Exploring European Regional Trade

BSE Working Paper 1384
December 2022 (Revised February 2023)

Marta Santamaría, Jaume Ventura, Uğur Yeşilbayraktar

Exploring European Regional Trade *

Marta Santamaría, Jaume Ventura and Uğur Yeşilbayraktar

February 6, 2023

Abstract

We use the new dataset of trade flows across 269 European regions in 24 countries constructed in Santamaría et al. (2020) to systematically explore for the first time trade patterns within and across country borders. We focus on the differences between home trade, country trade and foreign trade. We document the following facts: (i) European regional trade has a strong home and country bias, (ii) geographic distance and national borders are important determinants of regional trade, but cannot explain the strong regional home bias and (iii) the home bias is heterogeneous across regions and seems to be driven by political regional borders.

*Marta Santamaría: University of Warwick. Jaume Ventura: CREI, Universitat Pompeu Fabra and Barcelona School of Economics. Uğur Yeşilbayraktar: Universitat Pompeu Fabra and Barcelona School of Economics. E-mail: marta.santamaria@warwick.ac.uk, jventura@crei.cat, ugur.yesilbayraktar@upf.edu. We thank Jeffrey Frankel, Natalie Chen, Linda Tesar and other participants in the ISOM 2022 Conference for their valuable comments. Jaume Ventura acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program Advanced Grant (693512 - GEPPS), the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in RD (CEX2019- 000915-S) and from the Generalitat de Catalunya through the CERCA and SGR Programme (2017-SGR-1393).

1 Introduction

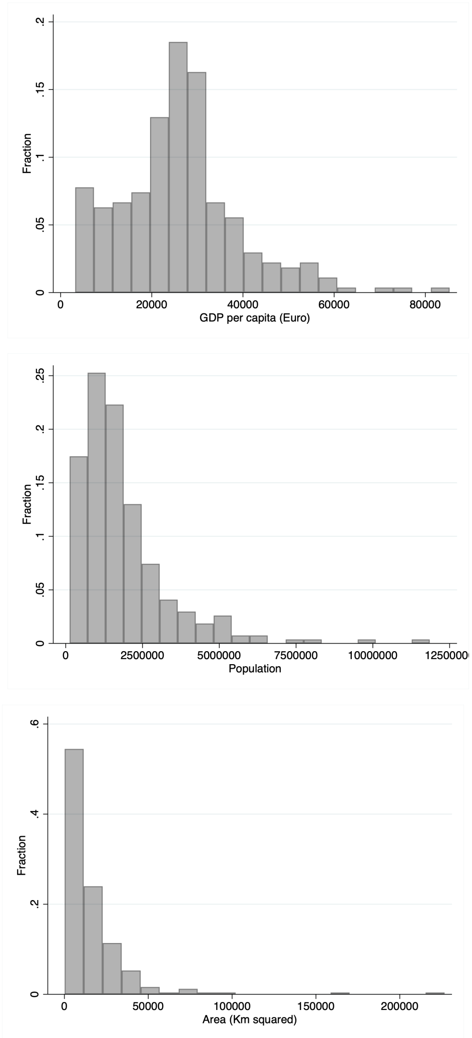
How do regions trade with each other? We know much about trade across countries thanks to the availability of detailed customs data. We know much less about trade within countries. In this paper, we use the dataset we constructed in Santamaría et al. (2020) to systematically explore for the first time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

Europe is a great laboratory to explore regional trade flows. One reason is that Europe is large, as it contains more than 500 million people and it produces about 20 percent of world GDP. Another reason is that European regions exhibit a lot of heterogeneity, as shown in Figure 1 using data from 2011 (the starting period of our dataset). The top panel shows the distribution of per capita GDP, which ranges from a low of 3,200 euros in Northwestern Bulgaria to a high of 85,330 euros in Central London. The middle panel shows the distribution of populations, which ranges from a low of 126,761 inhabitants in Valle d'Aosta to a high of 11,852,851 inhabitants in Île de France. The bottom panel shows the distribution of geographical areas, which range from a low of 160 Km^2 in Brussels to a high of 226,716 Km^2 North/East Finland.

The dataset constructed in Santamaría et al. (2020) is based on the European Road Freight Transport survey which collects data on truck shipments of goods in agriculture, manufacturing and mining. Thus, the dataset covers trade in goods by road, which according to Eurostat is about half of all European trade in goods. The dataset covers 269 regions from 24 European countries between 2011 and 2017 disaggregated into 12 different industries. An important aspect of this dataset is that it allows us to measure trade flows both across and within regions. Thus, for each year/industry, we have a complete matrix of bilateral trade including the diagonal entries.

The first and more salient aspect of European regional trade is that it has a strong home and country bias. Consider a shipment originating from a randomly selected European region. The probability that this shipment has a destination inside the origin region (i.e. home trade) is 40 percent. The probability that this shipment

Figure 1: Heterogeneity across European regions



Panel (a) plots the distribution of GDP per capita in the 269 European regions of our dataset, panel (b) plots the population in the regions in our dataset and panel (c) plots the distribution of areas in square kilometers of the 269 regions in our dataset.

has a destination outside the origin region but inside the country of the origin region (country trade) is 41 percent. The probability that this shipment has a destination outside the country of the origin region (foreign trade) is therefore only 19 percent. To evaluate these numbers, one must recognize that the size of the destination markets is quite different. The home market is smaller than the country market, and the latter smaller than the foreign market. When we correct for size,[1] we find enormous differences in the magnitudes of these types of trade. In particular, home, country and foreign trade are 469.5, 11.22 and 0.44 times what one would predict knowing only the sizes of the origin and destination markets.

The second salient aspect of European regional trade is the importance of geographic distance and national borders. The ranking of home > country > foreign trade suggests that these factors are important. Foreign trade involves sellers and buyers that are farther away and do not share the same government. Both of these factors are known to have negative consequences for trade. We show that a parsimonious gravity model that uses only national borders and distance can explain about two-thirds of the variation in European regional trade. Obviously, a model with these elements is designed to create a bias towards home and country trade. But there is more to this. The importance of borders generates a small-country effect, namely, that regions in small countries trade more within and outside their country. The importance of geographical distance generates a remoteness effect, namely, that regions that are geographically remote should trade more with other regions inside their country, and less with regions outside. We observe that both the small-country and remoteness effects are present in the European regional data.

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

A third salient aspect of European regional trade is that the strong home bias in trade cannot be explained by geographical distance and national borders. There are

[1]That is, by dividing by the product of the sizes of the origin and destination markets.

few observations of home trade, 269 out of 73,361, but these observations stand out for their size since they add to 40 percent of all trade. To determine the source of this home bias, we exploit a special feature of the data. Due to government structure differences, in some countries the regions in our dataset are only statistical regions created for the purpose of sharing data with Eurostat, while in other countries the regions in our dataset coincide with political divisions with different levels of self government. This allows us to test whether the home bias effect emerges in all regions, or whether the home bias effect emerges only when it coincides with political borders. We separate region-pairs by the type of border that divides them, either statistical or political, and show that it is the latter and not the former that exhibit a large home bias in trade. Thus, it seems that a large part of the home bias is driven by political borders.

There is an abundance of papers that use the gravity framework to study trade flows. Head and Mayer (2014) provide an extensive review of this literature and the improvements in the methods since being introduced by Tinbergen (1962). Due to the scarcity of data at the subnational level, most of these studies have focused exclusively on international trade. Among the most notable exceptions are papers that use the commodity flow survey to study intranational trade flows in the United States such as Hillberry and Hummels (2008) and Coughlin and Novy (2012). There exist also other papers that look at intranational trade in other countries, for instance Head et al. (2002) for France, Nitsch and Wolf (2013) for Germany and Wrona (2018) for Japan. All these papers focus exclusively on intranational trade. One contribution of this paper is to provide an integrated view of intranational and international trade, and their interactions for Europe, which includes 24 countries and 269 regions.

Our findings suggest that political borders, both national and regional, are an important determinant of trade. Thus our paper is closely related to a large literature that aims at measuring border effects. The seminal papers in this literature are McCallum (1995), Anderson and Van Wincoop (2003), Chen (2004). Two recent papers that also focus on Europe are Santamaría et al. (2020), from which we borrow the data, and Head and Mayer (2021). The final contribution of this paper is to show that border effects apply to political borders but not statistical ones.

