for 53% of the total transaction value (not shown in the Figure).

The following analysis is based mainly on two variables from the BBVA transaction level data: the transaction value in Euro and the distance between the buyer and the seller in kilometres. Details with respect to the data sets used and required calculations are provided below.

For all of the following results, an important caveat applies: data is restricted to customers of BBVA and to transactions that were enabled using the BBVA credit cards. Results are therefore representative for other Spanish customers only to the extent that these are similar in terms of credit card usage.

2.1. Domestic e-commerce

2.1.1. Regional data and empirical strategy

The analysis of domestic trade patterns uses an aggregation of all individual purchase operations between any two out of the 19 (NUTS-2) Spanish regions for a given class of merchant. BBVA classifies all merchants registered in Spain into 17 distinct categories, namely Bars and Restaurants, Contents, Home, Fashion, Transportation, Leisure, Food, Sports and Toys, Tech, Travel, Health, Auto, Wellness and Beauty, Hotel Service, Property Service, Hyper, and Bank. Thus, one exemplary observation is the sum of all purchases of customers registered in Cataluña from merchants that are selling Fashion and are registered in Madrid. This implies a potential maximum of 6 137 observations (19*19 region pairings for 17 product categories). In reality, BBVA registered 4 760 non-zero transactions. Zero-trade flow observations are manually added to the data to yield a total of 6 137 observations. The geographic distance for each observation is calculated as simple average over all great circular distances between the client's and the merchant's address as registered in the Spanish banking system. For the manually added observations (zero transaction flows), distances are approximated by the distance between the largest cities in each region.[8]

The data is used to run variants of the following simple gravity equation:[9]

$$ln\, t_{cms} = \alpha + \beta_1\, ln\, GDP_c + \beta_2\, ln\, GDP_m + \beta_3\, ln\, \widehat{dist}_{cms} + \beta_5 H_{cm} + \boldsymbol{\gamma G_{cms}} + \varepsilon_{cms}\,,$$

where $\ln t_{cms}$ is the natural logarithm of the total transaction volume between client region (c) and merchant region (m) for a given product category (s). Because the gravity regressions are specified in logs, the manually created zero-trade flow observations are dropped from the model when using linear estimation techniques. The estimation will therefore in many cases rely on the Poisson regression model proposed by Silva and Tenreyo (2006), using a pseudo maximum likelihood algorithm. $\ln GDP_{\mathrm{c}}$ and $\ln GDP_{\mathrm{m}}$ are the logs of client region and merchant region GDP respectively. $\widehat{dist}_{\mathrm{cm}}$ is the (arithmetic) mean distance between the registered addresses of all individual clients and merchants contributing to the aggregate transaction flow. H_{cm} is an indicator variable equal to 1 if client and merchant are registered in the same Spanish region and captures an inverse of the domestic border effect for trade between Spanish regions. The literature has estimated this border effect to be negative (calling it a home bias) and accordingly the sign of the coefficient is expected to be positive (*e.g.* Wolf, 2000). $\boldsymbol{G_{cms}}$ is a vector of client region, merchant region, or merchant type specific controls or fixed effects. ε_{cms} is the estimation residual, assumed to be random.

Adding client and merchant specific regional fixed effects is a standard solution to control for general equilibrium price differentials between regions or countries in cross-sectional gravity equations (Feenstra, 2016). These differentials reflect that trade between two regions or countries does not only depend on characteristics of the regions directly involved but also trade costs and barriers that each of the countries faces vis-a-vis any other country or region. Accordingly, they are usually referred to as *multilateral resistance* terms (Anderson and van Wincoop, 2003).

2.1.2. Regression results for domestic e-commerce

Table 1 shows results for the basic gravity model applied to the region and product category aggregate data. Column 1 is a basic linear regression where the value of online transactions between two specific regions in a certain product category is regressed on distance, client and merchant region GDP in logs and an indicator variable equal to one if client and merchant region are the same, capturing the home bias. Following the literature (*e.g.* Yotov, 2016), standard errors are clustered at the region-pair level. Because transaction values are expressed in logs in the linear model, all zero-trade observations are dropped from the model.

The results for Column 1 are in line with standard gravity results for traditional trade flows. Purchases (*i.e.* imports) decrease with distance to the trading partner and increase with the economic size (*i.e.* GDP) of both trading partners. The positive coefficient on the *same region* dummy implies that the transaction value between clients and merchants registered in the same region tends to be higher than the value of transactions for trade between clients and merchants registered in different regions of Spain (home bias). All results are highly statistically significant.

Column 2 adds category, client-region and merchant-region fixed effects to control for unobserved effects in particular regions or for specific product categories. The economic size variables are accordingly dropped from the model and the explanatory power of the model measured by R^2 rises from 50% to 73%. Both coefficients of interest remain highly statistically significant and with the expected sign.

Column 3 adds category specific client and merchant-region fixed effects. As the product category dimension can technically be treated as a panel dimension, this is analogue to the use of time specific exporter and importer fixed effects in panel level data with a time dimension. The use of interacted fixed effects in panel level data has been recommended in the literature as control for omitted price index terms, *i.e.* multilateral resistance (*e.g.* Hillberry and Hummels, 2003 and Yotov, 2016) and should accordingly be used also for the category-panel. The R^2 increases to 91%.

Both the coefficient on distance and the *same region* dummy remain highly statistically significant. The distance coefficient of -0.67 implies that a 10% increase in distance between the two trading partners lowers purchases by 6.7%. While a direct comparison is problematic due to the different data used, previous literature typically finds the effect of distance to be slightly larger. Coughlin and Novy (2012) for example find an effect of 10.7% when looking at panel level data for the US and ignoring the product category dimension. Wolf (2000) and Hillberry and Hummels (2003) also find effects closer to 10%. The coefficient on the *same region* dummy (0.64) implies that the value of transactions within a region is on average 1.9 ($e^{0.64}$) times higher than the value of transactions between regions. The measured border effect is in between the corresponding effects obtained in Hillberry and Hummels (2003) [1.55] and both Wolf (2002) and Coughlin and Novy (2012) [4.4].

Column 4 follows the literature, arguing that ordinary least square regressions can lead to biased results in log-linear gravity equations as all observations with zero-transaction value are dropped from the model irrespective of the information they may contain (*e.g.* Santos Silva and Tenreyro, 2006 and Yotov *et al.* 2016). In column 4, the regression is therefore repeated using the Poisson Pseudo Maximum Likelihood (PPML) estimator that has been suggested by Santos Silva and Tenreyro (2006). This increase the sample size to 5 194 as several observations containing zero-transaction value are now included in the estimation.[10] As the results in Column 4 show, this reduces the negative effect of distance to 4.3% while the *same region* dummy ceases to be significant. This implies that while distance from the merchant still seems to matter for domestic e-commerce on average, evidence for the border effect in Spanish regions is mixed.

Column 5 uses the sector dimension to better understand how the border effect might vary for different product categories. Specifically, the estimates in column five provide separate *same region* coefficients for each product category, while still controlling for distance. Interestingly, the border effect varies significantly for different product categories. Specifically, for merchants categorised as *Hyper*, i.e. supermarkets, clients appear to be more than 18 times ($e^{2.9}$) as likely to purchase from within their home region. Not surprisingly, significant border effects are also found for purchases in the category of *Bars and Restaurants* [factor 3.9]. Other categories where clients appear to have a clear preference for purchases from their home region include *Sports and Toys* and *Hotel Services* [3.1] as well as *Food* [2.8], *Leisure* [2.6] and *Transportation* [1.7]. There appears to be no border effect for categories including *Fashion*, *Contents* or *Health*. There is a large positive effect for the *Bank* category, implying that clients are more than 17 times as likely to purchase from merchants in other regions rather than merchants in their own region. This result is probably a data artefact given that virtually all merchants involved in *bank* transactions are registered exclusively in the Basque Country.

These results illustrate how gravity estimation applied to e-commerce transactions yields very heterogeneous effects with respect to border effects for different product categories. The estimates gain plausibility given that much of the heterogeneity is in line with expectations. Given the statistical and economic significance of these differential effects, Column 6 allows the distance effect to vary across product categories while controlling for an average border effect. As the results show, the distance effect remains negative and significant across all product categories except for *Fashion*, where the p-value lies marginally above 10%, but substantially varies in size. As it turns out, distance is a particular important deterrent to transactions for the categories *Bank*, *Hyper* and *Property Services* (between -8.2% and -9.9%) while it is much less important for products such as *Travel*, *Fashion*, *Health* or *Tech* (between -1.6% and 2.4%). Interestingly, *Travel* and *Fashion* are the categories most frequently purchased by consumers. Taking these categories as a benchmark therefore seems to confirm that distance might be losing significance as a deterrent to intra-country trade in the era of e-commerce compared to the estimates obtained for offline trade in earlier studies.

Table 2 presents a different approach to illustrate heterogeneity in the domestic border effect. Column 1 repeats the simple baseline corresponding to Column 4 in Table 1 where the average border effect as measured by the *same region* dummy had been found to be statistically not different from zero. This changes when controlling for client region characteristics and their interaction with the *same region* dummy.

In particular, Column 2 confirms that for a region with an average level of education (a value of 48), the *same region* coefficient is close to zero ($3.86 - 0.08 * 48 = 0.02$), implying a marginal home bias factor of only 1.02. For a region with low value of education (at the 10^{th} percentile of the Spanish regional distribution, i.e. a value of 40), the effect is significantly higher, namely 1.93, implying that purchases from within the region are almost twice as likely as purchases from another region. For regions with a high value of education (90th percentile, *i.e.* 58) on the other hand the coefficient turns negative and the effect is 0.46, implying that purchases from within the own region are less than half as likely as purchases from another region. Accordingly, the estimates suggest that the online shopping behaviour with respect to domestic border effects varies significantly with the level of education in a given region, with costumers in regions with a high average level of educational attainment being much more likely to purchase from other regions.

Columns 3 to 5 repeat the exercise for several variables measuring moments of the dissemination of digital technologies among individuals and households in a given region. In Column 3 the percentage of individuals having used online banking from the 2016 annual report of the network society is used. The estimates imply that for a region at the mean (49%) the coefficient on the *same region* dummy is slightly negative, implying that purchases from within the domestic region are about 10% less likely than purchases from other regions (factor 0.9). For a region with a relatively high usage of online banking (90th percentile, *i.e.* 57%), cross-regional purchases are significantly more likely than within regional purchases (factor 0.4), whereas in regions with low usage of online banking (39%), within-regional purchases are 2.4 times more likely than purchases from other regions.

Column 4 shows that in regions with a low share of households connected to the internet (74%), costumers are roughly 1.65 times as likely to shop from within their own region, while the corresponding effect is reversed and the value is reduced to a factor 0.39 in regions with a high share of households with Internet connections (83%).

Column 5 confirms the previous results, suggesting that customers in regions with a low average frequency of internet usage (at least 5 times per week) are almost three times as likely to purchase from within their region whereas the purchases from other regions are close to four times (factor 0.26) as likely in regions with a high frequency of internet usage.

