Exploring European Regional Trade *

Marta Santamaría, Jaume Ventura and Uğur Yeşilbayraktar

June 17, 2022

1 Introduction

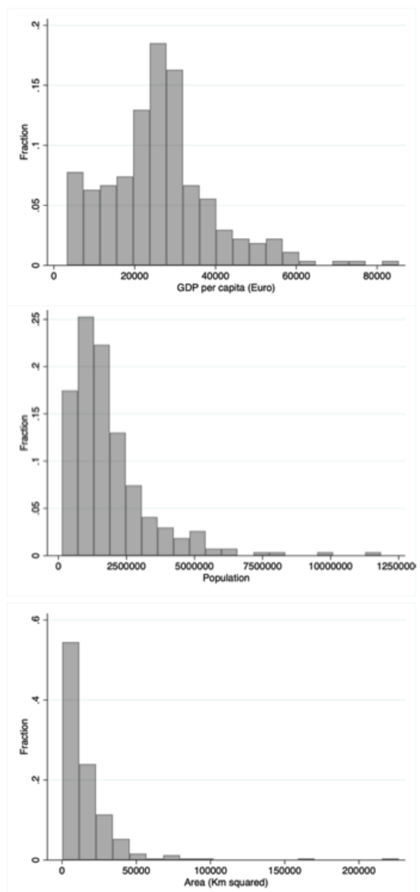
How do regions trade with each other? International trade flows across countries have been studied over decades thanks to the availability of detailed customs data. We know much less about trade patterns across and within regions due to the scarcity of data. In this paper, we use the dataset we constructed in Santamaría et al. (2020) to systematically explore for the first time trade patterns across and within European regions.

Europe is a great laboratory to explore regional trade flows. One advantage is that Europe is large, as it contains more than 500 million people and it produces about 20 percent of world GDP. Another advantage is that European regions exhibit a lot of heterogeneity, as shown in Figure 1 using data from 2011 (the starting period of our dataset). The top panel shows the distribution of incomes per capita GDP, which ranges from a low of 3,200 euros in Northwestern Bulgaria to high of 85,330 euros in Central London. The middle panel shows the distribution of populations, which ranges from

*Marta Santamaría: University of Warwick. Jaume Ventura: CREI, Universitat Pompeu Fabra and Barcelona GSE. Uğur Yeşilbayraktar: Universitat Pompeu Fabra and Barcelona GSE. E-mail: marta.santamaria@warwick.ac.uk, jventura@crei.cat, ugur.yesilbayraktar@upf.edu.

a low of 126,761 people in Valle d'Aosta to a high of 11,852,851 people in Île de France. The bottom panel shows the distribution of geographical areas, which range form a low of 160 Km^2 in Brussels to a high of 226,716 Km^2 North/East Finland.

Figure 1: Heterogeneity across European regions



The dataset constructed in Santamaría et al. (2020) is based on the European Road Freight Transport survey which collects data on truck shipments of goods in agriculture, manufacturing and mining. Thus, the dataset covers trade in goods by road. Thus, we neither cover trade in services nor trade in goods that uses another transportation mode. According to Eurostat, about half of all European trade in goods is done by road. The dataset covers 269 regions from 24 European countries between 2011 and 2017 disaggregated into 12 different industries. An important aspect of this dataset is it allows us to measure trade flows both across and within regions. Thus, for each year/industry, we have a complete matrix of bilateral trade including the diagonal entries.

The first and more salient aspect of European regional trade is that it has a strong home and country bias. Consider a shipment originating from a randomly selected European region. The probability that this shipment has a destination inside the origin region (i.e. home trade) is 40 percent. The probability that this shipment has a destination outside the origin region but inside the country of the origin region (country trade) is 41 percent. The probability that this shipment has a destination outside the country of the origin region (foreign trade) is therefore only 19 percent. To evaluate these numbers, one must recognize that the size of the destination markets is quite different. The home market is typically smaller than the country market, and the latter smaller than the foreign market. When we correct for size in the usual way,[1] we find enormous differences in the magnitudes of these types of trade. In particular, home, country and foreign trade are 469.5, 11.22 and 0.44 times what one would predict knowing only the sizes of the origin and destination markets.

The second salient aspect of European regional trade is the importance of geographic distance and national borders. The ranking of home > country > foreign trade suggests that these factors are important. Foreign trade involves producers and consumers that are farther away and do not share the same government. Both of these factors are known to have negative consequences for trade. We show that a parsimonious model that uses only

[1] That is, by dividing by the product of the sizes of the origin and destination markets. We explain how we do this in section 2.2.

national borders and the (log) of distance can explain about two-thirds of the variation in European regional trade. Obviously, a model with these elements is designed to create a bias towards home and country trade. But there is more to this. The importance of borders generates a small-country effect, namely, that regions in small countries trade more with both regions within and outside their country. The importance of geographical distance generates a remoteness effect, namely, that regions that are geographically remote in Europe should trade more with other regions inside their country, and less with regions outside. We observe that both the small-country and remoteness effects are present in the European regional data.

We consider increasingly sophisticated versions of the model that allow for more flexible specifications of distance and border effects. First, we allow for a variable elasticity of trade to distance. This does not make much of a difference, however. Second, we allow border effects to be different for region pairs that have a common language or currency. We find that both sharing a language and a currency reduce the border effect, with the impact of a common language being larger than that of a common currency. Finally, we estimate a different border effect for each country pair. We observe that the border effect is quite heterogeneous. Even though the data suggests that all these refinements are capturing aspects of the data, these more flexible and sophisticated versions of the distance and border effects do not much to the model's ability to explain the variation in the data.

A third salient aspect of European regional trade is that the strong home bias in trade cannot be explained by geographical distance and national borders. There are few observations of home trade, 269 out of 73,361, but these observations stand out for their size since they add to 40 percent of all trade. To determine the source of this home bias, we exploit a special feature of the data. Due to government structure differences, in some countries these NUTS2 and NUTS1 regions are only statistical regions created for the purpose of sharing data with Eurostat, while in other countries NUTS2 and NUTS1 regions coincide with political divisions with different levels of self government. This allows us to test whether the home bias effect emerges in all regions, or whether the home bias effect emerges only when it when coincides with political regional borders. The results strongly suggest that its is borders again that generate the home bias. When we separate statistical and political regions and show that it is the later and not the former that exhibit a large home bias in trade. Thus, it seems that regional borders matter as much or more than national borders.

We also decompose trade into intensive (average size of shipments) and extensive (average shipments per industry and number of industries) margins. First, these margins do not contribute equally to each type of trade. While 54% of the variance of home trade is explained by the intensive margin, the intensive margin explains 89.5% of the variance of foreign trade.

Finally, we use a matrix of social interactions from Bailey et al. (2020) to compare trade and social interactions. Using all trade interactions (originating at home, in the country or internationally) we see that trade interactions are correlated with social interactions. However, this is no longer true if we consider each type of trade (home, country and foreign) separately. Social interactions are positively correlated with country interactions but they are negatively correlated with home trade and barely correlated at all with foreign trade. The patterns turn out to be remarkably similar.

1.1 Related Literature

There is an abundance of papers that study trade across countries within the gravity framework. Head and Mayer (2014) provide an extensive review of this literature and the improvements in the methods since being introduced by Tinbergen (1962). Due to the scarcity of data at the subnational level, with the exception of Canada, McCallum (1995) and the US, Anderson and Van Wincoop (2003), most studies have focused on country-level trade flows. The aim of this strand of literature is to explore the factors that shape trade flows, with a special interest on the effect of distance and the effect of borders. We build on these previous studies and use the gravity theoretical framework to help guide our analysis. Our main contribution relative to previous work (Wei (1996); Chen (2004)) is that we use both the empirical gravity model (fixed regression model) and the structural gravity framework to explore how both methods perform as predictors of rich data on trade flows across Europe. In addition, we test the performance of these models to trade within-regions, across-regions and across-countries.

A special feature of our data is that our unit of observation is the region. We can measure regions trading with other foreign regions as well as trade flows across regions in the same country and within a given region. These dimensions of our European trade data allow us to explore a set of questions related to how goods move across subnational units. The number of papers that focus on regional trade flows has been relatively scarce, largely due to the unavailability of data. Nascent work on regional trade has been centred around US Hillberry and Hummels (2008) (Coughlin and Novy (2012)). [2] Two exceptions are Wrona (2018), that uses intra-regional data from Japan to estimate the border effect between East and West-Japan and Nitsch and Wolf (2013) that use a truck shipments survey from Germany to study the effects of the Inner German border in post-reunification trade flows. We contribute to this literature by providing a unified study of regional trade, it's obstacles and determinants, within and across countries, for 24 different European countries. Among the papers studying subnational barriers to trade, a few papers have hypothesised the inexistence of regional border

[2]Due to the availability of the US Commodity Flows Survey that documents trade flows at highly geographically disaggregated level.

effects on trade. A first study in this line was Head et al. (2002), showing that the mismeasurement of internal distances can lead to the estimation of "illusory" border effects. In the same spirit, Hillberry and Hummels (2008) claim that previous findings on the existence of a "state border effect" in the United States were not using the correct level of disaggregation in the data. Once the data is used at the zip-code level, with accurate measures of distance and dropping wholesale shipments, there is no State border effect in the US. We challenge this finding by showing that European regions display a home bias effect even when we use actual distances covered by the shipments. In countries were regions are administrative/political units, trade is reduced when moving across regions while in countries without subnational regions, trade flows freely within the country.

Finally, our findings uncover heterogeneous levels of international and domestic barriers faced by different European countries, contributing to the wide body of research devoted to the effects of the European Union (for a recent survey see (Head and Mayer, 2021). Previous work has studied the heterogeneity of trade barriers in Europe across industries (Chen and Novy, 2011). We contribute with a detailed analysis of trade flows at the region level, providing evidence of a large variation in the effects of distance, borders and other internal barriers across countries and regions in Europe. While the European integration process has been at work for decades, there is considerable heterogeneity in the integration level of regions, even for regions in the same country.

2 A first look at the data

In this section we describe our dataset and provide a first look at the patterns of regional trade in Europe. The bottom line is simple: regions trade with themselves much more than with other regions within the same country, and regions trade with regions within the same country much more than with regions in other countries. This ranking of home > country > foreign trade is not surprising, but the magnitude of the differences might be.

2.1 The dataset

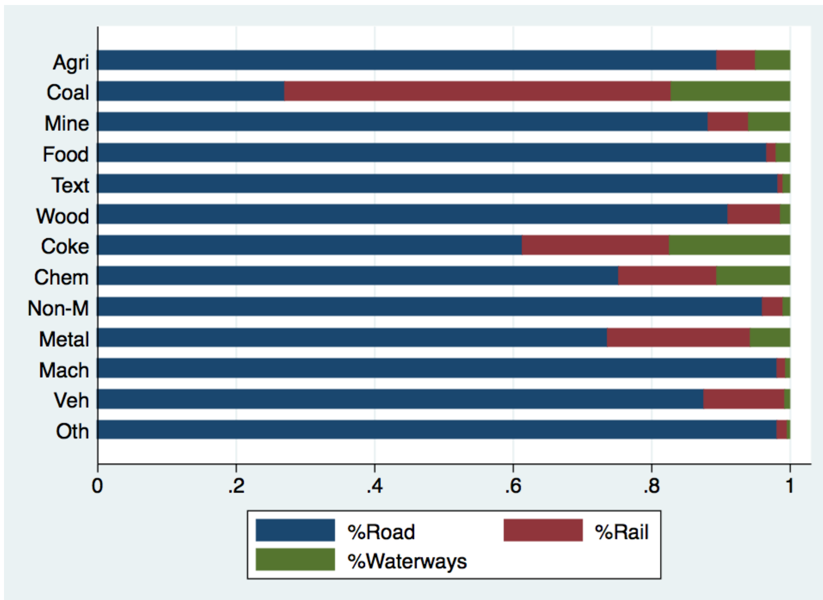
We use the dataset of regional trade flows across European regions constructed by Santamaría et al. (2020) using the European Road Freight Transport survey. This dataset covers trade in goods among 269 regions from 24 European countries between 2011 and 2017. This trade is disaggregated into 12 different industries that cover essentially all of agriculture, mining and manufacturing.

The European Road Freight Survey collects data on truck shipments. According to Eurostat, 70 percent of all European trade in goods is inland trade and 30 percent is sea trade. There are 13 categories in the European Road Freight Survey that cover all of agriculture, mining and manufacturing. Figure 2 shows the share of road trade in inland trade for each of these industries. Except for *Coal and lignite, crude petroleum and natural gas*, road trade is by far the most prevalent mode of inland transportation. This is why Santamaría et al. (2020) dropped this industry and the dataset contains the remaining 12 industries.

The European Road Freight Survey collects data adhering to the geographic divisions presented by the Nomenclature of Territorial Units for Statistics (NUTS) classification. The NUTS classification is a hierarchical system for dividing up the economic territory of the EU, the UK and the EFTA member countries for the purposes of collection, development and harmonisation of European regional statistics. Our regions are defined by the NUTS2 classification.

The dataset contains the value of goods shipped among all region pairs

Figure 2: Inland trade by industry



for all industries and years. Since trade flows vary little between 2011 and 2017, we use averages over the entire period and ignore the time dimension. In the main text of the paper, we always report the results obtained with an aggregate bilateral trade matrix that also averages across industries. In the appendix, we also report the results obtained using each of the 12 industry bilateral trade matrices. Whenever relevant, we discuss in the main text the most notable differences between the results obtained with the average matrix and the industry matrices.

Each of these matrices takes the following form:

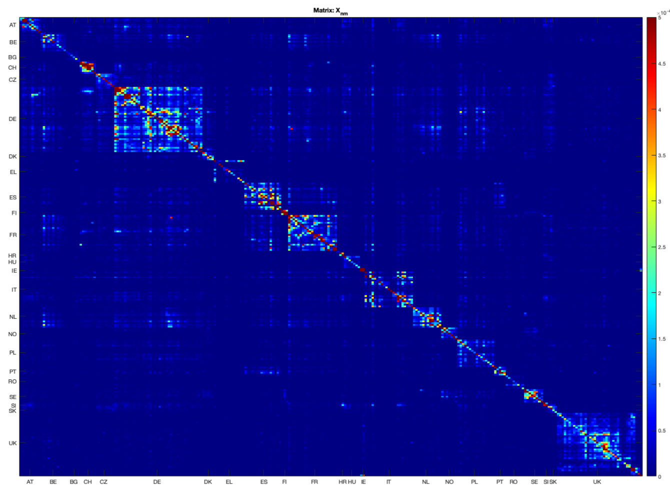
$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NN} \end{bmatrix} \tag{1}$$

where X_{nm} is the total value of shipments of goods from origin n to destination m. We measure shipments as a share of total shipments: ${}_{nm}X_{nm} = 1$. Thus, X_{nm} is the probability that a shipment of goods originates in n and has destination m.

Figure 3 shows a heat map of the matrix of bilateral trade. We refer to the entries in the main diagonal as *home* trade because they record trade within regions. Despite being a small set of entries (269 out of 72,361), each of them contains a lot of trade. Adding them, we find that home trade constitutes 40% of all European regional trade. We refer to the off-diagonal entries such that origin and destination regions are in the same country as *country* trade. Since regions within a country have been listed together, these entries can be identified in Figure 3 as the squares centered around the diagonal (without including the latter). Larger squares refer to countries with more regions, such as Germany or France. Smaller squares refer to countries with fewer regions such as Portugal or Ireland. Country trade entries tend to contain less trade than home trade entries. But there are many more country trade entries (4,958 out of 72,361) and, adding them, we find that country trade constitutes about 41% of all European regional trade. Finally, we refer to the remaining off-diagonal entries as *foreign* trade. We can identify these entries in Figure 3 as the off-diagonal entries outside the squares. Though most of the entries are foreign trade (67,134 out of 72,361), each of them contains little trade. This is why adding them we find that foreign trade constitutes only 19% of all European regional trade. There is therefore a strong bias towards home and country trade in our data.

The matrix in Figure 3 contains a fair amount of zeros. Not surprisingly, there are no zeros for home trade. But there are a few zeros for country trade: 157 out of 4,958 region pairs. And there are many more for zeros for foreign trade: 25,699 out of 67,134 region pairs. This distribution of zeros is also consistent with a strong home and country biases in European regional

Figure 3: Bilateral trade matrix for European regions



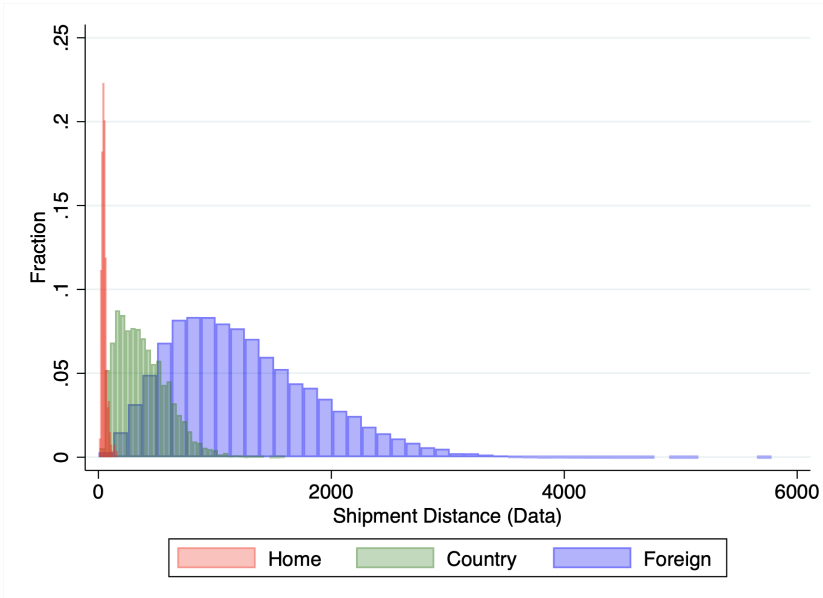
trade.