2 A first look at the data

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

2.1 The dataset

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

The European Road Freight Survey collects data 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 European Union, the United Kingdom 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 European Road Freight Survey collects data on truck shipments between European countries. One limitation of our data is that it covers trade by road only but not other modes of transportation. With respect to this, Eurostat reports that about 70 percent of all intra-european trade in goods is inland trade and 30 percent is sea trade, and that road trade accounts for about 75 percent of inland trade. To compute these numbers Eurostat uses only foreign trade. If these proportions are also similar for home and country trade, this means that we cover about 52 percent of all intra-european trade. There are 13 industries in the European Road Freight Survey that cover all of agriculture, mining and manufacturing. Except for one (*Coal* *and lignite, crude petroleum and natural gas*), road trade is by far the most prevalent mode of inland transportation. This is why Santamaría et al. (2020) dropped this industry and the dataset contains the remaining 12 industries.

The second limitation of our data is that it does not cover trade with non-European partners. To understand the implication of this, consider a shipment from China to Switzerland that goes through the port of Rotterdam. In country level statistics this would be recorded as a shipment from China to Switzerland. In our survey this would be recorded as a shipment from the Netherlands to Switzerland. This should not cause a problem for a researcher using this data as long as she is aware of this discrepancy.[2]

Our analysis should be interpreted narrowly as referring to trade by road among European regions. Since we have no dataset that is more comprehensive at this time, there is a natural temptation to extrapolate our results to all trade, that is, to trade that uses other means of transportation and/or involves origins and destinations outside of Europe. This type of extrapolation should be done with caution, as the following observations show. Road shipments might be used more frequently than sea shipping among region pairs that are geographically close than among region pairs that are geographically distant. Road shipments are used to trade most agricultural and manufacturing goods, but they are certainly not used to trade services. Regions at the periphery of Europe are likely to trade more with non-European partners and less with European partners. Regions that have ports are likely to serve as intermediaries for trade originating outside of Europe. All of these observations (and possibly more) must be weighted when one tries to extrapolate our findings to trade that is not covered in our dataset.

An important advantage of our dataset is that it includes trade of a region with itself. That is, it provides estimates of the diagonal elements of the matrix of bilateral trade that are produced with the same methodology that is used to produce estimates

[2]In any case, a simple calculation suggests that we are not overestimating intra-european trade across countries. Recall that 52% of all intra-european (foreign) trade is done by road. Thus, our dataset should contain about 52% of all intra-european (foreign) trade. When we add all foreign trade in our dataset we have about 44% of all intra-european (foreign) trade as reported by Eurostat. If anything, we are underestimating rather than overestimating foreign trade.

for the non-diagonal entries of the matrix. This is typically not the case in datasets that cover international trade. The absence of the diagonal elements in these datasets forces researchers to test theories only partially or, alternatively, to "fill in" the missing diagonal entries using methodologies that are not consistent with those used to build the rest of the matrix.

The dataset contains the value of goods shipped among all region pairs for all industries and years. We refer to the region where a shipment starts as the seller and to the region where the shipment arrives as the buyer. We do not know the identities of the specific parties involved in the shipments. Some of them might entail moving goods between establishments of a given firm, while others might entail moving goods from establishments of one firm to those of a different one. We do not know either how the parties obtained the goods and what they do with them. Some firms shipping goods might be the original producers of these goods, while other might be intermediaries. Some firms receiving the goods might be the final consumers of the goods, while others might be intermediaries. Having this additional information would be useful to test alternative trade theories, but it is not crucial to provide an accurate description of how goods flow within and across European regions.

Since these flows vary little between 2011 and 2017, we use averages over the entire period and ignore the time dimension. Here we mostly focus on the aggregate bilateral trade matrix that also averages across industries. Whenever relevant, we discuss the most notable differences between the results obtained with the average matrix and the industry matrices.

Each of these bilateral trade matrices takes the following form:

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

where X_{nm} is the total value of shipments of goods from origin n to destination m. We measure shipments as a share of total shipments: $\sum_n \sum_m X_{nm} = 1$. Thus, X_{nm}

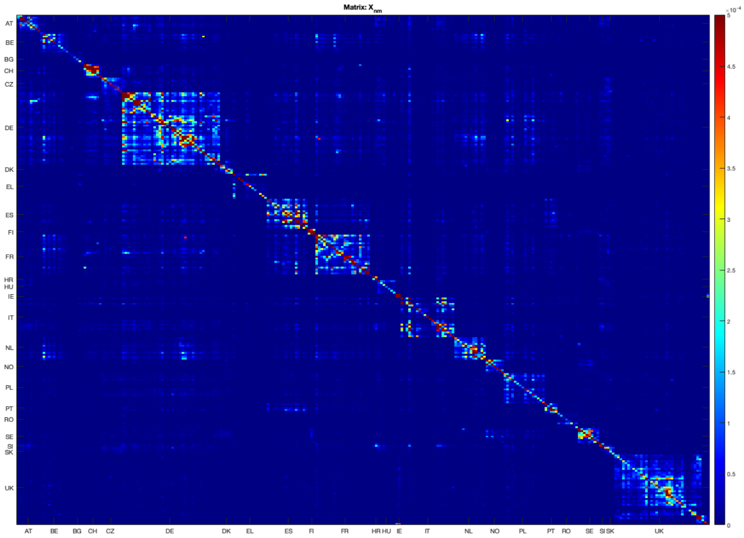
is the probability that a shipment of goods has origin n and destination m.

Figure 2 shows a heat map of the matrix of bilateral trade. We refer to the entries in the main diagonal as *home* trade because they record trade within regions. Despite being a small set of entries (269 out of 72,361), each of them contains a lot of trade. Adding them, we find that home trade constitutes 40 percent 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 2 as the squares centered around the diagonal (without including the latter). Larger squares refer to countries with more regions, such as Germany or France. Smaller squares refer to countries with fewer regions such as Portugal or Ireland. Country trade entries tend to contain less trade than home trade entries. But there are many more country trade entries (4,958 out of 72,361) and, adding them, we find that country trade constitutes about 41 percent of all European regional trade. Finally, we refer to the remaining off-diagonal entries as *foreign* trade. We can identify these entries in Figure 2 as the off-diagonal entries outside the squares. Though most of the entries are foreign trade (67,134 out of 72,361), each of them contains little trade. This is why adding them we find that foreign trade constitutes only 19 percent of all European regional trade. There is therefore a strong bias towards home and country trade in our data.

The matrix in Figure 2 contains a fair amount of zeros. Not surprisingly, there are no zeros for home trade. But there are a few zeros for country trade: 157 out of 4,958 region pairs. And there are many more for zeros for foreign trade: 25,699 out of 67,134 region pairs. This distribution of zeros is also consistent with a strong home and country biases in European regional trade.

What explains these biases? A prime suspect is distance. The distance traveled by shipments classified as home, country and foreign trade is not the same. Fortunately, the European Road Freight Survey survey provides the actual distance traveled by each individual shipment, including shipments within and across regions. Figure 3 shows the histograms for distance traveled for home, country and foreign trade separately. The average distance traveled for the different types of trade is 21.2 Kms, 223.0 Kms and 631.9 Kms, respectively. There is little overlap, for instance, between

Figure 2: Bilateral trade matrix for European regions



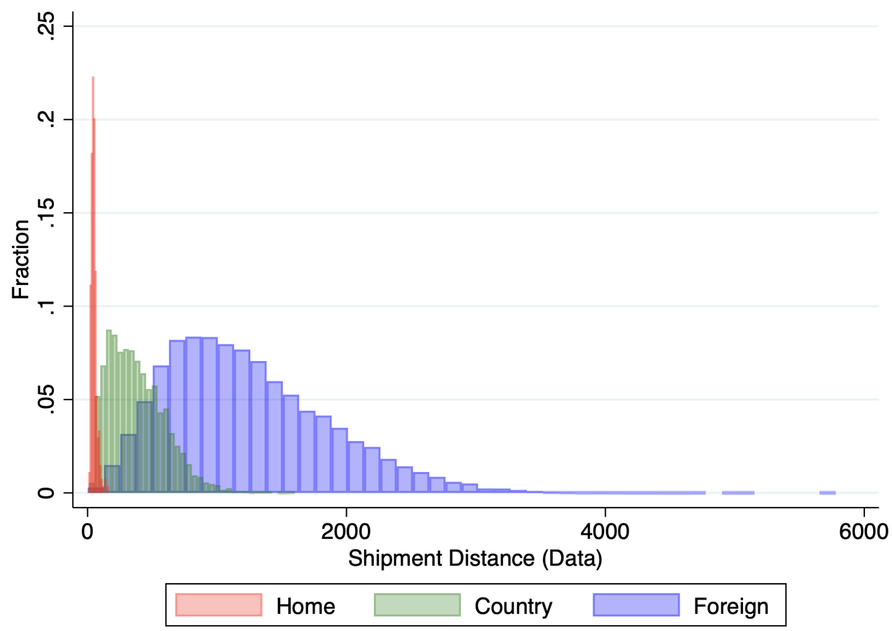
the histograms for home and foreign trade.