Finally, Column 6 uses the regional Consumer Price Index to determine whether regions with a relatively high consumption prices are more likely to purchase online from other regions. The estimates suggest that this is indeed the case but the effects are smaller than for the other variables. Accordingly, a relatively low cost region with a CPI at the 10th percentile (127.4) is more than twice as likely to purchase from within the region compared to other regions (effect of 2.1) whereas the likelihood of within-region purchases is only 52% higher in regions with relatively high prices (CPI of 133 at the 10th percentile of the distribution).

These estimates provide novel insights into the determinants of domestic e-commerce spending patterns that are in line with general expectations but have seldom been confirmed using econometric evidence. The following section intents to shed additional light on cross-border transactions.

2.2. Cross-border e-commerce

2.2.1. Cross-border data and empirical strategy

In this section the micro-level data is aggregated by merchant country, using all 19 Spanish regions as possible destinations. A disaggregation by merchant category is no longer viable as foreign merchants could not been classified by BBVA. In 2015, BBVA recorded transactions of Spanish costumers involving 115 different countries, plus Spain. Treating Spain as an additional merchant country for domestic transactions leads to a potential maximum of 2 204 (116*19) observations of which for 1 581 BBVA has observed positive transactions. The missing observations arise because not all Spanish regions have purchased products from the full set of countries. As before, missing observations are manually added with a transaction value of zero. Additionally, 66 countries are added for which there was no recorded trade flow into any of the Spanish regions but other gravity variables were available. This leads to a total of 3 458 possible observations.[11]

At the country level, distances for domestic purchases from Spain are derived as before using the geolocation of actual clients and merchants and forming the simple average great circle distance between all clients in a given region and all Spanish merchants. Distances with respect to foreign countries are calculated slightly differently, as in this case the registered address of the merchant is not available. Instead, the geographic distance between each client and the capital city of the merchant's country is used to calculate the simple average distance from each of the Spanish regions to each of the foreign countries. If no positive transaction flow was observed, the distance between the largest city in the region and the foreign capital was used instead.

Note that different from the regional exercise presented in the previous section, trade flows are now unidirectional. Thus, while costumers in a given client region can purchase from several merchant countries, merchant countries do not purchase from Spanish regions. This is different from traditional gravity models where most country pairs shows up twice in the data, with flows being recorded from country A to country B and vice versa. But as no clients of BBVA are registered in other countries in the data, only imports into a given Spanish region are observed

In the light of this data structure, aggregating purchases from within Spain irrespective of the particular Spanish merchant regions is preferred as this allows a treatment of Spain in analogy to other merchant countries. However, as purchases from within the client region are no longer distinguished from purchases from other Spanish regions, a *same region* dummy can no longer be identified. Additionally, all domestic transactions into a given Spanish region are basically treated equally, irrespective of the particular merchant region and most of its characteristics. But given that the following specifications focus on country characteristics rather than the regional (within-Spain) determinants of trade flows, this doesn't appear to be a major downside. For robustness, two additional datasets have been constructed, that allow distinguishing within-region from between-region transactions as well as to account for heterogeneous regional trade flows.

In a first variation, payment flows to Spanish merchants are further distinguished according to whether they go to the client's home region or some other Spanish region (*Rest of Spain*). Clients in a given Spanish region can accordingly purchase either from within their own region, from the *Rest of Spain* or from a foreign country. This variation is similar to the base data in that it also treats transaction flows as unidirectional, and therefore in analogy to purchases from foreign countries. The maximum number of possible observations is accordingly increased by 19 within-regional flows. On the other hand, *Rest of Spain* is no homogeneous country which is a slightly odd setting. For example, note that the GDP of *Rest of Spain* is constructed as the GDP of Spain minus the GDP of purchasing region and therefore varies by region.

In a second variation, all (bidirectional) regional trade flows within Spain (*e.g.* from Cataluña to Madrid and *vice versa*) are included and added to the unilateral payment flows to foreign countries. This implies that there are now 19*19 within-Spain transactions in addition to the international purchases. This structure uses more of the available information but is highly asymmetric as it mixes unidirectional and bidirectional flows. Additionally, Spanish merchant regions are very small when compared to foreign trade partners, significantly altering the size of some effects as will be shown below.

The variations of the data are used to run regressions of the following form:

$$ln\, t_{cm} = \alpha + \beta_1\, ln\, GDP_c + \beta_2\, ln\, GDP_m + \beta_3\, ln\, \widehat{dist}_{cm} + \beta_4 B_{cm} + \boldsymbol{\gamma G_{cm}} + \varepsilon_{cm}\,,$$

where the interpretation of most variables remains the same as above except for missing product category dimension. Furthermore, GDP_{m} is now country level GDP of Spain or foreign trade partners. The domestic *same region* (home market) dummy H_{cm} is replaced by a *foreign country* dummy B_{cm} equal to 1 if the merchant region is not Spain. This is equal to the *border dummy* that has been used in the previous literature to estimate the trade deterring effect of a border when comparing international to inter-regional trade (*e.g.* McCallum, 1995). $\boldsymbol{G_{cm}}$ resembles a vector of gravity controls that could be client region c, merchant country m or client-merchant pair specific. The vector captures classical gravity controls like a common language or joint membership in a free-trade agreement (here in particular the European Union), but also some novel controls detailed below. In most specifications, the vector also contains a set of client-region fixed effects. ε_{cm} is the estimation residual, assumed to be random.

An important downside of the data structure is that merchant country fixed effects cannot be added as they would be collinear with the *foreign country* dummy.[12] This is problematic as it prohibits the use of fixed effects to control for multilateral resistance terms in the international data but is not uncommon in the literature on border effects (*e.g.* Coughlin and Novy, 2012). Importantly, adding a time dimension would not solve this problem.

2.2.2. International e-commerce – towards a baseline regression

Table 3 develops a baseline regression for international e-commerce.[13] Column 1 runs the standard log-linear regression, implying that all zero trade flows are dropped from the model. The economic size controls (GDP) are positive, smaller than one and highly statistically significant in line with much of the literature. Distance is negative and highly significant, implying that an increase in distance by 10% leads to a decrease in trade value by slightly above 10.7%.[14] The *foreign country* dummy is -4.1 implying that domestic purchases are about 60 (=$e^{4.1}$) times larger than cross-border purchases, a very large effect compared to literature findings.[15]

Column 2 includes client region fixed effects, capturing any observed and unobserved effects that are constant for a given region and might interfere with the estimation. To the extent that client and merchant fixed effects would control for multilateral resistance terms in this setting, including client fixed effects will control for multilateral resistance with regard to purchasing regions in Spain (inward multilateral resistance).[16] The results in Column 2 are very similar to the previous results.

Column 3 estimates the model using PPML and thus allowing for the inclusion of zero trade flows. The number of observations rises significantly to 3 458 and the explanatory power of the model rises to 75%. Both the effect of distance and GDP of the merchant country are increasing in absolute size but remain highly statistically significant. The *foreign country* dummy ceases to be significant, implying that domestic and foreign purchases are equally likely.

Column 4 shows that the reason for this disappearance is that that the *foreign country* dummy hides a significant degree of heterogeneity. In particular, Column 4 adds a *European Union* dummy as well a dummy capturing *Common Currency* for client and merchant region, implying that the merchant country is part of the Euro area. Trade theory and previous literature find that a common market and a common currency are beneficial for trade, suggesting a positive coefficient on these variables (*e.g.* Frankel, 2010). While this is indeed the case for the European Union, the effect of the Euro at first sight appears to be counterintuitive, implying that a common currency reduces trade. This finding will be further discussed below. The explanatory power of the model rises to 95%.

The *foreign country* dummy is now highly statistically significant again. However, there are now several border effects implied by the model, and the interpretation of their economic significance has become more complex. Specifically, the coefficient of the *foreign country* dummy now resembles only the border effect with respect to trade with other Euro area members. As Spanish regions are coded as both, European Union and as pertaining to the Euro area, the coefficients on these variables are the same for both groups and drop out of a comparison. Accordingly, this particular border effect is close to 9.4 ($e^{2.24}$), implying that purchases from within Spain are almost ten times as likely as purchases from other Euro area members. For the border effect with respect to other EU members that do not use the Euro the differential effect of a *common currency* has to be taken into account. Specifically, compared to cross-border trade with other non-Euro EU members, domestic trade is only about 2.4 times higher.[17] Finally, compared to trade with other foreign countries that neither form part of the European Union nor are using the Euro, the implied border coefficient is 4.98, implying that domestic trade is about 145 times more likely than trade with these countries on average.[18]

These results show that there is large heterogeneity in the border effect for different sub-groups of trading partners. In particular, they show that the high average estimate presented in Column 2 is largely driven by countries outside of the European Union. In absolute terms the large size of the implied border effects, and in particular for countries outside of the European Union, are likely to be driven by at least three different factors. First, compared to overall trade patterns, cross-border e-commerce is still highly concentrated within the European Union, with relatively few purchases from other countries (see OECD, forthcoming). In our data, countries outside of the European Union account for at most 5% of total trade. If the controls of the model cannot fully explain this pattern, all remaining variation is likely to be picked up by the border dummy. Second, the border effect is known to be varying in country size and tends to be particularly large for smaller countries. Given that major partner countries outside of Europe are relatively large compared to Spain, the border effect as measured for Spain can be expected to be higher than estimates for other regions might suggest. Finally and related to that, previous literature has shown that a lack of proper controls for multilateral resistance terms is likely to lead to biased estimates with respect to the absolute size of the border effect (see Anderson and van Wincoop, 2002 and Feenstra, 2016). Given these caveats, the absolute size of the international border effects obtained from the model should be taken with a grain of salt.

Nevertheless, this paper argues that the finding of significant *differential effects* for different partner regions, that are in line with expectations seem to support the hypothesis of actual border effects driving the results. In this regard, it is important to understand why trade with non-Euro members of the European Union seems to be almost four times higher than trade with Euro area members. As it turns out, this can be explained by cross-border purchases being highly concentrated in Great Britain, accounting for over 60% of all transactions involving a foreign country. As Column 5 illustrates, controlling for a Great Britain effect turns the sign on common currency coefficient positive, which is what is to be expected according to the literature. Thus, the negative sign on *common currency* in Column 4 is a reflection of the particular importance of Great Britain for Spanish online consumers. On the other hand, both the EU effect and the effect of distance cease to be significant in Column 5.

In the case of the European Union, this is not surprising, given that the Euro Area and Great Britain jointly control for a very large part of the European common market, leaving only very limited residual variation to identify a European Union effect. By the same token, having three control variables that capture the specifics of the European Market implies that the residual variation used to identify the distance effect needs to rely on variation arising mostly from within different country groups, as the average difference in distance between the domestic market and foreign countries is likely to be captured by the *foreign* dummy, and the average difference in distance between European Union and other foreign countries is accounted for by a combination of the Great Britain, the EU membership and the common currency effect.

