Exploring European Regional Trade *

Marta Santamaría, Jaume Ventura and Uğur Yeşilbayraktar

June 10, 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 characterize for the first time trade patterns across and within European regions.

Europe is a great laboratory to explore trade flows across regions and countries. The first advantage of our data is that we can observe regions and countries of different economic, geographic and demographic sizes. In 2019, the poorest country is our data had a (nominal) GPD per capita of 9,828 USD, Bulgaria, while the richest, Switzerland, had a GPD per capita of 81,994 USD.[1] The smallest country in our data, Slovenia, is 20,273

*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.

[1]We report statistics from before 2020 to avoid capturing changes caused by the COVID-19 emergency.

square kilometers while the largest country in our data, France, is 551,695 square kilometers. Finally, the least populated country we observe is Slovenia with 2,077,837 inhabitants in 2018, while the largest country, Germany, had 83,124,418 inhabitants in 2018.

In addition to the great heterogeneity of countries and regions we can observe, Europe is an important economic region of the world. The 24 countries in our data account for 63% of the OECD members and for more than 20% of world GDP. The European Union, the largest economic union in the world, is remarkably important when we study international trade. It accounts for 27% of world trade, with intra-European trade alone accounting for 18% of world trade. Exports of European Union members to each other sum up to four trillion USD annually, larger than total Chinese exports to the rest of the world (3 trillion USD) and total US exports to the rest of the Exports of European Union members to each other sum up to 4.2 trillion USD in 2019, larger than Chinese exports to the rest of the world (2.5 trillion USD in 2019) and more than US exports to the rest of the world (1.6 trillion USD in 2019).[2].

Moreover, our dataset not only allows us to measure trade flows between European countries but trade flows happening between regions in the same country (including within-region) which are orders of magnitude larger than international flows, in both value and weight. Thus, our detailed and deep examination of trade flows within and across European regions will allow us to derive lessons about a large share of the global trading activity.

The paper is organized as follows. Section 2 examines the raw data shows its main features. 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

[2]EU statistics from Eurostat "https://ec.europa.eu/eurostat/documents/2995521/10624801/6-25032020-AP-EN.pdf/c65584e1-c88c-f038-aec1-bd1235c395e1". China and US statistics from WITS, World Bank: "https://wits.worldbank.org/countryprofile/"

shipment has a destination outside the origin region but inside the origin country (country trade) is 41 percent. The probability that this shipment has a destination outside the origin country (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,[3] we find enormous differences in the magnitudes of these types of trade. In particular, home, country and foreign trade are 470, 11 and 0.44 times what one would predict knowing only the sizes of the origin and destination markets.

Section 3 takes a more formal look at the data using the standard gravity model. A parsimonious model that uses only national borders and the (log) of distance can explain about two-thirds of the variation in European regional trade. We show how to reach this conclusion using the popular fixed-effects regressions (with a twist!), and also through a standard calibration exercise. Section 4 goes a bit deeper into the effects of borders and explores the extent to which these are heterogeneous. We find that indeed there is substantial heterogeneity. Borders in the periphery of Europe tend to be more important than in the center.

Section 5 looks at the home bias in trade. There are few observations of home trade, 269 out of 73,361, but these observations stand for their size. Recall that they add to 40 percent of all trade. The models analyzed in section 3 work well in explaining average trade, but perform less well in explaining home trade. Why is home trade so large? We show in this section that, to a large extent, this is due to presence of regional political borders. 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

[3]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.

regional borders matter as much or more than national borders.

Section 6 examines an additional margin that may explain the heterogeneity in magnitude of home, country and foreign trade. We decompose trade into intensive (average size of shipments) and extensive (average shipments per industry and number of industries) margins. First, these margins don't 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, section 7 uses 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 **?**. 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)). [4] 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 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

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

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 much more with themselves than with other regions within the same country, and regions trade much more with regions within the same country 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 be 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 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 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 shipments of goods from origin n to destination m. We measure shipments as a share of total shipments:

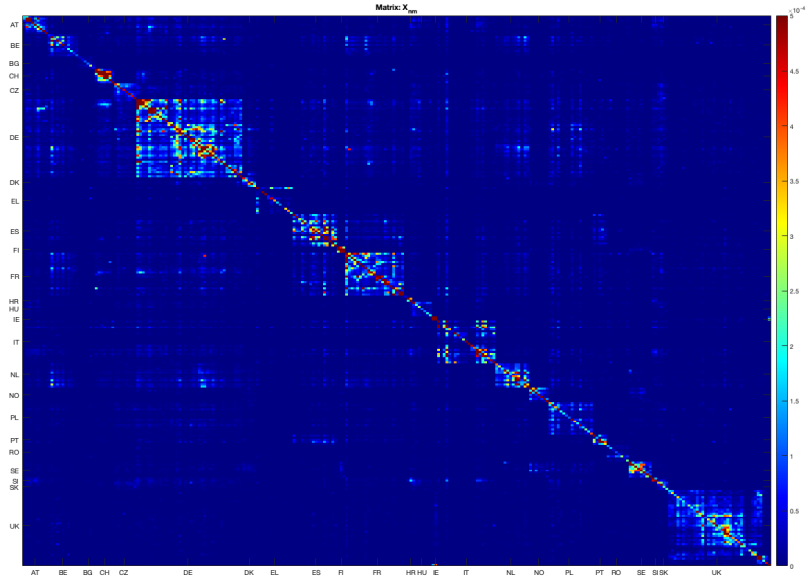
$$_{nm}X_{nm} = 1 \tag{2}$$

Thus, X_{nm} is the probability that a shipment of goods originates in n and has destination m.

Figure 1 shows a heat map of the matrix of bilateral trade. We classify the entries of this matrix into three distinct types, home, country and foreign 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 1 as the boxes centered around the diagonal (without including the latter). Larger boxes refer to countries with more regions, such as Germany or France. Smaller boxes 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 1 as the off-diagonal entries outside the boxes. 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 1 contains a fair amount of zeros. Naturally, there

Figure 1: Bilateral Trade Matrix: X_{nm}



are no zeros for home trade. But there are a few zeros for country trade: 157 out of 4958 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 consistent with a strong home and country biases in European regional trade.

What explains these biases? A prime suspect is distance, of course. The distance traveled by shipments classified as home, country and foreign trade is not the same. Fortunately, the ERFT survey provides the actual distance traveled by each individual shipment, including shipments within and across regions. Thus we can use actual distance and we are not forced to rely on estimates as it is customary in studies of bilateral trade at the country level.

As a short digression, it is instructive to compare actual distance with the estimates we would obtain using the methods used in these studies. The standard approach in the literature is to estimate the unknown actual dis-

tance traveled between two regions using the length of a straight line between the centres of the origin and destination regions. This approach, however, is not suitable to estimate the distance traveled by shipments within a region. Thus, the literature typically estimates the unknown actual distance traveled within a region using the area of the region (Nitsch (2000), Redding and Venables (2004)), Head and Mayer (2014)).[5]

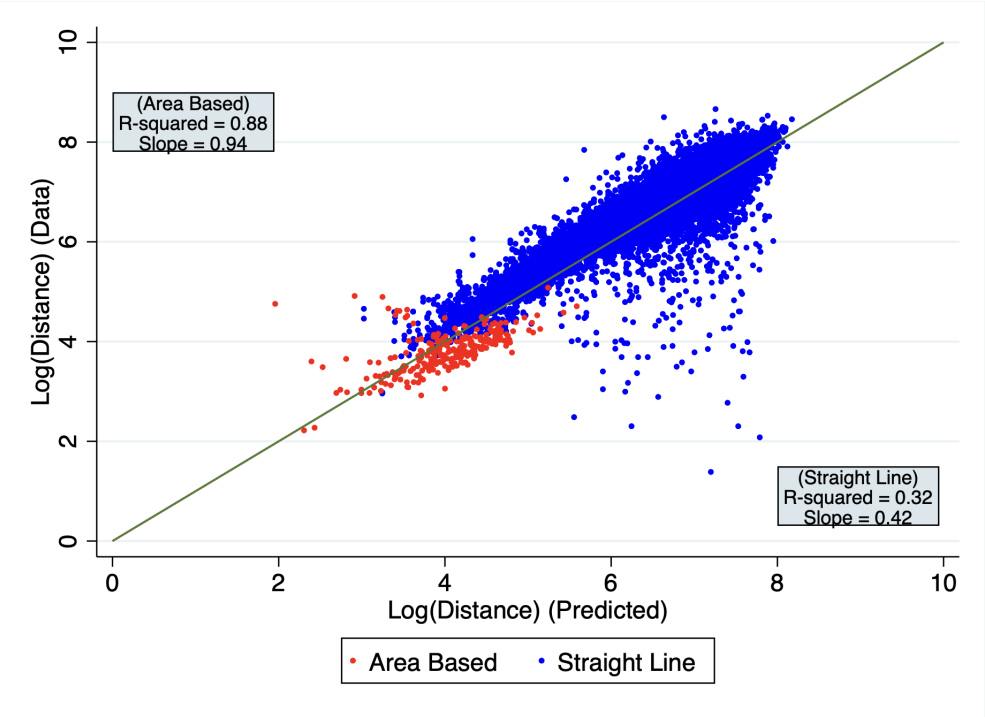
Figure 2 plots the actual (log) distance reported in the survey (vertical axis) with the (log) distance we would obtain using these methods (horizontal axis). To ease comparisons, we plot the 45^o line and use different colors for estimates across regions that use the straight-line measure and estimates within regions that use the area-based measure. The straight-line measure seems to explain only about a third of the variation in actual distances. Moreover, a regression of the actual distance on the predicted one delivers a slope coefficient of 0.42, suggesting that the straight-line measure tends to overestimate long distances and underestimate short ones. The area-based measure explains about almost 90 percent of the variation in actual distance. Moreover, the slope coefficient of a regression of actual distance on predicted one is about 0.95, indicating that this measure does not systematically overestimate or underestimate long and short distances.

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

[5] An alternative is to use the weighted average of the bilateral straight-line distances of the largest cities inside the region, using the size of the cities as weights Mayer and Zignago (2011). This approach seems reasonable for large geographic units, but it may be problematic when computing distances for small units such as regions since many have only one large city.

Figure 2: Actual vs Predicted Distance



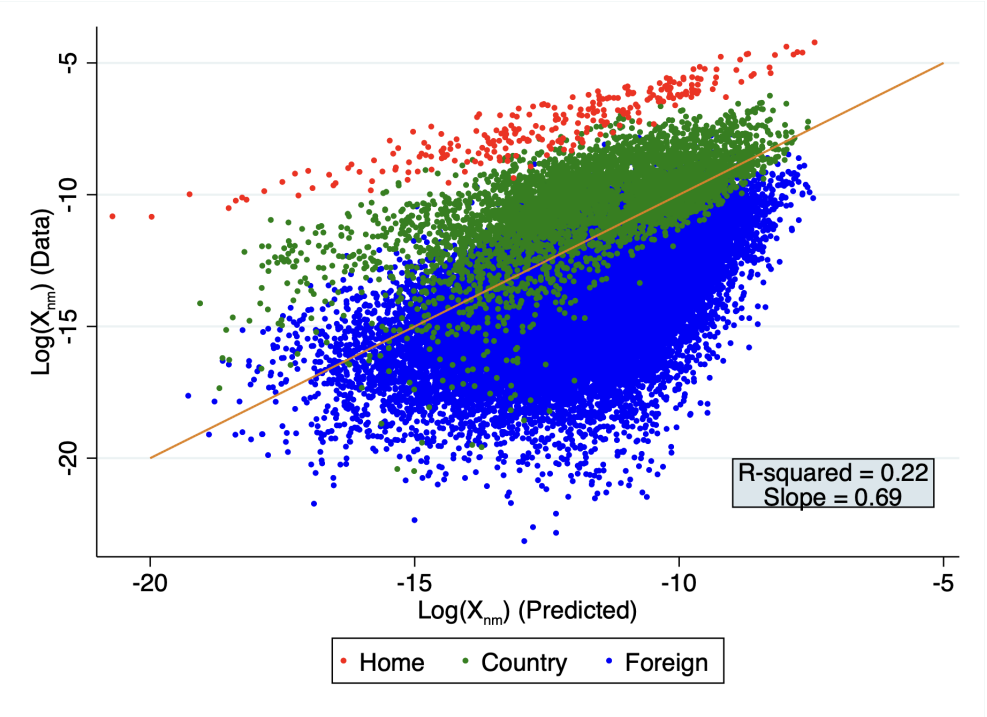
To answer these questions, we need a benchmark that is size free. To understand why, consider the case of Catalonia and La Rioja. The probabilities of a sale to the Basque Country by Catalonia and La Rioja are 0.000226 and 0.0000542, respectively. The probabilities of a purchase by the Basque Country from Catalonia and La Rioja are 0.0004281 and 0.0000601, respectively. Catalonia's 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 consider 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 consider this *independence benchmark* prediction: "the probability of a shipment from origin n to destination m should be $X_n^O X_m^D$." One can interpret this prediction as a theoretical assertion or as a forecast with limited information. Consider a theory that asserts 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. This implies that the events O_n and D_m are pairwise independent and, as a result, X_{nm} must be $X_n^O X_m^D$. Consider alternatively the problem of forecasting the trade probability between origin n and destination m, with information only about the sizes of these regions. With this limited information, the only rational forecast for X_{nm} is $X_n^O X_m^D$. Whether it is a theoretical assertion or a forecast with limited information, the independence benchmark encapsulates the notion 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 patterns of deviations from the independence benchmark to learn about the determinants of trading relationships.