2.2 Normalized market shares

Our goal is to understand the shape of 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?

To answer these questions, we need a benchmark that is size free. To see this, consider the case of Catalonia and La Rioja, two regions in Spain. The probabilities of a sale to the Basque Country, another region in Spain, for Catalonia and La Rioja are 0.000226 and 0.0000542, respectively. The probabilities of a purchase by Catalonia and La Rioja from the Basque country are 0.0004281 and 0.0000601, respectively. Catalonia's trade probabilities are one order of magnitude larger than those of La Rioja. Does this mean that Catalonia has a more intensive 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

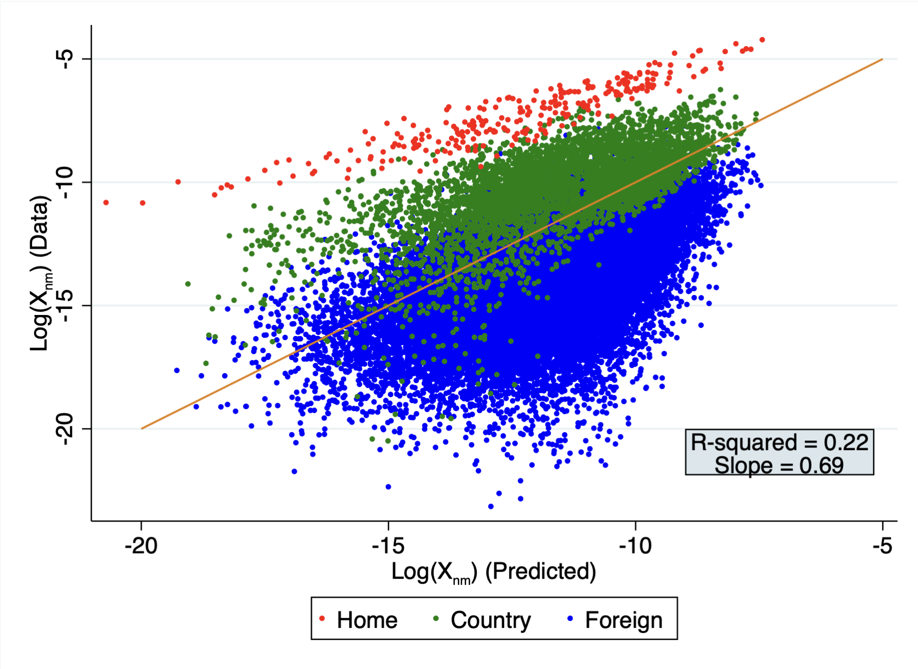
Figure 3: Home, country and foreign distances



almost inevitable that Catalonia trades more with the Basque Country than La Rioja. The size of origin and destination regions matters and we need to correct for this.

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

Figure 4: Actual vs predicted trade (log) probabilities



Thus, we can use deviations from this benchmark to learn about the role that these factors play in shaping trade relationships.

Figure 4 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. The result that size explains close to a quarter of the total variation in trade probabilities 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 4 is that more than three quarters of the variation in trade probabilities cannot be explained by size. This is the variation we care about. Home trade observations are located well above the 45 degree line, confirming that regions trade with themselves much more than what their sizes suggest. The same applies to country trade observations, although to a lesser extent. The counterpart is that most foreign trade observations are below the 45 degree line. European regions have intense trading relationships with themselves and with other regions within their country, and mild trade relationships with regions in other countries.

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

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

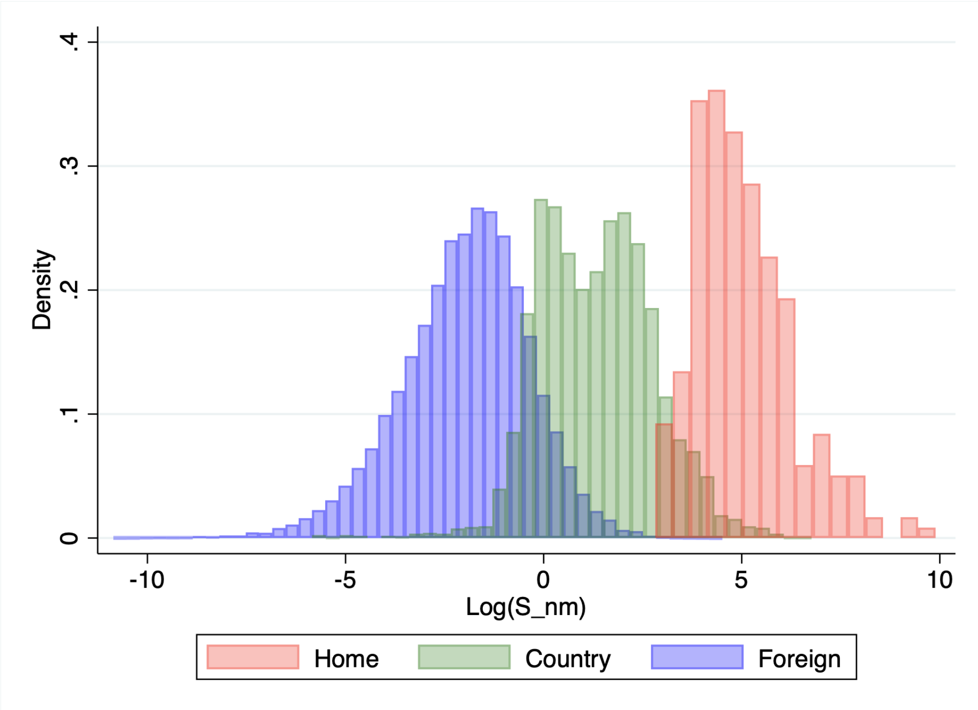
We refer to this measure as a normalized market share.[3] This measure corrects for the mechanical effect of size on trade and it has a very simple interpretation: if $S_{nm} = 2$ (0.5), shipments from origin n to destination m are twice (half) as large as one would be able to predict knowing only the sizes of the regions. Thus, S_{nm} is a size-free measure of how strong a trade relationship is.[4]

Figure 5 plots histograms of (log) normalized market shares for home, country and foreign trade. The average values for the different types of trade are 469.5, 11.22 and 0.44, respectively. 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!). More so, since we are using data on trade in goods and not trade in services.

[3]The reason is that S_{nm} has two alternative interpretations that suggest this name. First, S_{nm} is the share of origin n in destination market m, i.e., X_{nm}/X_m^D; normalized by the share of origin n in the entire European market, i.e., X_n^O. Second, S_{nm} is the share of destination m in origin market n, i.e., X_{nm}/X_n^O; normalized by the share of destination m in the entire European market, i.e., X_m^D. The World Bank uses a related measure for country trade named Trade Intensity Index (https://wits.worldbank.org/wits/wits/witshelp/Content/Utilities/e1.trade_indicators.htm.). This index normalizes probabilities by international trade instead of total trade, i.e., it does not include home trade.