It is important to highlight however, that the distance effect is not exclusively driven by the European Union or Great Britain, given that the p-value for a negative distance effect reaches 10.6%, and therefore almost significance again, after excluding the *common currency* effect (Column 6). The effect of a common currency is now partly reflected in the increase in the EU (including Euro area) effect and the decrease of the Great Britain effect that now incorporates part of the relatively lower trade volumes with respect to other Euro area member countries. The implied coefficients suggest that domestic trade is only little more than twice as likely as trade with Great Britain ($e^{0.755} = 2.13$) but more than 20 times as likely as trade with other European countries clearly illustrating the role of the Great Britain effect. Column 7 allows more of the variation that arises between the markets (domestic, EU and other foreign) to be captured by the distance effect by replacing the EU dummy, with a general dummy indicating a free trade agreement between Spain and its partner country, a variable that should be less correlated with geographic location. Not surprisingly, therefore, distance turns highly significant again.

Finally, Column 8 repeats the specification of Column 4, which will serve as the baseline for the following analysis, but excludes some large multinational firms that are likely to conflate the estimates. In particular, there are several firms in the data where the location of the financial headquarter that is registered by BBVA is likely not to be representative of actual shipping destinations. To name just some obvious candidates, transactions realised through PayPal or similar payment intermediaries are mostly captured as flows to the location of the intermediary rather than the location of the seller. Additionally, large online marketplaces like Amazon or AliExpress potentially have many distribution centres or local websites that are likely to be geographically separated from the financial headquarters. In part, the location of these headquarters is responsible for the high importance of countries like Great Britain, Ireland and the Netherlands. Compared to Column 4, recalculating transaction values excluding these firms leads to a slight increase in the explanatory power of the model and lowers the absolute size and significance of the coefficient for distance. Overall, however, the gravity specification for e-commerce is qualitatively robust to the exclusion of these firms.[19] More importantly, while there is little room for improvements in the explanatory power of the fully specified model that controls for a European Union effect, rerunning the specification in Column 3 with the modified dataset substantially increases the R^2 from 75% to 92%. This suggests that large multinational enterprises and international payment services can have a quite distortive effect when it comes to the explanatory power of the empirical gravity model using payment flows rather than actual flows of goods.

2.2.3. Robustness – alternative data aggregation

Table 2 provides some robustness results with respect to structure and aggregation of the underlying data. Column 1 repeats the baseline specification (4) from Table 1. In Column 2, *same region* purchases are separated from total purchases from Spain for each region. As mentioned before, this allows controlling for an internal border or home market effect. Accordingly, each Spanish region now has one more potential trading partner, namely itself. All variables of interest are statistically significant and with the expected sign, including the *same region* dummy that implies a very large home market effect (almost factor 2 000). One potential explanation for this very large effect is delivered by Coughlin and Novy (2016) who suggest that the size of the internal border effect is decreasing in the size of a region, because internal trade costs become relatively more important when small regions are aggregated into a larger region. In Column 2, each *same region* is very small in comparison to the only other region (i.e. *Rest of Spain*), implying that internal trade frictions in each region will also be small compared to the trade frictions associated crossing the border into the *Rest of Spain*. Accordingly, one would expect the border effect to be particularly large for each of these small regions. Additionally, in difference to the regional model presented earlier, the current model cannot make use of merchant region dummies, implying that omitted variable bias might also be partly responsible for the large estimates.[20]

The other variables remain relatively stable. In economic terms it is noteworthy, that the international border effect that compares domestic purchases with purchases from non-EU countries now would imply a factor of 99 [exp(3.9-1.4-(2.1))]. Furthermore, domestic trade is estimated to be about eight times larger than trade with other EU member countries. While the absolute size of these effects should not be taken at face value, the relative ordering remains stable.

In Column 3 Spain is further dissolved, providing 19 merchant regions for each of the 19 client regions apart from the foreign countries. All variables from the baseline regression remain statistically significant and with the expected sign, while the *same region* dummy ceases to be significantly different from zero. There are two possible explanations for this. First, relative to the previous specification each client region is now relatively large with respect to the other Spanish regions that have been dissolved from one aggregate *Rest of Spain* into several smaller regions. Following the agglomeration results, the increase in the relative size of the client's home region could be responsible for the diminishing internal border effects. Second, the result is also in line with Column 4 of Table 1 in the regional specification, showing that the internal border effect significantly varies by merchant category, with the effect being not statistically different from zero on average.

On the other hand, the international border effect becomes implausible large (factor of over six million) which, apart from the caveat with respect to the absolute size of the border effects mentioned earlier, appears to be driven to some extent by the highly asymmetric structure of the data. In particular, while the data provides information on bidirectional trade flows between each of the Spanish regions, cross-border purchases are exclusively unidirectional, with products arriving from other countries into Spanish regions but virtually no trade flow going out in return. Adding counterfactual random purchases for foreign countries such that region's imports and exports are more or less balanced overall significantly reduces the implied coefficient from 15.7 to 9.4, implying a reduction in the border effect to a factor 12 000 (Column 4). Some asymmetry however remains, given that no internal trade flows are generated for foreign countries.[21]

Columns 7 and 8 repeat Columns 2 and 3 respectively, dropping a number of large multinationals from the sample to account for the "online intermediary" effect. The results are qualitatively in line with the full sample and, importantly, deliver significant results and the expected sign for distance and the international border dummy.

2.2.4. Other factors that determine cross-border e-commerce

With this section the report aims to highlight that with e-commerce new determinants of cross-border trade are becoming important. E-commerce is different from traditional trade in at least two important dimensions. First, the transactions captured in the BBVA data are mostly business-to-consumer transactions (B2C) rather than the business-to-business and global value chain transactions that dominate traditional trade statistics. Second, in order for e-commerce to develop in a given country, several conditions have to be met. In particular, e-commerce firms usually require sufficient digital infrastructure, the availability of online payment methods or a business environment that enables innovative forms of doing business (see OECD, 2018). Table 5 and 6 present specifications that intent to capture whether the availability of these factors in merchant countries is helping to explain trade patterns in a digital world. In both tables, the base specification is further extended to include two additional controls that arise from the gravity literature, namely a *common official language* indicator and an indicator capturing whether the merchant country is *landlocked* or not. Typically, trade is expected to increase for countries with common official language and to decrease for landlocked countries.[22] While not statistically significant in all specification, the two dummies always have the right sign, supporting the literature findings, and do not interfere strongly with the main effects (Column 1).

In Table 5 all purchases are used for the estimation, whereas in Table 6 some large multinationals have been removed ("online intermediary" effect). Note that this increases the explanatory power of the model in all specifications. Interestingly, the overall effect of *distance* appears to be slightly larger on average for the sample including large multinational than for the sample excluding these firms. Overall, the value varies between -0.3 and -0.9 with an average of around roughly -0.7 in the full data set and roughly -0.5 in the restricted data set. While it is difficult to explain the difference between the two data sets, it is noteworthy that both estimates are below the literature benchmark of -10.7 reported in Disdier and Head (2008). This is also in line with the results from the regional model and suggests that effect of distance on trade might indeed have become slightly less important in a digital world. Yet, given the difficulty to compare results between different data sets, this report is careful not to put strong emphasis on this result, leaving a more detailed comparison to future research.

With respect to the factors that can be thought to determine the development of e-commerce in potential merchant countries, the discussion will go through Tables 5 and 6 in parallel. Importantly, as many of the considered factors are probably correlated with economic development, all regressions include a set of dummies that are intended to control for per-capita income differentials (in terms of purchasing power parity) between the different countries. In particular, the dummies capture whether a country belongs to the first, the second or the third quartile of all included countries in terms of per-capita income. The fourth quartile, with the highest per-capita income, is used as the base category. Most coefficients turn up negatively when they are statistically significant, implying that relatively less is purchased from countries that are not in the top-quartile of the income distribution.[23]

Controlling for income, Column 2 shows that the percentage of respondents (age 15+) who reports having a credit card in a given country, according to the World Bank Global Findex Survey, is positively associated with the level of cross-border purchases from that country in both datasets, suggesting that the dissemination of payment methods that can be used online have become an important determinant of e-commerce trade flows and the location of e-commerce sellers. Specifically, the estimated coefficient suggests that an increase in the dissemination of credit cards by one percentage point is associated with an increase in purchases from a given country by 6.4% and 4.5% respectively.

Column 3 assesses whether the uptake of digital technologies at the consumer side, here measured by the number of active mobile subscriptions per 100 inhabitants (ITU) in the merchant country, can help to explain the distribution of Spanish purchase flows across countries. The coefficient is not significant and a role for the uptake of mobile broadband technologies in the merchant country as a determinant of cross-border e-commerce cannot be confirmed. It should be noted, however, the per-capita income class of countries has been kept constant for this exercise. Dropping the income controls results in a positive coefficient that is significant with p-value of 0.87 in the restricted sample. Because it cannot be assessed whether this is due to the variable spuriously picking up the omitted income controls or due to an actual effect that had been overshadowed by the income controls, no clear result with regard to this variable can be achieved. Related alternative measures from the ITU database delivered similar results.

In Column 4 the model controls for digital infrastructure that might be important enablers of e-commerce from a seller perspective. Specifically, the number of secure internet servers per 1 million people (based on Netcraft) might be an important determinant for the creation of e-commerce firms in a given country and thus have an effect on the distribution of cross-border purchases. In both datasets the variable appears as a positive and significant (at the 5% and 10% level respectively) covariate to the value of online purchases from a given country. The coefficients suggest that an increase in the number of servers by 10% is associated with in an increase in purchases from a country by around 5% on average.

The hypothesis that motivates Column 5 is that a dynamic start-up scene might be vital for the development of successful e-commerce firms in countries. The specification controls for the total number of start-up procedures that are required to register a business, as provided by the World Bank Doing Business indicators. While the coefficient is negative, the model does not identify results that are statistically different from zero. Replacing the measure with a measure of start-up costs provides significant results but causes a warning of *overfitting* in the PPML convergence process and is therefore not shown. Results from the micro-level data regressions presented below nevertheless suggest that start-up procedures might play a role in determining cross-border e-commerce.

Column 6 adds the United Postal Union (UPU) Postal Reliability Index in order to control for the development of the domestic postal system in the merchant country. As cross-border e-commerce often involves the delivery of small packages, an efficient domestic postal system might help e-commerce sellers to initiate activity in the domestic market before delving into cross-border exports. The coefficients in both data sets turn out positive and significant at the 10% and 5% level respectively. The absolute size of the coefficient implies that an increase in the index (that varies between 0 and 100) by one point is associated with an increase in cross-border purchases for that country by about 13%, a remarkably strong effect if interpreted causally. An index measuring international logistics performance from the World Bank also delivers positive and significant results in the case of the restricted dataset.