What explains these biases? A prime suspect is distance. The distance traveled by shipments classified as home, country and foreign trade is not the same. Fortunately, the European Road Freight Survey survey provides the actual distance traveled by each individual shipment, including shipments within and across regions. Figure 4 shows the histograms for distance traveled for home, country and foreign trade separately. The average distance traveled for the different types of trade is 21.2 Kms, 223.0 Kms and 631.9 Kms, respectively. There is little overlap, for instance, between the histograms for home and foreign trade.

2.2 Normalized market shares

Our goal is to understand the shape of the trade probabilities contained in the matrix of bilateral trade. Which region pairs have strong trading relationships? Which ones have weak trading relationships? What are the

Figure 4: Home, country and foreign distances



factors that shape the trading relationship of a given region pair?

To answer these questions, we need a benchmark that is size free. To understand why, consider the case of Catalonia and La Rioja, two regions in Spain. The probabilities of a sale to the Basque Country, another region in Spain, for Catalonia and La Rioja are 0.000226 and 0.0000542, respectively. The probabilities of a purchase by the Basque Country for Catalonia and La Rioja are 0.0004281 and 0.0000601, respectively. Catalonia's trade probabilities are one order of magnitude larger than those of La Rioja. Does this mean that Catalonia has a stronger trade relationship with the Basque Country than La Rioja? This would be an absurd conclusion, we think, since Catalonia's population is 7.6 million while La Rioja's is 0.3 million. It is therefore almost inevitable that Catalonia trades more with the Basque Country than La Rioja. The size of origin and destination regions matters and we need to correct for this.

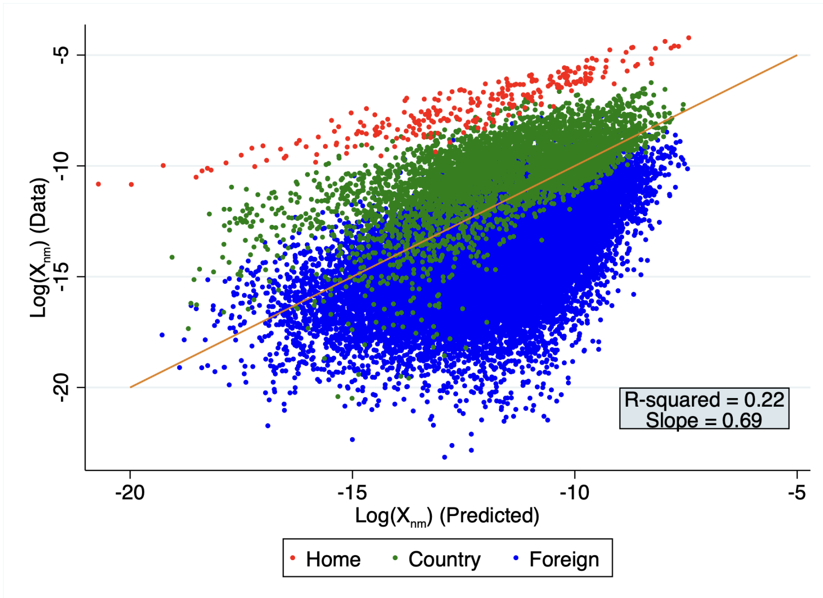
To determine how to correct for size, let us define two events: (i) $O_n =$

a shipment has origin n, and (ii) D_m = a shipment has destination m. The probability of these two events are $X_n^O \equiv \sum_l X_{nl}$ and $X_m^D \equiv_k X_{km}$, respectively. Let us now propose this independence benchmark: "the probability of a shipment from origin n to destination m should be $X_n^O X_m^D$." This benchmark essentially says that the events O_n and D_m are pairwise independent. One can interpret this benchmark as a theoretical assertion or as a forecast with limited information. A theory asserting that all sellers have the same probability of trading with a given buyer and all buyers have the same probability of trading with a given seller implies that $X_{nm} = X_n^O X_m^D$. If we only know the sizes of regions n and m, the best forecast for their trade probability is $X_{nm} = X_n^O X_m^D$. In both interpretations, the independence benchmark captures the idea that bilateral trade is independent of how far the trading partners are in terms of geographical distance, political institutions, factor endowments, tastes, and so on. Thus, we can use deviations from this benchmark to learn about the tole of these factors shaping trade relationships.

Figure 5 plots $\ln(X_{nm})$ against $\ln\left(X_n^O X_m^D\right)$. Not surprisingly, size shows its weight and pairs containing large regions trade more than pairs containing small regions. A simple regression of $\ln(X_{nm})$ on $\ln\left(X_n^O X_m^D\right)$ delivers an R-squared of 0.22 and a slope coefficient of 0.69. Thus, size explains about a quarter of the total variation in trade probabilities. This result is not very interesting, though, since this relationship is somewhat mechanical. How could the trade probabilities involving a given region not be related to the region's size, which is defined as the sum of the trade probabilities of the region?

What is really interesting about Figure 5 is that more than three quarters of the variation in trade probabilities cannot be explained by size. This is the variation we care about. Home trade observations are located well above the 45 degree line, confirming that regions trade with themselves much more than what their sizes suggest. The same applies to country trade observations,

Figure 5: Actual vs predictd trade (log) probabilities



although to a lesser extent. The counterpart is that most foreign trade observations are below the 45 degree line. European regions have strong trading relationships with themselves and with other regions within their country, and weak trade relationships with regions in other countries.

To make this idea precise, we measure the strength of the trade relationship for a region pair with the ratio of the actual trade probability and the trade probability predicted by the independence benchmark:

$$S_{nm} = \frac{X_{nm}}{X_n^O X_m^D} \tag{2}$$

Santamaría et al. (2021) refer to S_{nm} as a normalized market share.[3] This

[3]The reason is that S_{nm} has two alternative interpretations that suggest this name. First, S_{nm} is the share of origin n in destination market m, i.e., X_{nm}/X_m^D; normalized by the share of origin n in the entire European market, i.e., X_n^O. Second, S_{nm} is the share of destination m in origin market n, i.e., X_{nm}/X_n^O; normalized by the share of destination m in the entire European market, i.e., X_m^D.

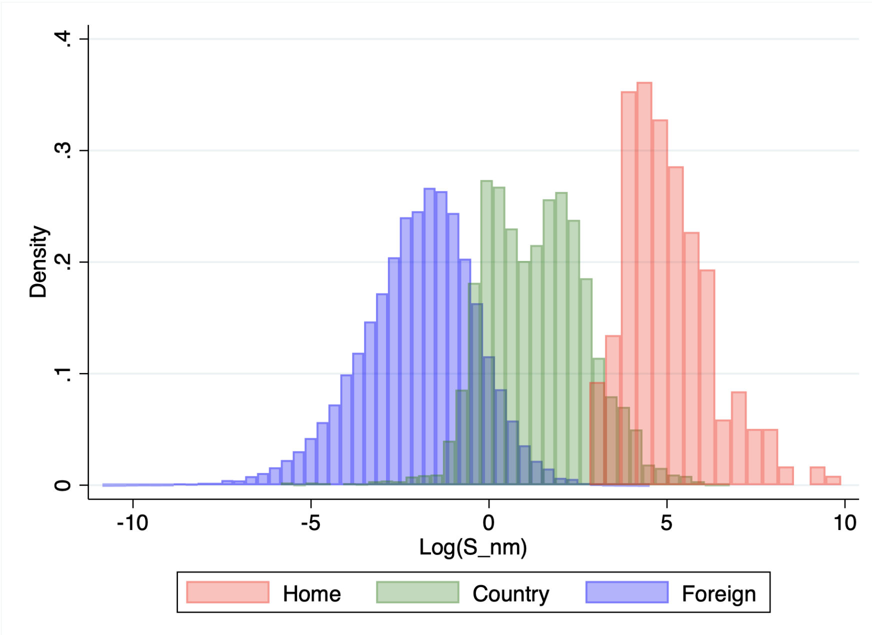
measure corrects for the mechanical effect of size on trade and it has a very simple interpretation: if $S_{nm} = 2$ (0.5), shipments from origin n to destination m are twice (half) as large as one would be able to predict knowing only the sizes of the regions. Thus, S_{nm} is a size-free measure of how strong a trade relationship is.[4]

Figure 6 plots histograms of normalized market shares separately for home, country and foreign trade. The average normalized market shares for the different types of trade are 469.5, 11.22 and 0.44. The distributions of normalized market shares for these types of trade have little overlap. The ranking home > country > foreign trade is not surprising. But the magnitude of the differences is (at least to us!), given that we are using data on trade in goods and not trade in services.

Finally, and just to whet the appetite for what is coming next, Figure 7 plots $\ln S_{nm}$ against the (log) of actual distance. It is apparent that the strength of trade relationships declines with distance. This surely helps explains part of the home and country biases in trade. But Figure 7 also shows that distance cannot be the single explanation for these biases. Within any given distance interval, we can observe the ranking of home > country >foreign trade. What else is going on? We turn next to a systematic examination of the data using the standard gravity framework.

[4]If we go back to the example of Catalonia and La Rioja, we find that normalized market shares for Catalonia are 2.83 (sales/exports) and 3.91 (purchases/imports) and for La Rioja 16.17 and 15.19. Catalonia and the Basque Country trade between three and four times more than one would predict given their sizes, but La Rioja and the Basque Country trade between fifteen and sixteen times more! Thus, it is La Rioja that has a stronger trade relationship with the Basque Country. One reason for this is that La Rioja is much closer geographically to the Basque Country than Catalonia is.

Figure 6: Home, country and foreign normalized market shares



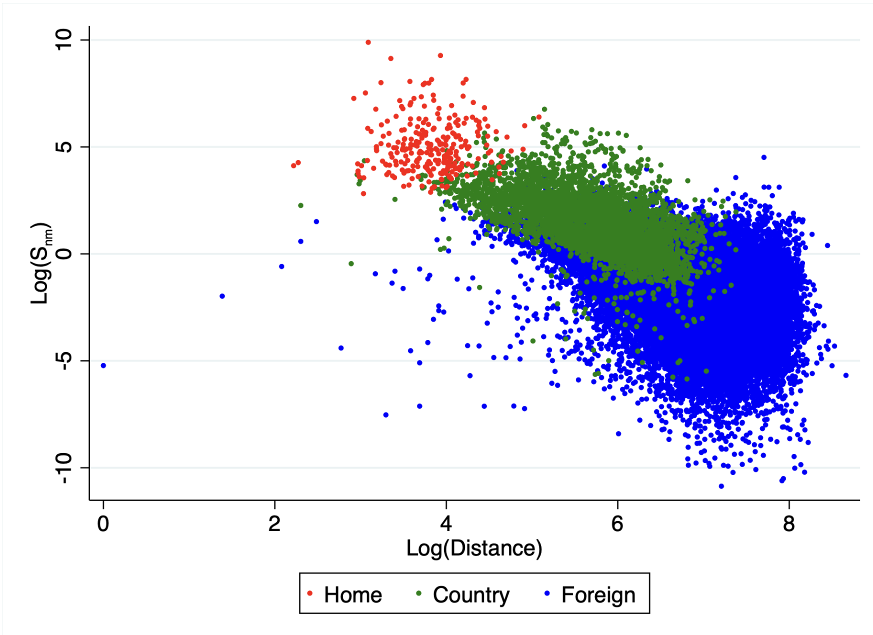
3 A gravity look at the data

Figure 8 shows the matrix of (log) normalized market shares. The goal of this section is to provide a parsimonious description of this matrix. To do this, we use the gravity framework to guide our search for patterns. The bottom line is simple again: using distance and borders we can explain about two thirds of the variation in (log) normalized market shares. To reach this conclusion, we explore a battery of increasingly flexible specifications for distance and border effects.

3.1 The gravity framework

The gravity framework provides a specific mathematical structure that adjusts trade probabilities to take into account distance, borders and other variables. Let M_{nm} be a measure of the cost of shipping goods from origin

Figure 7: (Log) normalized market shares and (log) distance



n to destination m. We refer to M_{nm} as bilateral market access. Gravity models postulate a bilateral market access function of this form:

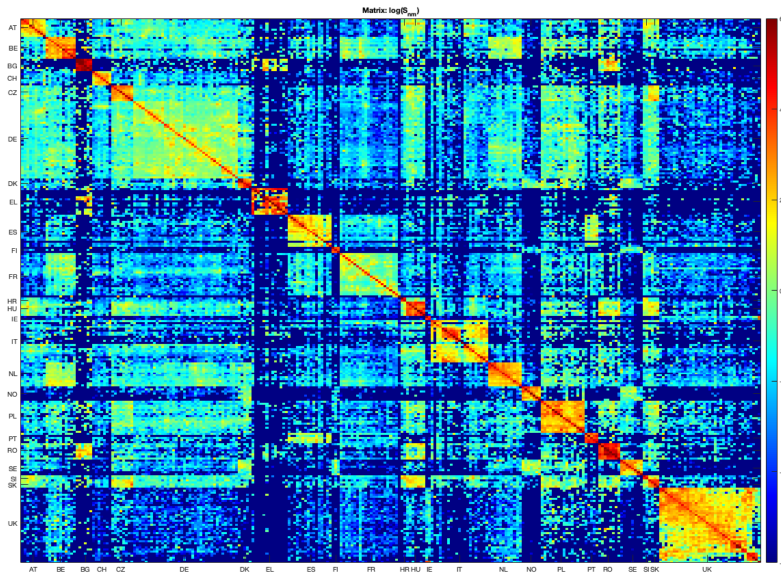
$$M_{nm} = \exp \left\{ {}_i \theta^i Z^i_{nm} \right\} \tag{3}$$

where $\{Z^i_{nm}\}$ is a set of bilateral variables that determine market access, $\{\theta^i\}$ is a set of theoretical coefficients. The set of bilateral variables typically contains the (log) distance traveled and a border dummy measuring whether the regions are in the same country or not. In many cases, other variables that might affect the costs of shipping goods are added such as dummies measuring whether the regions have a common language or currency.

The gravity framework consists of the following mathematical model:

$$X_{nm} = \frac{M_{nm}}{M^O_n M^D_m} X^O_n X^D_m \tag{4}$$

Figure 8: Bilateral matrix of (log) normalized market shares for European regions



which, alternatively, can be expressed in terms of normalized market shares as follows:

$$S_{nm} = \frac{M_{nm}}{M^O_n M^D_m} \tag{5}$$

where M^O_n and M^D_m is a set of numbers that satisfy the following restrictions:

$$1 =_m X^D_m \frac{M_{nm}}{M^O_n M^D_m} \tag{6}$$

$$1 =_n X^O_n \frac{M_{nm}}{M^O_n M^D_m} \tag{7}$$

We refer to M^O_n and M^D_m as origin and destination measures of average market access.[5] Equations (6) and (7) are not additional theoretical restrictions,

[5]The literature often refers to these terms as multilateral resistance terms or price levels, but labeling them as origin and destination measures of market access is more

but instead consistency requirements that ensure that probabilities add, i.e., $1 =_m X^D_m S_{nm}$ and $1 =_n X^O_n S_{nm}$.

It is well known that there is a large set of theoretical models that are consistent with the formulation of the gravity framework in Equations (5), (6) and (7) (Head and Mayer (2014)). Thus, all these models predict that the trade relationship of a region pair is strong if its bilateral market access is large relative to the average market access of origin and destination regions. Two gravity models that generate the same set of bilateral market access measures $\{M_{nm}\}$ predict the same matrix of bilateral trade. These models might be based on different sets of assumptions. For our purposes, however, these two models are observationally equivalent and they will be treated as a single model here.

3.2 An important example

We explore next a parsimonious version of the gravity model that offers a number of interesting insights and, as we shall show soon enough, it is also capable of explaining a substantial fraction of the variation in the matrix of (log) normalized market shares shown in Figure 7. In particular, let us assume the following bilateral market access function:

$$M_{nm} = \exp\{\sigma D_{nm} + \beta B_{nm}\} \tag{8}$$

where $\sigma, \beta \leq 0$. The variable $D_{nm} \geq 0$ is the (log) average travel distance between regions n and m. The variable B_{nm} is a dummy variable that takes value 0 if regions n and m belong to the same country, and takes value 1 otherwise. The coefficients σ and β measure the (negative) effect of distance and borders on bilateral market access, respectively.