Figure 3 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 3 is that three quarters of the

Figure 3: Actual vs Predicted: $\log(X_{nm})$



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

$$S_{nm} = \frac{X_{nm}}{X_n^O X_m^D} \tag{3}$$

In Santamaría et al. (2021), we referred to S_{nm} as a normalized market share.[6] This 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 origin and destination regions. Thus, S_{nm} is a size-free measure of how strong a trade relationship is.[7]

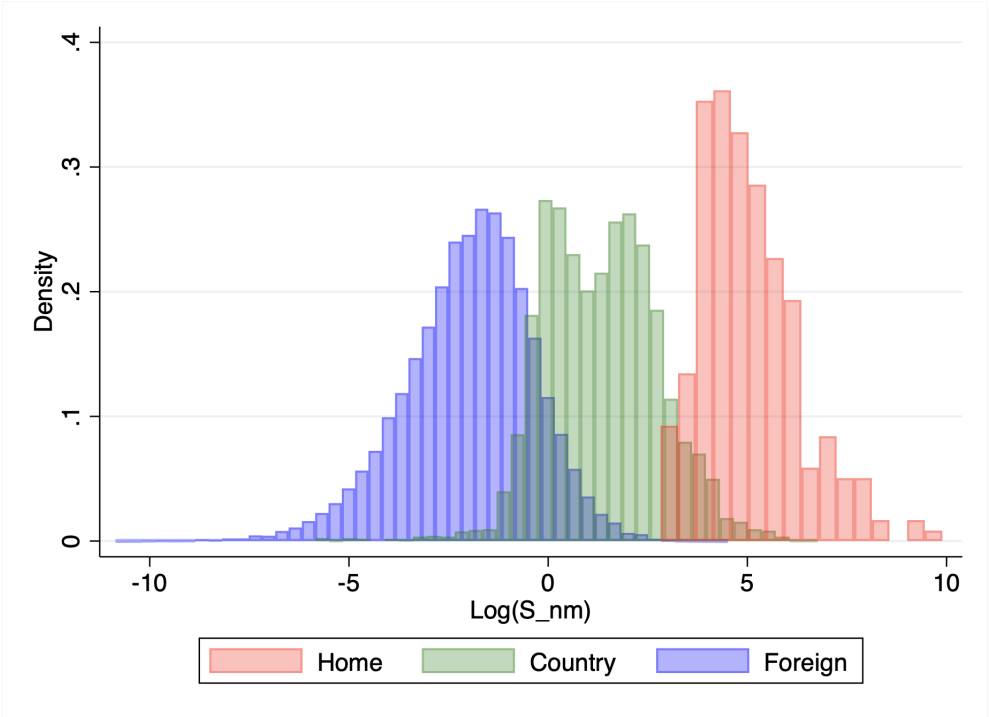
Figure 4 plots histograms of normalized market shares separately for home, country and foreign trade. The average normalized market shares for the different types of trade are 469.5, 11.22 and 0.44. The distributions of normalized market shares for these types of trade have little overlap. The ranking 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.

How do these biases vary across European regions? To answer this question, we compute normalized market shares for home, country and foreign trade for each region in our sample. To construct these shares we calculate for each region the probabilities of home trade (with itself), country trade

[6]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.

[7]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 might be that La Rioja is much closer geographically to the Basque Country than Catalonia is.

Figure 4: Normalized Market Shares



(with the rest of the regions within its country) and foreign trade (with the regions outside its country). We then normalize these trade probabilities by the sizes of the respective markets.[8] The results of these calculations are presented in map format in Figure 5 , 6 and 7.

[8]Formally, these shares are defined as follows:

$$S_{nH} = S_{Hn} = \frac{X_{nn}}{X_n^O X_n^D}$$

$$S_{nC} = \frac{\sum_{m \in C_n} X_{nm}}{X_n^O \sum_{m \in C_n} X_n^D} \quad \text{and} \quad S_{Cm} = \frac{\sum_{n \in C_m} X_{nm}}{X_m^D \sum_{n \in C_n} X_n^O}$$

$$S_{nF} = \frac{\sum_{m \in F_n} X_{nm}}{X_n^O \sum_{m \in F_n} X_n^D} \quad \text{and} \quad S_{Fm} = \frac{\sum_{n \in F_m} X_{nm}}{X_m^D \sum_{n \in F_n} X_n^O}$$

where C_n is the set of regions that belong to the country of region n (excluding region n itself), and F_n is the set of regions that do not belong to the country of region n.

Figure 5: Normalized Market Shares: Home

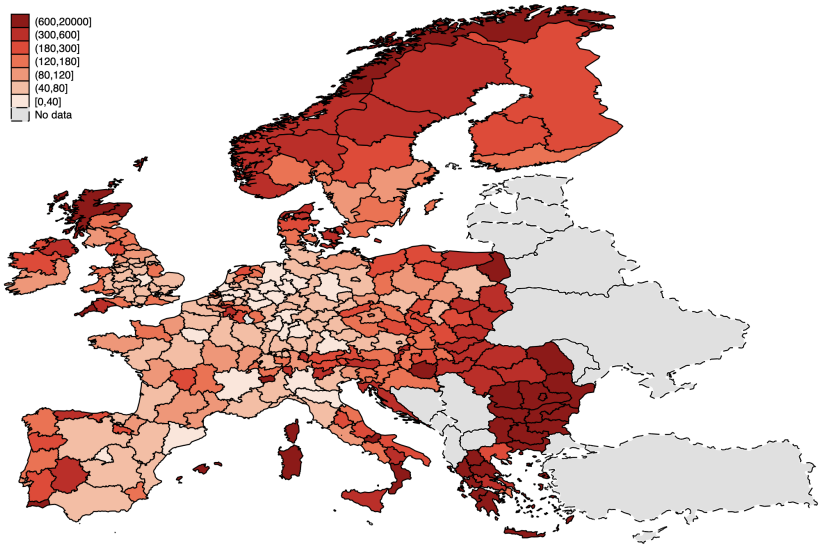
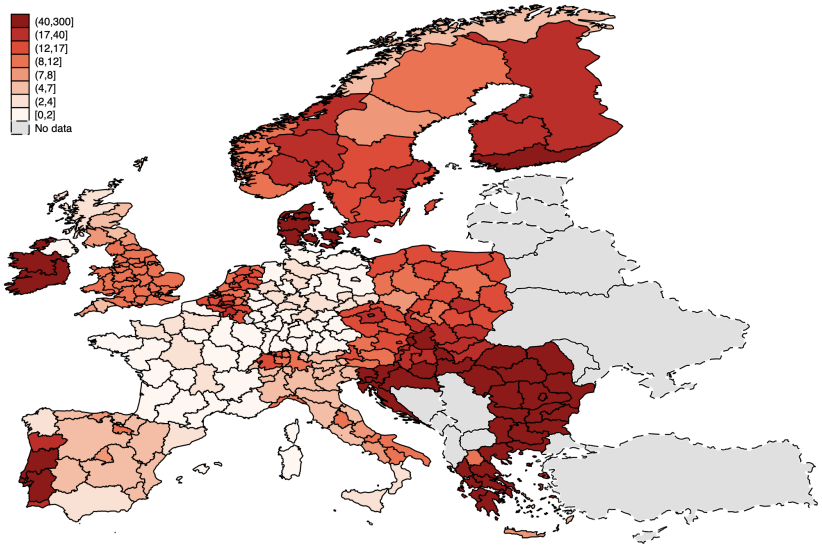


Figure 5 shows normalized market shares for home trade. A quick look at the map shows that home trade tends to be larger in remote regions, i.e., those at the periphery of Europe. We also see that Home trade tends to be lower in the more densely populated regions of Europe. We see this pattern at the European level in the so-called Blue Banana.[9] 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.

Figure 6 shows normalized market shares for country trade. Again, we see that country trade tends to be larger in remote countries. We also see that country trade tends to be smaller in large countries with many regions,

[9]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.

Figure 6: Normalized Market Shares: Country

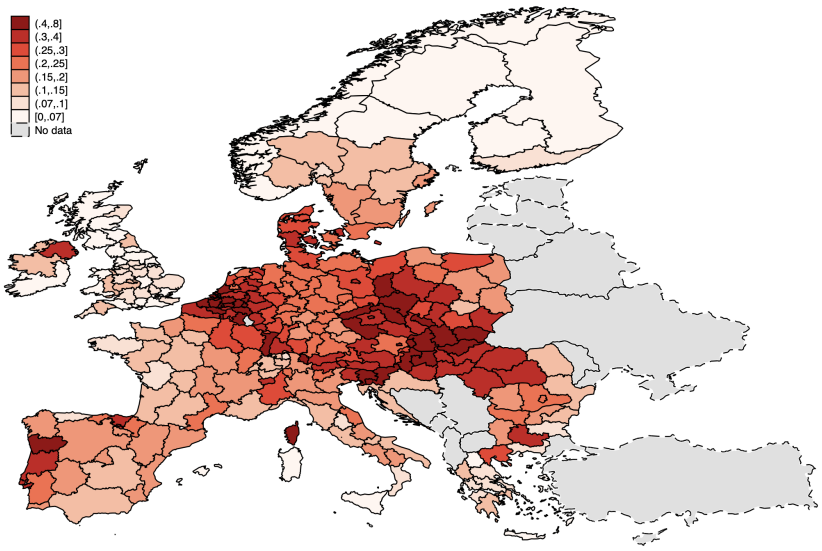


such as France or Germany.

Figure 7 shows normalized shares for foreign trade. Yet again, we find that remoteness matters but now in a reverse fashion. Foreign trade tends to be larger the more remote the region is. We also observe two large areas in Europe with a large foreign trade relative to the rest. The first one is located around Belgium and the Netherlands, and includes some French and German regions. The second area is located in the East of Europe and contains regions from Austria, Hungary, Poland, the Czech republic and nearby countries.

While looking at Figures 5 , 6 and 7, some readers might have noticed that some regions tend to have systematically large normalized market shares for all types of trade, while some regions tend to have systematically low shares. For instance, Darmstadt's (Germany) normalized market shares for home, country and foreign trade are 22.66, 1.67 and 0.2, respectively; while Central Macedonia's (Greece) are 294.1, 43.14 and 0.28. How is this possible? Darmstadt is a large region in a large country, while Central Macedonia is a small region in a small country. Darmstadt has a strong trading relationship

Figure 7: Normalized Market Shares: Foreign

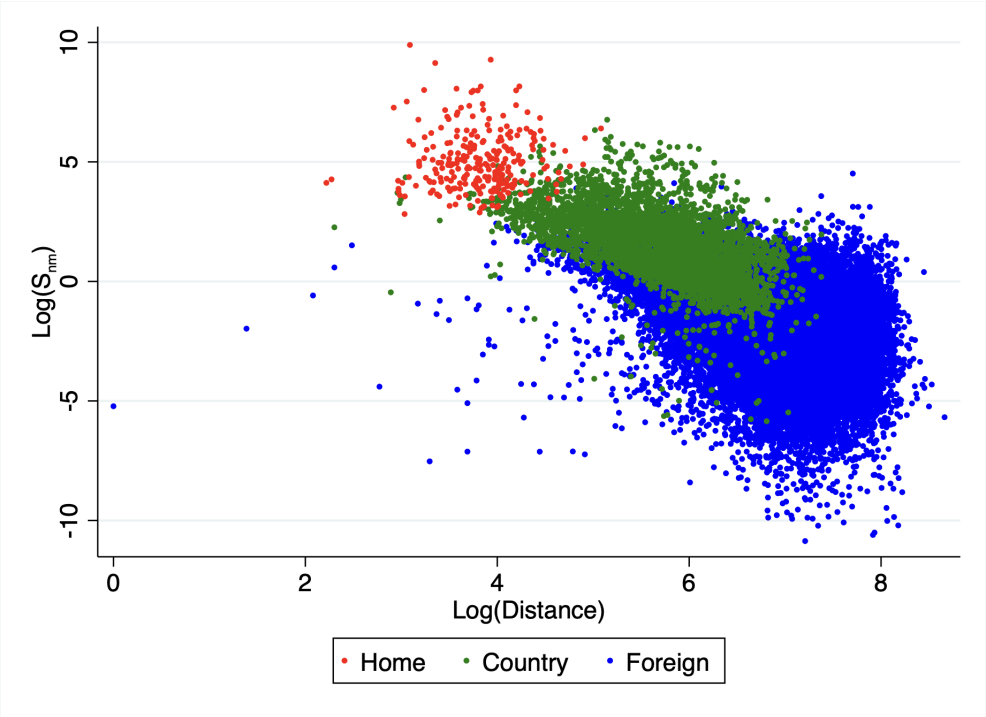


with many large regions (German regions including itself) which add to 29% of Europe. Central Macedonia has an even stronger trading relationship with a few small regions (Greek regions including itself) which add to 1% of Europe. It is not possible to have strong relationships with everyone (same as not every observation can be above average!) If you have strong relationships with *many large* regions, not only each of these relationships cannot be too strong but also the remaining relationships must be very weak. If you have strong relationships with only a *few small* regions, these relationships can be very strong and yet the remaining relationships do not have to be very weak.[10]

Finally, and just to whet the appetite for what is coming next, Figure 8 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 ex-

[10]The rules of probability imply that: $S_H \times X_H^D + S_C \times X_C^D + S_F \times X_F^D = 1$. In the case of Darmstadt, this calculation is $22.6618 \times 0.01751 + 1.6756 \times 0.27309 + 0.2053 \times 0.70941 = 1$. In the case of Central Macedonia, this calculation is $294.1044 \times 0.00131 + 43.1463 \times 0.00791 + 0.2758 \times 0.99077 = 1$.