The remaining columns control for different aspects of the legislative framework that might be relevant for the development of a dynamic e-commerce landscape in potential merchant countries. Columns 7 controls for the overall rule of law in a country and thus the overall confidence in and abidance with rules of society and, in particular, the quality of contract enforcement, property rights *etc.* (see World Bank Governance Indicators). The indicator varies from -2.5 to 2.5 and the highly significant and positive coefficient implies that an increase in performance by 0.5 is associated with an increase in cross-border purchases from that country by 85%, again a very large effect, highlighting that these correlations should not be taken lightly for causal effects. Column 8 considers regulatory quality from the same data base and finds an even larger effect, supporting the hypothesis that a wellfunctioning regulatory environment is positively linked to cross-border e-commerce performance in a statistically significant way.

Finally, Columns 9 and 10 use more specific indicators of the regulatory framework with respect to e-commerce transactions. In particular, Column 9 controls for the existence of a legal framework for electronic transactions or e-signatures and Column 10 controls for the existence of a legal framework for cybercrime prevention. The indicators stem from the UNCTAD Cyberlaw Tracker. While none of the coefficients is significant for the full sample, the existence of a legal framework with respect to electronic transactions and e-signatures and, statistically more significant, with respect to cybercrime prevention show up with a positive sign and statistically significant (at the 10% and 1% level respectively) in the restricted sample.[24]

3. Micro-level data

3.1. Micro-level data and empirical strategy

In this section, the previous analysis is complemented with data that remains much closer to the micro-level observations in the data. Specifically, transaction level data of individual bank customers is aggregated according to the geographic distribution of these transactions. Accordingly, one exemplary observation resembles the purchases of individual *i* in Spanish region *c* from all merchants in region *m*. Purchases for each client are aggregated over all merchants at the regional level because no information is available on foreign merchants' exact location and explanatory variables do not vary by merchant. Explanatory and control variables are then added as before, based on either the client region or the country level.

The final data is composed of 2 332 384 individuals, performing 5 368 219 transactions with regions or countries. The average amount of each transaction is EUR 76.55. Table 1 in Annex A shows characteristics of transactions by demographic subgroups. The results show that men tend to engage on average in more transactions than women. The average transaction value increases with age, which might be due to a higher average income. Furthermore, younger people tend to be more open to online shopping than older people, which might be related to a higher average digital affinity.

As in the macro-level regressions, we use two data settings. In the first setting, we consider only domestic purchases. This provides us with 3 319 282 micro-aggregate purchases flows, with each individual potentially purchasing from 19 different regions, including their own. However, if all individuals had the same purchasing patterns, the numbers suggest that individuals were shopping from 1.4 regions only on average. With respect to the international purchase data, Spain is again treated as a single country and all domestic purchases of individual *i* are compared with purchase from other countries. This data set consists of 5 193 798 observations implying that for equal purchase patterns, each individual would have purchased from 2.2 different countries on average, including purchases from Spain. If the data is restricted to exclude purchases from large multinationals, where the payment flows are not likely to match the trade flow, the data set is reduced to 3 782 133 observations, indicating that for many individuals, all cross-border purchases from a given country are related to these firms. This alone, is an important finding as it confirms the role of large players and intermediaries like Amazon, AliExpress or PayPal for cross-border e-commerce.

The variations of the data are used to run regressions of the following form:

$$ln\, t_{icm} = \alpha + \beta_1\, ln\, GDP_c + \beta_2\, ln\, GDP_m + \beta_3\, ln\, \widehat{dist}_{icm} + \beta_4 (B,H)_{icm} + \boldsymbol{\gamma G_{icm}} + \varepsilon_{icm}\,,$$

where the interpretation of most variables remains the same as above except for the added individual dimension. GDP_{m} is now either regional or country level GDP, depending on the setting. Furthermore, the data now incorporates a home market dummy (H_{icm}) and the *foreign country* dummy (B_{icm}) derived from variation at the individual level. $\boldsymbol{G_{icm}}$ resembles a vector of gravity controls that can be individual i, client region *c*, or merchant country *m* specific. Available controls at the individual level include the age and gender of the person. Other variables have not been available in the data provided by BBVA. The estimation residual ε_{icm} is assumed to be random.

Due to the size of the data set and the BBVA data infrastructure that prioritises data security over accessibility, both the use of individual client fixed effect and PPML estimation techniques was not feasible with the individual level data. Accordingly, the data set is confined to the use of the observed positive value transactions.

3.2. Regression results for domestic e-commerce

Table 7 presents regressions for regional transactions. In line with the macro-level approach and much of the gravity literature focusing on the domestic border effect or *home bias*, regresses the log of the purchase value on economic size,[25] distance and the *same region* dummy. Both the economic size indicators show up with the expected sign and are highly significant. The coefficient on the *same region* dummy is also positive and suggests that individuals purchases on average 27% more from within their home region rather than other regions ($e^{0.24}$). Surprisingly, the distance effect shows up with highly statistically significant positive sign, implying that purchase increase slightly by 0.33% when increasing distance between client and merchant by 10%.

One possible explanation for this result would be the dominance of certain regions as merchant location. Thus, if there is a large number of widely dispersed individuals all purchasing from relatively distant Barcelona or Madrid, one might indeed expect a positive sign for distance. To account for this and other unobserved heterogeneity between different regions, Column 2 adds client and merchant region fixed effects to the model. As the results change very little and distance remains with a positive sign, it can be concluded that the important economic hubs attached to certain regions are not driving the result.

Column 3 therefore adds an additional control that allows for heterogeneous distance effects for very close compared to distant transactions. Specifically, the specification adds a dummy equal to one if the distance between client and merchant is larger than 50km. The coefficient on distance now resembles the effect for purchases within a radius of 50km, whereas the interaction coefficient has to be added in order to obtain the distance effect for purchases from merchants that are further away. As it turns out, the interaction effect is negative and highly statistically significant. For purchases within the 50km radius, the model now suggests that purchase increase by 0.8% if increasing the distance between buyer and seller by 10%. On the other hand, beyond the 50km radius, increasing the distance by 10% leads to a *reduction* in purchases by roughly 0.3%. While this is a very small effect compared to the literature estimates, it is qualitatively in line with general trade theory. The same region effect is reduced to 16%.

Unfortunately, product level data at the micro-level was not available in the dataset provided by BBVA and it can therefore not be assessed to what extend the result of a positive coefficient for close-by purchases might be driven by particular product categories. As the analysis of Table 1 at the aggregate level has shown, the distance coefficient varies widely across product categories. In particular, some purchase categories, including large supermarkets (*Hyper*) as well as Bars and Restaurants tend to be highly local, with purchases decreasing significantly with increasing distance from the client. But this effect does not distinguish between far and close distances and might therefore hide significant heterogeneity in the effect.[26] In particular, for close range transactions, individual clients might substitute e-commerce for offline purchases only in cases where the merchant is beyond "walking distance". If the merchant is very close, customers might often find it easier to pass by directly instead of ordering online.[27] A deeper analysis of these issues must be left for future research. The findings are also in line with the ongoing transformation of e-commerce that now involves more and more inherently local transactions that are mostly driven by factors other than distance (see OECD, forthcoming).

Column 4 further uses the individual dimension of the data to control for client characteristics, and in particular age and gender. Both factors turn out to be highly statistically significant covariates of e-commerce purchases, with men purchasing 4.3% more on average than women and ten more years of age increasing purchases by 3.5%. All the main effects remain basically unaltered when controlling for age and gender.

3.3. Regression results for international e-commerce

Table 8 shows micro-regressions for the cross-border data, including purchases from the identified large multinationals. Compared to the regional data regression, the distance coefficient is slightly smaller yet still positive without controlling for close-range specific effects. Both client and merchant GDP enter the regression positively and are highly significant. The *foreign country* or international border effect is negative as expected and highly significant.

Column 2 allows the distance effect to vary for close by and distant merchants and confirms the findings from the regional level data. In particular, an increase in distance for close-by purchases by 10% increases purchases by about 1%. The effect is in the same ballpark compared to the regional estimates which is in line with the effect being driven from variation mostly at the domestic level. For purchase from further away than 50km, each 10% increase in distance decreases purchases by roughly 2.2%. This is still significantly smaller than comparable estimates from the literature and seems to confirm results from the aggregate regressions indicating a relatively smaller negative effect of distance on e-commerce transactions. While this could be explained by services purchased via e-commerce now often being *digitally delivered*, such as apps, more research is needed to confirm this hypothesis. In particular, as earlier research found distance to remain a significant predictor of cross-border consumption even for digital goods due to its association with cultural differences (see Cowgill and Dorobantu, 2014).

Column 3 shows that the results remain robust after controlling for client region fixed effects. A regression using *European Union* membership as a further control variable in the full data seemed to confirm the heterogeneous effects found in the aggregate data and also produced the distance effect and the interaction with the expected sign. The results are not shown because an apparent program failure led to a loss of over 500 000 observations, allowing no conclusive comparison between the different estimates.

Accordingly, from Column 4 onwards, the regressions have been run for the restricted data set excluding large multinationals or major payment operators. This reduces the number of observations from over five to less than four million. Column 4 repeats Column 1 with the restricted data. The results confirm the sign on most major explanatory variables except for merchant country GDP that enters into the regression with a negative effect. On the other hand, the absolute size of all other variables increases significantly.

Column 5 shows that these results do not vary significantly when controlling for *European Union* and *Euro Area* membership of the merchant country, which is now possible. Both coefficients indicate a positive and highly significant effect. This seems to suggest that the effect of *Great Britain*, that was confounding the effect of a common currency in the aggregate regressions, is less pronounced for individual level transactions. The negative coefficient on merchant country GDP decreases significantly in absolute terms but remains negative, indicating that Spanish credit card clients of BBVA are less likely to purchase from economically large countries. The fact the size of the coefficient is reduced after controlling for European trading partners seems to imply that to some extent the effect was driven by these countries. Nevertheless, the effect remains negative in all of the following regressions. Whether this is due to the specifics of the micro-level data or more broadly related to the patterns of cross-border e-commerce observed for Spanish costumers is not going to be further addressed in this paper.

Column 6 adds the minimum distance control and confirms the heterogeneity of the distance effect after excluding large multinationals. The estimates suggest that an increase in distance between client and merchant by 10% within a radius of 50km increases purchases by 0.9%, but reduces purchases by 0.54% once outside of the 50km radius. The estimates further suggest that individuals are about 2.6 times more likely to purchase domestically than from other Euro member countries ($e^{0.95}$) and roughly 2.8 times more likely to purchase from other European Union, non-Euro member countries ($e^{0.95+0.14}$). Finally, domestic purchases are about 3.5 times more likely than purchases from non-European Union members. These estimates are not implausible and might suggest that the micro-level estimates are less affected by the lack of controls for multilateral resistance terms.

Column 7 shows that the results with respect to the border effects do not change qualitatively after controlling for client region fixed effects. On the other hand, the distance effect implied by the coefficients in Column 7 remains positive even beyond the 50km radius, implying that increasing the distance between client and merchant by 10% leads to a slight increase in purchases by 0.1%.

Column 8 shows that increasing the minimum radius to 75km, returns the negative coefficient for distance, implying that an increase by 10% reduces trade between the average individual client and a given merchant country by 0.75%, slightly larger but still close to the previous results. The border effects remain in the same ballpark that came out of specification 6.