Figure 9 contains three theoretical matrices of (log) normalized market

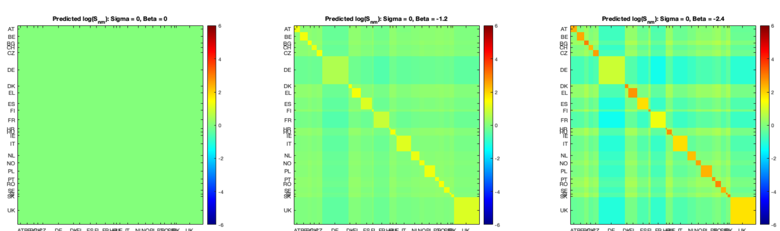
transparent.

shares produced with this model. In all of them, we set $\sigma = 0$ so that:

$$M_{nm} = \begin{cases} 1 & \text{if } B_{nm} = 0 \\ e^{\beta} & \text{if } B_{nm} = 1 \end{cases} \tag{9}$$

From left to right, these matrices assume that $\beta = 0$, $\beta = -1.2$ and $\beta = -2.4$, respectively. Thus, we start from the independence benchmark with all (log) normalized market shares equal to zero on the left, and then increase the border effect in two steps as we move right. As the border effect becomes stronger, bilateral market access for region pairs in different countries shrinks. As a result, average market access for all origin and destination regions also shrinks. Crucially, this shrinkage is larger for regions within small countries than for regions within large ones.[6] The reason, of course, is that the costs of trade have increased more for the former than for the latter.

Figure 9: Borders and trade



These observations lead to two important theoretical predictions. The first one is that, as the border effect becomes stronger, country/home trade grows and foreign trade shrinks. This generates squares centered along the diagonal with high-trade entries inside them and low-trade entries outside. The second theoretical prediction is that, as the border effect becomes stronger, regions in small countries experiment more growth of country/home trade and less shrinkage of foreign trade. This small-country effect (which is due

[6]By the size of the country, we mean the sum of the sizes of its regions.

exclusively to the differential change in average market access) creates a specific source of heterogeneity and it has a very simple intuition. If you have above-average trade relationships with many/large regions (i.e. large country), not only each of these relationships cannot be too much above average but also the remaining relationships must be well below average. If you have above-average trade relationships with few/small regions (i.e. small country), these relationships can be well above average and yet the remaining relationships do not have to be much below average.

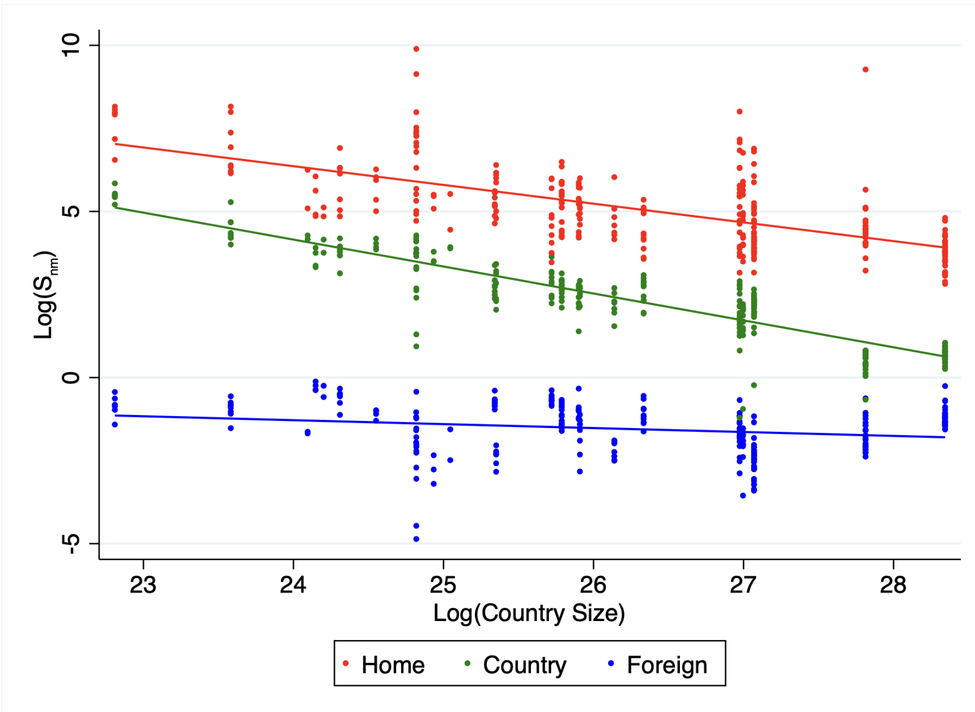
Figure 10 plots actual (log) normalized market shares against country size, using different colors for home, country and foreign trade. Not surprisingly, we see again that home/country trade is larger than foreign trade, which is consistent with the first theoretical prediction. Much more interesting is that regions in small countries have larger (log) normalized market shares that regions in large countries. This can be seen when we compare (log) normalized market shares within each type of trade. Clearly, the small-country effect is present in the European regional trade data.

Figure 11 shows three additional theoretical matrices of (log) normalized market shares produced with the model. In all of them, we set $\beta = 0$ so that:

$$M_{nm} = e^{\sigma D_{nm}} \tag{10}$$

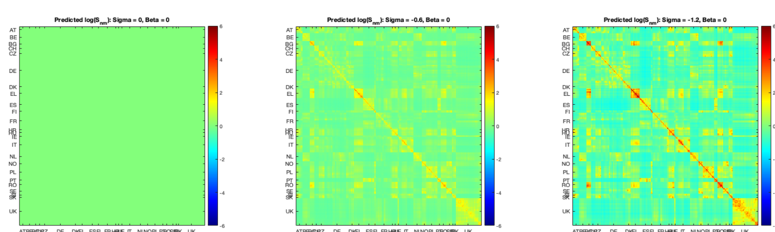
From left to right, these matrices assume that $\sigma = 0$, $\sigma = -0.6$ and $\sigma = -1.2$, respectively. Thus, we start with the independence benchmark again, and then increase the cost of distance in two steps as we move right. As the distance effect becomes stronger, bilateral market access for all region pairs shrink. This shrinkage is larger for region pairs that are far away from each other. As bilateral market access shrink, average market access for all origin and destination regions also shrink. Now, this shrinkage is larger for regions that are remote within Europe than for regions that are central. The reason,

Figure 10: (Log) normalized market shares and country size



again, is that the costs of trade have increased more for the former than for the latter.

Figure 11: Distance and trade



These observations lead to two theoretical predictions. The first one is again that, as the distance effect becomes stronger, country/home trade grows and foreign trade shrinks. The reason is that regions in different countries are far away from regions in the same country (recall Figure 4). This generates again squares centered along the diagonal with high-trade entries inside them and low-trade entries outside. An interesting novelty is that now trade is not homogeneous inside these squares. In particular, there is more trade in the diagonal than in the rest of these squares since regions are closer to themselves than to other regions within the same country. The second theoretical prediction prediction is that remote regions experiment more growth of country/home trade and more shrinkage of foreign trade. This remoteness effect creates a second specific source of heterogeneity, which is also quite intuitive.

Figure 12 plots actual (log) normalized market shares against an index of remoteness.[7] A quick look at the figure shows that (log) normalized market shares for home and country trade do indeed grow with remoteness, while (log) normalized shares for foreign trade shrink. The remoteness effect is also present in European regional trade data.

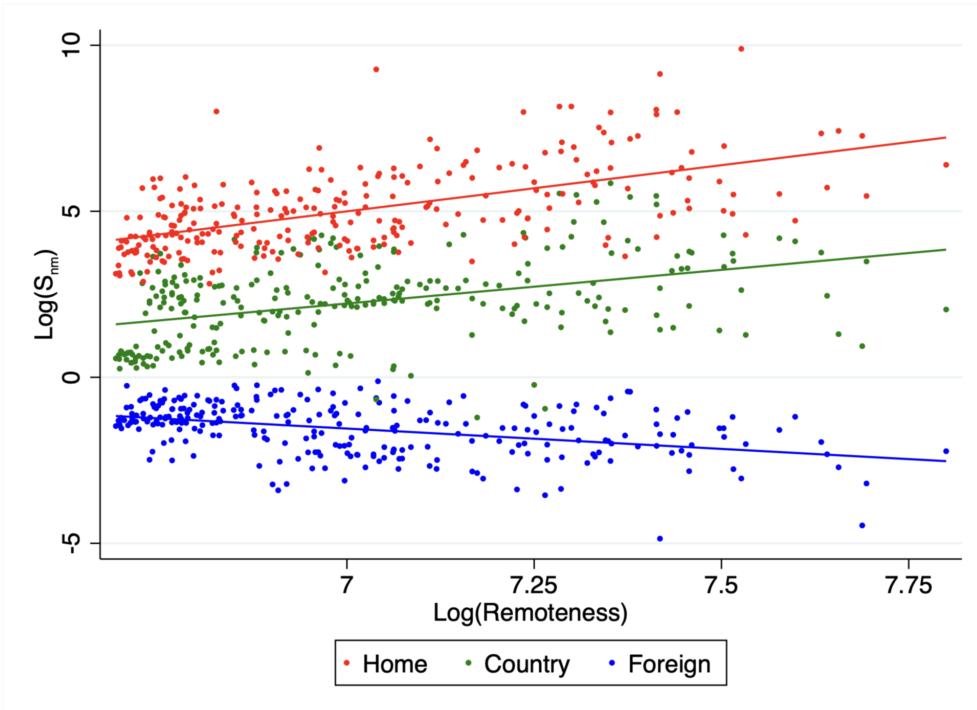
Armed with these intuitions, we search next for the combination of σ and β that provides the best fit of this model to the data. To do this, we define a two-dimensional grid over σ and β. For each point in the grid, we compute: (i) a complete set of bilateral market access measures $\{M_{nm}\}$; (ii) a complete set of origin/destination average market access measures $\left\{M_n^O\right\}$ and $\left\{M_m^D\right\}$; and (iii) the matrix of predicted (log) normalized market shares. We then choose the values of σ and β that minimize the distance between the matrices of actual and predicted (log) normalized market shares.[8] This procedure leads us to choose $\sigma = -1.3$ and $\beta = -2.4$. Figure 13 shows how sensitive is the fit of the model to changes in parameter values.

Figure 14 plots the actual matrix of (log) normalized market shares in the left panel and the matrix of predicted (log) normalized market shares in

[7]This index is the average distance to all other regions in Europe.

[8]To minimize the distance we use as a criterion the Frobenious norm.

Figure 12: (Log) normalized market shares and remoteness



the right panel. Even though there are differences across the two matrices, it seems that the parsimonious model discussed here captures some of the most important patterns in the matrix. To reinforce this message, we plot in Figure 15 plots the entries of these matrices against each other. More formally, we find that this parsimonious model explains about 48 percent of the variation in (log) normalized market shares. If we examine each type of trade separately, we find that the model explains 53 percent of the variation in home trade, 55 percent of the variation in country trade, and 20 percent of the variation in foreign trade.

We could perhaps provide an even better description of the data by using more flexible and realistic formulations of the border and distance effects. With this idea in mind, we turn next to the popular fixed-effects regressions.

Figure 13: Sensitivity analysis

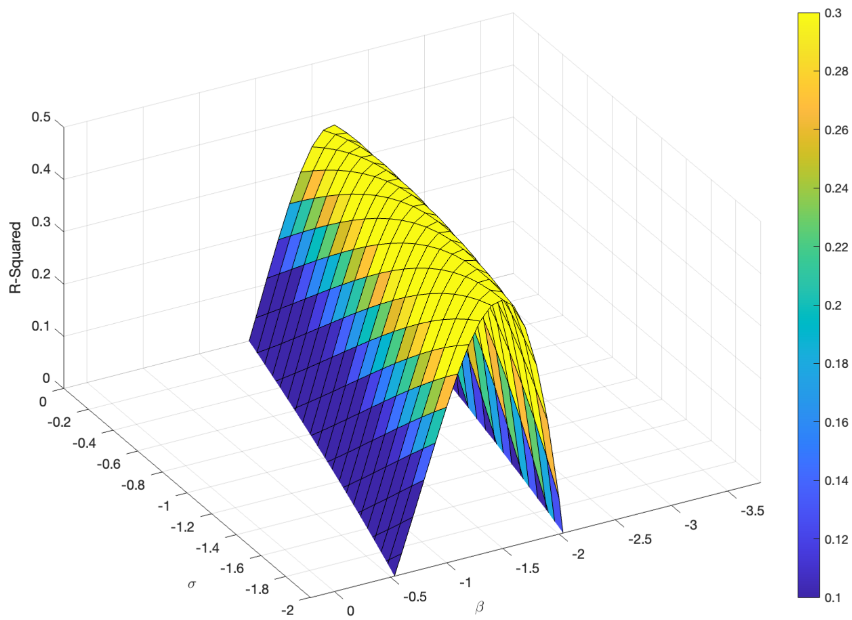


Figure 14: Actual vs predicted matrices of (log) normalized market shares

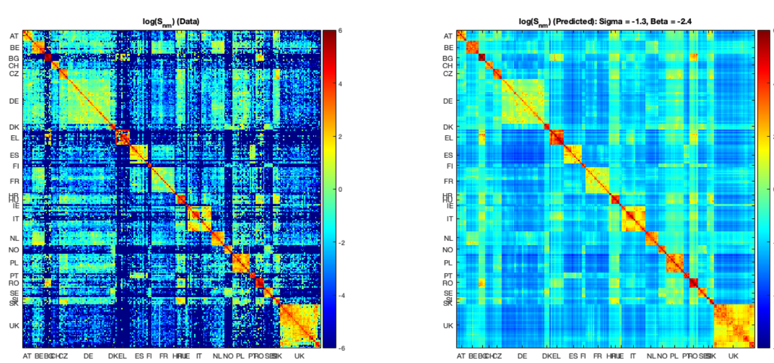
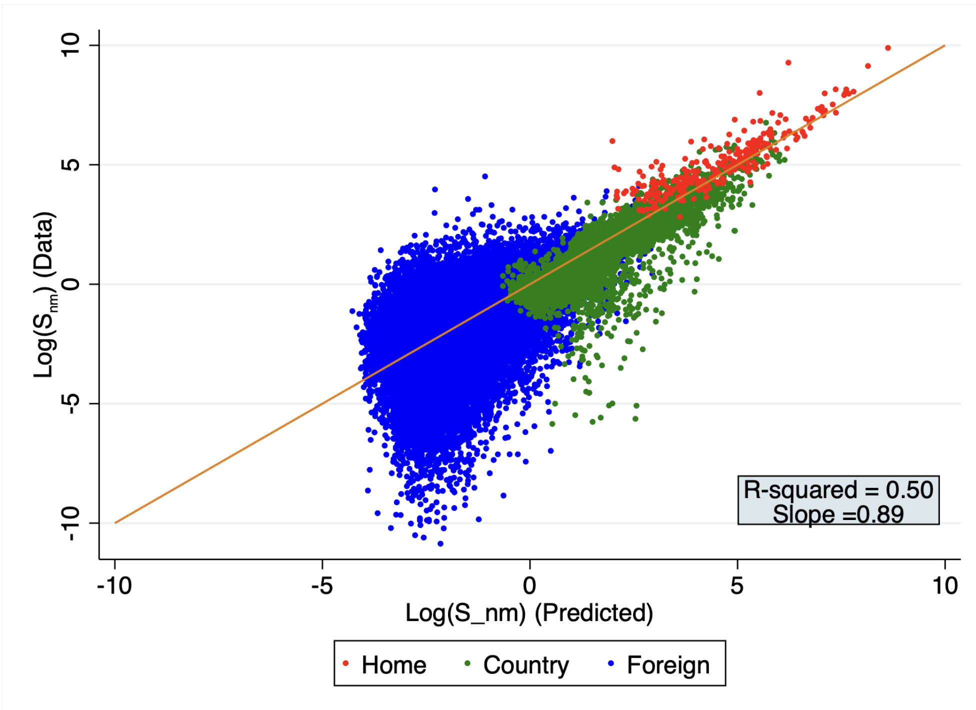


Figure 15: Actual vs. predicted (log) normalized market shares



3.3 Fixed-effects regressions

Consider a gravity model that generates a set of bilateral access measures $\{M_{nm}\}$ for all region pairs in our sample. With these measures at hand, we could compute the deviations between actual and predicted probabilities:[9]

$$\ln S_{nm} - \ln \left(\frac{M_{nm}}{M_n^O M_m^D} \right)$$

Showing that the model provides a good explanation of the data is then equivalent to showing that these deviations are small and random, i.e., attributable to sampling error and not to model misspecification.

Unfortunately, this simple and appealing empirical strategy is not possi-

[9]Knowing $\{M_{nm}\}$, we can compute $\{M_n^O\}$ and $\{M_m^D\}$ using Equations (6) and (7).

ble. Typically, gravity models provide the set of variables $\{Z_{nm}^i\}$ that enter into the bilateral market access function M_{nm}, but they do not provide specific values for their coefficients $\{\theta^i\}$. Thus, we can compute neither the set of bilateral market access measures $\{M_{nm}\}$, nor the set of origin and destination market acess measures $\left\{M_n^O\right\}$ and $\left\{M_m^D\right\}$.