Figure 8: Normalized Market Shares vs Distance



plains part of the home and country biases in trade. But Figure 8 also shows that distance cannot be the single explanation for these biases. Within any given distance interval, country trade is always larger than foreign trade. And home trade is even larger. What else is going on? We turn next to a systematic examination of the data using the standard gravity framework.

3 A gravity look at the data

The gravity framework has dominated research on international trade for the last couple of decades. In this section, we use this framework to organize our exploration of the European regional trade data. First, we estimate a series of fixed-effects regressions to test whether some popular gravity models provide

a good explanation of the data. Second, we perform a simple calibration exercise using a parsimonious version of the model to find the best set of paramater values to be used in applications using this data.

3.1 The gravity framework

The gravity framework provides a specific mathematical structure that adjusts trading probabilities to take into account distance and other variables. Let M_{nm} be a measure of the cost of shipping goods from origin n to destination m. We refer to M_{nm} as bilateral market access. Gravity models postulate a bilateral market access function of this form:

$$M_{nm} = \exp\left\{{}_i\theta^i Z^i_{nm}\right\} \tag{4}$$

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 share a language, a currency, or membership into an international agreement such as the European Union or the Schengen treaty.

The gravity framework consists of the following mathematical model:

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

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

$$S_{nm} = \frac{M_{nm}}{M^O_n M^D_m} \tag{6}$$

where M_n^O and M_m^D is a set of numbers that satisfy the following restrictions:

$$1 =_m X_m^D \frac{M_{nm}}{M_n^O M_m^D} \tag{7}$$

$$1 =_n X_n^O \frac{M_{nm}}{M_n^O M_m^D} \tag{8}$$

We refer to M_n^O and M_m^D as origin and destination measures of average market access.[11] Equations (7) and (8) are not additional theoretical restrictions, but instead consistency requirements that ensure that probabilities add, i.e., $1 =_m X_m^D S_{nm}$ and $1 =_n X_n^O S_{nm}$.[12]

It is well known that there is a a large set of theoretical models that are consistent with the formulation of the gravity framework in Equations (6), (7) and (8) (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. If two gravity models generate the same set of bilateral market access measures $\{M_{nm}\}$, they both predict the same matrix of bilateral trade. These models might reach this prediction using different arguments and offer different interpretations of the coefficients. These models might also generate conflicting predictions for other aspects of the data, such as the microeconomic patterns of consumption and production. For our purposes, however, these two models are observationally equivalent and they will be treated as a single model here.

[11] The literature often refers to these terms as multilateral resistance terms or price levels. We think that labeling them as origin and destination measures of market access is more transparent.

[12] These are the restrictions we used in an earlier footnote describing the trade of Darmstad and Central Macedonia.

3.2 Fixed-effects regressions

Let us 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 can compute the deviations between actual and predicted probabilities:[13]

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

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 possible. 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 a related fixed-effects regression:

$$\ln S_{nm} = \phi_n^O + \phi_m^D +_i \theta^i Z_{nm} + u_{nm} \tag{9}$$

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 $\ln M_m^D$.

To show that the model provides a good explanation of the data, we need to show that (i) the estimates of the theoretical paramters $\{\theta^i\}$ are significative and have the correct theoretical sign; (ii) the R-squared ot the

[13] Knowing $\{M_{nm}\}$, we can compute $\left\{M_n^O\right\}$ and $\left\{M_m^D\right\}$ using Equations (7) and (8).

regression is high enough; (ii) the estimates of the fixed effects are consistent with the origin and destination market access measures that are implied by our estimates of the theoretical parameters $\{\theta^i\}$.[14] If these requirements are not met, we conclude that the gravity model considered does not provide a good explanation of the data.

Note that, using $\ln S_{nm}$ as the dependent variable and not $\ln X_{nm}$ as is customary, allows us to impose stricter requirements on the proposed model. Using $\ln S_{nm}$ as the dependent variable means that the R-squared of regression (9) is not inflated by the mechanical correlation between size and trade probabilities. Thus, it is more difficult to obtain a high R-squared here than in previous studies that use $\ln X_{nm}$. Using $\ln S_{nm}$ as the dependent variable also allows us to perform an additional check that was not possible in previous studies that use $\ln X_{nm}$. Since we can interpret the fixed effects as estimates of the theoretical origin and destination market access measures, it is possible to check whether they are consistent with the estimated parameters.[15]

Table 1 shows the results of estimating the fixed-efffects regression for four alternative gravity models. Column (1) shows a parsimonious model that includes only a border dummy and a simple measure of distance that assumes a constant elasticity of trade to distance:

$$D_{nm} = \sigma \ln T_{nm} \tag{10}$$

where T_{nm} are Kms traveled. Despite the small number of (coarse) variables, this model explains almost two-thirds of the variation in trade probabilities.

[14] That is, we can use the estimated parameter values $\{\theta^i\}$ to produce estimates of the bilateral market access measures $\{M_{nm}\}$ and use these to construct estimates of $\{M_n^O\}$ and $\{M_m^D\}$ using Equations (7) and (8). These estimates should be consistent with the estimates of the region origin/destination fixed-effects.

[15] 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.

This is especially remarkable given that we have eliminated the effects of size using (log) normalized market shares instead of (log) trade probabilities.

The estimated effects of borders and distance are significative and economically large. The estimated coefficient for the border dummy means that, controlling for distance, being separated by a national border reduces bilateral trade to $\exp\{-2834\} \times 100 = 9$ percent of the independence 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 distance are very helpful to predict deviations from the independence benchmark.

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, $b = 1, ..., B$:

$$D_{nm} = \begin{cases} \sigma_1 \ln T_{nm} & \text{if } 0 < T_{nm} \leq T_1 \\ \sum_{i=2}^{b} (\sigma_{i-1} - \sigma_i) \ln T_{i-1} + \sigma_i \ln T_{nm} & \text{if } T_{b-1} < T_{nm} \leq T_b \end{cases} \tag{11}$$

Naturally, this second specification collapses to the first one if $\sigma_b = \sigma$ for all $b = 1, ..., B$. Using this flexible specification does not seem to have a noticeable effect on the results. The R-squared and the border coefficient are not affected. The effect of distance seems a bit weaker though. Although now the elasticity is not constant, we can still determine this effect by plotting the effect of distance on trade for both specifications. Figure 9 shows our estimates for the two distance functions in Equations (10) and (11). The constant-elasticity specification is always above the varaible-elasticity specification. The difference however does not seem to be large.

In columns (3) and (4) we add four dummy variables that take value one if the regions involved share a language, a currency or membership in the European Union and the Schengen treaty, respectively. Column (3) uses the

Table 1: Gravity: Fixed Effects Regressions

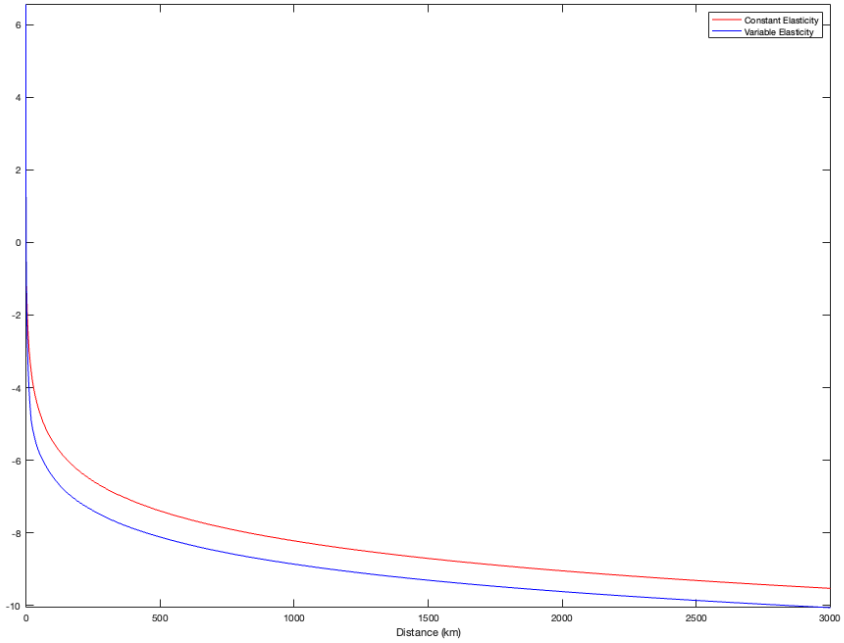
	(1) Log(S_nm)	(2) Log(S_nm)	(3) Log(S_nm)	(4) Log(S_nm)
Border Effect	-2.384***	-2.340***	-1.426***	-1.387***
	(0.260)	(0.243)	(0.157)	(0.153)
Log(Dist)	-1.190***		-1.052***	
	(0.0668)		(0.0576)	
Common Language			0.764***	0.783***
			(0.0881)	(0.0921)
Both EU			1.826***	1.749***
			(0.386)	(0.397)
Common Currency			0.190	0.176
			(0.127)	(0.124)
Both Schengen			1.205***	1.231***
			(0.252)	(0.249)
Origin FE	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes
Distance Bins	No	Yes	No	Yes
Observations	46505	46505	46505	46505
R^2	0.610	0.611	0.630	0.631

Standard errors in parentheses

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

constant-elasticity distance function, while the Column (4) use the variable-elasticity distance function. This does not seem to produce any substantial difference here either. Three out of four of the dummy variables are significative and consistent with theoretical predictions. The only dummy variable that does not seem to matter is the common-currency dummy. Interestingly, the border effect declines. One (admittedly loose) interpretation is that part

Figure 9: Distance Functions: Constant vs Variable Elasticity



of the border effects captured in columns (1) and (2) works through these variables.

Figure 10 shows whether the estimated fixed-effects are consistent with origin and destination measures of market access. This Figure has four panels, one for each model. Each panel plots our estimates of the fixed effects (x-axis) against the values of origin and destination market access recovered from our estimates (y-axis). Overall, the two alternative ways of estimating origin/destination market access measures seem to deliver similar results.

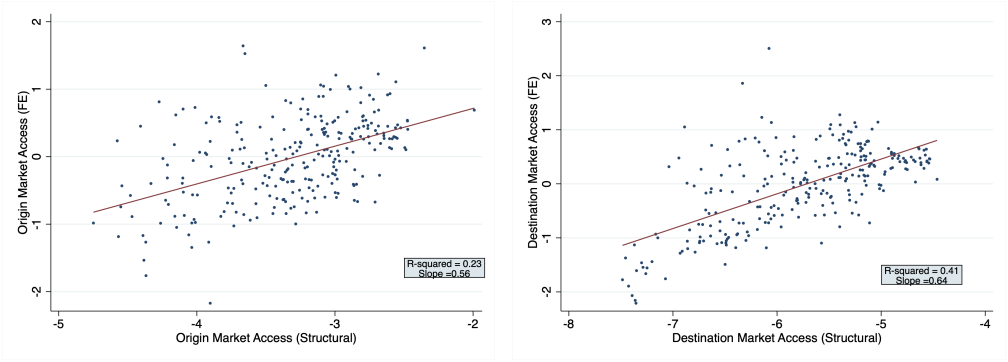


Figure 10: Market Access: Structural Estimates vs Fixed Effects (Simple Model)

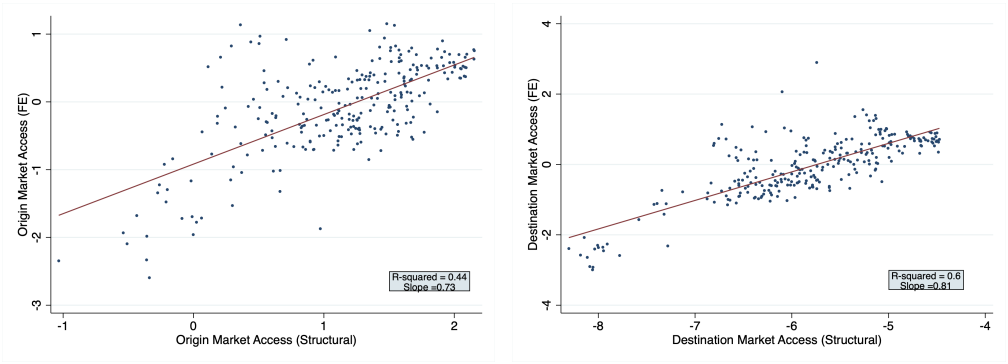


Figure 11: Market Access: Structural Estimates vs Fixed Effects (Full Model)

3.3 A calibration exercise

We often use simple, parsimonious models, to perform counterfactuals. Let us assume that distance and borders are not going to change much over the time horizon for which the counterfactual is going to be performed. Assume also that we want to use a gravity model that contains these two elements that delivers the following bilateral market access function:

$$M_{nm} = \kappa \exp\left\{\sigma \ln T_{nm} + \beta B_{nm}\right\} \tag{12}$$

where $\kappa > 0$ and $\sigma, \beta \leq 0$. The variable $T_{nm} \geq 0$ is the 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 country borders on bilateral market access, respectively. The parameter κ is chosen to ensure that our measures of bilateral market access add to one:

$$\kappa^{-1} =_n \, _m \exp\{\sigma \ln T_{nm} + \beta B_{nm}\} \tag{13}$$

This normalization does not affect any of the results but it simplifies some of the expressions that follow.

The question we ask here is: how much does such a parsimonious gravity model explain of European regional trade? To achieve this, which are the best values for σ and β? We refer to these coefficients as the distance and border effects.

We start by showing the effects of this parameters in a simple simulation. Figure 12 shows the effects of distance on normalized market shares.