Column 9 is mainly for comparison with the results in the following table, where the *common currency* dummy was not available in the data.[28] The results should be compared to Column 7 and show that the effects with respect to distance remain relatively similar with and without the *common currency* control. Naturally, the border effects are now differently composed, with the *European Union* dummy providing an average estimate among countries using or not using the Euro. The implied economic effects are slightly larger, suggesting that domestic purchases tend to be 4.1 times larger than purchases from non-European Union members and 3.2 times larger than purchases from other members of the European Union.

Table 9 adds some additional control variables to the model. Specifically, Column 1 repeats Column 9 but adds the gender and age controls as well as a dummy indicating landlocked countries and a dummy indicating a common official language. Both individual characteristics are again positive and highly statistically significant in line with the regional level results and confirm that older male individuals are likely to purchase more online. Both the landlocked and the language dummy are highly significant and show the expected sign. The border dummies are similar to the previous estimates and the effect of distance turns negative and highly statistically significant beyond the 50km radius.

In Column 2 to 4 some of the merchant country characteristics used at the aggregate level are added to the model and confirm that individuals tend to purchase more from countries with a high level of credit card dissemination. Additionally, the two measures for the cost and the complexity of start-up procedures now show up with a negative sign and highly statistically significant, indicating that a business environment with low burdens for start-ups is positively associated with e-commerce purchases from that country.

4. Conclusions

The analysis of online payment data that has been presented in this report suggests that the gravity model of trade has substantial explanatory power when it comes to domestic and cross-border e-commerce. Typical trade cost determinants like distance or internal and international borders are confirmed to have a substantial impact on the patterns of e-commerce. The estimated effects however suggest that the role of distance might have become more complex in the digital era, with overall smaller effects on average, likely to be driven by heterogeneity in the type of products associated with each payment flow. Additionally, increasing distance between client and merchant can have a positive effect on trade in a very close radius around the client (here 50km) that highlight the potential existence of substitution effects between offline and online commerce for short distance transactions.

The analysis also reveals how large multinationals and payment intermediaries can distort the picture that the gravity model provides with respect to online transactions. Additionally, novel explanatory variables suggest that factors related to the dissemination of digital technologies, education, price level differences at the client region level, or secure servers, availability of credit cards, regulatory quality and ease of start-up procedures at the merchant country level are significantly associated with the geographic patterns e-commerce.

More research is required to better understand the encountered effects, in particular with respect to distance and the heterogeneity of effects between different product categories as well as short- and long-distance purchases and to confirm these results for different data sets. The analysis presented in this report is only a first steps towards exploring the potential that new sources of data provide for policy analysis in the age of digitalisation, where official data sources still provide a crucial benchmark but increasingly require a complementation with private sector data to address the flood of new issues arising with the digital transformation.

References

Anderson, James E and Eric van Wincoop (2003), “Gravity with gravitas: A solution to the border puzzle”, *American Economic Review*, 2003, Vol. 93, Issue 1, pp. 170–192.

Blum, Bernardo S., Avi Goldfarb (2006), “Does the internet defy the law of gravity?”, *Journal of International Economics*, Vol. 70, Issue 2, pp. 384-405, https://doi.org/10.1016/j.jinteco.2005.10.002.

Cairncross, Frances (1997), *The Death of Distance: How the Communications Revolution Will Change Our Lives*, Watertown, MA: Harvard Business School Press.

Coughlin, Cletus C. & Dennis Novy (2013), “Is the International Border Effect Larger than the Domestic Border Effect? Evidence from US Trade”, *CESifo Economic Studies*, CESifo, Vol. 59, Issue 2, pp. 249-276, June.

Cowgill, Bo, Cosmina Dorobantu (2014), *Worldwide Gravity in Online Commerce*, mimeo.

Cowgill, Bo, Cosima Dorobantu (2016), *The US-Canada Border Effect: Evidence from Online Commerce*, mimeo.

Coughlin, Cletus C., Dennis Novy (2012), “Is the International Border Effect Larger than the Domestic Border Effect? Evidence from US Trade”, *CESifo Economic Studies*, Vol. 59, Issue 2/2013, pp. 249-276, doi:10.1093/cesifo/ifs002.

Coughlin, Cletus C., Dennis Novy (2016), “Estimating Border Effects: The Impact of Spatial Aggregation”, *CEPR Discussion Paper 11226*, Centre for Economic Policy Research, Discussion Paper Series, ISSN 0265-8003.

Disdier, Anne-Célia and Keith Head (2008), “The Puzzling Persistence of the Distance Effect on Bilateral Trade”, *Review of Economics and Statistics*, Vol. 90, Issue 1. MIT Press.

Feenstra, Robert. C. (2002), “Border Effects and the Gravity Equation: Consistent Methods for Estimation”, *Scottish Journal of Political Economy*, Vol. 49, pp. 491–506, doi:10.1111/1467-9485.00244.

Feenstra, R. C. (2016), *Advanced International Trade: Theory and Evidence*, 2nd Edition, Princeton University Press.

Frankel, Jeffrey (2010), “The Estimated Trade Effects of the Euro: Why Are They Below Those from Historical Monetary Unions among Smaller Countries?”, in NBER book *Europe and the Euro*, Alberto Alesina and Francesco Giavazzi (Editors), National Bureau of Economic Research, University of Chicago Press.

Freund, Caroline and Diana Weinhold (2004), “The effect of the Internet on international trade”, *Journal of International Economics*, Vol. 62, Issue 1, pp. 171-189, https://EconPapers.repec.org/RePEc:eee:inecon:v:62:y:2004:i:1:p:171-189.

Gomez-Herrera, Estrella, Bertin Martens and Geomina Turlea (2014), “The drivers and impediments for cross-border e-commerce in the EU”, *Information Economic and Policy*, 2014, Vol. 28, Issue C, pp. 83-96.

Hillberry, R., D. Hummels (2003), “Intra-national Home Bias: Some Explanations”, *the Review of Economics and Statistics*, Vol. 85, Issue 4, pp. 1089-1092.

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas (2009), "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre", *American Economic Journal: Microeconomics*, Vol. 1, Issue 1, pp. 53-74.

INE (2014), *Encuesta sobre Equipamiento y Uso de Tecnologias de Informacion y Comunicación en los hogares 2014*, Instituto Nacional de Estadistica.

Lendle, A., M. Olarreaga, S. Schropp and P.-L. Vézina (2016), "There Goes Gravity: eBay and the Death of Distance", *Economic Journal*, Vol. 126, pp. 406–441. doi:10.1111/ecoj.12286.

McCallum, John (1995), "National borders matter: Canada-US regional trade patterns", *American Economic Review*, 1995, Vol. 85, Issue 3, pp. 615–623.

Morley, Clive, Jaume Rosselló, Maria Santana-Gallego (2014), "Gravity models for tourism demand: theory and use", *In Annals of Tourism Research*, Volume 48, pp. 1-10, ISSN 0160-7383, https://doi.org/10.1016/j.annals.2014.05.008.

OECD (forthcoming), *A Dynamic E-Commerce Landscape: Developments, Trends, and Business Models*, OECD Publishing, Paris.

OECD (2011), *OECD Guide to Measuring the Information Society 2011*, OECD Publishing Paris, https://dx.doi.org/10.1787/9789264113541-en.

Ramos, Raul (2016), "Gravity models: A tool for migration analysis", *IZA World of Labor 2016*, Vol. 239, doi: 10.15185/izawol.239.

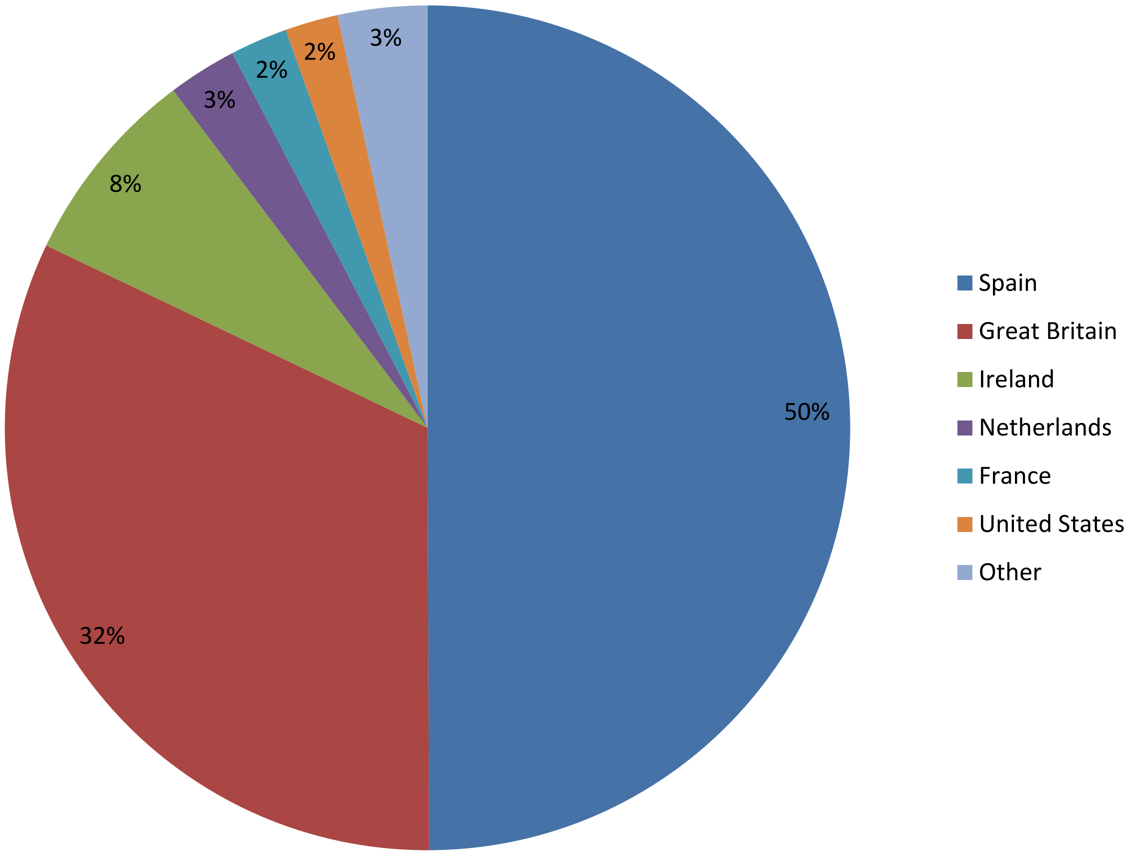
Santos Silva, João and Silvana Tenreyro (2006), "The Log of Gravity", *The Review of Economics and Statistics*, Vol. 88, Issue 4, pp. 641-658.

Wolf, Holger (2000), "Intra-national Home Bias in Trade", *Review of Economics and Statistics*, Vol. 82. No. 4, MIT Press, http://www.jstor.org/stable/2646651.

Wooldridge, Jeffrey M. (2012), *Introductory Econometrics: A Modern Approach*, South-Western College Pubs, 5th edition.