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

Figure 5: Home, country and foreign normalized market shares

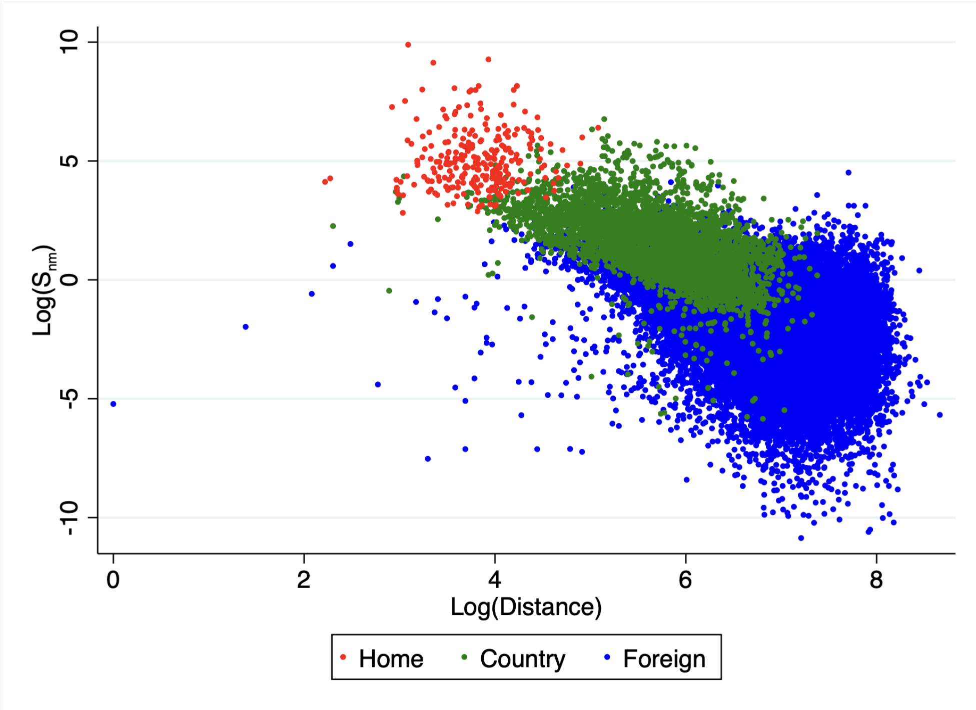


Finally, and just to whet the appetite for what is coming next, Figure 6 plots the (log) normalized market shares against the (log) of actual distances. It is apparent that the strength of trade relationships declines with distance. This surely helps explain part of the home and country biases in trade. But Figure 6 also shows that distance cannot be the single explanation for these biases. Within any given distance interval, we can observe the ranking of home > country > foreign trade. What else is going on? We turn next to a systematic examination of the data using the standard gravity framework.

3 A gravity look at the data

Figure 7 shows the matrix of (log) normalized market shares. The goal of this section is to provide a parsimonious description of this matrix. To do this, we use the gravity framework to guide our search for patterns. The bottom line is simple again: using

Figure 6: (Log) normalized market shares and (log) distance



distance and borders we can explain about two thirds of the variation in (log) normalized market shares. To reach this conclusion, we explore a battery of increasingly flexible specifications for distance and border effects.

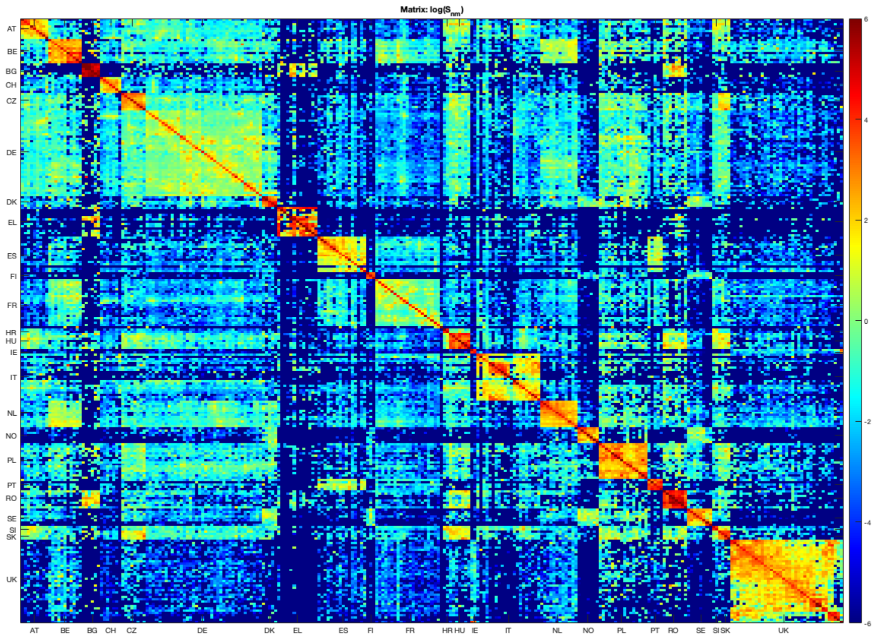
3.1 The gravity framework

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

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

where $\{Z^i_{nm}\}$ is a set of bilateral variables that jointly determine market access and $\{\theta^i\}$ is a set of theoretical coefficients. The set of bilateral variables typically contains

Figure 7: Bilateral matrix of (log) normalized market shares for European regions



a distance variable and a border dummy measuring whether the regions are in the same country or not. In many cases, other variables that might affect the costs of shipping goods are added such as dummies measuring whether the regions have a common language or currency.

The gravity framework consists of the following mathematical model:

$$X_{nm} = \frac{M_{nm}}{M_n^O M_m^D} X_n^O X_m^D \tag{4}$$

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

$$S_{nm} = \frac{M_{nm}}{M_n^O M_m^D} \tag{5}$$

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

$$1 = \sum_m X_m^D \frac{M_{nm}}{M_n^O M_m^D} \tag{6}$$

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

We refer to M_n^O and M_m^D as origin and destination measures of average market access.[5] Equations (6) and (7) are not additional theoretical restrictions, but instead consistency requirements that ensure that probabilities add, i.e., $1 = \sum_m X_m^D S_{nm}$ and $1 = \sum_n X_n^O S_{nm}$.

It is well known that there is a large set of theoretical models that are consistent with the formulation of the gravity framework in Equations (5), (6) and (7) (See Head and Mayer (2014)). 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.

3.2 An important example

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

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

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

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

Figure 8: Borders and trade

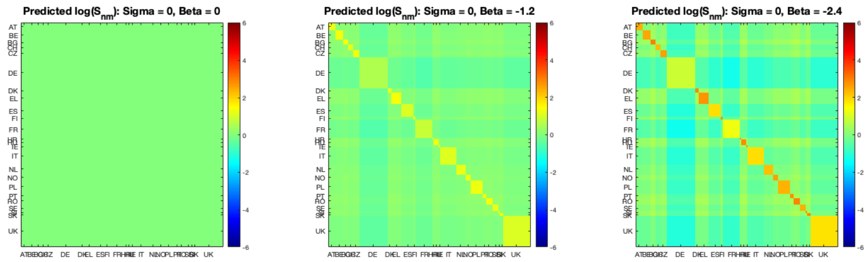


Figure 8 shows three theoretical matrices of (log) normalized market shares produced with this model. In all of them, we set $\sigma = 0$ so that:

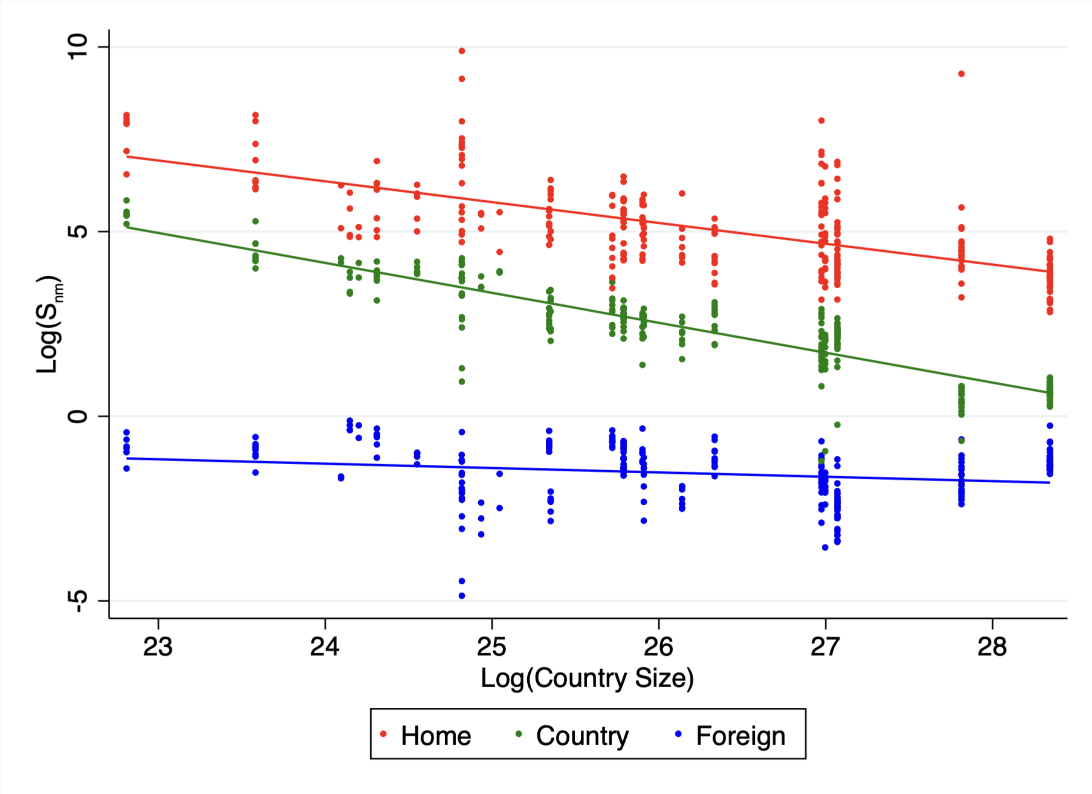
$$M_{nm} = \begin{cases} 1 & \text{if } B_{nm} = 0 \\ e^{\beta} & \text{if } B_{nm} = 1 \end{cases} \tag{9}$$