Figure 12: Structural Gravity: Borders and Trade

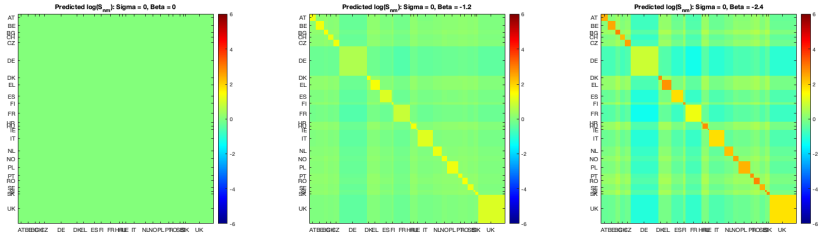


Figure 12 shows the effects of borders on predicted trade. From the left to the right, we see how the matrix of bilateral normalized market shares changes as the border effect grows. In all the matrices, we assume that there is no cost of distance, i.e. $\sigma = 0$. Without a border effect, the matrix

contains only ones and therefore the coloring is homogeneous. As we increase the border effect to -1.2, which is half of the value obtained in the previous section, we see already the country boxes appear. These boxes become quite clear as we raise the border effect to -2.4, which is the value we estimated in the previous section.

Figure 13: Structural Gravity: Distance and Trade

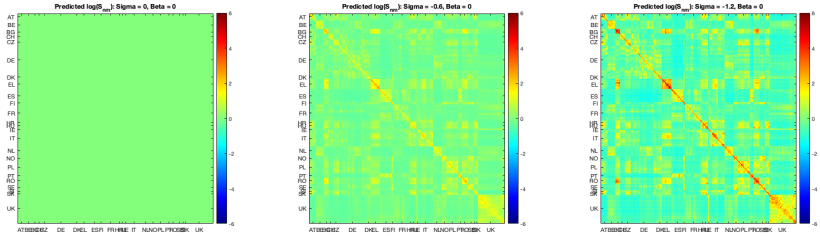


Figure 13 shows the effects of distance on predicted trade. Again, we see from left to right how the matrix of bilateral normalized market shares changes as the cost of distance grows. In all the matrices, we assume that there is no border effect, i.e., $\beta = 0$. Without a cost of distance, the matrix contains only ones and therefore the coloring is homogenous. As we increase the cost of distance to -0.6, which is half of the value obtained in the previous section, we see already some patterns in the matrix. The first one is that country boxes in the diagonal appear, since country trade observations are closer than foreign trade observations. In addition, we observe other boxes appear far away from the diagonal. These boxes contain foreign trade observations of countries that are nearby. As we further increase the cost of distance to -1.2, which is what we estimated in the previous section, these patterns become sharper.

Armed with these intuitions, we are ready now to calibrate this model. We find the best combination of σ and β to fit European regional trade with the model following these steps:

1. We define a two-dimensional grid over σ and β.

2. For each point of the grid, we compute:

   (a) a complete set of bilateral market access measures $\{M_{nm}\}$,

   (b) a complete set of origin and destination market access measures $\left\{M_n^O\right\}$ and $\left\{M_m^D\right\}$,

   (c) the matrix of predicted normalized market shares.

3. We choose the values of σ and β that minimize the distance between the matrices of actual and predicted market shares.

The result of this exercise leads us to choose $\sigma = -1.3$ and $\beta = -2.4$. These calibrated values are very close to those we obtain in column (1) of Table 1. Using these values, we construct the matrix of (log) normalized market shares. 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 the right panel. Figure 15 plots the elements of these matrices against each other. The two matrices are close and exhibit similar patterns.

A simple analysis of these matrices shows that the calibrated model explains about 48 percent of the total variation in normalized market shares. If we examine each of the types of trade separately, we observe that the calibrated 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.

Finally, we analyze how sensitive is the fit of the model to changes in parameter values. Figure 16 shows how the fit of the model declines as the values for σ and β depart from the chosen ones.

Figure 14: Normalized Market Shares: Data vs Predicted

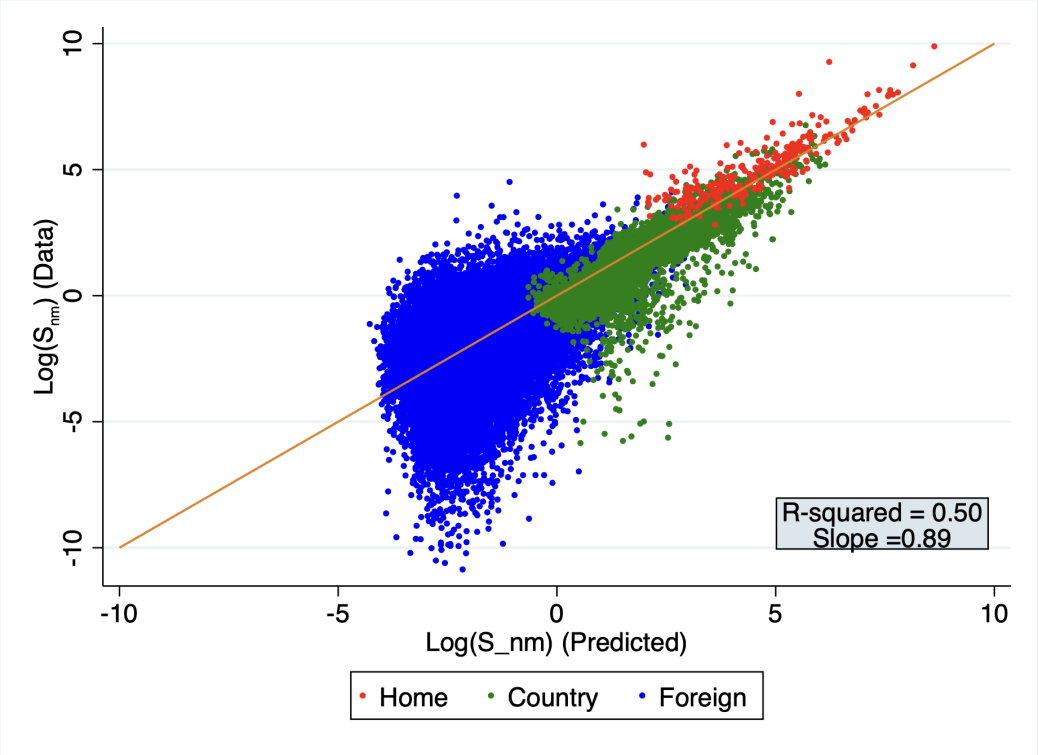


Figure 15: Matrix of Normalized Market Shares: Data vs Predicted

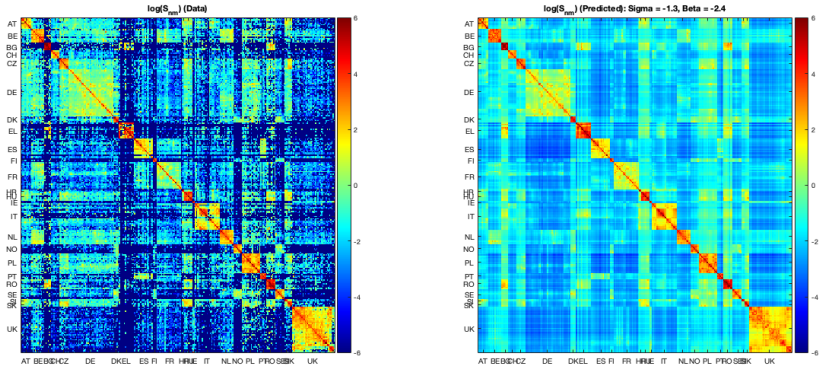
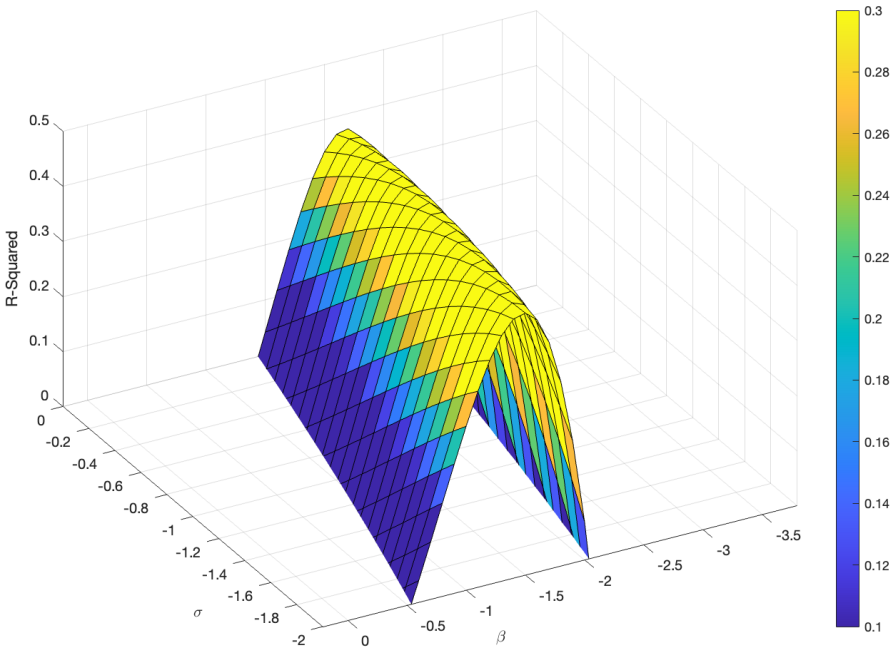


Figure 16: Structural Gravity: Explanatory Power



4 Are all borders the same?

We have seen that borders matter. We now ask ourselves whether trade flows that happen across borders are reduced in a comparable way across the European geography. Our goal in this section is to provide a comprehensive picture of the (endogenous and exogenous) barriers that regions in a given country face, when trading with foreign partners. We will capture these frictions with a border dummy, that we will allow to vary by country or country-pair.

As an example, consider the region of Catalonia in Spain and our size-corrected measure of trade, the normalised market share, S_{nm}. If we com-

puted the average difference between the average country market share and the average foreign market share for Catalonia, we would have a summary of the average barriers to international trade in Catalonia vis-a-vis Europe. We could make the same calculation with the northern neighbour region of Catalonia, the French region of Langued'oc-Rousillon. Would the difference in average country and foreign market shares be the same for both regions? How does this difference vary across European countries? We can also explore whether the factors that change at the border between two countries are different across country-pairs. For instance, take the region of Catalonia in Spain. Is the reduction in normalised market shares that we see when Catalonia trades with a French region (relative to a domestic one) the same reduction that we see when Catalonia trades with a Portuguese or Italian region? Even controling for size and distance, there may be factors that we omit from the gravity model and that vary the degree by which trade flows that cross an international border are different from within country trade flows.

To investigate this question, we run two additional sets of gravity regressions where we explore the heterogeneity of the border dummy. The first regression follows our main specification but expands the regression by adding two sets of border dummies. The first set is a summation across all origins of the border dummy interacted with an indicator that takes value 1 to capture the country in which the origin region is located. The second set of dummies would be a similar sum but for destination countries. Denote the countries in our sample with the index ν for origin and μ for destinations. There are 24 countries in our data, such that ν=1,...24. And μ=1,...,24. In

particular, we run the following regression:

$$lnS_{nm} = \sum_{\nu} \beta_{\nu} Border_{nm} \times I(c_n = \nu) + \sum_{\mu} \beta_{\mu} Border_{nm} \times I(c_m = \mu) + ln(Dist_{nm}) + \phi_n^O + \phi_m^D + u_{nm}, \tag{14}$$

where c_n and c_m are the origin and destination countries. For instance, for trade between Langued'oc-Rousillon and Catalonia, we would have a gravity regression with two border dummies, one interacted with a dummy for Spain and another interacted with a dummy for France. This interaction of dummies for the origin would be "activated" when the S_{nm} contains two regions in different countries and one of them is in Catalonia. Its coefficient β_{ES} would capture the average reduction that Catalonia (or any other region in Spain) experiences when trading with other foreign countries relative to trading with itself. We then take the average over the origin and destination components to get the average barriers to foreign trade that regions face in each country. Figure 17 presents the results for our 24 countries.

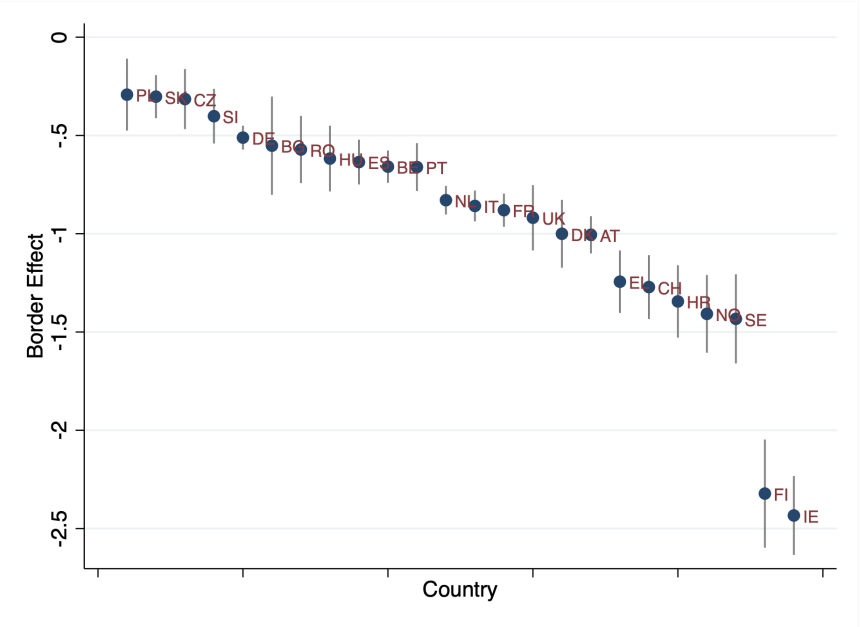


Figure 17: Borders dummy coefficient by country

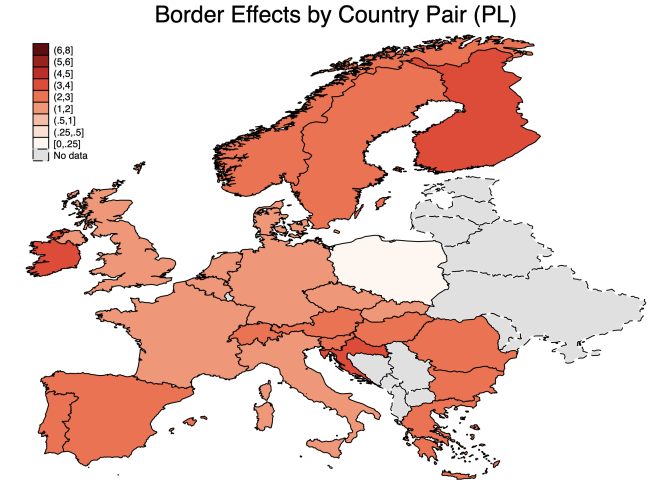
We find that borders matter everywhere. All border coefficients are negative and significant. As we were anticipating, the difference in normalised market shares between foreign and domestic (country and home) pairs is heterogeneous across countries. The country with the largest difference in trade between foreign and domestic trade is Ireland, with a border coefficient of almost -2.5. In Ireland, foreign market shares are 8% of the domestic market shares. The country with the smallest border coefficient is Poland, with a value of -0.35. In Poland, foreign normalised market shares are 70% of the value of domestic market shares. Most of the differences between country and foreign market shares are between -0.35 and -1.5, except for Finland and Ireland. These countries have some common characteristics, they are very peripheral in Europe. Apart from them, the rest of countries align in an intermediate range and there is no clear correlation with size (demographic, economic or geographical) and with other characteristics.