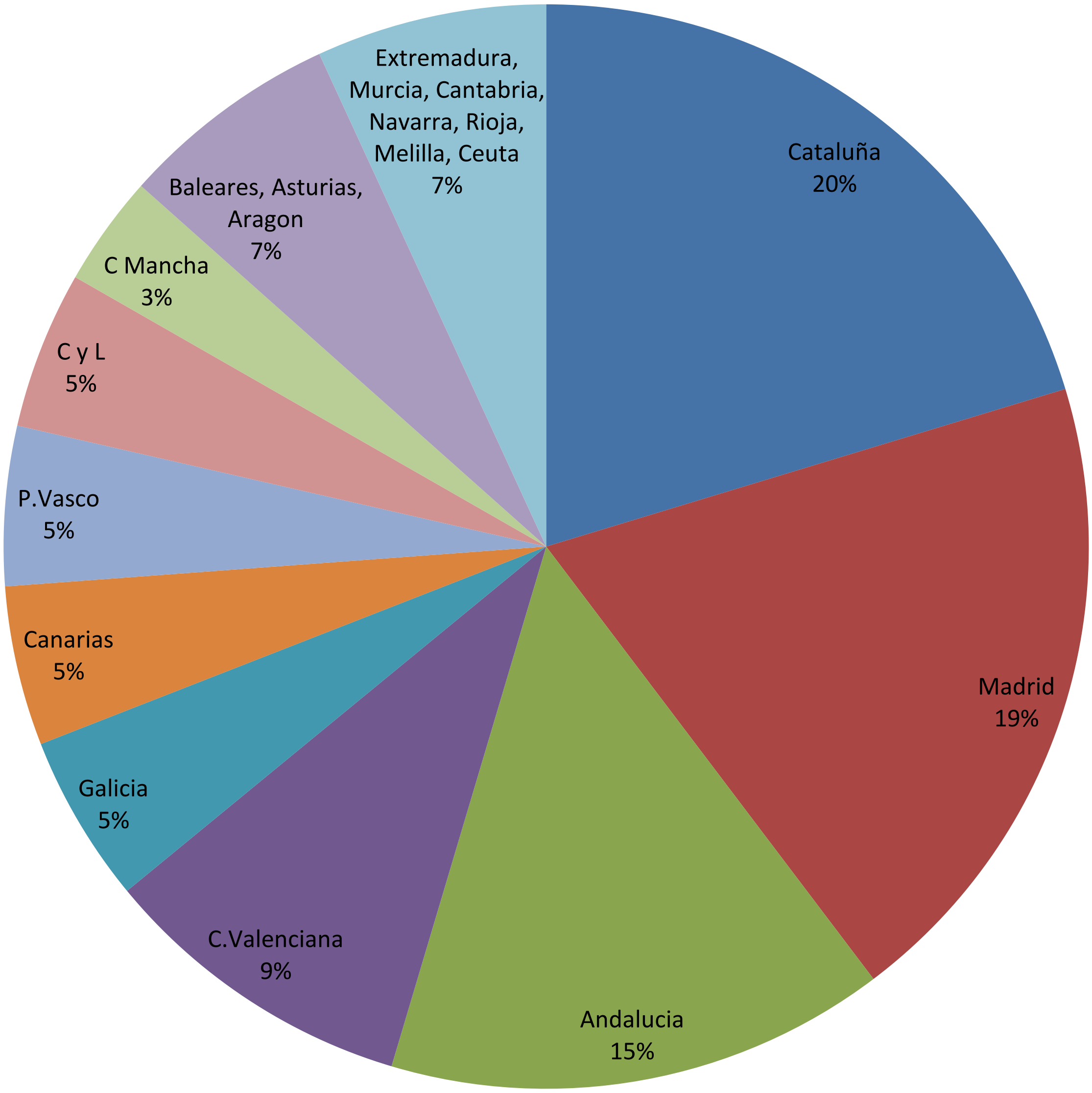
Yotov, V. Yoto, Roberta Piermantini, José-Antonio Monteiro, and Mario Larch (2016), *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, WTO.

Figure 1. Online payments by destination country



Source: BBVA data.

Figure 2. Source of payment flows by region



Source: BBVA data.

Table 1. Domestic e-commerce

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	PPML	PPML	PPML
log(Distance)	-0.367***	-0.549***	-0.665***	-0.432***	-0.308***	
	(0.0410)	(0.0454)	(0.0333)	(0.0673)	(0.0614)	
1 = Same Region	1.334***	0.891***	0.639***	-0.00719		0.259*
	(0.148)	(0.149)	(0.122)	(0.178)		(0.147)
log(Merchant GDP)	1.518***					
	(0.0274)					
log(Client GDP)	0.760***					
	(0.0213)					
Variable:					**Same Region**	**log(Distance)**
Variable * Bars & Restaurants					1.361***	-0.610***
					(0.350)	(0.0923)
Variable * Contents					0.289	-0.330***
					(0.206)	(0.0755)
Variable * Home					-0.123	-0.154*
					(0.186)	(0.0859)
Variable * Fashion					-0.103	-0.188
					(0.226)	(0.115)
Variable * Transportation					0.536*	-0.448***
					(0.282)	(0.0910)
Variable * Leisure					0.951***	-0.375***
					(0.319)	(0.0770)
Variable * Food					1.020***	-0.492***
					(0.193)	(0.0690)
Variable * Sports & Toys					1.129***	-0.639***
					(0.200)	(0.0689)
Variable * Tech					0.124	-0.238***

					(0.206)	(0.0656)
Variable * Travel					-0.319	-0.166**
					(0.205)	(0.0662)
Variable * Health					0.00261	-0.201***
					(0.157)	(0.0647)
Variable * Auto					-0.0673	-0.486***
					(0.161)	(0.0873)
Variable * Wellness & Beauty					-0.226	-0.418***
					(0.213)	(0.0792)
Variable * Hotel Services					1.118***	-0.565***
					(0.194)	(0.0738)
Variable * Property Services					-0.131	-0.824***
					(0.402)	(0.124)
Variable * Hyper					2.897***	-0.914***
					(0.509)	(0.118)
Variable * Bank					-2.817**	-0.945***
					(1.151)	(0.0797)
FX: Client, Merchant, Sector	No	Yes	No	No	No	No
FX: Client*Sector, Merchant*Sector	No	No	Yes	Yes	Yes	Yes
Cluster Variable	Distance	Distance	Distance	Distance	Distance	Distance
Observations	4 760	4 760	4 760	5 194	5 193	4 919
R-squared	0.503	0.726	0.909	0.937	0.956	0.952

Note: Product categories have been determined and defined by BBVA. Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Domestic e-commerce: client region characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PPML	PPML	PPML	PPML	PPML	PPML
log(Distance)	-0.432***	-0.520***	-0.528***	-0.583***	-0.503***	-0.404***
	(0.0673)	(0.0626)	(0.0627)	(0.0605)	(0.0640)	(0.0529)
1 = Same Region	-0.00719	3.861***	4.789***	12.34***	12.02***	8.401***
	(0.178)	(0.488)	(0.869)	(1.200)	(1.619)	(2.835)
Same Region * Education (Client)		-0.0795***				
		(0.0102)				
Same Region * Online Banking (Client)			-0.0990***			
			(0.0185)			
Same Region * HH with Internet Connection (Client)				-0.157***		
				(0.0155)		
Same Region * Frequency of Internet Usage (Client)					-0.148***	
					(0.0196)	
Same Region * CPI (Client)						-0.0629***
						(0.0216)
FX: Client*Sector, Merchant*Sector	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Variable	Distance	Distance	Distance	Distance	Distance	Distance
Observations	5 194	5 304	5 258	5 304	5 304	5 304
R-squared	0.937	0.952	0.951	0.957	0.943	0.940

Note: Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 3. Cross-border baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	OLS	OLS	PPML	PPML	PPML	PPML	PPML	PPML
log(Distance)	-1.071***	-1.089***	-1.715***	-0.569**	-0.0340	-0.230	-0.653***	-0.345*
	(0.256)	(0.263)	(0.379)	(0.233)	(0.145)	(0.143)	(0.190)	(0.193)
log(Merchant GDP)	0.813***	0.819***	1.144***	1.178***	0.659***	0.738***	0.799***	0.941***
	(0.105)	(0.106)	(0.296)	(0.248)	(0.205)	(0.191)	(0.167)	(0.221)
log(Client GDP)	0.760***							
	(0.0400)							
1 = Foreign Country	-4.072***	-4.020***	-0.181	-2.242***	-3.136***	-3.025***	-2.684***	-2.616***
	(0.774)	(0.789)	(1.092)	(0.714)	(0.498)	(0.568)	(0.681)	(0.571)
1 = EU				4.104***	0.876	2.616***		3.367***
				(0.555)	(0.541)	(0.528)		(0.526)
1 = Common Currency				-1.355**	2.240***			-0.455
				(0.666)	(0.816)			(0.677)
1 = Great Britain					4.396***	2.270***	2.502***	
					(0.664)	(0.441)	(0.417)	
1 = Free Trade Agreement							1.613***	
							(0.468)	
Sample	Base	Base	Base	Base	Base	Base	Base	Restricted
FX: Client	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Variable	Merchant	Merchant	Merchant	Merchant	Merchant	Merchant	Merchant	Merchant
Observations	1 581	1 581	3 458	3 420	3 420	3 420	3 420	3 420
R-squared	0.479	0.487	0.750	0.951	0.979	0.978	0.969	0.967

Note: Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 4. Cross-border - robustness

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	PPML	PPML	PPML	PPML	PPML	PPML
log(Distance)	-0.569**	-0.690***	-0.457**	-0.651***	-0.581***	-0.392*
	(0.233)	(0.213)	(0.202)	(0.223)	(0.211)	(0.225)
log(Merchant GDP)	1.178***	1.148***	1.346***	0.799***	0.928***	1.213***
	(0.248)	(0.221)	(0.227)	(0.134)	(0.196)	(0.228)
1 = Same Region		7.593***	-0.144	-0.402	5.970***	0.0970
		(1.969)	(0.672)	(0.764)	(1.650)	(0.784)
1 = Foreign Country	-2.242***	-2.099***	-12.46***	-7.150***	-2.246***	-11.33***
	(0.714)	(0.648)	(2.653)	(1.597)	(0.531)	(2.607)
1 = EU	4.104***	3.899***	4.480***	3.706***	3.042***	3.589***
	(0.555)	(0.519)	(0.540)	(0.588)	(0.502)	(0.517)
1 = Common Currency	-1.355**	-1.387**	-1.290*	-1.457**	-0.509	-0.406
	(0.666)	(0.670)	(0.679)	(0.737)	(0.676)	(0.657)
Sample	Base	Base/Home	Regions	Regions	Base/Home	Regions
Restricted	No	No	No	No	Yes	Yes
Counterfactual Purchases	No	No	No	Random	No	No
FX: Client	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Variable	Merchant	Merchant	Merchant	Merchant	Merchant	Merchant
Observations	3 420	3 439	3 762	7 980	3 439	3 762
R-squared	0.951	0.899	0.874	0.784	0.901	0.822