To solve this problem, we can estimate the following fixed-effects regression:

$$\ln S_{nm} = \phi_n^O + \phi_m^D +_i \theta^i Z_{nm} + u_{nm} \tag{11}$$

where ϕ_n^O and ϕ_m^D are region fixed effects and u_{nm} is an error term that is assumed to be orthogonal to the regressors. The idea behind this regression is to allow the data to choose the parameters $\{\theta^i\}$ that give the model the best chance to explain the data. The estimates of the fixed effects are then interpreted as our estimates of $\ln M_n^O$ and $\ln M_m^D$.[10]

Table 1 shows the results of estimating regression (11) for six different gravity models. Column (1) shows the parsimonious model that we used in the previous subsection. In particular, there is a border dummy B_{nm} and a measure of distance given by:

$$D_{nm} = \sigma \ln T_{nm} \tag{12}$$

where T_{nm} are Kms traveled. This specification of the distance effect assumes a constant elasticity of trade to distance which is equal to σ. In column (2), we use a more general distance function that allows for the elasticity of trade to distance to vary across distance brackets or bins, $b = 1, ..., B$:

$$D_{nm} = \begin{cases} \sigma_1 \ln T_{nm} & \text{if } 0 < T_{nm} \leq T_1 \\ \sum_{i=2}^{b} (\sigma_{i-1} - \sigma_i) \ln T_{i-1} + \sigma_i \ln T_{nm} & \text{if } T_{b-1} < T_{nm} \leq T_b \end{cases} \tag{13}$$

[10] Recovering origin and destination market access measures from a fixed-effects regression is much more difficult when the dependent variable is $\ln X_{nm}$. See Fally (2015) for a discussion of this problem.

Table 1: Gravity: Fixed Effects Regressions

	(1) Log(S_nm)	(2) Log(S_nm)	(3) Log(S_nm)	(4) Log(S_nm)	(5) Log(S_nm)	(6) Log(S_nm)
Border dummy	-2.384***	-2.340***				
	(0.260)	(0.243)				
Border / common language / common currency dummy			-1.530***	-1.491***		
			(0.189)	(0.185)		
Border / common language / different currency dummy			-1.799***	-1.742***		
			(0.228)	(0.221)		
Border / different language / common currency dummy			-2.267***	-2.242***		
			(0.183)	(0.171)		
Border / different language / different currency dummy			-2.777***	-2.744***		
			(0.221)	(0.208)		
Border dummies for each country pair	No	No	No	No	Yes	Yes
Distance (constant-elasticity)	-1.190***		-1.071***		-1.006***	
	(0.0668)		(0.0607)		(0.0712)	
Distance (variable-elasticity)	No	Yes	No	Yes	No	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46505	46505	46505	46505	46505	46505
R^2	0.610	0.611	0.623	0.624	0.666	0.668

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Equation (13) collapses to Equation (12) if $\sigma_b = \sigma$ for all $b = 1, ..., B$. Thus, allowing for a variable elasticity of trade to distance cannot worsen the fit of the model.

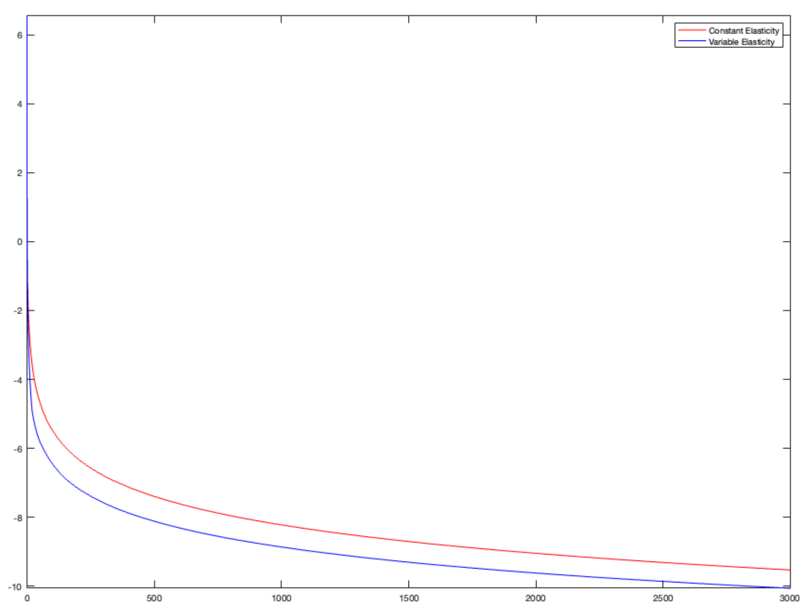
Column (1) shows that the parsimonious model explains almost two-thirds of the variation in trade probabilities. This is especially remarkable given that we have eliminated the effects of size using (log) normalized market shares instead of (log) trade probabilities.[11] Border and distance effects are significative, economically large and not far away from those that we found in the calibration exercise above. The estimated coefficient for the border dummy means that, controlling for distance, a national border reduces bilateral trade to $\exp\{-2.384\} \times 100 = 9.21$ percent of the independence

[11] We have estimated all the regressions in Table 1 using $\ln X_{nm}$ as the dependent variable instead of $\ln S_{nm}$. All the coefficients of bilateral variables remain unchanged up to the third decimal. Since the size correction is now picked up by the fixed effects, now to be interpreted as $\ln\left(\frac{X_n^O}{M_n^O}\right)$ and $\ln\left(\frac{X_m^D}{M_m^D}\right)$, the R-squared of the regressions is a bit inflated. Going from Column (1) to (6) the R-squared starts at 0.681 and grows up to 0.729.

benchmark. The estimated coefficient for distance implies that, controlling for borders, a one percent increase in distance traveled reduces bilateral trade by 1.19 percent with respect to the independence benchmark. Clearly, borders and distances can predict deviations from the independence benchmark.

Column (2) shows that using a variable elasticity of trade to distance does not affect much the results. The R-squared and the border coefficient are essentially the same. Figure 16 plots the effect of distance on trade for the two distance functions in Equations (12) and (13) using the estimates of the regressions in columns (1) and (2). The constant-elasticity specification is always above the variable-elasticity one, indicating that the former might be overestimating the effects of distance on trade. But the difference does not seem to be large.

Figure 16: Constant vs. variable elasticity distance functions



Columns (3) and (4) allow for some heterogeneity in the border effect.

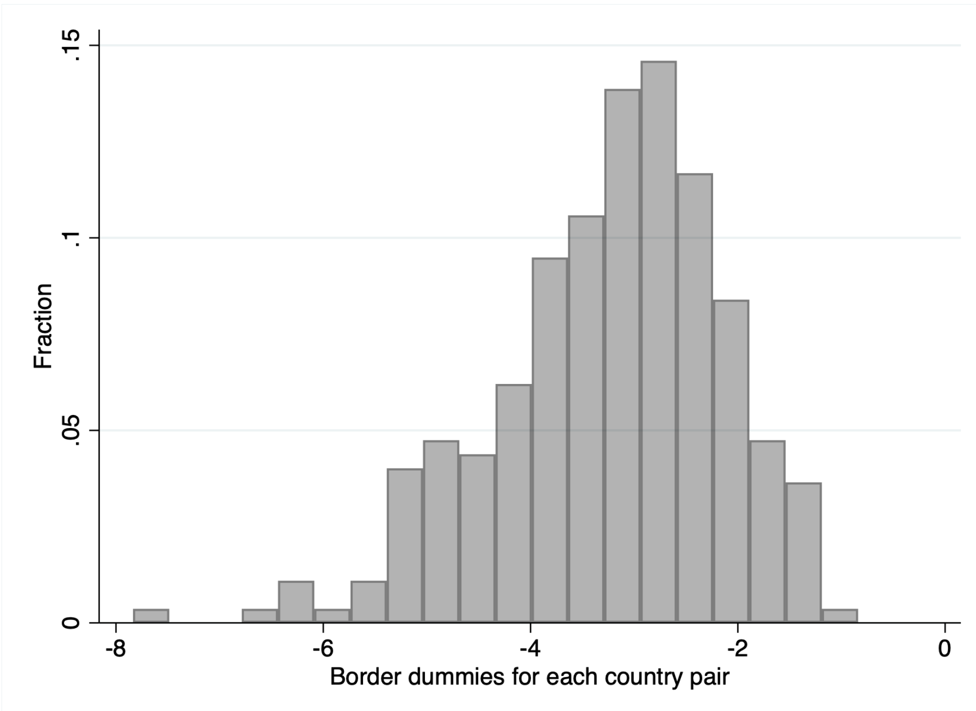
In particular, the border effect is allowed to depend on whether the regions involved have a common language and currency. The idea is that sharing a language and/or a currency facilitates trade and reduces the border effect. Using this flexible specification of the border effect raises the R-squared of the regression only marginally. Interestingly, we see that the distance effect is a bit smaller now since the estimated elasticity of trade to distance is -1.071. Again, there is not much difference between the constant- and variable-elasticity especifications for the distance effect.

The most interesting result in Columns (3) and (4) is that indeed the border effect depends on whether the region pair shares a language and/or a currency. At one extreme, a national border separating a region pair that shares both language and currency reduces bilateral trade to $\exp\{-1.491\} \times 100 = 22.52$ percent of the independence benchmark. At the other extreme, a national border separating a region pair that shares neither language nor currency reduces bilateral trade to $\exp\{-2.744\} \times 100 = 6.43$ percent of the independence benchmark. The estimated coefficients suggest that not sharing a language is more deleterious to trade than not sharing a currency, even though both variables seem to matter.

Columns (5) and (6) estimate different border effect for each country pair. That is, we allow the French-Spanish border to have different effects than the Finish-Spanish or the Irish-British borders. Since there are 24 countries in our sample, we are estimating 276 different border effects. This is the most flexible specification of the border effect so far. Yet, we find that the R-squared of the regression increases only marginally. The distance effect is reduced even further as the estimated elasticity of trade to distance is now -1.006. We confirm again that using the constant- or the variable-elasticity specifications of distance does not make much of a difference.

We do find however that there is substantial heterogeneity in border effects. Figure **??** and Figure **??** show this.

Figure 17: Histogram of country pair dummies



4 The home bias in trade

In our previous exploration of the effects of borders and distance on trade, we have been considering that all trade flows within a country are the same. However, as documented in previous sections, the trade that happens within each of the 269 regions is orders of magnitude larger than the rest, accounting for 40% of intra-European flows in our data. We now explore how large is this difference by adding a Home bias dummy to our gravity estimation. This Home bias dummy variable takes value one if the observation is the normalised market share of a region with itself, and value zero otherwise. It captures the differences in trade flows between shipments within a region and shipments of a region with other domestic regions.

Table 2 shows the same fixed-effects regressions that we saw in Table

??, including this additional variable. There are three key takeaways from these tables. First, the home bias variable is large and significant. Across all columns the coefficient is positive and comparable in size with the border effect. Focusing on our extended model in columns (3)-(4), the average normalised market share of a region with itself is between 3.5 (exp(1.271)) and 8.7 (exp(2.166)) times larger than the average normalised market share between two different regions in the same country, controlling for distance and other political variables. The second thing to notice is that the method we choose to control for distance makes a large difference. When we use the actual distance, estimating a constant elasticity of distance, the home bias coefficient is around 1. When we add a distance function that allows the elasticity of distance on trade to vary by distance brackets, we estimate an even larger home bias coefficient of around 2. Finally, the distance effects we estimate adding the home bias coefficient is slightly lower showing how both variables capture the very localised nature of trade. The R-squared of the regressions does not change much after we introduce the home-bias dummy. This reflects the fact that home trade has a very small number of observations in the overall trade matrix. Failing to fit those is not severely penalized in the tests we performed above.

In this section, we want to learn more about the factors that determine home trade. First, we explore which regional characteristics are correlated with a large home bias. Second, we explore whether the comparatively higher trade flows that we observe within regions are a genuine Home bias in trade or an artefact of the data aggregation process.

4.1 Determinants of the home bias in trade

Figure 18 shows the spatial distribution of Home normalised market shares in Europe. The first striking pattern is how heterogeneous these Home market

Table 2: Gravity: Fixed Effects Regressions

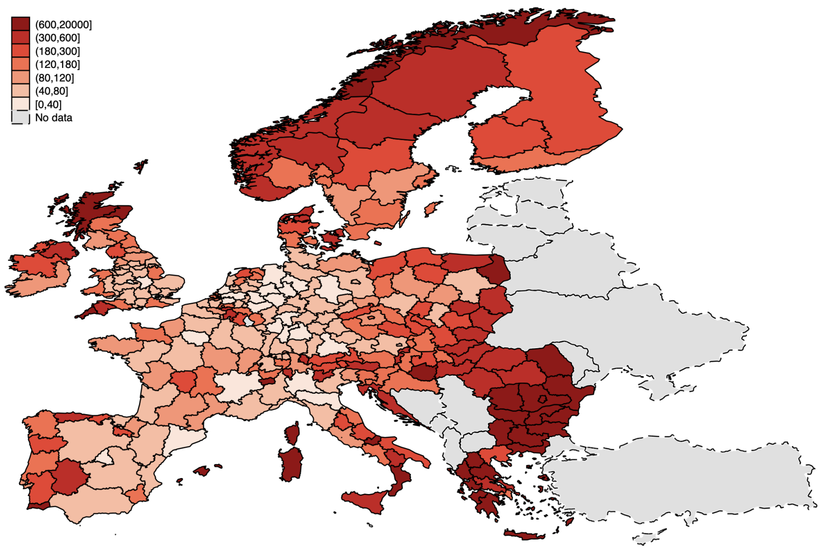
	Unique Border dummy				Country-pair border dummy	
	(1)	(2)	(3)	(4)	(5)	(6)
	Log(S_nm)	Log(S_nm)	Log(S_nm)	Log(S_nm)	Log(S_nm)	Log(S_nm)
Border dummy	-2.380***	-2.321***				
	(0.261)	(0.241)				
Border / common language / common currency			-1.499***	-1.466***		
			(0.182)	(0.179)		
Border / common language / different currency			-1.763***	-1.726***		
			(0.228)	(0.218)		
Border / different language / common currency			-2.265***	-2.217***		
			(0.176)	(0.165)		
Border / different language / different currency			-2.782***	-2.729***		
			(0.222)	(0.208)		
Home Bias	1.013***	2.079***	1.271***	2.166***	1.424***	2.233***
	(0.259)	(0.409)	(0.218)	(0.352)	(0.184)	(0.289)
Distance (constant-elasticity)	-1.150***		-1.016***		-0.903***	
	(0.0689)		(0.0604)		(0.0670)	
Distance (variable-elasticity)	No	Yes	No	Yes	No	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46505	46505	46505	46505	46505	46505
R^2	0.611	0.613	0.625	0.627	0.669	0.671

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

shares are. The region with smallest home share has a market share of 40 (40 times larger that the predicted share under the independence assumption), while the region that trades the most with itself has a market share of up to 20,000 (for instance the most remote regions in Greece). The second important pattern is that geography seems to play an important role. For instance, regions in countries in the periphery of Europe, like Greece and Bulgaria in the south and Norway and Sweden in the north tend to have higher home shares. Island regions also have higher home shares. And, interestingly, within-country geography is also correlated with these patterns: regions that are in the periphery of a country display higher home shares than more central regions (ie. regions in the south of Italy and Portugal, in the west of Spain and in the north of the UK and Denmark have higher home shares than the rest of the country).

Economic variables appear to be relevant as well. First, notice that re-

Figure 18: Normalized Market Shares: Home



gions in large economies like Germany, France or England have lower home shares. We also see that home trade tends to be lower in the more densely populated regions of Europe. We see this pattern at the European level in the so-called Blue Banana.[12] We also see this pattern within some countries that are outside the Blue Banana. For instance, Madrid and Catalonia have the lowest shares in Spain, while Warsaw and Athens have the lowest shares in Poland and Greece.

We aim to explore which factors predict this variation in home market shares by regressing the home market share on different geographic and economic variables. Table 3 shows regressions of home shares on a number of region characteristics, and country fixed effects.

[12]The Blue Banana is a corridor of highly urbanized land spreading over Western and Central Europe. It stretches approximately from North West England through the English Midlands across Greater London to the European Metropolis of Lille, the Benelux states with the Dutch Randstad and Brussels and along the German Rhineland, Southern Germany, Alsace-Moselle in France in the west and Switzerland (Basel and Zürich) to Northern Italy (Milan and Turin) in the south.

Column 1 reports the results using the following geographical variables: distance in the region (based on average distance covered by shipments within the region), Remoteness in Europe (weighted average of bilateral distance to each European region), Island dummy and mountain region dummy. Except for internal distance, the other geographical variables have relevant explanatory power. As we could see in the map, more remote regions, island regions and mountain regions have higher home shares. These simple geographical variables explain a substantial 41.3% of the variation in home market shares. Geography does not only impede international trade but also domestic trade.

Column 2 adds economic variables: presence of ports, motorway density, population, share of employment in manufacturing and in the public sector, the share of people with secondary education (or higher) and the share of foreign born population. The introduction of economic variables reduces the magnitude of the geographic variables' coefficients, since economic variables will be correlated with geographic features. The economic variables are very relevant too. Motorway density reduces the home market share, showing that infrastructure helps overcome geographical obstacles. As we observed in the map, the most populated regions have lower market shares. In addition, the economic structure matters substantially: regions with high manufacturing shares and regions with more educated populations have low home shares, showing they are more open to trade. Regions with high government spending also have lower home market too which could indicate that they are important regions in the country's administration. Finally, regions with more migrants have lower home market shares (they participate more in trade), which confirms previous findings on the importance of networks in trade (Combes et al., 2005). The introduction of these economic determinants increases the R-squared to 79.9% (by 38.2%).