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

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

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

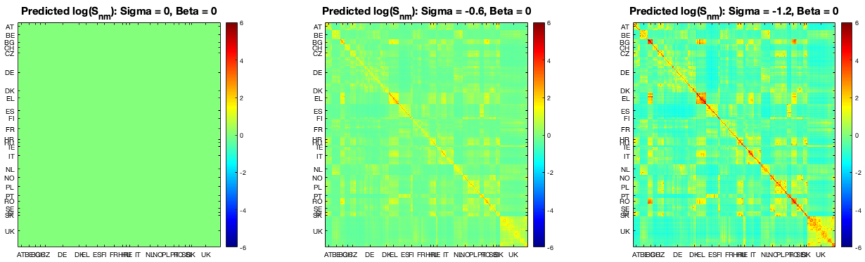
Figure 9: (Log) normalized market shares and country size



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

To explore whether the data is line with these theoretical predictions, Figure 9 plots actual (log) normalized market shares against country size, using different colors for home, country and foreign trade. Not surprisingly, we see again that home/country trade is larger than foreign trade, which is consistent with the first theoretical prediction. More interesting is that regions in small countries have larger (log) normalized

Figure 10: Distance and trade



market shares than regions in large countries. This can be seen when we compare (log) normalized market shares within each type of trade. Clearly, the small-country effect is present in the European regional trade data.

Figure 10 shows three additional theoretical matrices of (log) normalized market shares produced with the model. In all of them, we set $\beta = 0$ so that:

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

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

These observations lead to two additional theoretical predictions. The first one is again that, as the distance effect becomes stronger, country/home trade grows and foreign trade shrinks. The reason is that regions in different countries are far away from regions in the same country (recall Figure 3). This generates again squares centered along the diagonal with high-trade entries inside them and low-trade entries outside. An interesting novelty is that now trade is not homogeneous inside these squares. In particular, there is more trade in the diagonal than in the rest of these squares since regions are closer to themselves than to other regions within the same country. The second theoretical prediction is that remote regions experiment more growth in country/home trade and more shrinkage of foreign trade. The reason again is that average market access shrinks more for regions that are remote within Europe than for regions that are more central. This remoteness effect creates a second specific source of heterogeneity, which is also quite intuitive.

To check whether the data is consistent with these theoretical predictions, Figure 11 plots actual (log) normalized market shares against an index of remoteness.[7] A quick look at the figure shows that (log) normalized market shares for home and country trade do indeed grow with remoteness, while (log) normalized market shares for foreign trade shrink. The remoteness effect is also present in European regional trade data.

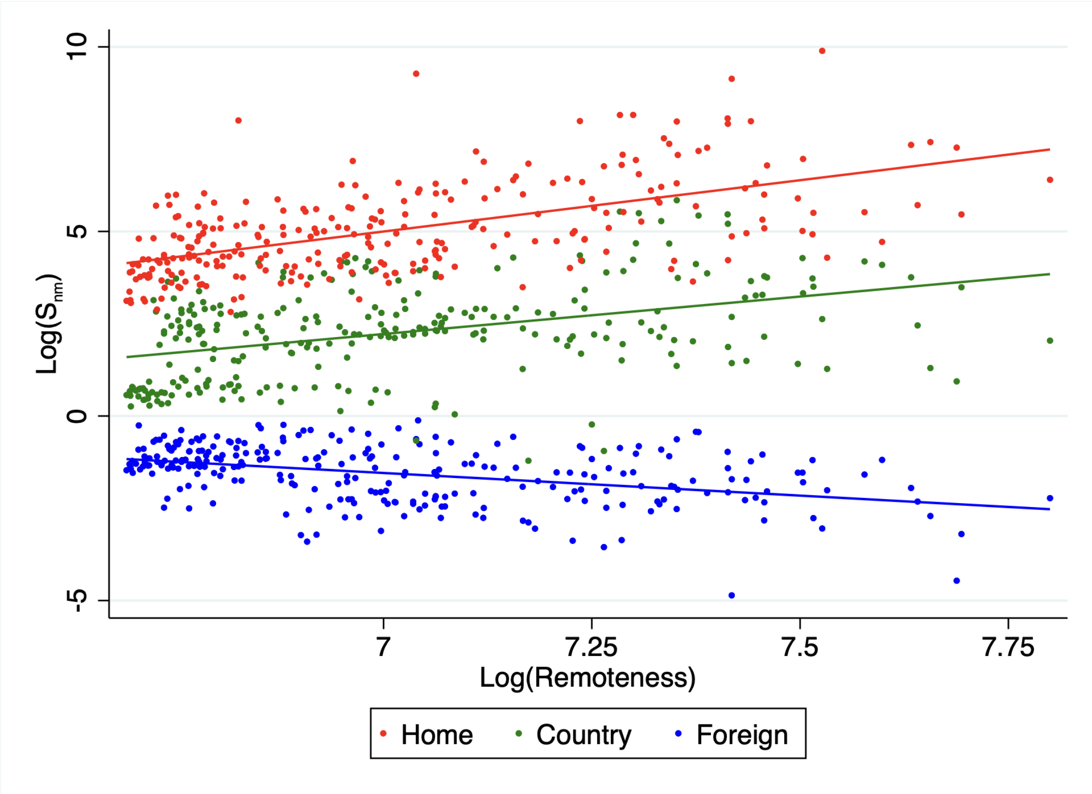
Armed with these intuitions, we search next for the combination of σ and β that provides the best fit of this model to the data. To do this, we define a two-dimensional grid over σ and β. For each point in the grid, we compute: (i) a complete set of bilateral market access measures $\{M_{nm}\}$; (ii) a complete set of origin/destination average market access measures $\left\{M_n^O\right\}$ and $\left\{M_m^D\right\}$; and (iii) the matrix of predicted (log) normalized market shares. We then choose the values of σ and β that minimize the distance between the matrices of actual and predicted (log) normalized market shares.[8] This procedure leads us to choose $\sigma = -1.3$ and $\beta = -2.4$. Figure 12 shows how sensitive is the fit of the model to changes in parameter values.

Figure 13 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. Even though there are differences across the two matrices, it seems that the parsimonious model discussed here captures some of the most important patterns in the data. To reinforce this message, Figure 14 plots the entries of these matrices against each other.

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

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

Figure 11: (Log) normalized market shares and remoteness



3.3 Fixed-effects regressions

Next, we estimate the following fixed-effects regression:

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

where ϕ_n^O and ϕ_m^D are region fixed effects and u_{nm} is an error term that is assumed to be orthogonal to the regressors. The idea behind this regression is to allow the data to choose the parameters $\{\theta^i\}$ that give the model the best chance to explain the data. The estimates of the fixed effects are then interpreted as our estimates of $\ln M_n^O$ and $\ln M_m^D$.[9]

[9] 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.