Note: Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 5. Cross-border – explaining patterns of e-commerce – full sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(12)
VARIABLES	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML
log(Distance)	-0.604**	-0.501*	-0.535**	-0.669**	-0.591**	-0.909**	-0.727**	-0.714**	-0.646**	-0.648**
	(0.245)	(0.301)	(0.250)	(0.280)	(0.264)	(0.430)	(0.295)	(0.311)	(0.259)	(0.259)
log(Merchant GDP)	1.097***	0.690**	0.895***	0.940***	1.032***	0.939***	1.046***	1.210***	1.022***	1.020***
	(0.270)	(0.301)	(0.293)	(0.312)	(0.312)	(0.286)	(0.305)	(0.268)	(0.316)	(0.318)
1 = Foreign Country	-2.144***	-1.486*	-2.217***	-2.397***	-2.754***	-1.536	-2.774***	-3.931***	-2.064***	-2.061***
	(0.721)	(0.846)	(0.820)	(0.847)	(0.883)	(0.956)	(0.749)	(0.603)	(0.739)	(0.740)
1 = EU	3.989***	3.660***	3.486***	3.281***	3.121***	3.146***	2.933***	1.958***	3.502***	3.492***
	(0.571)	(0.630)	(0.622)	(0.713)	(0.650)	(0.891)	(0.822)	(0.713)	(0.729)	(0.738)
1 = Common Currency	-1.379**	-0.603	-1.558**	-1.236**	-1.212**	-1.413**	-0.731	0.337	-1.415**	-1.416**
	(0.661)	(0.599)	(0.673)	(0.493)	(0.592)	(0.556)	(0.602)	(0.671)	(0.646)	(0.646)
1 = Landlocked	-1.623*	-1.874*	-1.918**	-1.829*	-0.797	-2.319**	-1.568	-1.047	-1.604	-1.608
	(0.946)	(1.083)	(0.886)	(0.944)	(1.174)	(1.033)	(0.991)	(0.857)	(0.977)	(0.979)
1 = Common Language	0.282	1.922***	1.406	2.202**	2.072**	4.501**	2.105**	1.028	1.113*	1.133*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(12)
	(0.848)	(0.674)	(0.957)	(0.977)	(0.905)	(1.801)	(0.987)	(1.013)	(0.621)	(0.620)
Credit card (% age 15+)		0.0644***								
		(0.0225)								
ln(mobile Broadband subs. per100)			-0.108							
			(0.666)							
Secure Internet Servers (per 1 million)				0.485*						
				(0.290)						
Start-up procedures (number)					-0.284					
					(0.199)					
Postal Reliability (UPU)						13.10*				
						(7.693)				
Rule of Law							1.682***			
							(0.403)			
Regulatory Quality								3.585***		
								(0.782)		
Legal: transactions/e-signature									1.567	
									(1.091)	
Legal: cybercrime prevention										1.987

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(12)
										(1.222)
1st Income Quartile		-1.176	-4.316***	-0.644	-2.673	1.013	0.762	4.716**	-2.876*	-2.766*
		(1.863)	(0.872)	(2.798)	(1.667)	(2.161)	(1.563)	(2.206)	(1.679)	(1.607)
2nd Income Quartile		0.616	-2.361***	-0.0826	-1.102	-0.266	1.220	4.675***	-1.708*	-1.753*
		(1.452)	(0.501)	(1.801)	(1.308)	(1.191)	(0.966)	(1.652)	(0.932)	(0.941)
3rd Income Quartile		-0.642	-2.397***	-1.404	-1.811**	-2.275**	0.120	1.710	-1.332**	-1.336**
		(1.082)	(0.632)	(1.095)	(0.893)	(0.983)	(0.988)	(1.276)	(0.656)	(0.660)
Sample	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base
Restricted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FX: Client	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Var.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.
Observations	2 736	2 717	3 116	3 325	3 287	3 154	3 401	3 401	3 287	3 287
R-squared	0.947	0.969	0.955	0.951	0.955	0.940	0.950	0.967	0.946	0.946

Note: Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 6. Cross-border – explaining patterns of e-commerce – restricted sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML
log(Distance)	-0.358*	-0.404	-0.411	-0.490**	-0.389*	-0.721**	-0.574**	-0.652**	-0.399*	-0.402*
	(0.209)	(0.277)	(0.253)	(0.218)	(0.218)	(0.291)	(0.229)	(0.276)	(0.225)	(0.225)
log(Merchant GDP)	0.866***	0.571**	0.691***	0.727***	0.794***	0.713***	0.807***	0.978***	0.768***	0.764***
	(0.236)	(0.263)	(0.240)	(0.244)	(0.249)	(0.225)	(0.248)	(0.225)	(0.256)	(0.257)
1 = Foreign Country	-2.544***	-1.920**	-2.431***	-2.783***	-2.965***	-1.899***	-3.074***	-3.770***	-2.476***	-2.471***
	(0.578)	(0.752)	(0.863)	(0.690)	(0.682)	(0.681)	(0.600)	(0.482)	(0.574)	(0.574)
1 = EU	3.309***	2.860***	2.727***	2.461***	2.481***	2.403***	2.048***	1.315**	2.698***	2.677***
	(0.547)	(0.606)	(0.644)	(0.595)	(0.603)	(0.744)	(0.697)	(0.640)	(0.628)	(0.637)
1 = Common Currency	-0.470	-0.0816	-0.587	-0.379	-0.439	-0.565	0.0508	0.852	-0.520	-0.522
	(0.676)	(0.568)	(0.800)	(0.507)	(0.581)	(0.523)	(0.582)	(0.585)	(0.667)	(0.667)
1 = Landlocked	-1.573*	-1.817**	-1.715**	-1.844**	-1.060	-2.288***	-1.729**	-1.227*	-1.620*	-1.626**
	(0.835)	(0.915)	(0.676)	(0.760)	(0.912)	(0.788)	(0.879)	(0.737)	(0.828)	(0.827)
1 = Common Language	-0.185	1.415**	1.176*	1.480*	1.278*	3.743**	1.145	0.466	0.628	0.646
	(0.707)	(0.627)	(0.620)	(0.779)	(0.701)	(1.522)	(0.808)	(0.835)	(0.523)	(0.525)
Credit card (% age 15+)		0.0446**								

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		(0.0211)								
ln(mobile Broadband subscr. per100)			0.452							
			(1.326)							
Secure Internet Servers (per 1 million)				0.518**						
				(0.264)						
Start-up procedures (number)					-0.211					
					(0.156)					
Postal Reliability (UPU)						12.87**				
						(6.286)				
Rule of Law							1.625***			
							(0.380)			
Regulatory Quality								3.027***		
								(0.673)		
Legal: transactions/e-signature									2.006*	
									(1.040)	
Legal: cybercrime prevention										2.591***
										(0.992)
1st Income Quartile		-2.679*	-3.967**	-1.352	-3.708***	0.0798	-0.382	2.763	-3.918***	-3.773***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		(1.533)	(1.728)	(2.350)	(1.312)	(1.839)	(1.255)	(1.786)	(1.315)	(1.260)
2nd Income Quartile		-0.483	-1.900	-0.278	-1.657*	-0.511	0.719	3.353**	-2.175***	-2.209***
		(1.206)	(1.246)	(1.552)	(0.966)	(0.977)	(0.812)	(1.352)	(0.729)	(0.726)
3rd Income Quartile		-1.004	-1.928***	-1.021	-1.608**	-1.872**	0.379	1.588	-1.440***	-1.446***
		(0.883)	(0.590)	(0.958)	(0.691)	(0.824)	(0.817)	(1.059)	(0.528)	(0.530)
Sample	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base
Restricted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FX: Client	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Var.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.	Merch.
Observations	3 420	2 717	3 116	3 325	3 287	3 154	3 401	3 401	3 287	3 287
R-squared	0.967	0.975	0.968	0.970	0.971	0.968	0.970	0.976	0.967	0.966

Note: Cluster Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 7. Micro level regressions - regional

	(1)	(2)	(3)	(4)
VARIABLES	OLS	OLS	OLS	OLS
log(Distance)	0.0328***	0.0358***	0.0802***	0.0831***
log(Distance) * 1(Dist>50km)			-0.1074***	-0.1103***
1 = (Distance > 50km)			0.3526***	0.3622***
1 = Same Region	0.2393***	0.2746***	0.1531***	0.1529***
log(Merchant GDP)	0.2206***			
log(Client GDP)	0.0313***			
Client Age				0.0035***
Client Gender				0.0429***
FX: Client, Merchant	No	Yes	Yes	Yes
Cluster: Client, Merchant	Yes	Yes	Yes	Yes
Observations	3 319 282	3 319 282	3 319 282	3 319 282
Akaike	6 134 461	5 960 184	5 955 756	5 944 219

Note: Robust or cluster robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 8. Micro level regressions – cross-border e-commerce

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
log(Distance)	0.0154***	0.1063***	0.1068***	0.1553***	0.1781***	0.0919***	0.0901***	0.0882***	0.0908***
log(Distance) * I(Dist>50km)		-0.317***	-0.3236***			-0.1461***	-0.0794***	-0.1626***	-0.0889***
1 = (Distance > 50km)		1.7101***	1.7872***			1.414***	1.0877***	1.5972***	1.1394***
log(Merchant GDP)	0.0301***	0.0505***	0.0508***	-0.1238***	-0.0147***	-0.0417***	-0.0385***	-0.0475***	-0.0771***
log(Client GDP)	0.019***	0.0231***		0.0699***	0.0751***	0.0718***			
1 = Foreign Country	-0.4096***	-0.1213***	-0.1217***	-1.2956***	-1.2189***	-0.9549***	-1.0629***	-0.9675***	-1.1146***
1 = EU					0.5397***	0.1528***	0.2887***	0.1466***	0.2578***
1 = Common Currency					0.1784***	0.1439***	0.1358***	0.1257***	
Min. Dist.	No	No	No	No	No	50km	50km	75km	50km
Sample	Full	Full	Full	Restricted	Restricted	Restricted	Restricted	Restricted	Restricted
FX: Client	No	No	Yes	No	No	No	Yes	Yes	Yes
Cluster Variable	No	No	Region	No	No	No	Region	Region	Region
Observations	5 193 797	5 193 797	5 193 797	3782133	3 782 133	3 782 133	3 782 133	3 782 133	3 782 133

Akaike	11 847 061	11 738 336	11 715 277	8 625 498

Note: Robust or cluster robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 9. Micro level regressions – cross-border e-commerce – restricted sample