We now explore whether country-pairs also face different implied barriers to trade, by estimating the border coefficient interacted with an origin country-destination country dummy. We estimate the following regression:

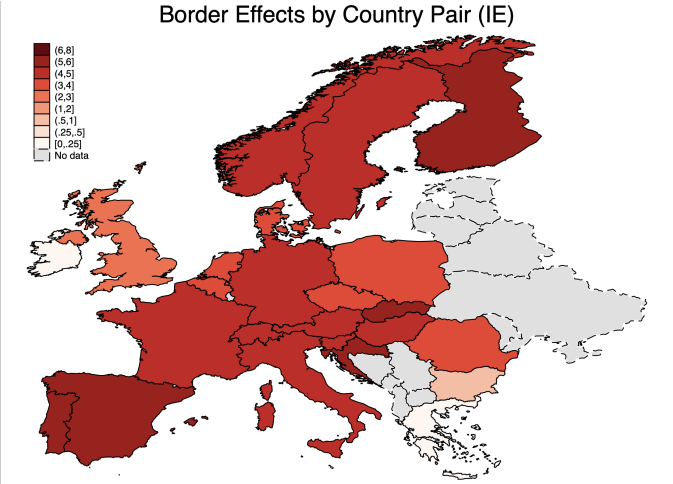
$$\begin{aligned} lnS_{nm} = \textstyle\sum_{\mu}\sum_{\nu}\beta_{\nu\mu}Border_{nm} \times I(c_n = \nu \& c_m = \mu) + \\ ln(Dist_{nm}) + \lambda' X_{nm} + \phi_n^O + \phi_m^D + e_{nm}, \end{aligned} \tag{15}$$

where X_{nm} is a vector of controls that includes the Own region dummy, and the political variables (common language and currency and membership of EU and Schengen). This regression provides us with one border coefficient for each country in Europe (a total of 24*23=552). In order to visualise part of the richness of the data, we can plot the border coefficients in a map. Figure 18 shows the border of Belgium and Ireland, the two countries at the extremes of the distribution, with the rest of countries in Europe.

As we already saw in the average effects, Poland has low borders with everyone and Ireland has wide borders. Distance seems to matter, since border coefficients are darker with further away region, showing how the identification of the distance and border effects separate is challenging (as we showed in previous sections). But there is substantial heterogeneity across trade partners. For Poland, trade flows less easily into Switzerland and Austria than into other neighbours. For Ireland, the main trade partner is the UK but trade flows are also relatively more intense with Belgium, the Netherlands, Poland and the Czech Republic. The patterns are as rich as the data. We report the figure for each country in our dataset in the Appendix.



A) Poland



B) Ireland

Figure 18: Heterogeneity by Country-pairs

Notes: The figure shows the estimated border coefficient of Poland (Panel A) and Ireland (Panel B) with its international partners. The color shading represents the value of the coefficient, with darker shades representing stronger borders.

5 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 are 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 labeled "Own region". The Own region dummy variable takes value one if the observation is home trade, and value zero otherwise.

Table 2 shows the same fixed-effects regressions that we saw in Table 1, including this additional variable. There are two key takeaways from this table. First, the home bias variable is large and significative. Focusing on our extended model in columns 3 and 4, the average normalised market share of a region with itself is between 3.79 (exp(1.333)) and 10 (exp(2.310)) times larger than the average normalised market share between two different regions in the same country, controlling for distance and other political variables. Notice that this magnitude is at least as large as the border coefficient. This indicates that the gravity models consider in the previous section grossly underestimate home trade. Second, the coefficients of other variables and the R-squared of the regressions do 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 matrix. Failing in 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 what characteristics make regions to have a large home bias, and which characteristics have the opposite effect. Second, we try to explore whether the comparatively higher trade flows that we observe within regions are a genuine Home bias in trade or an artefact of the

Table 2: Gravity: Including the Ownregion Dummy

	(1) Log(S_nm)	(2) Log(S_nm)	(3) Log(S_nm)	(4) Log(S_nm)
Border Effect	-2.380***	-2.321***	-1.393***	-1.360***
	(0.261)	(0.241)	(0.149)	(0.146)
Log(Dist)	-1.150***		-0.994***	
	(0.0689)		(0.0569)	
Ownregion	1.013***	2.079***	1.333***	2.310***
	(0.259)	(0.409)	(0.188)	(0.323)
Common Language			0.794***	0.780***
			(0.0872)	(0.0907)
Both EU			1.747***	1.749***
			(0.385)	(0.383)
Common Currency			0.190	0.178
			(0.125)	(0.124)
Both Schengen			1.244***	1.272***
			(0.249)	(0.251)
Origin FE	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes
Distance Bins	No	Yes	No	Yes
Observations	46505	46505	46505	46505
R^2	0.611	0.613	0.632	0.634

Standard errors in parentheses

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

data aggregation process.

5.1 Determinants of the home trade

Figure 5 showed the spatial distribution of home normalised market shares in Europe. The region with smallest home shares had a market share of 40 (40 times larger that the predicted share in the independence framework), while the regions that trade the most with themselves have market shares of up to 20,000 (for instance remote regions in Greece). Geography seems to play a role. For instance, regions at the geographical boundaries of Europe tend to have higher home shares. Regions that are geographically small, also display higher shares. Economic variables seem relevant too. Notice that regions in large economies like Germany, France or England have lower home shares.

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

Column 1 reports the results using the following geographical variables: Internal distance in the region (based on geographic area), 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. More remote regions, island regions and mountain regions have higher home shares. These simple geographical variables explain 41.3% of the variation in home market shares.

Column 2 adds economic variables: presence of ports, motorway density, population, share of employment in manufacturing and services, the share of secondary education and the share of foreign born population. The introduction of economic variables reduces the magnitude of the geographic variables' coefficients. The economic variables are important too. Motorway density and population reduce the home market share. Both high manufacturing regions and regions with high government spending have high market

shares. More educated regions have higher home market shares and regions with more migrants have lower home market shares. The introduction of these economic determinants increases the R-squared by 38.2%.

Finally, column 3 adds country fixed effects. Comparing column 2 and 3 allows us to see whether differences in home market shares come from within country differences or across country differences. The within-country specification (column 3) has a higher R-squared of almost 90%, showing that country differences are important. 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. And larger regions in terms of population or with a larger share of manufacturing have lower home market shares. These results 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).

5.2 Government structure and home trade

In this section we exploit our rich dataset to contribute to the debate on the existence of an "own region" or "home bias" effect in trade. Previous studies have suggested that part of the "own region" 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 avail-

Table 3: Home Bias: Determinants

	(1) Home	(2) Home	(3) Home
Log(Regional internal distance)	-0.131	-0.0316	-0.0912
	(0.120)	(0.0991)	(0.138)
Log(European Remoteness)	2.459***	1.355***	1.603***
	(0.288)	(0.201)	(0.454)
Island Region	1.800***	0.892**	0.965***
	(0.509)	(0.374)	(0.333)
Mountain Region	0.336***	0.172**	0.198**
	(0.119)	(0.0735)	(0.0831)
Major Port Region		-0.143	-0.131
		(0.129)	(0.105)
Motorway Density		-7.546***	-7.242***
		(1.356)	(1.823)
Log(Population)		-0.786***	-0.735***
		(0.0554)	(0.0732)
Share of Emp. (Manuf.)		-9.917***	-9.822***
		(1.165)	(1.867)
Share of Emp. (Public)		-15.80***	-0.107
		(1.658)	(3.945)
Sh. Secondary or tertiary educ		1.375***	-1.479
		(0.386)	(0.941)
Share Migrant Pop.		-2.608***	-0.588
		(0.581)	(0.789)
Country FE	No	No	Yes
Observations	269	265	265
R^2	0.413	0.795	0.890

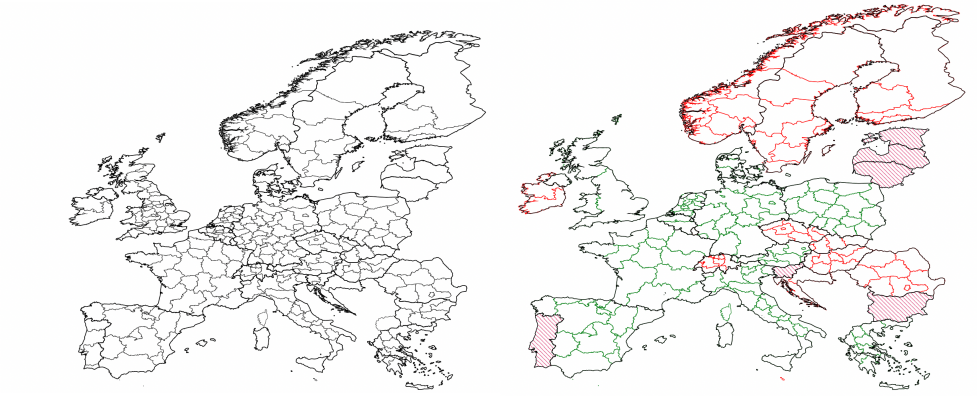
Standard errors in parentheses

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

able. This 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 CFS at zip code level and find that the state border effect does not survive once you control for actual distance of the shipment using ziplevel data. Second, the data is aggregated after collection, the levle of aggregation of spatial units can create a"spatial attenuation effect" (Coughlin and Novy, 2021). Due to the existence of spatial frictions, the more micro regions we combine, the larger the inside-region frictions will be, and the weaker the border effect we estimate. This would make us estimate weaker border effects for large countries and stronger border effect for small countries.

We seek to explore this question by using a specific feature 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 own region 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 highest 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 own region effect. Due to government structure differences, 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



Statistical (NUTS2) regions **Political Regions**

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 region at a lower level of desaggregation. 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 own region effect emerges in all countries, regardless of the level of the actual political borders, or whether we only estimate the own region 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 di-

vided in 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
Romania	–	8
Slovakia	–	4
Switzerland	–	7
Norway	–	7
Hungary	–	8
Croatia	–	2
Czech Republic	–	8

   Here we have 9 countries: Finland (5 NUTS2 units), Romania (8), Slovakia (4), Sweden (8), Switzerland (7), Norway (7), Hungary (8), Croatia (2), Czech Republic (8). 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 Laenders 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 own region 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:

$$\ln(\mathrm{S}_{nm}) = \beta_1 OwnRegion(NUTS1)_{nm} + \beta_1 OwnRegion(NUTS2)_{nm} + \gamma log(dist_{nm}) + \delta_n + \eta_m + e_{nm}, \quad (16)$$

where the outcome variable is the normalised trade market share. The variable $OwnRegion(NUTS2)_{nm}$ indicates a low-level regional border and takes value 1 when n and m are in the same NUTS2 region. The variable $OwnRegion(NUTS1)_{nm}$ 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 sim-

ilarly to shipments across NUTS2 regions. By contrasts, 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 Own region 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 an own region 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 1-4 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 (Novy et al, Hummels and Hllberry).

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

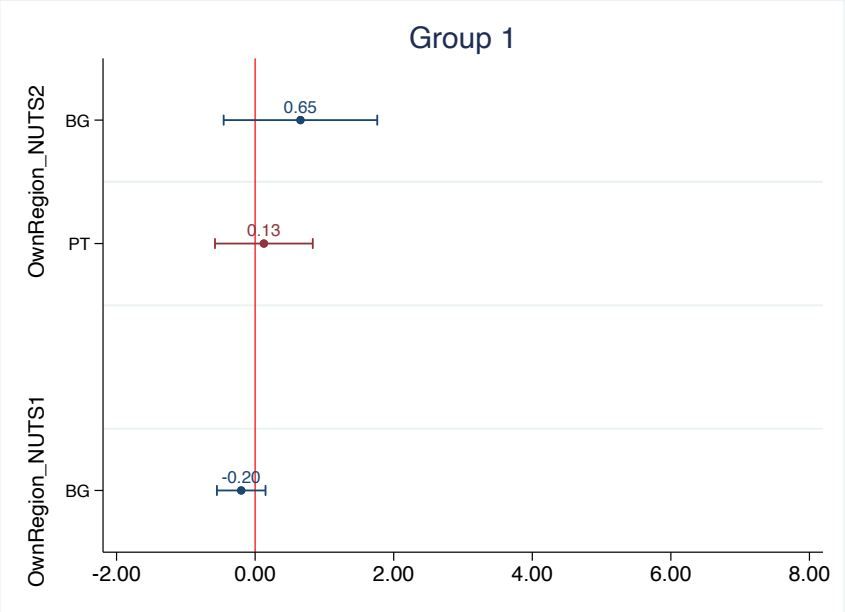


Figure 20: Own region: Group 1

Countries with regional divisions at a lower level of disaggregation This is one of the case 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 own region 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 own region patterns. We find, as expected, that there is no own region

effect at the level of NUTS1 (some of these countries only have one NUTS1 region, so coefficient cannot be estimated).