Figure 12: Sensitivity analysis

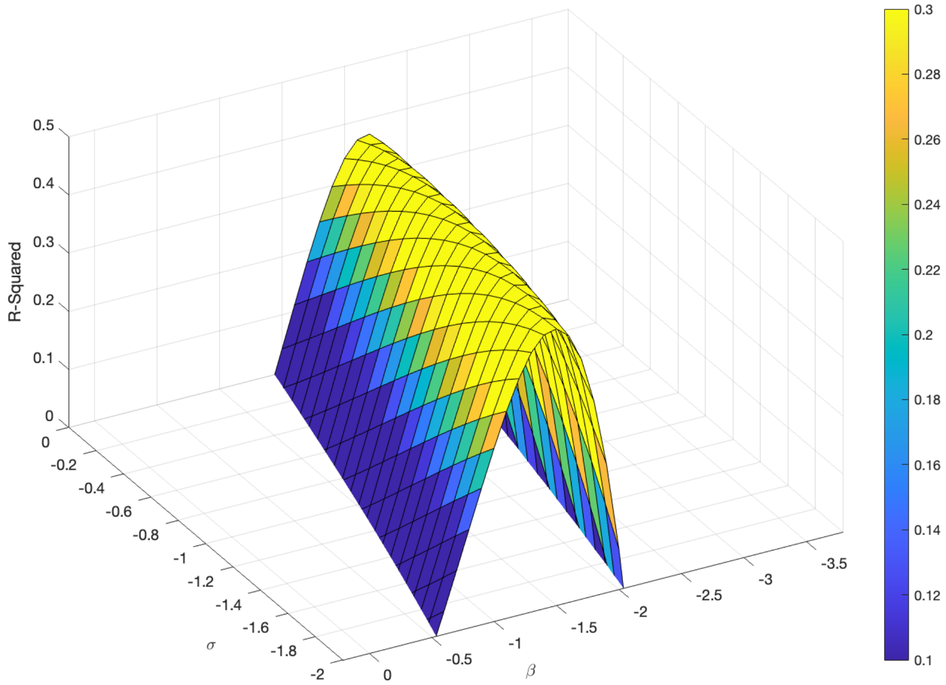


Figure 13: Actual vs predicted matrices of (log) normalized market shares

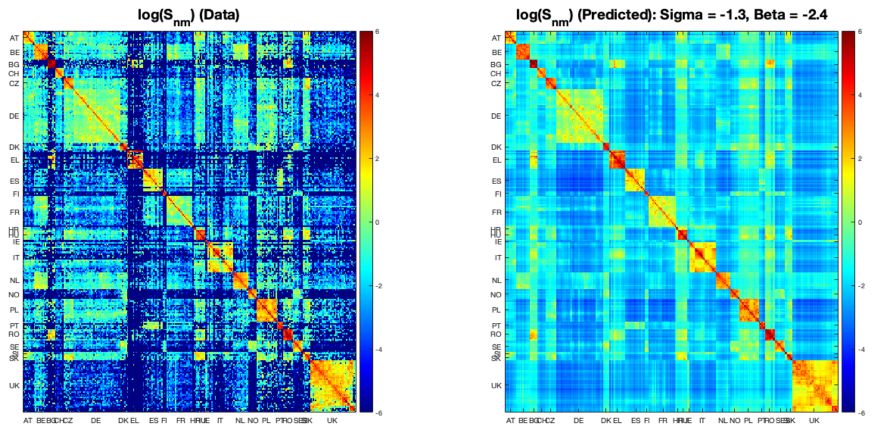


Figure 14: Actual vs. predicted (log) normalized market shares

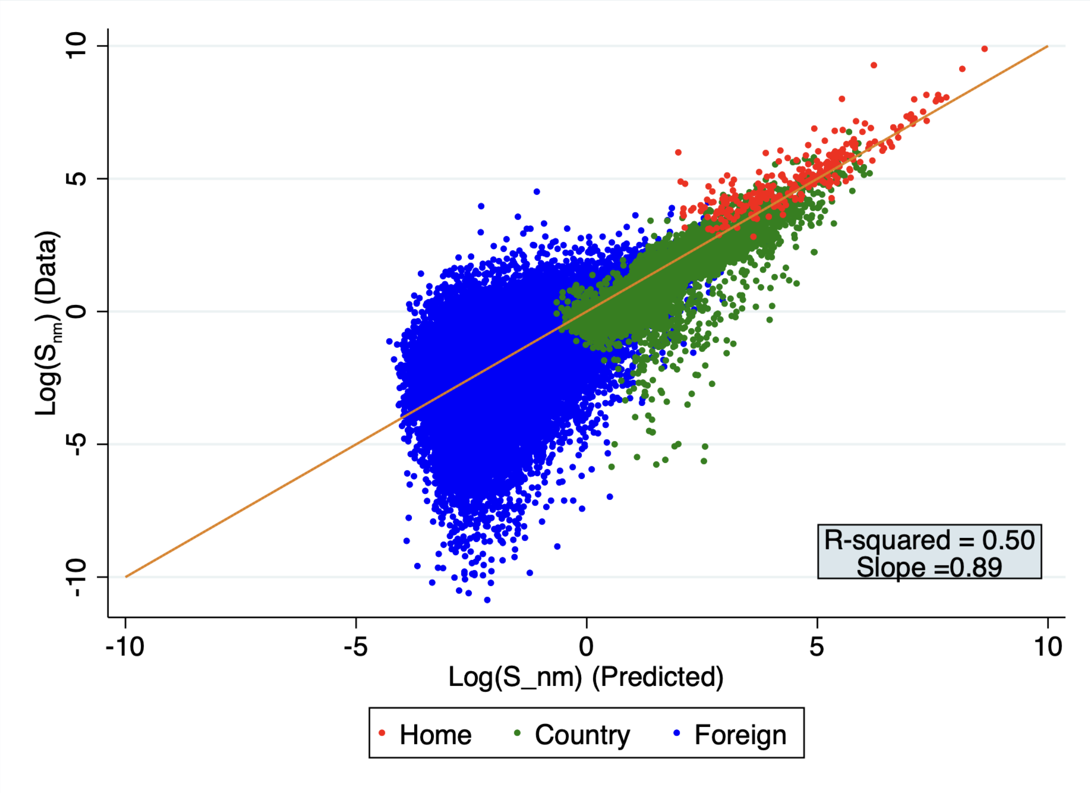


Table 1 shows the results of estimating regression (11) for six different gravity models. Column (1) shows the parsimonious model that we used in the previous subsection. In particular, there is a border dummy B_{nm} and a measure of distance D_{nm} which is the (log) average kilometers travelled from n to m. This specification therefore assumes a constant elasticity of trade to distance.

Column (1) shows that the parsimonious model explains almost two-thirds of the variation in trade probabilities. This is especially remarkable given that we have eliminated the effects of size using (log) trade normalized market shares instead of (log) trade probabilities.[10] Border and distance effects are significative, economically

[10]We have estimated all the regressions in Table 1 using $\ln X_{nm}$ as the dependent variable instead of $\ln S_{nm}$. All the coefficients of bilateral variables remain unchanged up to the third decimal. Since the size correction is picked up by the fixed effects, now to be interpreted as $\ln\left(\frac{X_n^O}{M_n^O}\right)$ and

Table 1: Gravity: Fixed Effects Regressions

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

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.
* $p < .1$, ** $p < .05$, *** $p < .01$

large and not far away from those that we found in the calibration exercise above. The estimated coefficient for the border dummy means that, controlling for distance, a national border reduces bilateral trade to $\exp\{-2.384\} \times 100 = 9.21$ percent of the independence benchmark. The estimated coefficient for distance implies that, controlling for borders, a one percent increase in distance traveled reduces bilateral trade by 1.19 percent with respect to the independence benchmark. Clearly, borders and distances can predict deviations from the independence benchmark.

In Column (2) we use a more general distance function that allows for the elasticity of trade to distance to vary across distance brackets.[11] The results are very similar. The R-squared and the border coefficient are essentially the same.

Columns (3) and (4) allow for some heterogeneity in the border effect. In particular, the border effect is allowed to depend on whether the regions involved have

$\ln\left(\frac{X_m^D}{M_m^D}\right)$, the R-squared of the regressions is a bit inflated. Going from Column (1) to (6) the R-squared starts at 0.681 and grows up to 0.729.

[11]In particular, we estimate a distance function that allows for different elasticities, one for each

a common language and currency.[12] The idea is that sharing a language and/or a currency facilitates trade and reduces the border effect. Using this flexible specification of the border effect raises the R-squared of the regression only marginally. Interestingly, we see that the distance effect is a bit smaller now since the estimated elasticity of trade to distance is -1.071. Again, there is not much difference between the constant- and variable-elasticity specifications for the distance effect.