	(1)	(2)	(3)	(4)
VARIABLES	OLS	OLS	OLS	OLS
log(Distance)	0.0928***	0.0929***	0.0928***	0.0929***
log(Distance) * I(Dist>50km)	-0.0948***	-0.0974***	-0.0964***	-0.1069***
1 = (Distance > 50km)	1.1885***	1.2034***	1.1981***	1.2591***
log(Merchant GDP)	-0.0845***	-0.1214***	-0.0855***	-0.122***
1 = Foreign Country	-0.1965***	-0.3483***	-0.2072***	-0.6167***
1 = EU	0.2359***	0.1958***	0.218***	0.0971***
1 = Landlocked	-0.2931***	-0.5255***	-0.2902***	-0.3764***
1 = Common Language	1.1963***	1.2481***	1.2007***	1.001***
Client Age	0.0021***	0.0021***	0.0021***	0.0021***
Client Gender	0.0481***	0.0456***	0.0481***	0.0485***
Credit card (% age 15+)		0.0064***		
Start-up procedures (number)			-0.0071***	
Cost of business start-up procedure				-0.04***
FX: Client	Yes	Yes	Yes	Yes
Cluster Variable	Region	Region	Region	Region
Observations	3 782 133	3 782 133	3 782 133	3 782 133
Akaike	8 482 914	8 477 065	8 482 862	8 462 504

Note: Robust or cluster robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Annex A. Appendix

Table A.1. Product classification (in Spanish)

Merchant Category Key	Description
es_auto	Automoción
es_barsandrestaurants	Bares y restaurantes
es_contents	Libros, prensa y revistas
es_fashion	Moda
es_food	Alimentación
es_health	Salud
es_home	Hogar
es_hotelservices	Alojamiento
es_hyper	Hipermercados y grandes superficies
es_leisure	Ocio y entretenimiento
es_otherservices	Otros servicios
es_propertyservices	Inmobiliaria
es_sportsandtoys	Deportes y juguetes
es_tech	Tecnología
es_transportation	Transporte
es_traveles_transportation	ViajesTransporte
es_wellnessandbeautyes	Cuidado y belleza
es_transportation	Transporte
es_travel	Viajes
es_wellnessandbeauty	Cuidado y belleza

Note: Assignment by BBVA.

.

Table A.2. Demographic characteristics

	Female		Male	
Age	Average distance (in km)	Average transaction value (in EUR)	Average distance (in km)	Average transaction value (in EUR)
Less than 30	1 104.82	55.8	1 195.16	58.03
Between 30 and 44	979.99	66.04	1 095.57	76.72
Between 45 and 64	940.62	82.58	1 024.11	96.83
More than 65	884.73	98.42	979.19	112.52

Source: BBVA Database.

Table A.3. Variables and sources

Value	Online transaction value (in EUR)
Distance	Great circle distance between bank registered address of BBVA client and registered merchant address (Spain) or capital of foreign country (in kilometers)
log(Merchant GDP)	World Bank national accounts data, and OECD National Accounts data files.
log(Client GDP)	Foreign country: World Bank national accounts data, and OECD National Accounts data files.
	Spanish autonomous regions: provided by BBVA
1 = Free Trade Agreement	Sourced from the CEPII GeoDist database.
1 = EU	Sourced from the CEPII GeoDist database.
1 = Common Currency	Sourced from the CEPII GeoDist database.
1 = Landlocked	Sourced from the CEPII GeoDist database.
1 = Common Language	Sourced from the CEPII GeoDist database.
Credit card (% age 15+)	Sourced from the World Bank Findex on Financial Inclusion and enotes the percentage of respondents aged above 15 years who report having a credit card.
In(mobile Broadband subscr. per100)	Activemovilebroadbandsubscrip: Active mobile broadband subscriptions per 100 inhabitants – ITU Database
Secure Internet Servers (per 1 million)	Secure servers are servers using encryption technology in Internet transactions. Data sourced from Netcraft (http://www.netcraft.com/) and World Bank population estimates.
Start-up procedures (number)	Startupproceduresto register Start-up procedures to register a business (number), Start-up procedures are those required to start a business, including interactions to obtain necessary permits and licenses and to complete all inscriptions, verifications, and notifications to start operations. Data are for businesses with specific characteristics of ownership, size, and type of production. World Bank, Doing Business project (http://www.doingbusiness.org/).
Cost of business start-up procedures (% of GNI per capita)	Sourced from the World Bank Doing Business Project. Denotes the cost to register a business is normalised by presenting it as a percentage of gross national income (GNI) per capita.
Postal Reliability (UPU)	Postal reliability index. Source: UPU Database
Rule of Law	Rule of Law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5. Source: Worldwide Governance Indicators
Regulatory Quality	Regulatory Quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5. Source: Worldwide Governance Indicators
Legal: transactions/e-signature	Does the country have a legal framework for electronic transactions/e-signature? – Unctad Cyberlaw
Legal: cybercrime prevention	Does the country have a legal framework for cybercrime prevention? – Unctad Cyberlaw
Client gender	Dummy of gender. Source: BBVA Data set
Client age	Age of client. Source: BBVA Data set
Client education	Percentage of achieved secondary education level and upper of client region. Source OCDE 'Education at a Glance' and Instituto Nacional de Estadística (INE) (Tempus dataset)
Education (Client)	Percentage of achieved secondary education level and upper of client region. Source OCDE 'Education at a Glance' and Instituto Nacional de Estadística (INE) (Tempus dataset)
Online Banking (Client)	Percentage of population of client region using e-banking MIET (2016) for Spanish regions
Frequency of Internet Usage (Client)	Percentage of population of client region using internet. INE (2015) for Spanish regions

End Notes

[1]The product categories, associated with each merchant, are assigned by BBVA. Table 10 in the Appendix provides an overview of the classification (only available in Spanish). *Contents* mostly relates to digital subscriptions for newspapers, magazines etc.

[2] See Gomez-Herrera *et. al.* (2014) for a similar approach.

[3] Hortaçsu *et al.* (2009) find effects of around -1% for each 10% increase in distance for trade on eBay.

[4] The rest of the paper will use the terms payment flows, online purchases, online transactions, e-commerce or trade interchangeably. It is important to keep in mind that the actual data analysis relies exclusively on payment flows.

[5] The person's identity is never revealed, including in the micro data that was used by BBVA Research. Only researchers from BBVA had access to the anonymised micro data.

[6] The data are aggregations of single real-time transactions.

[7] Due to country specific legislation, particular countries could not be identified in the data. These countries have been excluded from the analysis but potentially account for a substantial part of online transactions. For instance, the data does not contain transactions to merchants in Germany.

[8] Ignoring the category dimension the data yields 353 out of 361 (19*19) possible observations. Missing data is related to the two autonomous cities Melilla (six observations) and Ceuta (two observations), located on the north coast of Africa. These two cities also contribute substantially to the missing data at the sectoral level and can explain why many observations are dropped after controlling for merchant-category and client-category fixed effects (see below).

[9] The setup is similar to Hortaçsu *et al.* (2009), Table 3, Model IV.

[10] The number of observations is still below the potential total of 6 137. This is because the option *strict* had to be involved to obtain convergence with the interacted region-category fixed effects, which leads the program to drop several observations. The model can be estimated with all observations if a regression analogue to column 2 is invoked where regional and category fixed effects are not interacted. The results are qualitatively similar to column 4, with the *same region* dummy being insignificant and a distance coefficient of -0.33.

[11] As the BBVA micro-level data in principle captures the full population of transactions, the missing data points are actual zeros in most cases. For Germany, missing data points are due to regulatory reasons and the observation is accordingly treated as a missing rather than a zero trade flow.

[12] In particular, while each foreign country potentially shows up 19 times in the data as a merchant country, in all these cases the *foreign dummy* is equal to one. As there is no variation of the dummy at the merchant country level, the border effect would be swallowed by the fixed effects.

[13] In all the following regressions for the aggregate data, standard errors are clustered at the merchant country level, as clustering at the level of country pairs would imply a cluster size of one.

[14] This happens to be the exact value that Disdier and Head (2008) find in a meta-analysis of 1 467 estimates and is thus something like a literature benchmark.

[15] The corresponding effect was found to be 22 in McCallum (1995), but there is a large literature arguing that McCallum's effect already overestimates the true effect (*e.g.* Anderson and van Wincoop 2002, Feenstra, 2016). See below for more on the absolute size of the border effect.

[16] These are the average trade costs for imports that a given Spanish region is facing *vis-à-vis* the world that are not captured by the distance or any other variable in the model. The corresponding term for trade costs that exporting merchants are facing *vis-à-vis* the world (outward multilateral resistance) are *not* controlled for in the following specifications.

[17] This is the difference between the coefficients for Spain (4.1-1.36) and the coefficients for other EU members who are not using the Euro (4.1-2.24), *i.e.* 0.88. The border effect is then $e^{0.88} = 2.41$.

[18] (4.1-1.36) – (-2.41).

[19] Spanish law does not allow publishing the list of excluded firms. Overall, the exclusions imply a reduction in the total value represented in the data to 78.7% with respect to the full sample. In particular, the share of Great Britain, Netherlands and Ireland in total e-commerce transactions is substantially reduced from 42% to 27%. This implies for example that a repetition of Column 7 leads to a reduction in the coefficient of the Great Britain dummy from 2.5 to 1.5, implying a significantly smaller economic size of the effect (from a factor 12.2 to a factor 4.5).

[20] Note that the border effect is smaller but still very large when excluding cross-border trade flows in the regression, implying that the size of the *same region* effect is not driven by the international margin of the data (the coefficient is reduced to 6.6 still implying an effect of factor 735).

[21] Hortaçsu *et al.* (2009) also find very large implied border effects with a factor ranging from 400 to 49 000, depending on whether country or province data is used even after controlling for merchant and client region fixed effects.

[22] There are several other variables that are typically included in gravity specifications including colonial ties or adjacency (*e.g.* Yotov *et al.* 2016). In the case of Spain, colonial ties are highly correlated with the common language dummy. We include language rather than colonial ties as it seems to be a relatively more important control with respect to e-commerce. The main effects are not altered by the inclusion of a *colonial ties* dummy. *Contiguity* is not used as in the case of Spain the effect is largely captured by the *EU* dummy and as it would add a layer of complexity to the interpretation of the international border effects.

[23] As the number of observations varies between specifications according to the added variable, the quartile ranges are defined for each variable separately.

[24] Available variables indicating the existence of a legal framework for data protection and privacy online as well as consumer protection in e-commerce that were also available in the database delivered no statistically significant results.

[25] It should be noted that the choice of the controls for economic size at the micro-level is not evident and in particular with respect to the individual client, whose economic weight is probably captured better by the regional GDP per capita rather than regional GDP. Alternative estimates suggest that the coefficients on the main variables are not altered significantly by replacing GDP with per capita GDP in the regressions. As most regressions control for client region fixed effects, the difference should not matter in most regressions and client region GDP was used merely for convenience.

[26] Due to the aggregation of client specific purchases to regional purchases, close-region purchases are mostly lost for identification in the regional averages for which product specific data was available.

[27] An example could involve food delivery services that might only be invoked if the merchant is not situated right around the corner.

[28] Because the regressions have been run over a longer period of time, the availability of control variables has slightly changed from one regression model to the other. As the data access is currently decentralised across BBVA a reload of the variables was forgone for the sake of simplicity.