Finally, column 3 adds country fixed effects. Looking at the map of home shares, we saw some patterns that emerged across countries. In particular, some countries have common levels of home market shares across all their regions. Bulgaria, Romania and Greece have regions with high market shares, while France and Germany have regions with low market shares. How much of the explanatory power of our variables is picking up differences across countries? The within-country specification (column 3) allows us to explore this. This specification with country fixed effects has a higher R-squared, of almost 90%, showing that country differences are important and not captured by our set of variables. Indeed, running a regression using only country fixed effects explains 56% of the variation in home market shares. However, this means that within country differences still explain almost half of the variation. Consistent with this finding, most variables remain significant after the introduction of country fixed effects. In particular, regions that are more geographically remote or have a lower highway density in their country, have higher home shares than other regions in the same country. And larger regions in terms of population or with a larger share of manufacturing have lower home market shares. However, the effects of public employment, education levels and migration levels are no longer significant. These variables were likely picking up differences across European countries. The findings from these regressions suggest that home trade seems to be driven by factors similar to the ones that drive trade openness: access to markets (geographic variables), size of internal demand (population) and the composition of the economy (share of manufacturing). However, there are still some unexplained patterns in the data: why do some countries have more open regions than others? Which other region-level variables are important to understand home bias in trade?

Table 3: Home Bias: Determinants

	(1) Home	(2) Home	(3) Home
Log(Distance)	-0.0171	0.229**	-0.0266
	(0.145)	(0.0904)	(0.187)
Log(European Remoteness)	2.345***	1.353***	1.551***
	(0.265)	(0.194)	(0.466)
Island Region	1.872***	0.915**	0.988***
	(0.509)	(0.364)	(0.328)
Mountain Region	0.304**	0.154**	0.193**
	(0.118)	(0.0722)	(0.0831)
Major Port Region		-0.197	-0.127
		(0.129)	(0.107)
Motorway Density		-6.379***	-6.510***
		(1.179)	(1.454)
Log(Population)		-0.819***	-0.758***
		(0.0488)	(0.0590)
Share of Emp. (Manuf.)		-10.48***	-10.01***
		(1.174)	(1.905)
Share of Emp. (Public)		-16.84***	-0.410
		(1.634)	(3.917)
Sh. Secondary or tertiary educ		1.511***	-1.399
		(0.398)	(0.903)
Share Migrant Pop.		-2.287***	-0.386
		(0.500)	(0.702)
Country FE	No	No	Yes
Observations	269	265	265
R^2	0.410	0.799	0.890

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

4.2 Government structure and home trade

In this section we exploit our rich dataset to contribute to the debate on the existence of a "home bias" effect in trade. Previous studies have suggested that part of the "home bias" effect, the high within-region trade flows observed relative to cross-regional flows, can be a statistical artefact created by data aggregation (Hillberry and Hummels (2008), Coughlin and Novy (2021)). These papers highlight the potential mismeasurement of trade at short distances since finely disaggregated data is not generally available. This data measurement issue creates two problems. First we cannot see smaller geographical units within states or regions, making it difficult to control well for distance within regions. Hillberry and Hummels (2008) deal with this by using the Commodity Flow Survey (CFS) for the US and find that the state border effect does not survive once you control for actual distance of the shipment using zip-code level data. Second, since the data is aggregated after collection, the level of aggregation of spatial units can create a“spatial attenuation effect” (Coughlin and Novy, 2021). Due to the existence of spatial frictions, the larger number of regions we combine, the larger the inside-region frictions will be, and the weaker the border effect we will estimate. This would make us estimate weaker border effects for large countries and stronger border effects for small countries.

We seek to explore this question by using some unique features of our dataset. The first advantage of our data is that we have 24 different countries that can serve as laboratories in which we can explore the size of the home bias effect. The second advantage of the data is a peculiarity of the data collection and harmonization process. Since our shipment data is collected and provided by Eurostat, Eurostat has established a single EU-wide spatial structure. All countries that provide Eurostat with data must provide a division of their territories into three sub-national type of regions: NUTS3 regions, the lower level, NUTS2 regions, the intermediate level and NUTS1 regions, the top sub-national level. The NUTS classification provides us with size-comparable regional divisions across countries. In addition to this, this classification also offers an opportunity to test the robustness of the home bias effect. Due to government structure differences, in some countries these NUTS2 and NUTS1 regions are only statistical regions created for the purpose of sharing data with Eurostat, while in other countries NUTS2 and NUTS1 regions coincide with political divisions with different levels of self-government.

Figure 19: Statistical and Political borders in Europe

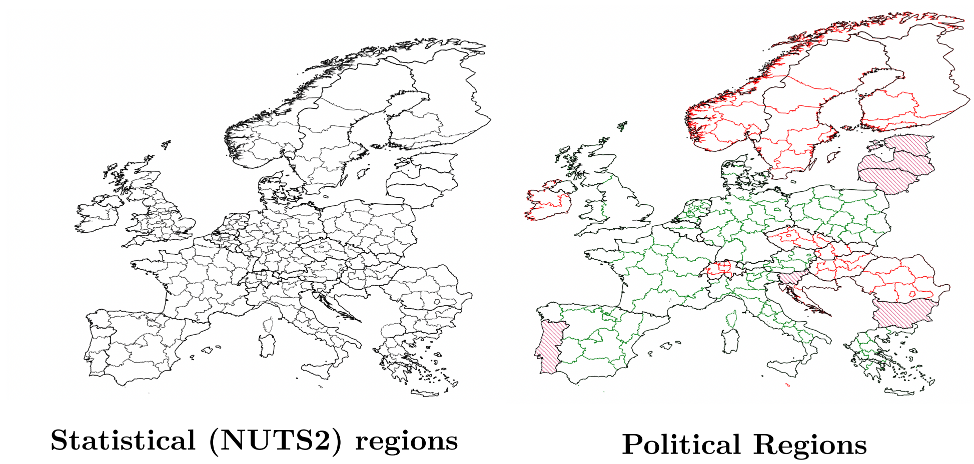


Figure 1 provides an illustration of the type of variation we can exploit in the data. The first panel shows the borders of the 269 regions in our dataset. The second panel shows the political borders that divide these regions. The dark black lines represent country borders. The green lines, represent regional political borders that coincide with statistical NUTS1 or NUTS2 regions. Out of our 24 countries, we have 12 countries in this group. The red lines represent statistical regions that group political regions at a lower level of disaggregation. This is the case in 9 countries. Finally, there are five countries with no political borders above the municipal level. These five countries are represented with the red diagonal pattern. Our strategy will be to test whether the home bias effect emerges in all countries, regardless of the level of the actual political borders, or whether we only find a home bias effect when it coincides with political regional borders.

To recap, we have the following country groups regarding the structure of the regional government.

1. Countries with no regional governments. These countries are not divided into regional governments at any level.

Country	#Regions	#NUTS2
Portugal	0	5
Bulgaria	0	6
Slovenia	0	2

2. Countries with data at an aggregation level that is coarser than the regional level. For these countries, NUTS2 units are larger than regions:

Country	#Regions	#NUTS2
Finland	> 5	5
Romania	> 8	8
Slovakia	> 4	4
Switzerland	> 7	7
Norway	> 7	7
Hungary	> 8	8
Croatia	> 2	2
Czech Republic	> 8	8

One example is Switzerland that is divided into 26 Cantons, that have government autonomy, but that collects data for Eurostat by aggregating the cantons into 7 regions. We do not observe their administrative regions since they are at a lower level than NUTS2.

3. Countries with data at an aggregation level that coincides with the regional level. For these countries, NUTS2 units are regions:

Country	#Regions	#NUTS2
Austria	9	9
Denmark	5	5
France	22	22
Greece	13	13
Ireland	3	3
Italy	21	21
Netherlands	12	12
Poland	16	16
Spain	16	16

4. Countries with data at an aggregation level that is finer than the regional level. These countries have regional governments at the level of NUTS1 regions, that are aggregates of the NUTS2 regions:

Country	#Regions	#NUTS2
UK	4	39
Germany	16	39
Belgium	3	11

In the case of Germany and Belgium, both NUTS2 and NUTS1 regions correspond to regional borders (provinces and regions in Belgium, government regions and Landers in Germany). In the case of the United Kingdom, the nations coincide with the NUTS1 regional division but the NUTS2 regions are purely statistical.

Based on this, we define the following groups of countries. Group 1 are countries without regions, Group 2 are countries with regions at a lower level than NUTS2 so we cannot identify its borders in the data. Group 3 are countries in which regions coincide with NUTS2 regions and finally group 4 are countries with regions at NUTS1 level, so we are able to see sub-regions within this upper level regions.

We exploit this heterogeneity in country configurations to estimate the home bias effect of the NUTS2 region and of the NUTS1 region. For this purpose we do not use any data on international trade, we only use trade flows that stay within the country. For each country we run the following regression:

$$\ln(\mathrm{S}_{nm}) = \beta_1 HomeBias(NUTS1)_{nm} + \beta_1 HomeBias(NUTS2)_{nm} + \gamma log(dist_{nm}) + \delta_n + \eta_m + e_{nm}, \quad (14)$$

where the outcome variable is the normalised trade market share. The variable $HomeBias(NUTS2)_{nm}$ indicates a low-level regional border and takes value 1 when n and m are in the same NUTS2 region. The variable $HomeBias(NUTS1)_{nm}$ indicates a high-level regional border and takes value 1 when n and m are in the same NUTS1 region. The coefficients of interest are β_1 and β_2, that are estimated by comparing with flows that cross NUTS1 and NUTS2 borders. We have the following hypothesis about the sign and significance of β_1 and β_2.

1. For countries of group 1 (no regional borders), we expect $\beta_1 = 0$ and $\beta_2 = 0$, since trading across or within NUTS regions does not imply crossing any government/administrative border. This means that, controling for distance, shipping within or across NUTS regions should be the same.

2. For countries of group 2 (regional borders lower than NUTS2), we expect $\beta_1 = 0$, and $\beta_2 \geq$ but small. The reason is that some shipments within NUTS2 may cross a border, so they may behave similarly to shipments across NUTS2 regions. By contrast, shipments across NUTS1 regions will be the same as shipments across NUTS2 regions since, as in the group 1, NUTS1 regions do not coincide with any borders.

3. For countries of group 3 (regional borders are at NUTS2 level) we expect $\beta_1 = 0$ and $\beta_2 > 0$. In this group the NUTS2 regions coincide with the regional government borders. Therefore, we expect shipments across the NUTS2 border to be less common than within NUTS2 borders. We expect $\beta_1 = 0$ since in these countries NUTS1 region do not have any administrative significance so we do not expect to see a home bias effect at this level.

4. For Germany and Belgium, in group 4, we expect $\beta_1 > 0$ and $\beta_2 > 0$. Both countries have political borders coinciding with administrative units at NUTS1 and NUTS2 level, so we expect a home bias effect in the data.

5. For the UK, in group 4, we expect $\beta_1 > 0$ and $\beta_2 = 0$. The UK only has administrative divisions, nations, at NUTS1 level. So we expect the border to appear at NUTS1 but not at NUTS2 level.

Figures 21-24 present the country-specific border effects, by country group. We can see that all of our hypothesis hold in the data, with the exception of very few countries, showing that governments play an important role in driving the border effects within countries. These results suggest that border effects exist within countries and are not caused by the aggregation of data into large geographic units (Hillberry and Hummels (2008), Coughlin and

Novy (2021)).

Countries with no regional divisions We find, as expected, no home bias effect for these two countries (Slovenia is omitted due to lack of enough observations, since it only has two regions).

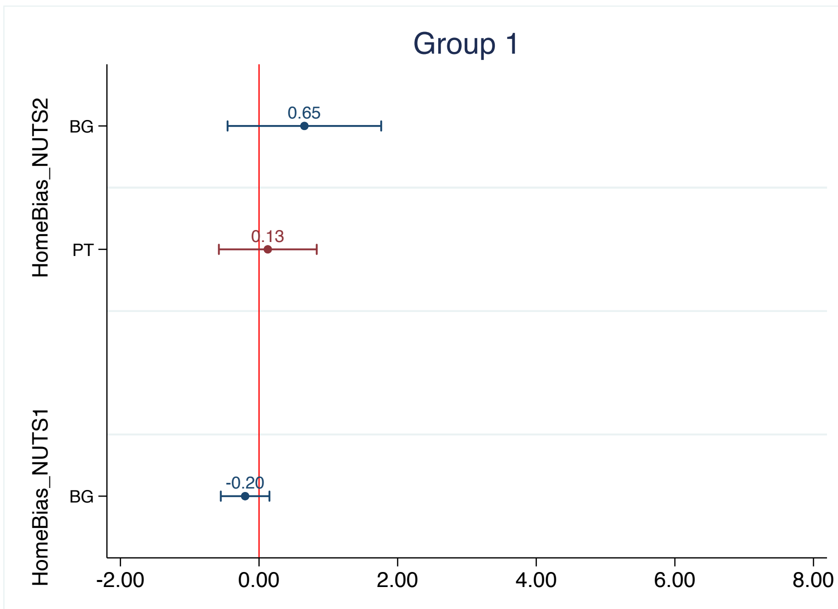


Figure 20: home bias: Group 1

Countries with regional divisions at a lower level of disaggregation This is one of the cases that would be of concern according to the previous literature as suffering from "spatial attenuation effect": we know there are regional borders within the NUTS2 observed regions. By aggregating different regions together inside each NUTS2, we are increasing the spatial frictions within-region, which would lead to an underestimation of the home bias effect. However, the effect could still be there because part of the trade flows within NUTS2 regions are pure within-region flows, while all of the cross-NUTS2 flows must be crossing some border. So we expected to see some positive effects, as is confirmed by the figure. Coefficients are larger than for group 1 and for two countries they are significant. This should not be the case for the NUTS1 regions. These regions are collections of lower level regional divisions for statistical purposes, so trade flows should not exhibit any home bias patterns. We find, as expected, that there is no home bias effect at the level of NUTS1 (some of these countries only have one NUTS1 region, so coefficient cannot be estimated).

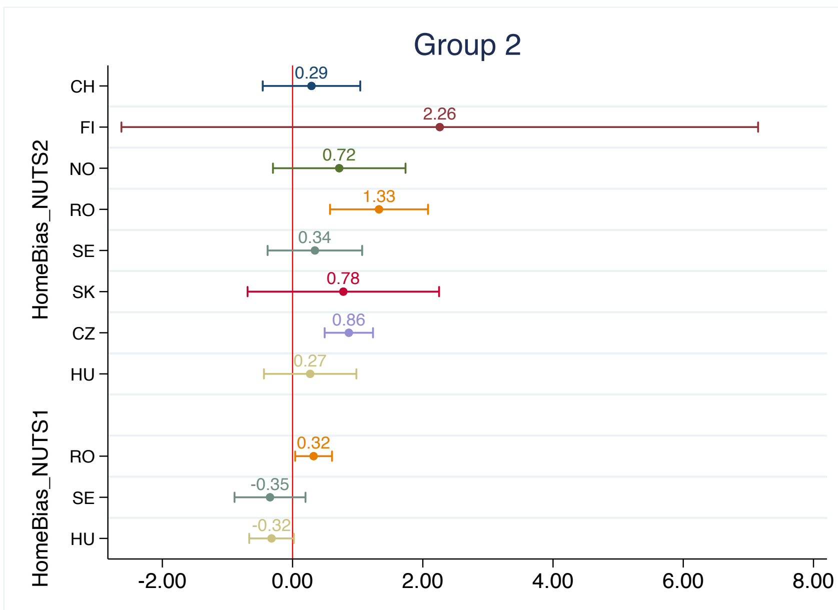


Figure 21: home bias: Group 2

Countries with regional divisions at the level of the statistical regions For these countries we expected a positive coefficient at NUTS2 level (that coincides with the regional border) and a zero for the NUTS1 regions, since these are statistical divisions. These predictions are confirmed in virtually all of the countries in the group. Particularly interesting is the comparison of the home bias (NUTS2) effect in Netherlands and Poland, both a magnitude of 0.4. These two countries are very different in size and in number of regions (Netherlands 12, Poland 16) but the home bias estimated is the same, showing that this strategy is helping us overcome the spatial aggregation problems mentioned above.

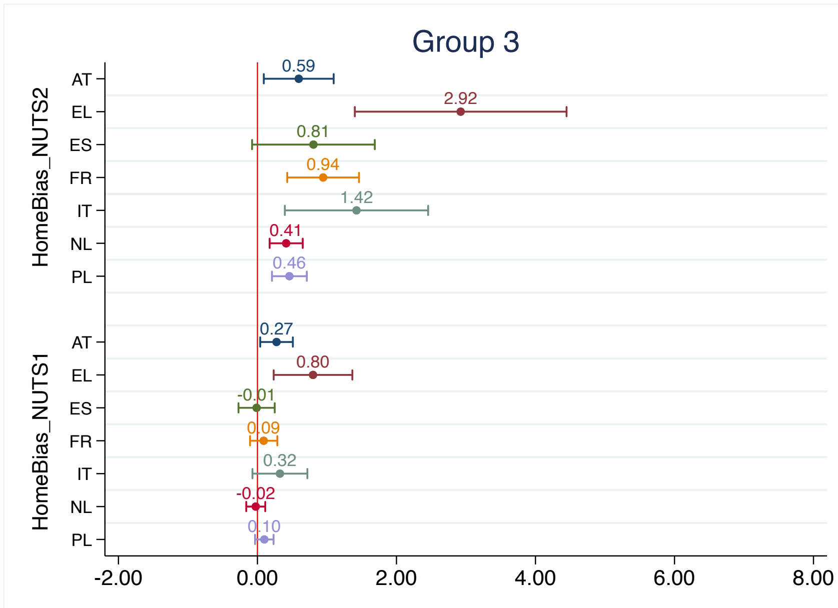


Figure 22: home bias: Group 3

Countries with borders at a level higher than the statistical regions For these countries, the predictions are confirmed. Germany and Belgium have home bias effects at both levels of divisions, with magnitudes that are remarkably close given the size difference between the two countries. For the United Kingdom, as we expected, we find no home bias effect at the statistical division level (NUTS2) but a home bias effect at the nation level (NUTS1).