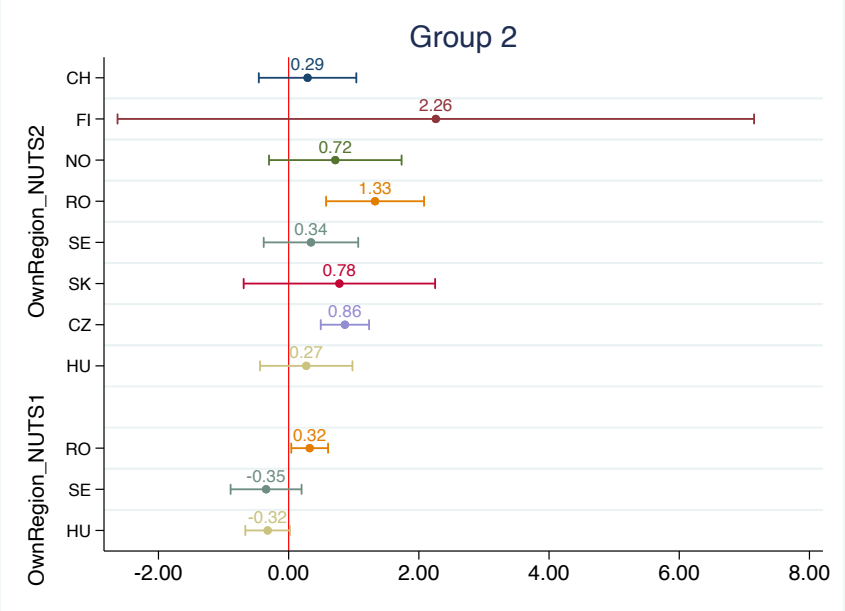


Figure 21: Own region: Group 2

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 own region (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 own region estimated is the same, showing that this strategy is helping us overcome the spatial aggregation problems mentioned above.

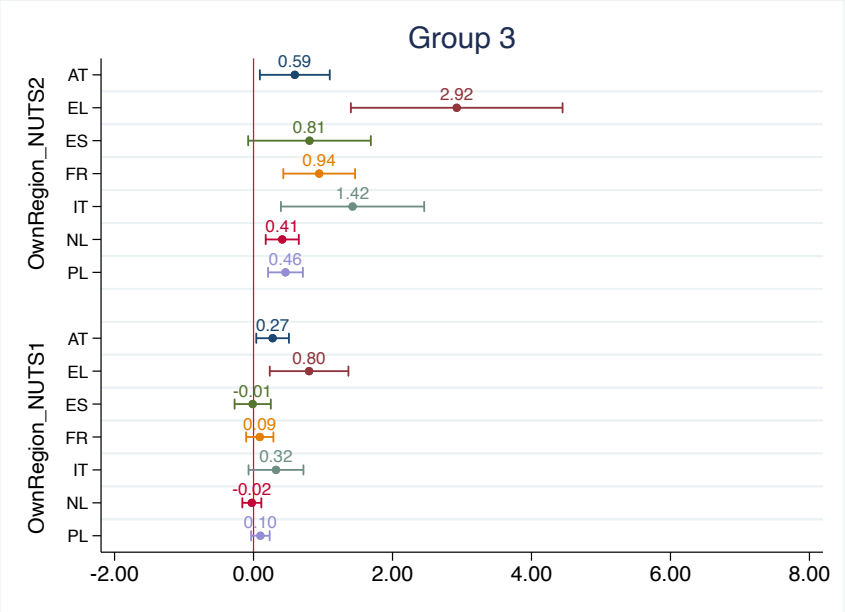


Figure 22: Own region: Group 3

Countries with borders at a level higher than the statistical regions For these countries, the predictions are confirmed. Germany and Belgium have own region 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 own region effect at the statistical division level (NUTS2) but a own region effect at the nation level (NUTS1).

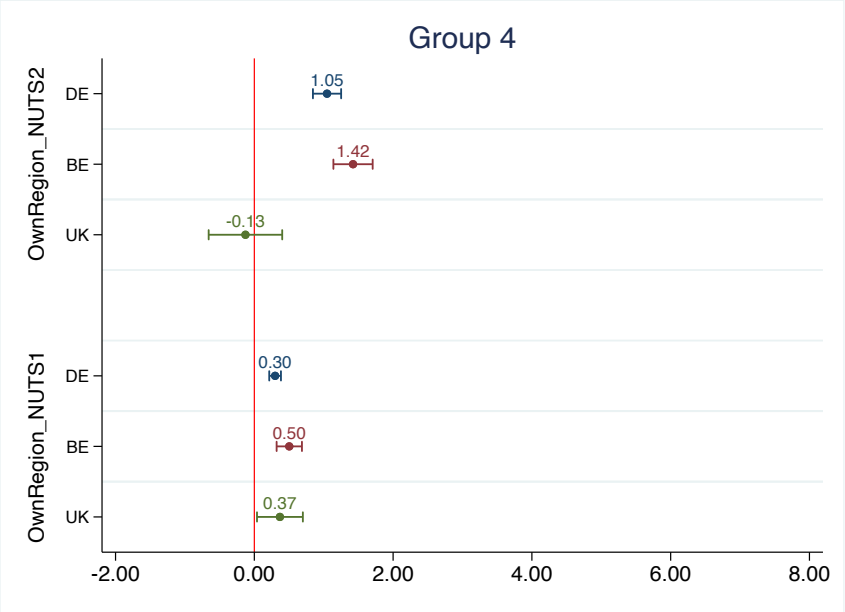


Figure 23: Own region: Group 4

These results partially confirm that aggregating political regions could alter the estimated own region 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 own region 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

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. 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.[16]

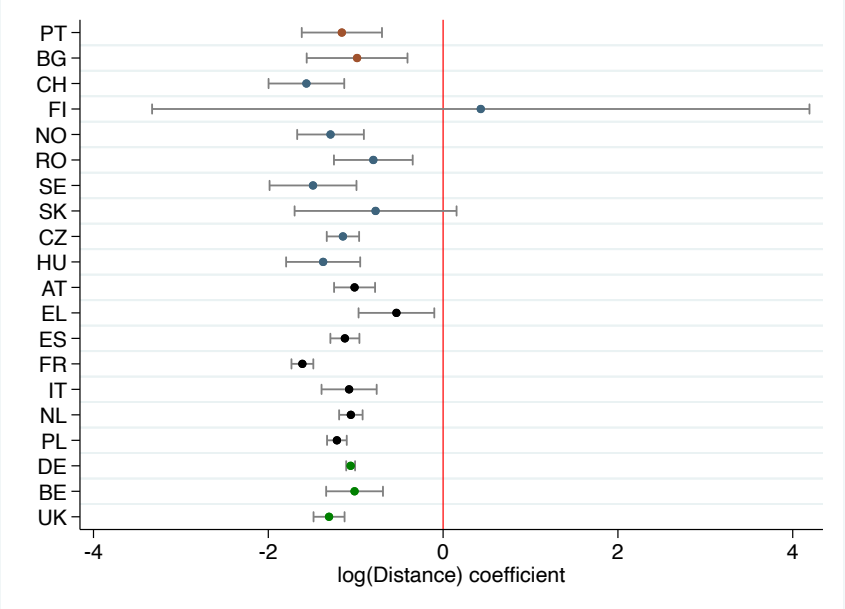


Figure 24: Distance-elasticity of domestic trade

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

6 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 Hummels and Hillberry (2008) that 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 additional information about shipments to understand the margins by which home, country and foreign trade are different. In other words, is home trade large because shipments happen in all industries or because more shipments are send in each industry, or more kilos are shipped by shipment?

Table 4: Summary statistics: Shipment level

	Home	Country	Foreign
	Mean	Mean	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 divide shipments by the border that they cross and report some summary statistics of our data by home (within region), country (within country

but across regional borders) and foreign (across national borders) type. Table 4 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.

One limitation of the microdata is that we do not observe values. To prevent any price imputation to affect 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:

$$Q_{n,m} = N_{nm}\overline{Q}, \tag{17}$$

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:

$$Q_{nm} = N^I_{n,m} \bar{S}_{n,m} \overline{Q} \tag{18}$$

Taking logs we get:

$$lnQ_{nm} = lnN^I_{n,m} + ln\bar{S}_{n,m} + ln(\overline{Q}) \tag{19}$$

Notice that equation 19 is an identity that holds for every region pair in every industry. It indicates that bilateral shipments can be decomposed into two extensive margin components, the number of industries and the number of shipments per industries, and one intensive margin component, the quantity shipped per shipment. Since this is an exact decomposition, we can use 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 13.9% and 31.7% 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 $\ln N^I_{n,m}$	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

6.1 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 Own region 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 19. In particular we estimate the following regression:

$$ln(Y_{nm}) = \beta Border_{nm} + ln(Dist_{nm}) + \phi_n^O + \phi_m^D + u_{nm}, \tag{20}$$

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)
Own Region	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 Own region effect. The next two columns, (2) and (3), report the results from using as outcome variables the two extensive margins: total number of industries shipped between n and m, and 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 "Own region" effect is negative in the number of industries and positive in the case of average shipment per industries. 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 intensive margin: 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. 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 commodities 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. Both the 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 comparable. Finally, since they only see domestic shipments, we will focus on the difference between our estimate

of the effect of "Own region" and their estimate of the effect of "Own state". The "Own region" 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 that the mechanisms that drive the internal border effects may differ in the two settings. A more detailed comparison country by country could tell us more about the differences and similarities between US and Europe.

7 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:

$$SocialConnectednessIndex_{nm} = \frac{FB_connections_{nm}}{FB_Users_n \times FB_Users_m} \tag{21}$$

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:

$$Social_{nm} = \frac{SCI_{nm}/(\sum_n SCI_{nm})}{\sum_m SCI_{nm}/(\sum_m \sum_n SCI_{nm})} \tag{22}$$

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 that have 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 re-

gressing the trade market share S_{nm} on the social market share $Social_{nm}$, for within region, within country and foreign 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 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. The third panel paints a different picture, social interactions are positively correlated with economic interaction for region pairs in the same country. However, they 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.

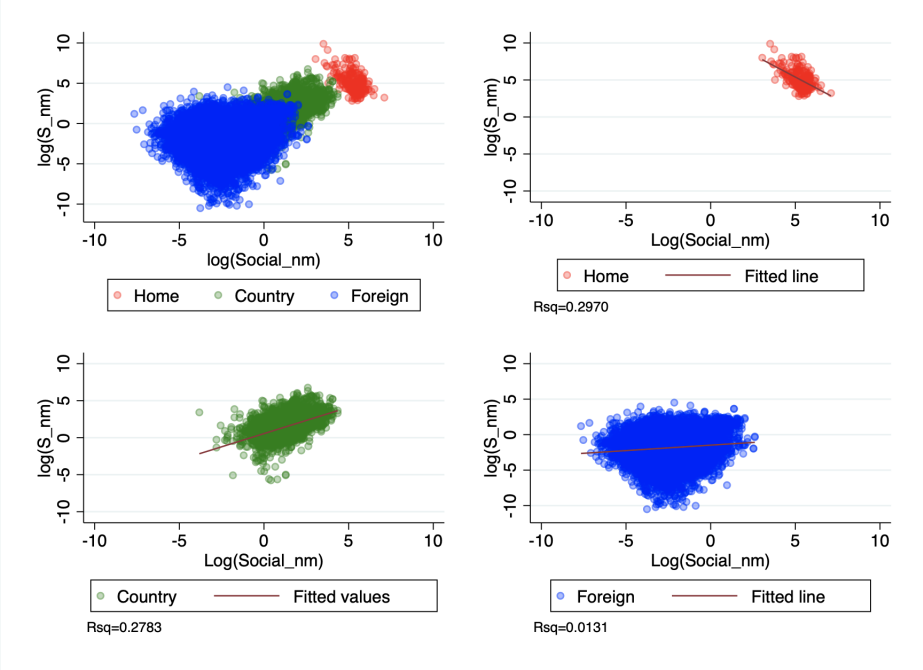


Figure 25: Economic interactions and trade interactions

7.1 Do social and economic interactions face similar frictions?

We have seen that social interactions are correlated with trade interactions in a pattern 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 gravity estimation equations from section 3.

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. Interestingly, gravity variables have a higher explanatory power for social interactions than for trade interactions between regions.

Table 7: Trade and Social interactions

	(1) Log(S_nm)	(2) Log(Social_nm)	(3) Log(S_nm)	(4) 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)
Own Region	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

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

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 Own region effect. On the other hand, distance is associated with higher reductions in trade interactions, $\log(S_{nm})$, than with 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.

8 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.

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).