The results in Columns (3) and (4) indicate that indeed the border effect depends on whether the region pair shares a language and/or a currency. At one extreme, a national border separating a region pair that shares both language and currency reduces bilateral trade to $\exp\{-1.49\} \times 100 = 22.52$ percent of the independence benchmark. At the other extreme, a national border separating a region pair that shares neither language nor currency reduces bilateral trade to $\exp\{-2.744\} \times 100 = 6.43$ percent of the independence benchmark. The estimated coefficients suggest that not sharing a language is more deleterious to trade than not sharing a currency, even though both variables seem to matter.

Columns (5) and (6) estimate different border effect for each country pair. That is, we allow the French-Spanish border to have different effects than the Finish-Spanish or the Irish-British borders. Since there are 24 countries in our sample, we are estimating 276 different border effects. This is the most flexible specification of the border effect so far. Yet, we find that the R-squared of the regression increases only marginally. The distance effect is reduced even further as the estimated elasticity of trade to

---

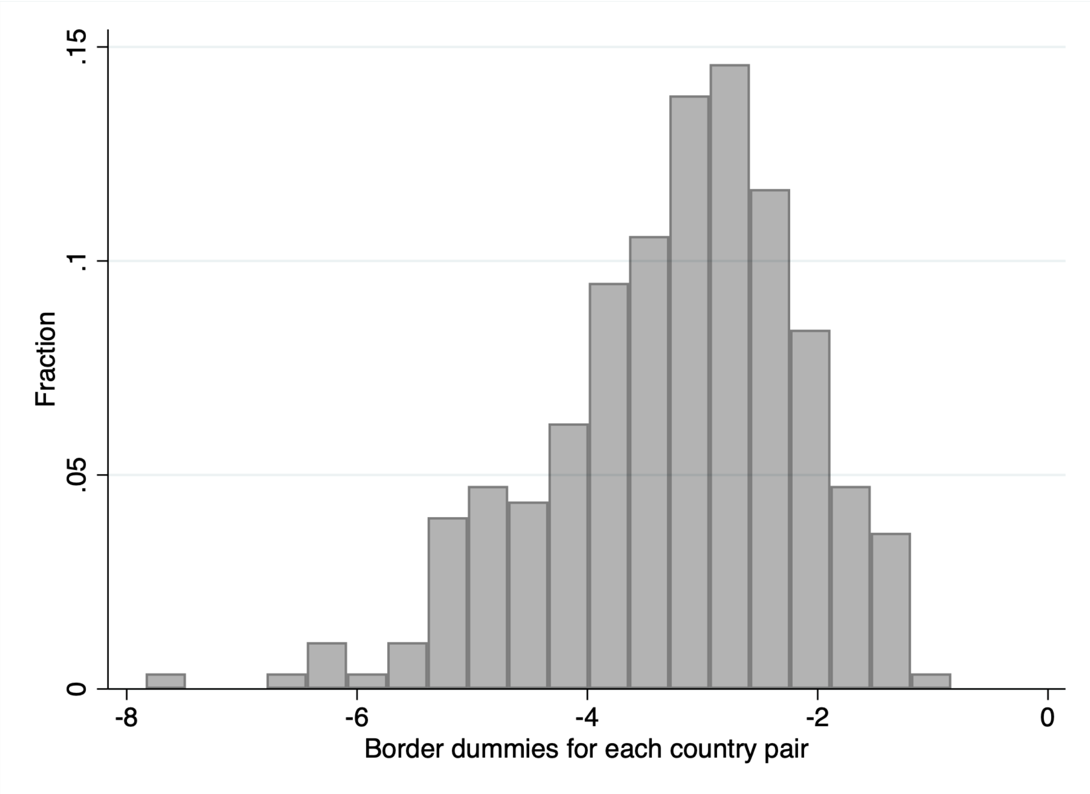
bin $b = 1, ..., B$:

$$\sigma(D_{nm})D_{nm} = \begin{cases} \sigma_1 \ln T_{nm} & \text{if } 0 < T_{nm} \leq T_1 \\ (\sigma_1 - \sigma_2)\ln T_1 + \sigma_2 \ln T_{nm} & \text{if } T_1 < T_{nm} \leq T_2 \\ (\sigma_1 - \sigma_2)\ln T_1 + (\sigma_2 - \sigma_3)\ln T_2 + \sigma_3 \ln T_{nm} & \text{if } T_2 < T_{nm} \leq T_3 \\ \vdots & \vdots \\ \sum_{b=1}^{B-1}(\sigma_b - \sigma_{b+1})\ln T_b + \sigma_B \ln T_{nm} & \text{if } T_{B-1} < T_{nm} \leq \infty \end{cases} \tag{12}$$

where $\sigma(D_{nm})$ is the implicit elasticity which varies as a function of distance. We typically find that σ_1 is larger in absolute value than the constant-elasticity estimate, and our estimates for σ_2, σ_3, ... are very close to -1.

[12] The language dummy captures whether two regions share (at least) one language. This variable was collected in Santamaría et al. (2020) using maps of language areas, and is computed at the region-pair level.

Figure 15: Histogram of country pair dummies



distance is now -1.006. We confirm again that using the constant- or the variable-elasticity specifications of distance does not make much of a difference. The estimates of border effects for each country pair show substantial heterogeneity. Figure 15 and Table 8 in the Appendix show this.

Zero trade flows are prevalent in our data, specially among foreign trade pairs. Our procedure so far has not taken this into consideration. To determine whether the inclusion of zeros changes our results, we follow the literature and report the estimates using the Poisson Pseudo-Maximum Likelihood (PPML) estimator (see Silva and Tenreyro (2006)). The results are shown in the Table 9 in the Appendix. The estimates are quite similar to those obtained with OLS and reported in Table 1. The main difference is that, in Columns (3)-(4) not sharing a currency now is more important than not sharing a language.

Table 2: Gravity: Fixed Effects Regressions

	(1) Agri	(2) Mining	(3) FBT	(4) Textiles	(5) Wood	(6) Coke Pet
Border Effect	-1.698***	-1.191***	-2.426***	-0.991***	-1.656***	-0.728***
	(0.191)	(0.209)	(0.142)	(0.147)	(0.104)	(0.182)
Distance (constant-elasticity)	-1.174***	-1.884***	-1.006***	-0.494***	-1.065***	-1.458***
	(0.0932)	(0.193)	(0.0650)	(0.0938)	(0.0480)	(0.169)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20226	10072	27764	11428	21348	6870
R^2	0.672	0.798	0.699	0.554	0.660	0.718

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.
* $p < .1$, ** $p < .05$, *** $p < .01$

We can compare the magnitude of our results to those of previous studies. Using a structural gravity approach, Mayer et al. (2019) find that belonging to the Single Market is estimated to triple trade (exp{1.177}= 3.24). In North-America, Anderson and Van Wincoop (2003) estimate that provinces in Canada trade 5 times more with themselves than with a neighbouring state in the United States. Our findings show that being in the same country increases trade by between 4 (exp{1.466}=4.3) and 16 (exp{2.782}= 16.1) times.

Tables 2 and 3 show the results for our baseline fixed-effects regressions for each industry individually. Our estimation shows some heterogeneity across industries. The first observation is that this model retains a high explanatory power for all industries, with the R-squared ranging between 0.554 and 0.798. The second observation is that the border effect is also substantial for all industries. It ranges from -0.728 (Coke and Petroleum) to -2.426 (Food, Beverage and Tobacco). For most industries (8 out of 12) it is between -1.4 and -1.8, slightly smaller than the average coefficient we obtained in Table 1. The third and final observation is that the distance coefficient varies substantially across industries, ranging from -0.494 to -1.884. For most industries this coefficient is close to -1, which is close to the average coefficient that we obtained in Table 1.

Table 3: Gravity: Fixed Effects Regressions

	(1) Chem	(2) Non-Metal	(3) Metal	(4) Machinery	(5) Vehicles	(6) Other
Border Effect	-1.619***	-1.860***	-1.610***	-1.810***	-1.674***	-1.422***
	(0.144)	(0.131)	(0.123)	(0.146)	(0.151)	(0.147)
Distance (constant-elasticity)	-1.005***	-1.388***	-0.914***	-0.640***	-0.570***	-0.603***
	(0.0581)	(0.0998)	(0.0603)	(0.0820)	(0.0753)	(0.0678)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24073	16764	22527	22368	20014	16100
R^2	0.633	0.766	0.623	0.586	0.566	0.565

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.
* $p < .1$, ** $p < .05$, *** $p < .01$

4 The home bias in trade

In our previous exploration of the data, we have treated all trade flows within the same country in the same way. However, we have shown in Section 2 that home trade is orders of magnitude larger than country trade, accounting for 40 percent of intra-European flows in our data. We now explore how large is this difference by adding a home-bias dummy to our gravity estimation.