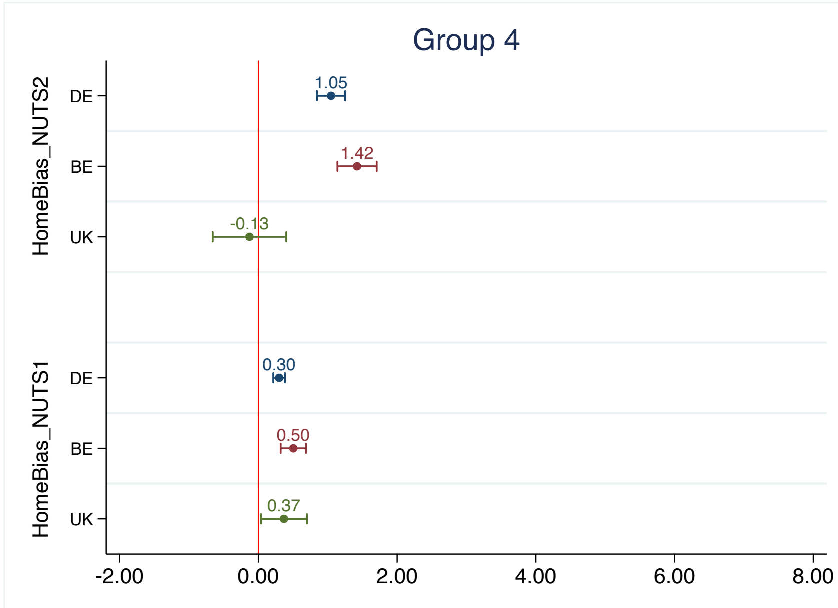


Figure 23: home bias: Group 4

These results partially confirm that aggregating political regions could alter the estimated home bias effect. When the data aggregation does not coincide with the political border, this results in an attenuation effect (as for countries in group 2). However, the existence of data on regional flows and distances at the right level of disaggregation confirm the existence of an home bias effect. We find own region effects coinciding with political subnational divisions. The magnitudes range from 0.41 to 1.33, and the effect is heterogeneous. Compared to the effect of international borders, the effect of the home bias is smaller but quite relevant.

What about other internal barriers to trade? Our data allows us to compare the effect of distance as a barrier to domestic trade. Figure 24 displays the coefficient of (log) distance in each of the countries. In all countries, except Finland, the coefficient on distance is negative and in most countries is precisely estimated. The magnitudes are around -1, but there is some heterogeneity across countries. One remarkable finding is that the distance elasticity does not seem to correlate strongly with the type of government structure at the country level, or with the size of the country or the number of regions. For instance, Norway, Poland, Czech Republic, Netherlands, Italy, Austria and Spain have distance elasticities that are extremely similar, while the countries are very heterogeneous. But some heterogeneity remains. We believe that understanding these differences would be an interesting avenue for future research. One of our contributions is to make these estimates on the home bias effect and the distance effect within country available to researchers that want to calibrate models and perform counterfactuals that take into account differences across countries.[13]

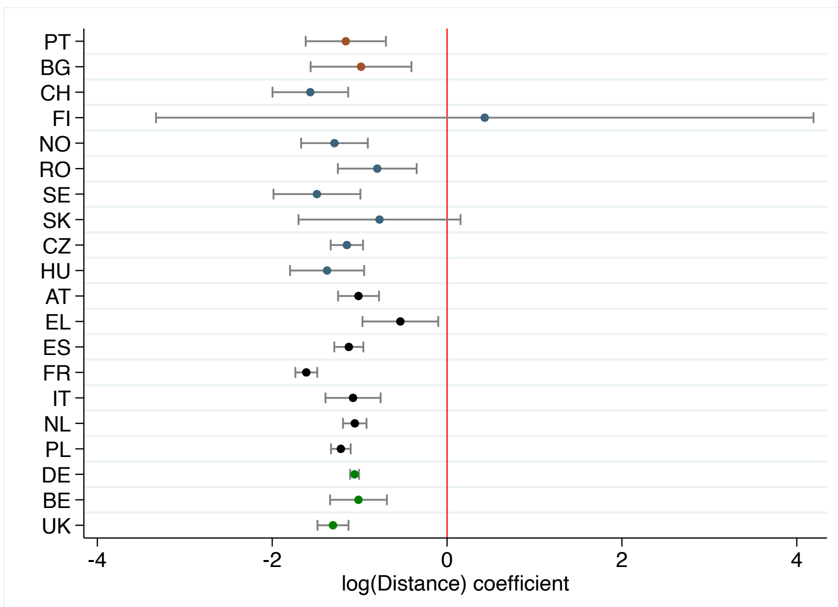


Figure 24: Distance-elasticity of domestic trade

[13] The datasets are available in the authors websites to download freely.

5 Intensive and extensive margins of trade

Home, country and foreign trade flows are very different in magnitude, and are not equally explained by the standard gravity model. In this section we explore whether the variation in each of these trade types comes from the extensive or the intensive margin. We follow Hillberry and Hummels (2008) to decompose aggregate trade into different intensive and extensive margins. For this purpose we use our data in the most disaggregated form: the shipment level. Our dataset contains almost 14,500,000 observations at the shipment level for which we know the origin, destination, quantity, industry, year and distance. In this section we use granular information about shipments to understand the margins by which home, country and foreign trade are different. The aim of the exercise if to capture which margins are important to explain the magnitudes and variance of home, country and foreign trade. For instance, is home trade large because shipments happen in all industries or because more shipments are send in each industry, or more kilos are sent in each shipment? Understanding these differences can also give us a hint about the margin that is affected by the existence of international and regional borders.

Table 4: Summary statistics: Shipment level

	Home Mean	Country Mean	Foreign Mean
Total weight shipped (Millions)	144.21	11.48	1.77
Total number of shipments	10168.92	901.36	105.97
Total industries shipped	11.90	11.15	6.79
Mean shipments per industry	849.65	76.08	10.58
Average weight per shipment	14562.49	14863.75	15929.85
Observations	9445992	3924440	1110031

We start by dividing shipments by type of trade (home, country and foreign) and report some summary statistics. Table ?? reports the average kilos shipped and total shipments between two regions in our dataset, as well as the total number of industries that are shipped. Each column reports the statistics for one type of trade flow. The aggregate level patterns reported previously are clear also at the lowest level of disaggregation: home flows are orders of magnitude larger than the other two, both in terms of kilos and in number of shipments. In terms of industries, the picture is different: on average 11 out of 12 industries are traded both within region and across regions, but only 7 industries are traded across national borders between the average region-pair. The last two rows report two intensive margin components: the observed shipments per industry and kilograms per shipment. Home trade flows are characterised by more than 10 times more shipments than country trade, and almost 85 times more shipments than foreign trade. By contrast, the weight per shipment is on average 15 tons. These statistics points to international borders reducing the extensive margin of number of industries, while regional borders reducing the intensive margin component of number of shipments, while none of the borders really affect the average kilos shipped per trip.[14]

One limitation of the microdata is that we do not observe values. To prevent that any price imputation affects the shipments we observe in the survey, we perform our analysis in terms of quantities shipped for this decomposition analysis. As a first step, let us consider the decomposition of the aggregate trade (kilos) shipped between two locations m and n as $Q_{n,m}$, that can be expressed as:

$$Q_{n,m} = N_{nm}\overline{Q}, \tag{15}$$

where N_{nm} is the total number of unique shipments and $\overline{Q}$, is the average weight per shipment. A unique shipment is defined as observing at least one observation per origin region$\times$ industry $\times$ destination region.

We can decompose the number of shipments further into the number of industries shipped, $N^I_{n,m}$ and the average shipments per industry, $\bar{S}_{n,m}$. In this way, we can write aggregate trade as the product of three components:

$$Q_{nm} = N^I_{n,m}\bar{S}_{n,m}\overline{Q} \tag{16}$$

Taking logs we get:

$$lnQ_{nm} = lnN^I_{n,m} + ln\bar{S}_{n,m} + ln(\overline{Q}) \tag{17}$$

Notice that equation 17 is an identity that holds for every region pair in every industry. It indicates that bilateral shipments can be decomposed into one extensive margin component, the number of industries and two intensive margin components, the number of shipments per industry and the quantity shipped per shipment. Since this is an exact decomposition, we can use

[14]This is intuitive since we use truck level data and 15 tons is the average capacity of a goods-shipping truck.

the properties of the Ordinary Least Squares estimator to decompose the contribution of each of the components to the variance of bilateral shipments by regressing each of the components on lnQ_{nm} (Bernard et al, 2021).

The first row in table 5 reports the decomposition using all trade flows, while the next three rows report the results of the decomposition for each type of trade. When we consider all types of shipments, the component that drives most of the variance is the quantity of kilos per shipment (84.4%), while the number of industries and shipments only contribute 8.8% and 6.8% respectively.

The relative importance of each margin is different when we compare home, country and foreign flows. The variance in home trade comes mostly from average kilos per shipment (54.4%) but the relative importance is much smaller than in aggregate trade, where 84.4% of the variation was attributed to this factor. Across regions, half of the variation in trade with home comes from average kilos per shipment and a bit more from the other two sources. In particular, the role of mean shipments per industry is much more important in home trade, accounting for 31.7% of the variance, while the importance of the number of industries margin is less important (13.9%). The variance for country trade across domestic pairs looks slightly more similar to the aggregate decomposition. Variation in the average kilos per shipment account for 72.1% of the variation, while the other two margins explain the remaining variation with similar shares (mean shipments per industry explains 13.3%, while number of industries explains 14.6%). Finally, foreign trade variation is almost completely by variation in mean kilos per industry (89.5%), while variation in the number of industries explains 6.6% and, finally, the mean shipments per industry explains only 3.9%.

This decomposition reinforces the notion that we have been highlighting throughout the paper of how different is home trade (within the region) from domestic and specially foreign trade. The extensive margin of trade, here represented by the number of industries a region trades with itself and the number of shipments that are made by each industry, explains almost half of the variation of home trade that we see across regions (45.6% of the variation). By contrast, the intensive margin explains most of the variation in flows between regions (both within countries and across countries).

Table 5: Bilateral shipments decomposition (ln Q_{nm})

Contribution to var(ln Q_{nm})	Num. of Industries ln$\mathrm{N}^I_{n,m}$	Mean shipments per Industry ln $\bar{S}_{n,m}$	Mean kg per shipment ln($\overline{Q}$)
All shipments	.088	.068	.844
Home	.139	.317	.544
Country	.146	.133	.721
Foreign	.066	.039	.895

5.1 Borders effects at the extensive and intensive margin

There are important differences in the shipping pattern of goods within regions, across regions and across countries. We now seek to understand which margins are more responsive to the main trade barriers we consider in this paper: borders and distance. In addition, and following our finding on the home bias in trade, we also explore how these margins change with the home bias indicator. To this end we run a gravity equation using all the region-pairs in our data but using as additional outcome variables the different intensive and extensive margin components in equation 17. In particular we estimate the following regression:

$$ln(Y_{nm}) = \beta Border_{nm} + ln(Dist_{nm}) + \phi^O_n + \phi^D_m + u_{nm}, \tag{18}$$

where we will use as outcome variable Y_{nm}: total quantity shipped (lnQ_{nm}) total shipments (lnShipments$_{nm}$), total number of industries shipped ($lnN^I_{n,m}$), average shipments per industry ($ln\bar{S}_{n,m}$) and average shipment size per industry, in kilograms, ($ln\overline{Q}$).

Table 6: Extensive and Intensive Margins

	Dependent variable				
	Tot kg	Tot shipments	Tot number Ind	Mean ship.	Mean kg
Border	-1.820***	-1.791***	-0.867***	-0.924***	-0.0288***
	(0.00826)	(0.00648)	(0.00405)	(0.00433)	(0.00511)
log(Distance)	-1.252***	-1.133***	-0.566***	-0.567***	-0.119***
	(0.00419)	(0.00329)	(0.00205)	(0.00220)	(0.00259)
Home bias	1.536***	1.886***	-0.647***	2.533***	-0.349***
	(0.0258)	(0.0202)	(0.0126)	(0.0135)	(0.0159)
Or FE+ Dest FE	Yes	Yes	Yes	Yes	Yes
Observations	189769	189769	189769	189769	189769
R^2	0.706	0.775	0.660	0.706	0.111

Table 6 displays the results. The first two columns use total kilograms and total shipments shipped from region n to region m. The effect of borders, distance and home bias have similar impacts on both variables: Borders and distance are associated with lower kilograms and less shipments, while there is a positive home bias effect. The next columns, (3) and (4) report the results from using as outcome variables the extensive margin (total number of industries shipped between n and m) and the intensive margin component of the average number of shipments per industry. Both margins are negatively associated with borders and distance, in a very similar way. However, these two margins have different implications for home trade. The "home bias" effect is negative in the number of industries (extensive margin) and positive in the case of average shipment per industries (intensive margin). This finding suggests that trade flows across a pair of regions n and m happen over a small set of industries if n=m, but the number of shipments per industry is larger. Finally, the last column reports the effects on the pure intensive margin component: the average kilograms per shipment. The most noticeable thing of column (5) is that the coefficients are smaller in magnitude and the R-squared is much smaller than the rest of the columns. Borders, distance and home bias (together with origin and destination fixed effects) only explain 11% of the variation in the size of shipments. Even if the variation we explain of average size shipment is small, we see that shipments are smaller across borders and within regions, and smaller with distance.

Let us compare these results with the findings of Hillberry and Hummels (2008) for intra state and cross-states shipments in the United States. The effect of distance is remarkably similar in the two studies along all margins.[15] The coefficient of distance on total trade in Europe is -1.252 and -1.348 for the US, while the effect of distance on total shipments is -1.133 in Europe and -1.292 for the US. The effect of distance on the number of industries is -0.566 in the European data and -0.653 in the US data. Finally, an increase of 1% in distance reduces the mean kilos per shipment by 0.349% in Europe and 0.44% in the United States. The two datasets used in both studies are very similar so it is not surprising that we find similar effects of distance, but it confirms that these datasets are very comparable even if the institutional setting and geographic disaggregation of the data is different between the United States and Europe. Finally, since Hillberry and Hummels (2008) only observe domestic shipments, we will focus on the difference between our estimate of the effect of "home bias" and their estimate of the effect of "Own state". The "home bias" coefficient on total kilos and number of shipments is larger and more significant in the European data than the "Own state" in the US data. However, in the European data shipments within regions happen in less industries (column 3) and have an average smaller size per shipment (column 5), while the opposite is true in the US. These differences show that regional borders in Europe are not comparable to US internal borders, and

[15]We compare our result to the 3-digit zip code level since we do not have data as disaggregated as the 5-digit zip code in the US CFS dataset.

that the mechanisms that drive the internal border effects may differ in the two settings. Finally, A more detailed comparison country by country could tell us more about the differences and similarities between US and Europe.

6 Trade and social interactions

Our granular regional trade dataset has allowed us to present a unified picture of how regions trade within themselves and with each other. In this last section we compare our data on economic interactions to a similar regional-level dataset of social interactions. The Social Connectedness Index (Bailey et al., 2020) measures the intensity of social connectedness between locations using an anonymized snapshot of active Facebook users and their friendship networks. The data is constructed by assigning users to locations based on their information and activity on Facebook, including the stated city on their Facebook profile, and device and connection information. For Europe, the $SocialConnectednessIndex$ is available at the NUTS2 region level, allowing us to construct a matrix of social connections and compare it to our matrix of trade flows.

Formally, the Social Connectedness Index between two locations i and j is defined as:

$$SocialConnectednessIndex_{nm} = \frac{FB\_connections_{nm}}{FB\_Users_n \times FB\_Users_m} \tag{19}$$

The freely available measures of the Social Connectedness Index Bailey et al. (2020) are scaled within each dataset to have a maximum value of 1,000,000,000 and a minimum value of 1.

We use this dataset to construct a Social connectedness market share, equivalent to our normalised market shares measures (S_{nm}). The Social connectedness market share between region n and region m is given by:

$$Social_{nm} = \frac{SCI_{nm}/(\sum_n SCI_{nm})}{\sum_m SCI_{nm}/(\sum_m \sum_n SCI_{nm})} \tag{20}$$

This measure can be interpreted in the same was as our trade market shares, $Social_{nm} = 1$ means that n and m have the social connections that would be expected if social links were created under the independence assumption (would be equal to the product of the "Social potential" of each region). If $Social_{nm} > 1$, n and m are more connected than the independence assumption would predict, and viceversa for the case when $Social_{nm} < 1$.

Do regions with a high intensity of economic interactions also display a high intensity of social interactions? To compare the full distribution, we transform the two measures by taking logs and we plot one against the other. We use different colors to indicate whether the interactions are within-region, within-country or across-country. Figure 25 reports the results.

Panel 1 presents the scatter plot of the two measures. The first thing to notice is that the measures are positively correlated: region pairs with strong economic links also have strong social links. The second thing to remark is that a large part of this positive correlation seems to be driven by the borders. A region with itself has more social interactions and more economic interactions while a region with a foreign region has low intensity of economic interactions and low intensity of social interactions.