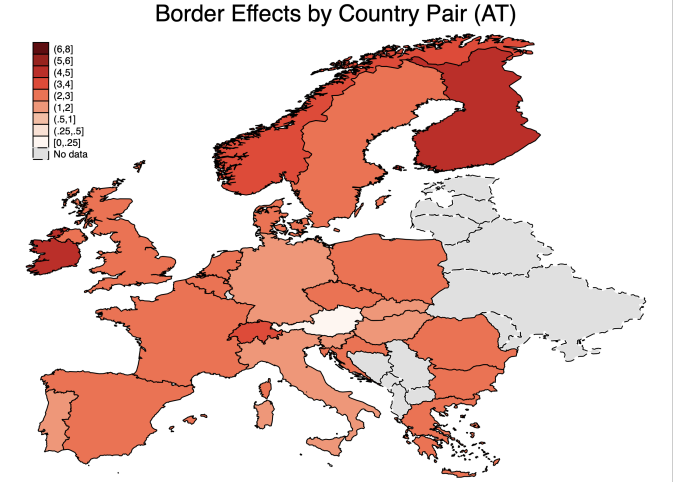
Wei, S.-J. (1996). Intra-national versus international trade: how stubborn are nations in global integration? Technical report, National Bureau of Economic Research.

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

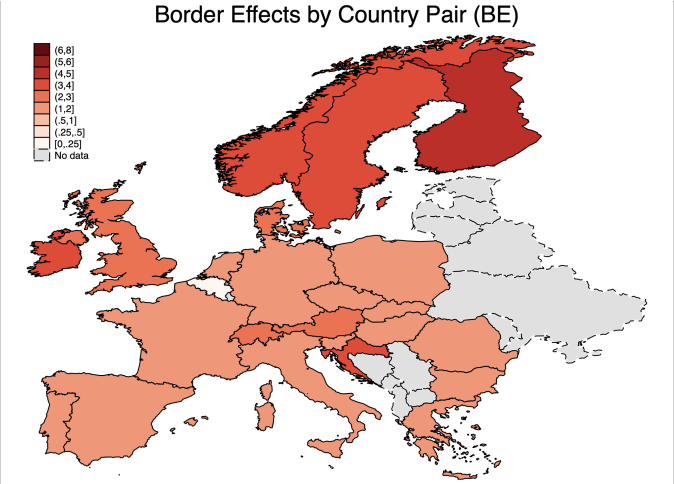
Part I

Appendix

A Additional Figures and Tables

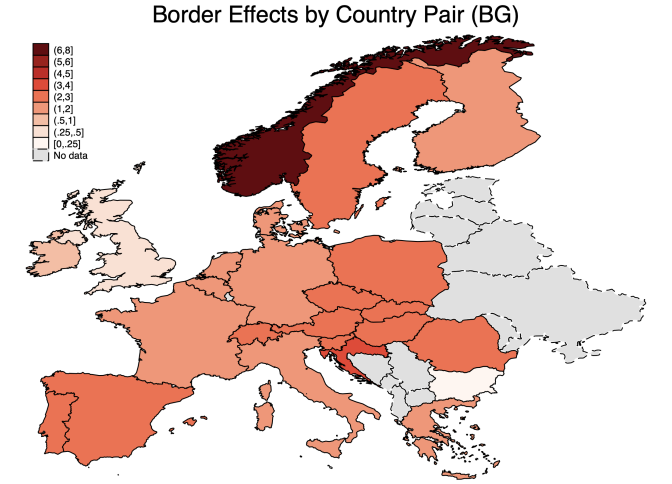


A) Austria

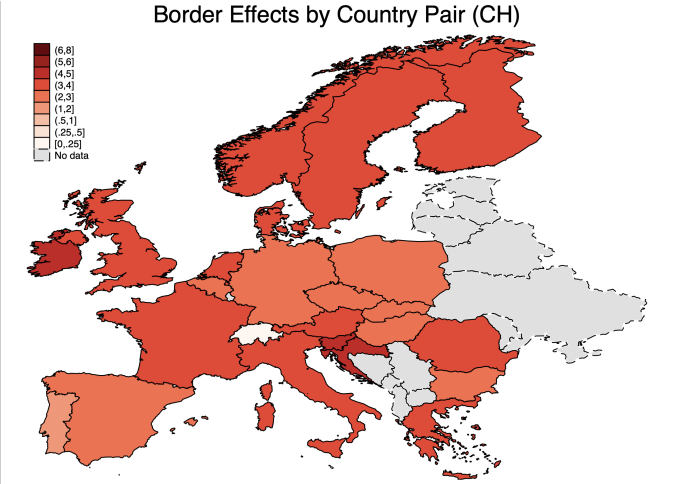


B) Belgium

Figure 26: Heterogeneity by Country-pairs

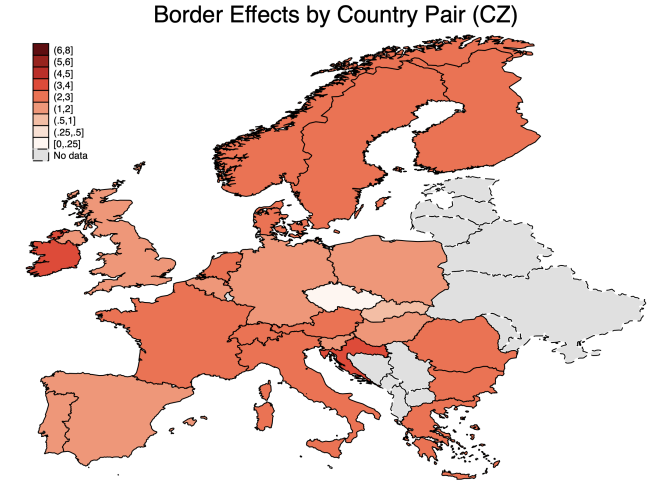


A) Bulgaria

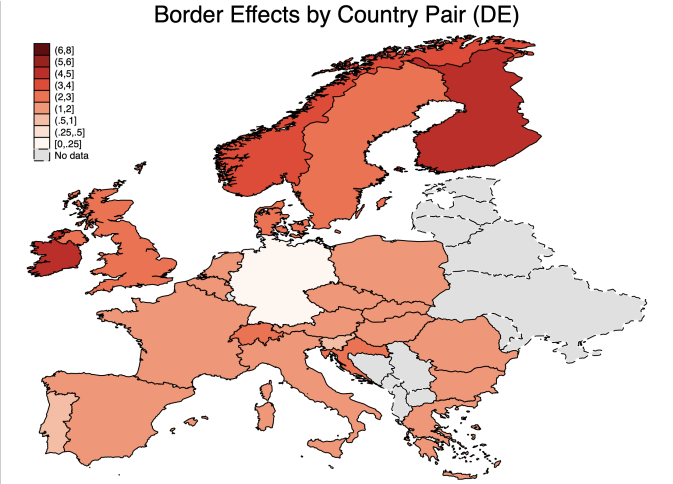


B) Switzerland

Figure 27: Heterogeneity by Country-pairs

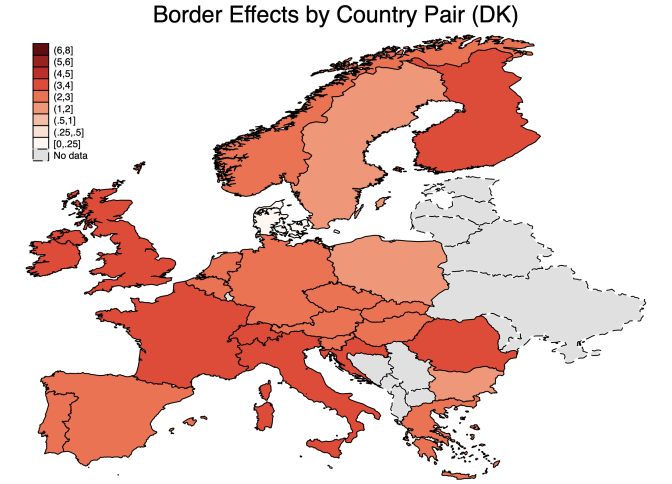


A) Czech Republic

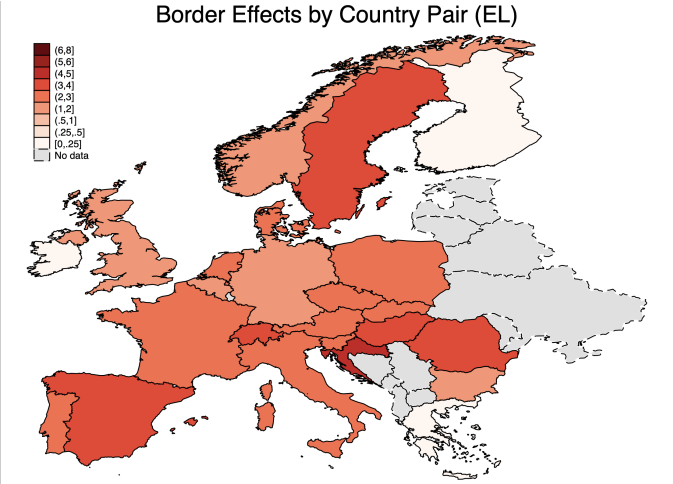


B) Germany

Figure 28: Heterogeneity by Country-pairs

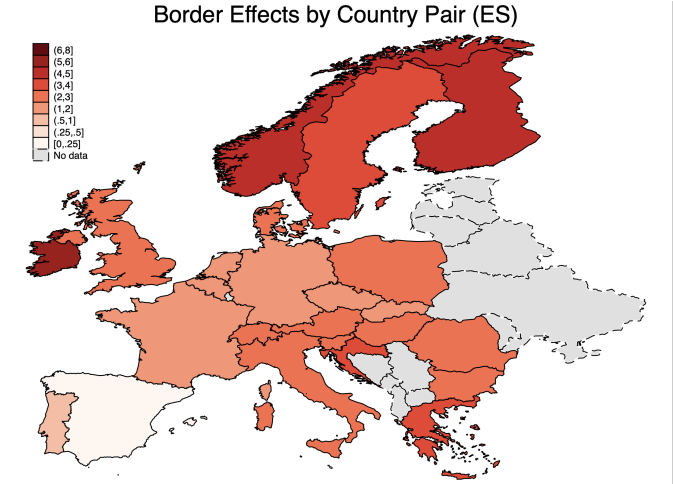


A) Denmark

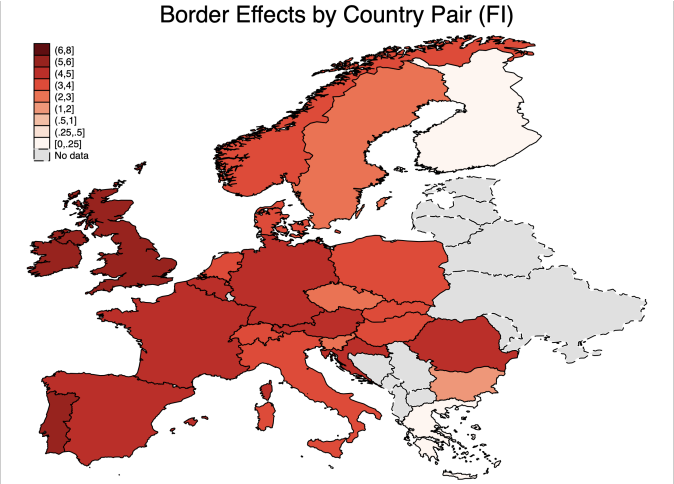


B) Greece

Figure 29: Heterogeneity by Country-pairs

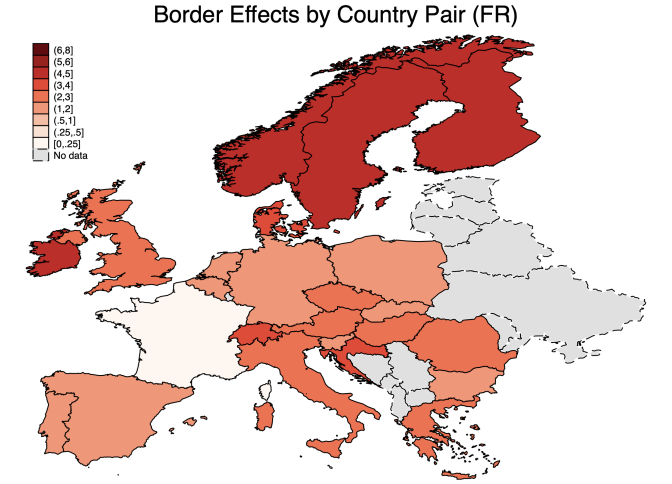


A) Spain

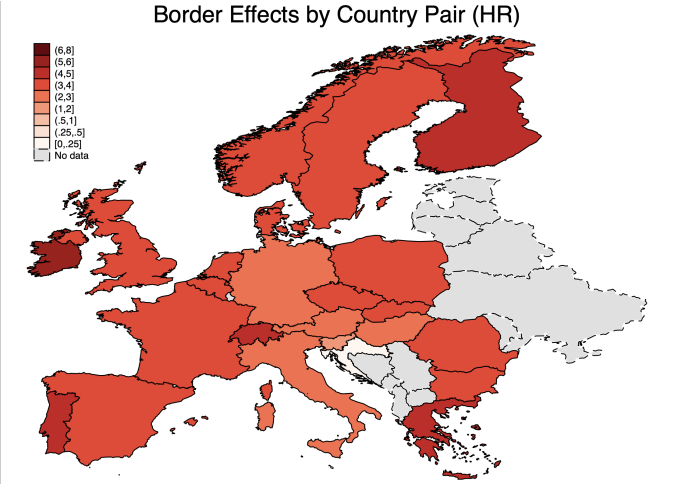