We now explore whether the positive correlation between the measures holds conditional on the borders. To do this we plot the fitted line of regressing the trade market share S_{nm} on the social market share $Social_{nm}$, for home, country and foreign interactions separately. Panels 2, 3 and 4 report the results. We find that social interactions are negatively correlated with trade interactions within a region. This seems to suggest that regions that trade largely with themselves are regions with less home online social connections. One potential explanation could be low GDP per capita or low rates of urbanization (that as we have seen are correlated with home bias) are correlated with a less intense use of online platforms. The third panel paints a different picture, social interactions are positively correlated with economic interaction for region pairs in the same country. And the slope is actually very close to one. Within a country, regions that are special trade partners also seem to be special facebook partners. However, social and trade interactions only seem to be weakly positively correlated when we look at foreign interactions. This means that, while it seems that social connectedness could be correlated with or driven by similar frictions as trade interactions, they do not help explain the home bias in trade (ie., the high values of trade flows that we observe in the data) or the variation in foreign normalised market shares. Further research and a country-level analysis could help us understand these findings.

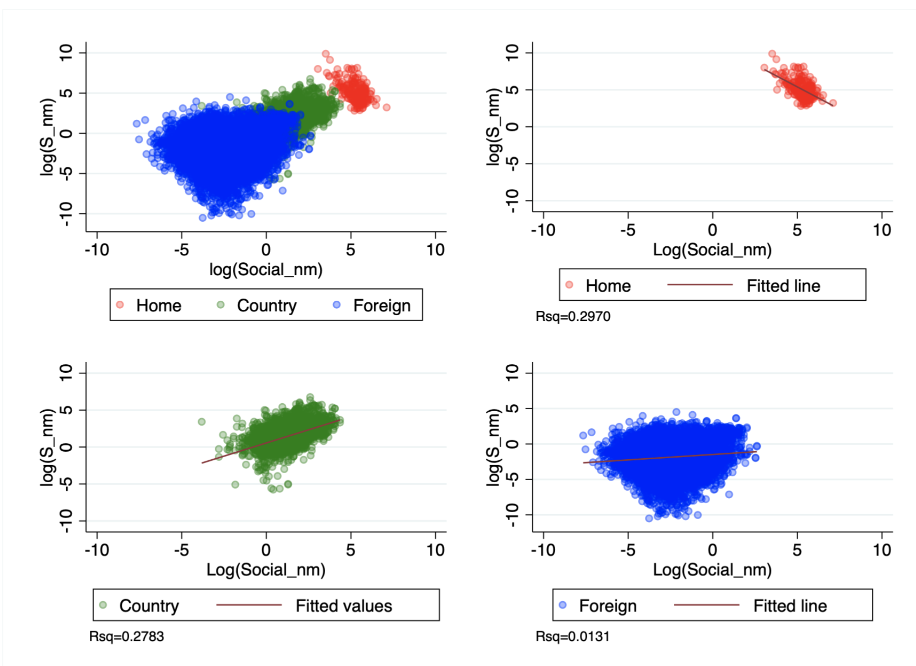


Figure 25: Economic interactions and trade interactions

6.1 Do social and economic interactions face similar frictions?

We have seen that social interactions are correlated with trade interactions in a way that is shaped heavily by borders. How are social interactions correlated with other variables such as distance? We now estimate a gravity equation in social interactions and compare the results to our (baseline) gravity estimation equations from section 3.

Table 7: Trade and Social interactions

	(1)	(2)	(3)	(4)
	Log(S_nm)	Log(Social_nm)	Log(S_nm)	Log(Social_nm)
Border Effect	-2.380***	-2.873***	-1.393***	-2.119***
	(0.0292)	(0.0269)	(0.0304)	(0.0349)
Log(Dist)	-1.150***	-0.645***	-0.994***	-0.560***
	(0.0177)	(0.0145)	(0.0176)	(0.0141)
Home Bias	1.013***	2.914***	1.333***	3.021***
	(0.0902)	(0.0811)	(0.0883)	(0.0794)
Common Language			0.794***	0.931***
			(0.0269)	(0.0257)
Both EU			1.747***	0.130
			(0.119)	(0.122)
Common Currency			0.190***	0.204***
			(0.0261)	(0.0270)
Both Schengen			1.244***	-0.536***
			(0.0560)	(0.0641)
Observations	46505	34449	46505	34449
R^2	0.611	0.803	0.632	0.817

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 7 reports the results. Columns 1 and 2 report the results of the parsimonious gravity regression using as dependent variables $\log(S_{nm})$ and

$\log(Social_{nm})$. Columns 3 and 4 present the results from the extended model for both measures, but the results are very similar. Interestingly, gravity variables have a higher explanatory power for social interactions than for trade interactions between regions. For both variables, most of our included regressors are important predictor variables. Focusing on columns 3 and 4 we see that the strength of different variables is not homogeneous. Borders are associated with higher reductions in social connections, $\log(Social_{nm})$, than with $\log(S_{nm})$, and we observe a similar pattern for the Home Bias effect. On the other hand, distance is associated with higher reductions in trade interactions, $\log(S_{nm})$, than in social interactions. This pattern seems to be consistent with the fact that facebook connections, although normally following real-life interactions, are costless to establish across space. This is not true for trade relationships, that face substantial distance costs. Common languages have positive and similar effects explaining both measures, while EU membership seems to matter only for trade. Finally, being part of the Schengen area increases trade interactions but reduces the probability of social interactions.

7 Concluding remarks

To be added

References

Anderson, J. E. and E. Van Wincoop (2003). Gravity with gravitas: a solution to the border puzzle. *American economic review 93*(1), 170–192.

Bailey, M., A. Gupta, S. Hillenbrand, T. Kuchler, R. Richmond, and J. Stroebel (2020). International trade and social connectedness. *Journal of International Economics*, 103418.

Chen, N. (2004). Intra-national versus international trade in the European Union: Why do national borders matter? *Journal of International Economics 63*(1), 93–118.

Chen, N. and D. Novy (2011). Gravity, trade integration, and heterogeneity across industries. *Journal of international Economics 85*(2), 206–221.

Combes, P.-P., M. Lafourcade, and T. Mayer (2005). The trade-creating effects of business and social networks: Evidence from France. *Journal of International Economics 66*(1), 1–29.

Coughlin, C. C. and D. Novy (2012). Is the International Border Effect Larger than the Domestic Border Effect? Evidence from U.S. Trade. CEP discussion papers, Centre for Economic Performance, LSE.

Coughlin, C. C. and D. Novy (2021, November). Estimating Border Effects: The Impact Of Spatial Aggregation. *International Economic Review 62*(4), 1453–1487.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of International Economics*, Volume 4, pp. 131–195. Elsevier.

Head, K. and T. Mayer (2021). The united states of europe: A gravity model evaluation of the four freedoms. *Journal of Economic Perspectives 35*(2), 23–48.

Head, K., T. Mayer, et al. (2002). *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade*, Volume 1. Citeseer.

Hillberry, R. and D. Hummels (2008). Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review 52*(3), 527–550.

Mayer, T. and S. Zignago (2011). Notes on cepii's distances measures: The geodist database.

McCallum, J. (1995). National borders matter: Canada-US regional trade patterns. *The American Economic Review 85*(3), 615–623.

Nitsch, V. (2000). National borders and international trade: evidence from the European Union. *Canadian Journal of Economics/Revue canadienne d'économique 33*(4), 1091–1105.

Nitsch, V. and N. Wolf (2013, February). Tear down this wall: on the persistence of borders in trade. *Canadian Journal of Economics/Revue canadienne d'Ã©conomique 46*(1), 154–179.

Redding, S. and A. J. Venables (2004, January). Economic geography and international inequality. *Journal of International Economics 62*(1), 53–82.

Santamaría, M. A., J. Ventura, and U. Yeşilbayraktar (2020, December). Borders within europe. (28301).

Tinbergen, J. . (1962). *Economic geography and international inequality.* New York, NY: Twentieth Century Fund.

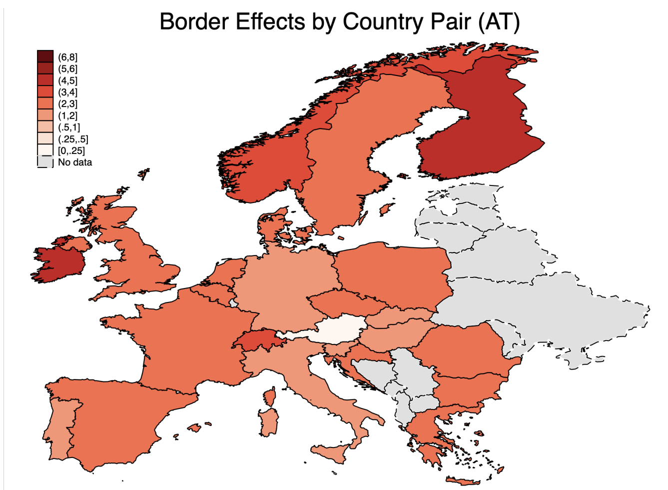
Wei, S.-J. (1996). Intra-national versus international trade: how stubborn are nations in global integration? Technical report, National Bureau of Economic Research.

Wrona, J. (2018). Border effects without borders: What divides japan's internal trade? *International Economic Review 59*(3), 1209–1262.