B) Finland

Figure 30: Heterogeneity by Country-pairs

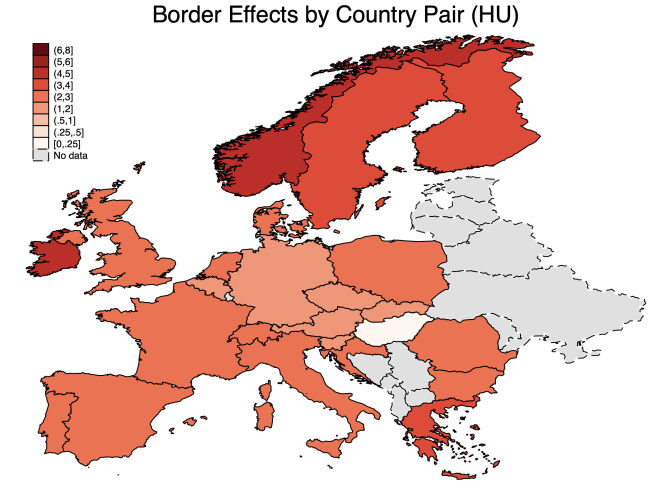


A) France

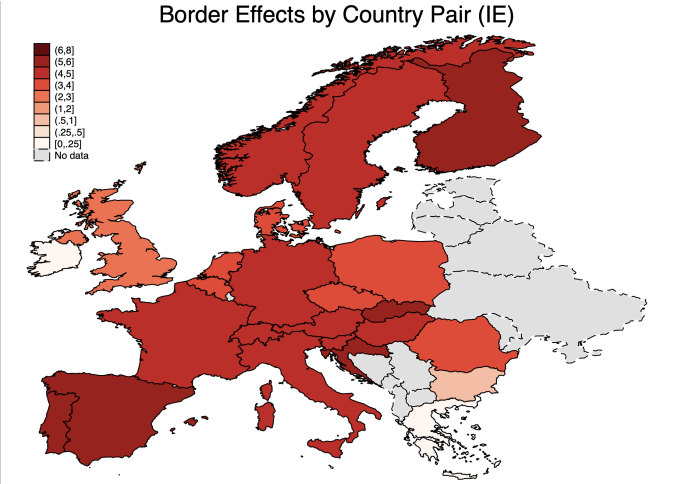


B) Croatia

Figure 31: Heterogeneity by Country-pairs

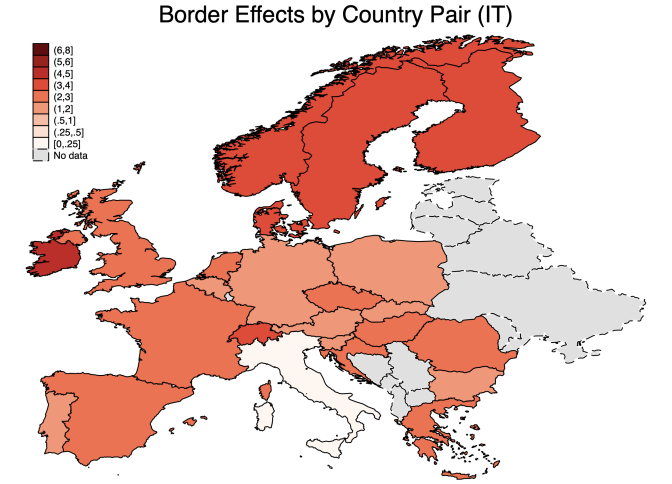


A) Hungary

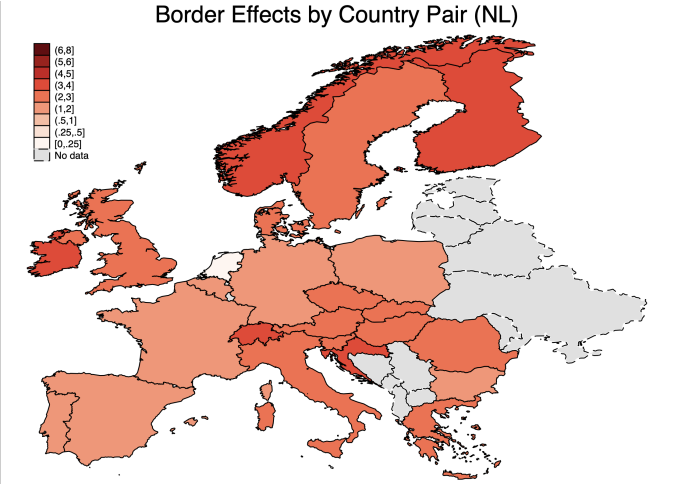


B) Ireland

Figure 32: Heterogeneity by Country-pairs

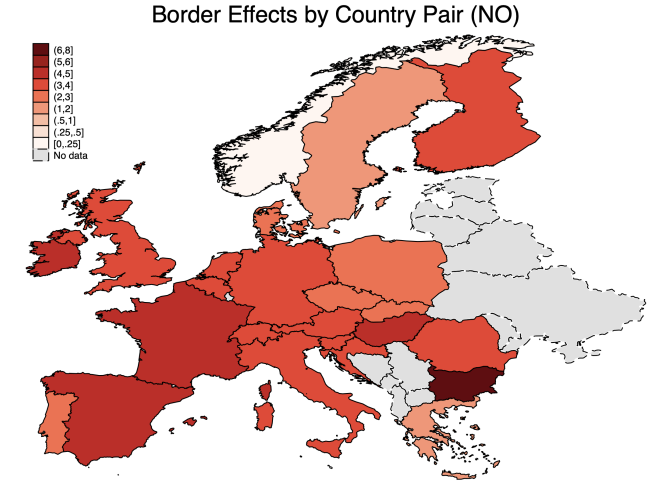


A) Italy

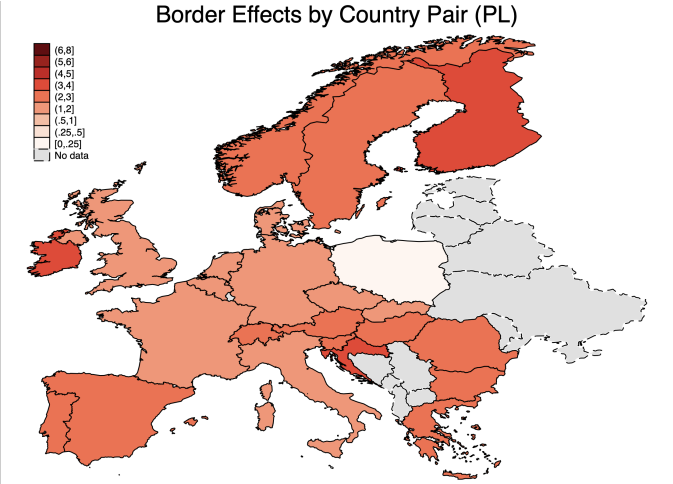


B) The Netherlands

Figure 33: Heterogeneity by Country-pairs

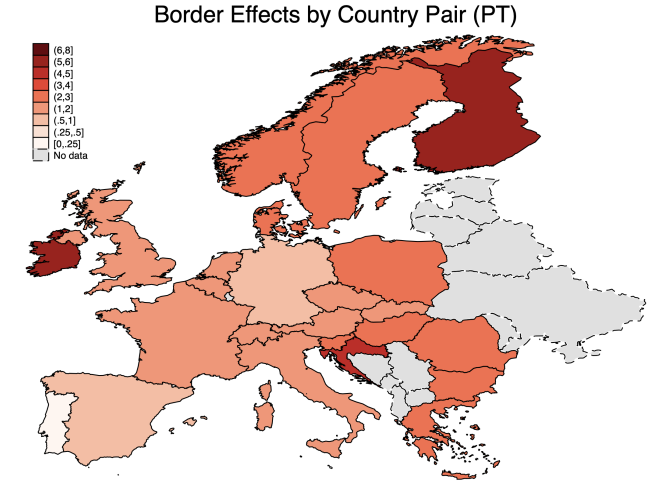


A) Norway

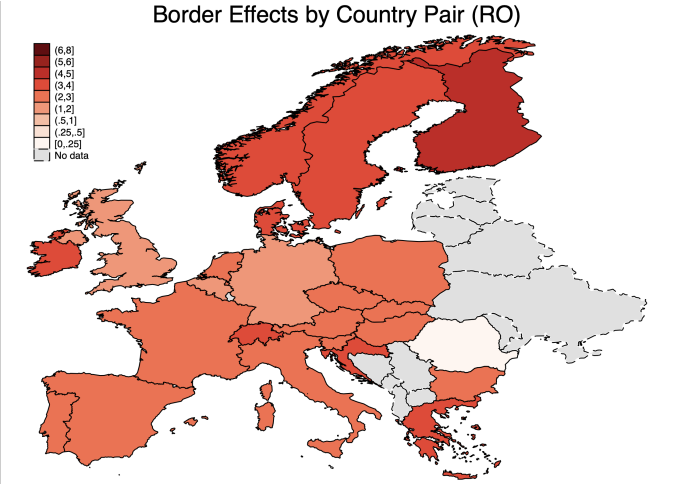


B) Poland

Figure 34: Heterogeneity by Country-pairs

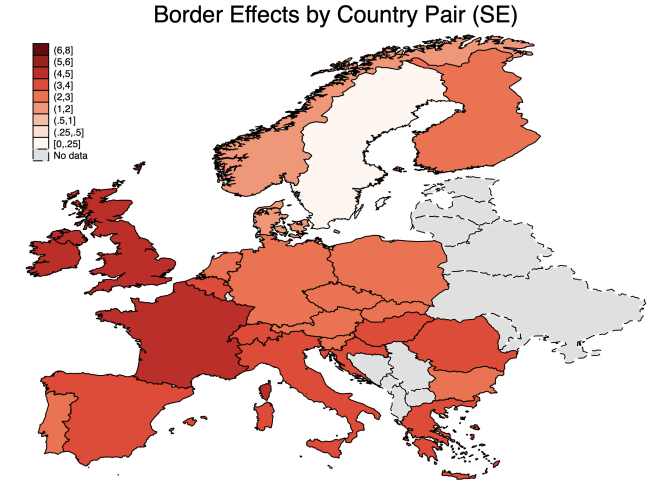


A) Portugal

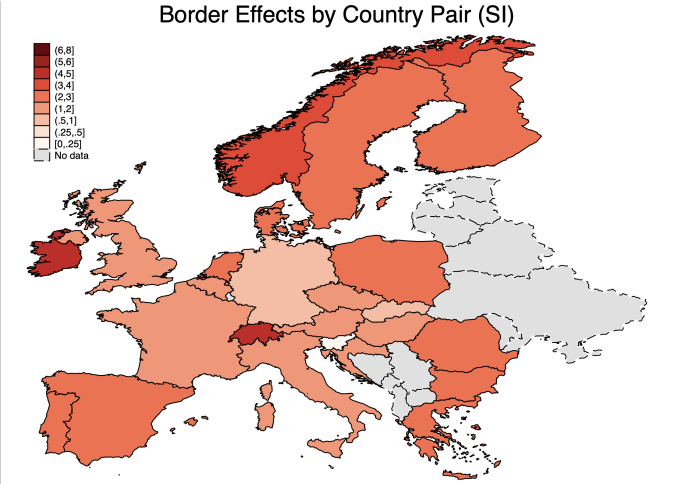


B) Romania

Figure 35: Heterogeneity by Country-pairs

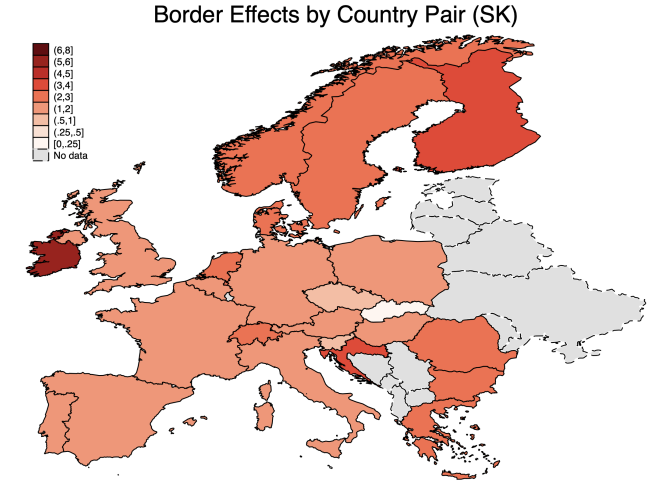


A) Sweden

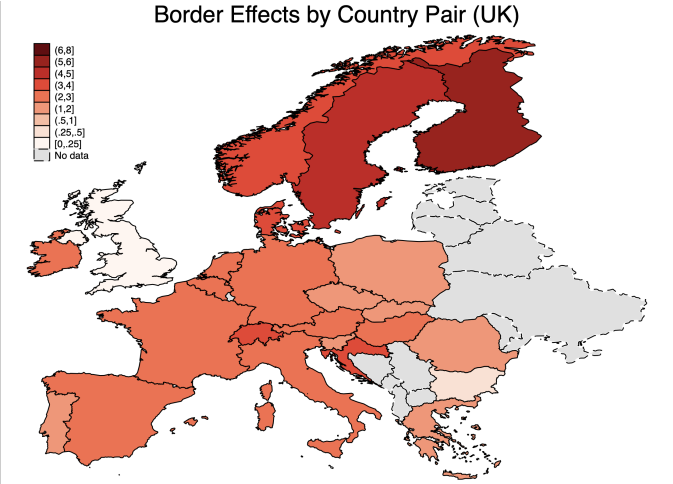


B) Slovenia

Figure 36: Heterogeneity by Country-pairs



A) Slovakia



B) United Kingdom

Figure 37: Heterogeneity by Country-pairs