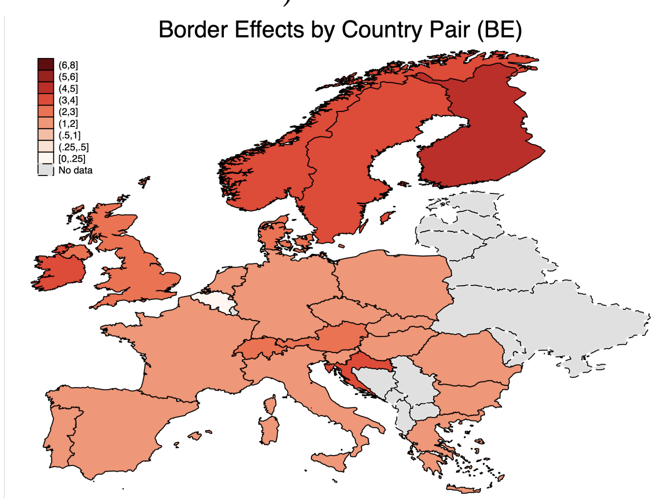
Part I

Appendix

A Additional Figures and Tables

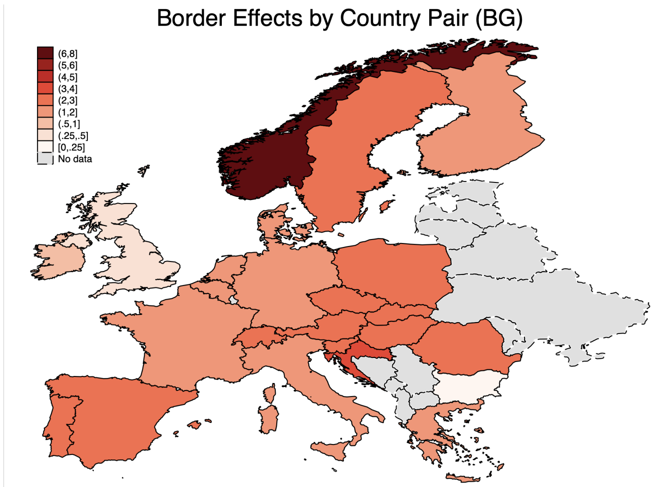


A) Austria

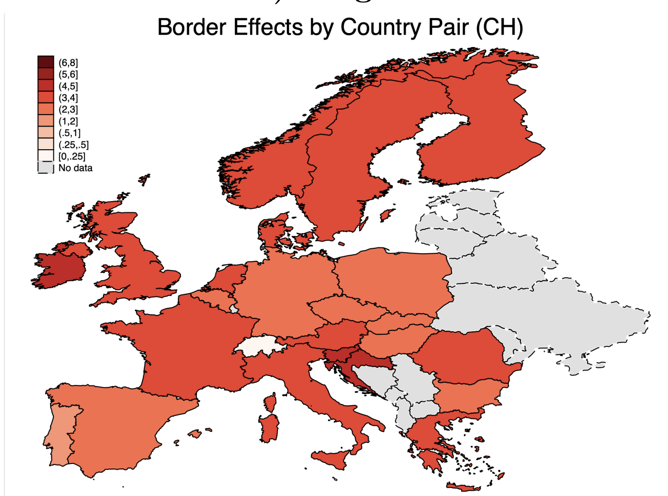


B) Belgium

Figure 26: Heterogeneity by Country-pairs

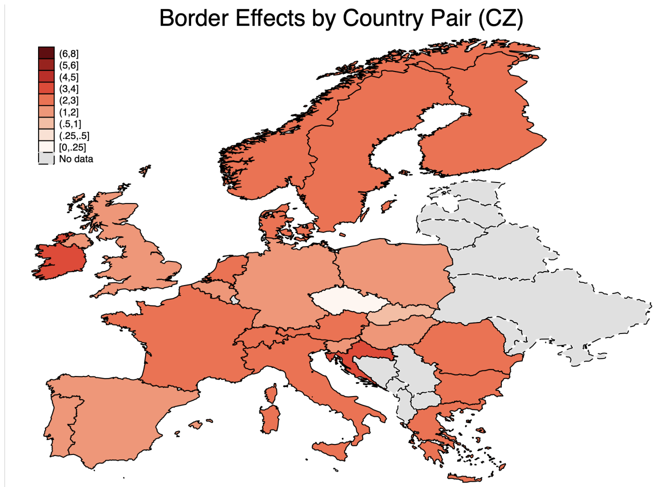


A) Bulgaria

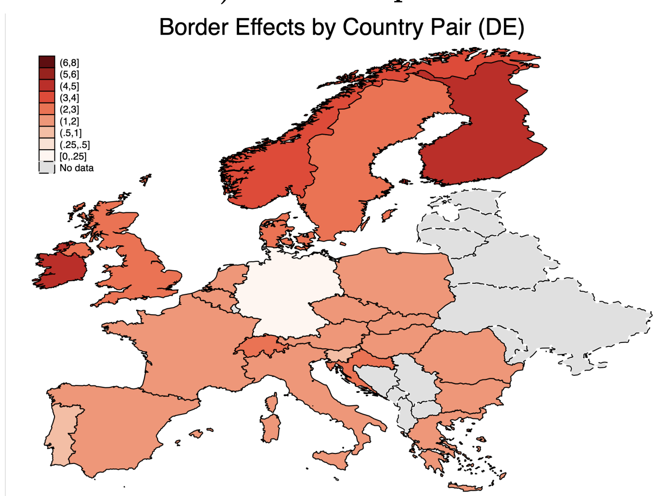


B) Switzerland

Figure 27: Heterogeneity by Country-pairs

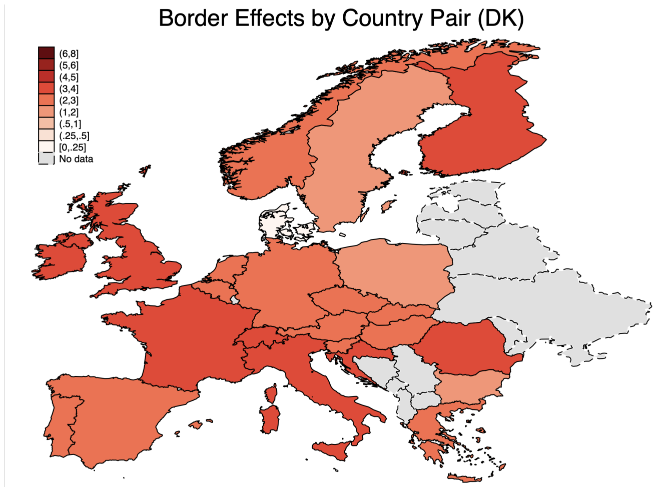


A) Czech Republic

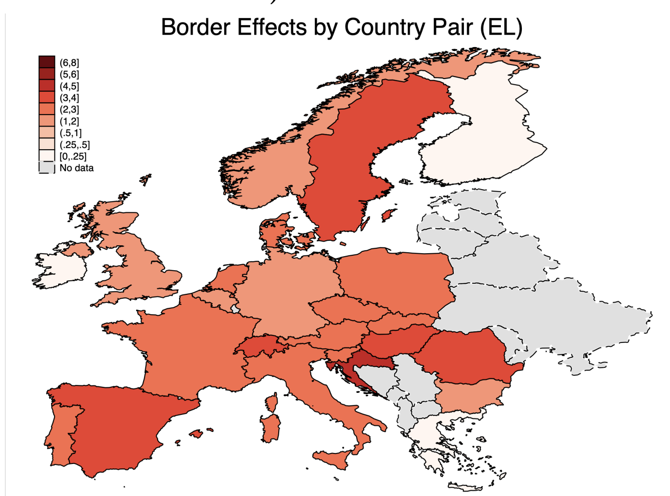


B) Germany

Figure 28: Heterogeneity by Country-pairs

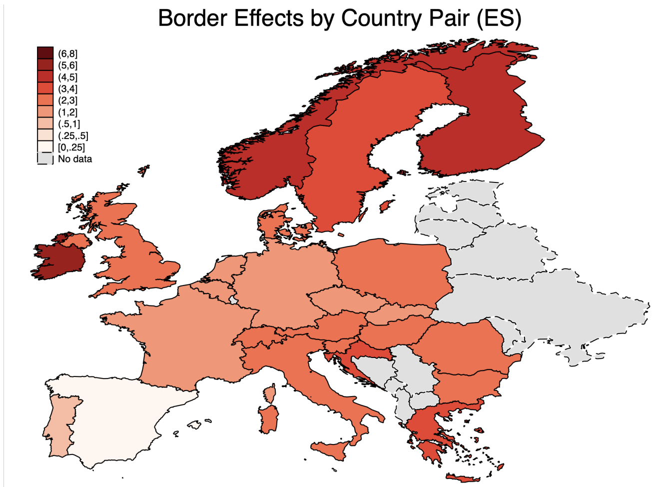


A) Denmark

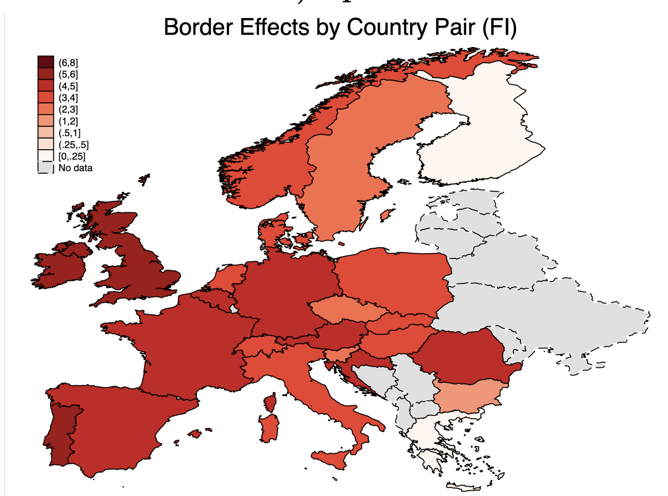


B) Greece

Figure 29: Heterogeneity by Country-pairs

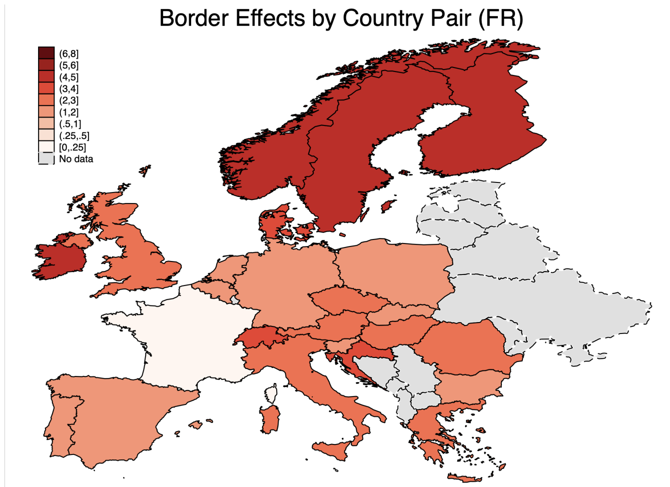


A) Spain

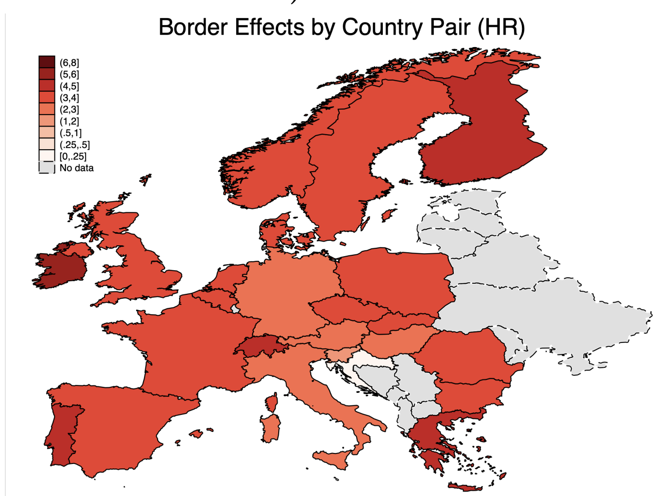


B) Finland

Figure 30: Heterogeneity by Country-pairs

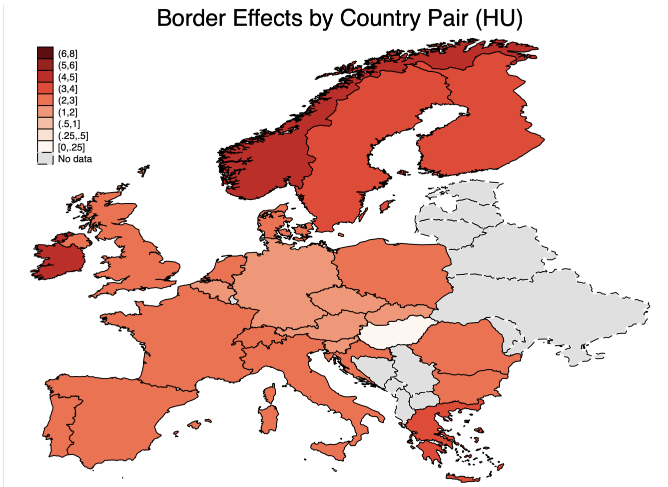


A) France

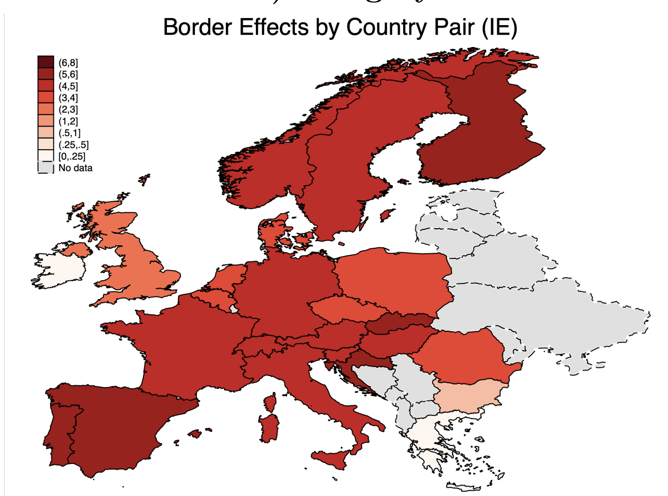


B) Croatia

Figure 31: Heterogeneity by Country-pairs

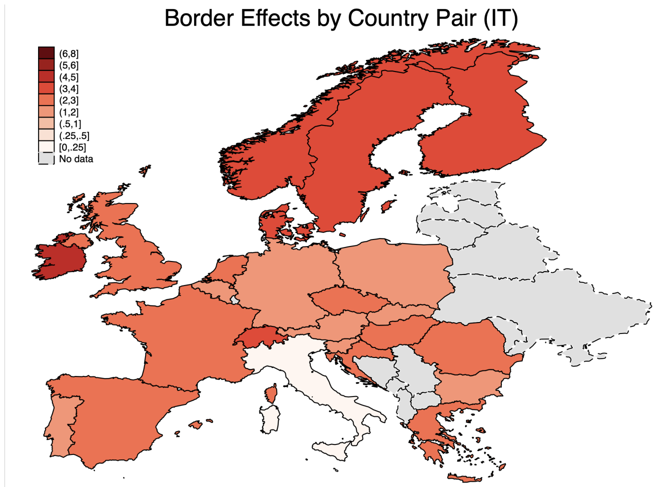


A) Hungary

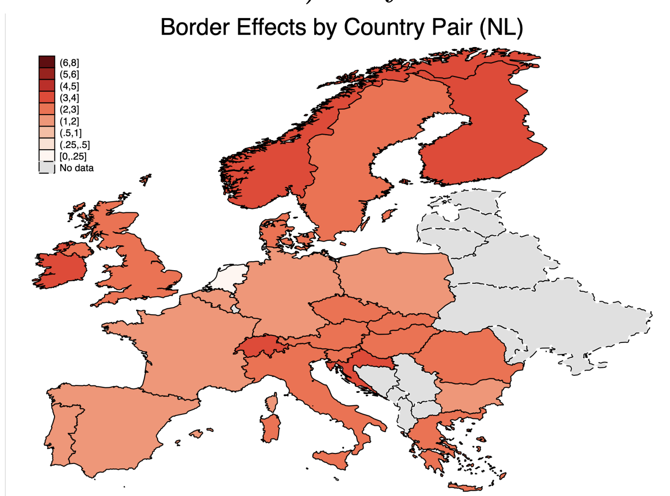


B) Ireland

Figure 32: Heterogeneity by Country-pairs

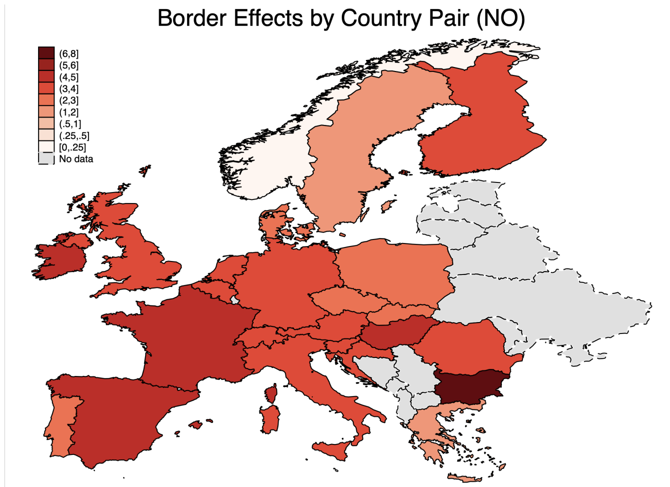


A) Italy

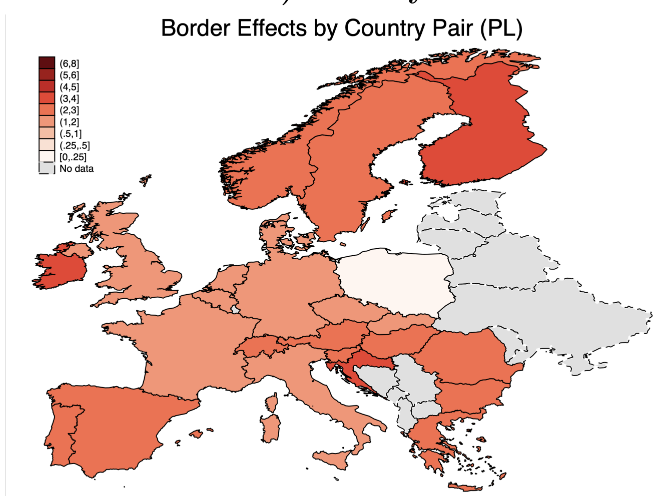


B) The Netherlands

Figure 33: Heterogeneity by Country-pairs

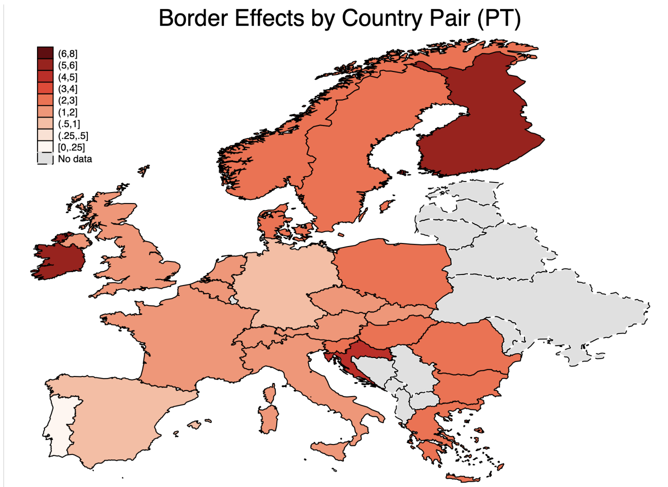


A) Norway

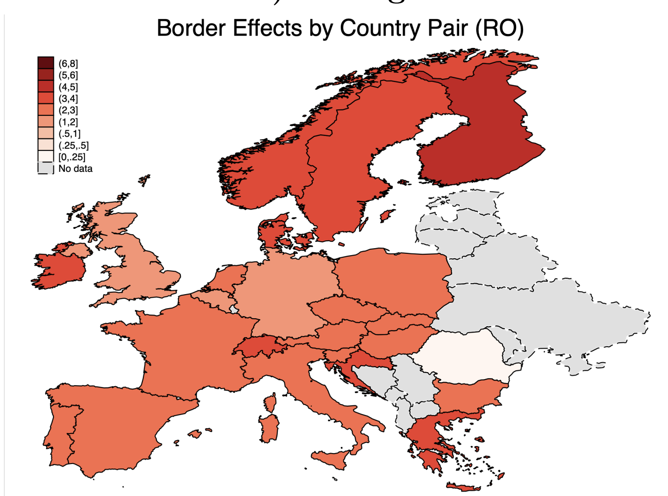


B) Poland

Figure 34: Heterogeneity by Country-pairs

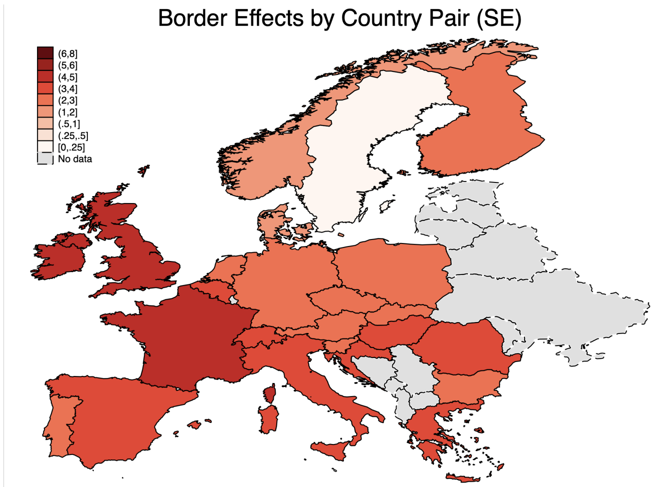


A) Portugal

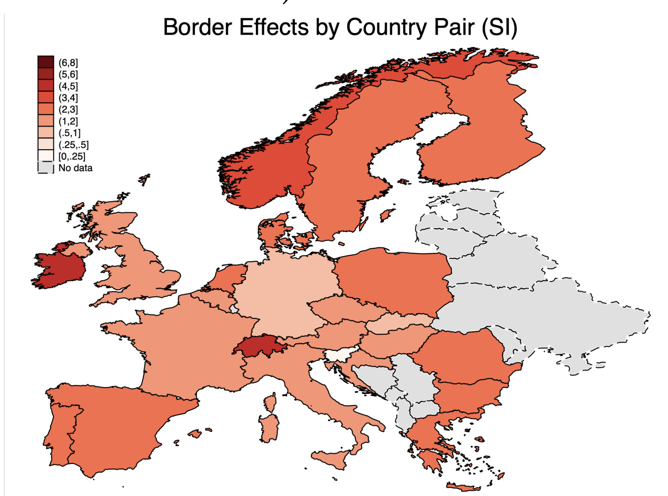


B) Romania

Figure 35: Heterogeneity by Country-pairs

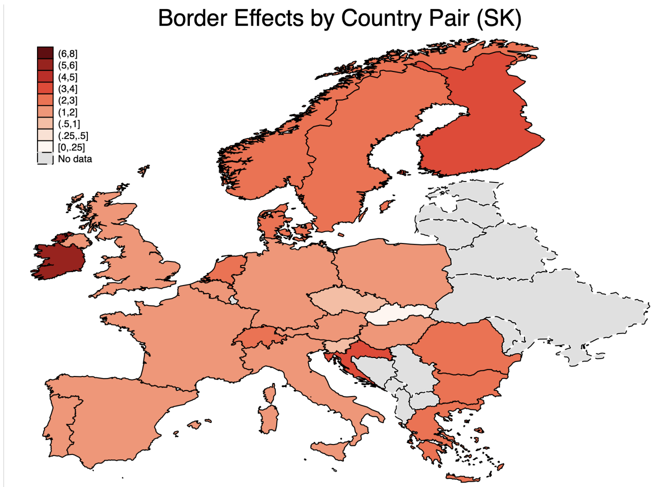


A) Sweden

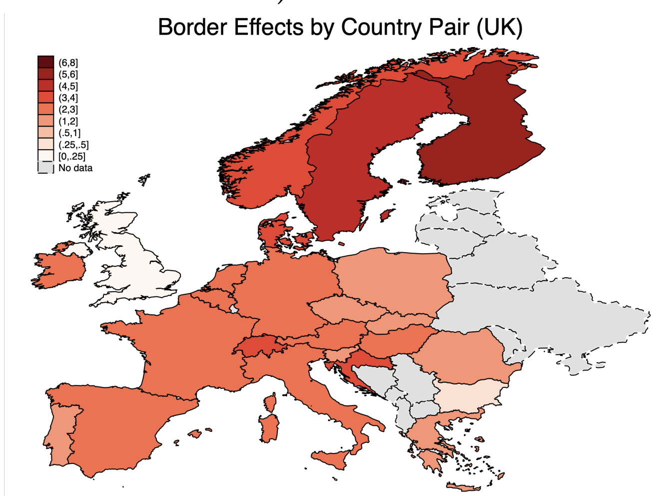


B) Slovenia

Figure 36: Heterogeneity by Country-pairs



A) Slovakia



B) United Kingdom

Figure 37: Heterogeneity by Country-pairs