Table 4 shows the same fixed-effects regressions that we saw in Table 1, including this additional variable. There are three key takeaways. First, the coefficient on the home-bias dummy is positive, large and significant.[13] Focusing on our extended model in columns (3)-(4), the average market share of a region with itself ranges from $\exp\{1.013\} \times 100 = 275$ and $\exp\{2.233\} \times 100 = 932$ percent larger than the average market share between two different regions in the same country, controlling for distance. Second, the R-squared of the regressions does not change much after we introduce the home-bias dummy. This reflects the fact that home trade has a very small number of observations in the overall trade matrix. Failing to fit those is not severely penalized in the tests we performed above. Third, our estimates of the

[13]Note that this coefficient increases substantially when we use the variable-elasticity distance function. The reason is that we typically find that at short distances the trade cost is higher. Therefore, a larger dummy is needed to explain the home bias in the data.

Table 4: Gravity: Fixed Effects Regressions

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

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.
* $p < .1$, ** $p < .05$, *** $p < .01$

border effect and distance do not change much as a result of adding the home-bias dummy.[14]

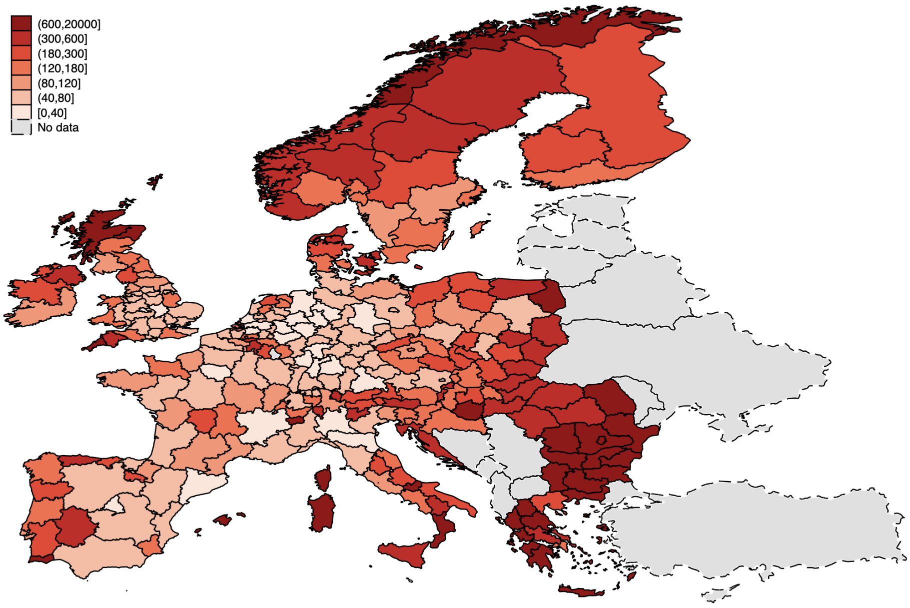
What explains this strong home bias in trade? To make progress in answering this question, we perform two exercises. First, we explore which regional characteristics are correlated with a large home bias. Second, we separate statistical and political regions and show that it is the latter and not the former that exhibit a large home bias in trade.

4.1 Correlates of the home bias in trade

Figure 16 shows the spatial distribution of the home market share in Europe. The first striking pattern is how heterogeneous these shares are across regions. They range from a low of 40 to a high of about 20,000. The map also shows that geography plays an important role. Regions in the periphery of Europe, like Greek and Bulgarian

[14]As a robustness check, we report the results using PPML in Tables 9 and 10 in the Appendix.

Figure 16: Normalised Market share: Home



regions in the South and Norwegian and Swedish regions in the North tend to have higher home trade. Island and mountainous regions also have higher home trade. Interestingly, within-country geography also plays a role: regions in the periphery of a country display higher home trade than more central regions. For instance, regions in the south of Italy and Portugal, in the west of Spain and in the north of the UK and Denmark have higher home trade than the rest of the country.

Interestingly, we see that home trade tends be lower in more densely populated regions of Europe. We see this pattern at the European level in the so-called Blue Banana.[15] We also see this pattern within some countries that are outside the Blue Banana. For instance, Madrid and Catalonia have the lowest home trade in Spain, while Warsaw and Athens have the lowest home trade in Poland and Greece.

Table 5 shows regressions of home trade on a number of regional characteristics,

[15]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.

and country fixed effects. Column (1) reports the results using the following geographical variables: distance, remoteness plus island and mountain region dummies. All these variables are significant, except for distance. This formally confirms that remote regions, island regions and mountainous regions have higher home trade. These simple geographical variables explain 41 percent of the variation in the home market share.

Column (2) adds economic variables: presence of ports, motorway density, population, share of employment in manufacturing and in the public sector, the share of population with at least secondary education and the share of foreign-born population. The introduction of economic variables reduces the coefficients of the geographic variables. All economic variables are significant except for presence of ports. Motorway density reduces the home market share, showing that infrastructure helps overcome geographical obstacles. As we observed in the map, the most populated regions also have lower home market shares. Economic structure also matters, regions with high manufacturing shares, larger governments, more educated populations and more migrants have lower home market shares. Adding all these economic variables raises the R-squared from 41 to 80 percent.

Column (3) adds country fixed effects. The R-squared increases to almost 90 percent, indicating that some of the variation in home trade has a country component. Some variables seem to be correlated with this country component since they now lose their significance and the magnitude of their coefficients is reduced: the share of employment in the public sector, the share of population with at least secondary eduction and the share of foreign-born population. However most of our variables remain significant and their coefficients are stable. This means that the country component does not explain all the variation in home market shares. To confirm this, we estimate a regression that includes only country fixed effects and find that explains 56 percent which is well below the 90 percent obtained in Column (3).

Finally, we explore industry heterogeneity in home bias correlates. We estimate the regression in Column (3) for each of our 12 industries and report the results in Tables 11 and 12 in the Appendix. For most of the industries the results align with the average findings reported here. The exceptions are Agriculture, Mining and

Table 5: Home Bias: Determinants

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

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.
* $p < .1$, ** $p < .05$, *** $p < .01$

Coke/Petroleum, for which remoteness does not play a role.

4.2 Government structure and regional borders

To learn more about the source of this home bias, we exploit a peculiarity of the data collection and harmonization process of our dataset. Since our shipment data is collected and provided by Eurostat, our units of observation are NUTS2 regions.

In some countries these NUTS2 regions are only statistical regions created for the purpose of sharing data with Eurostat. In other countries, however, they coincide with political divisions with different levels of self-government. This provides us with a unique opportunity to see the extent to which regional governments are behind this home bias in trade. In particular, we want to compare region pairs separated by statistical and political borders.

We work with two geographical classifications that partition our set of 24 countries into regions. The finer one is the NUTS2 classification that we have been using up to this point which includes 269 regions. The coarser one is the NUTS1 classification that includes 101 regions.

We define four dummies, to capture the different kinds of regional borders that within-country trade flows may cross. The first two dummies capture the presence of regional statistical borders: $SB1_{n,m}$ takes value 1 if there is a statistical border between n and m at the level of NUTS1, $SB2_{n,m}$ takes value 1 if there is a statistical border between n and m at the level of NUTS2. The second two dummies capture the presence of regional political borders: $PB1_{n,m}$ takes value 1 if there is a political border between n and m at the level of NUTS1, $PB2_{n,m}$ takes value 1 if there is a political border between n and m at the level of NUTS2. The omitted category will be home trade. Table 6 shows the value of these dummies for different countries, according to their political organization.

Table 6: Statistical and Political borders: classification

	Group 1	Group 2	Group 3	Group 4
Within NUTS2	No border	PB2=1	No border	No border
Across NUTS2	SB2=1, PB2=0	SB2=0, PB2=1	SB2=0, PB2=1	SB2=1 in UK, PB2=1 in BE and DE
Across NUTS1	SB1=1, PB1=0	SB1=1, PB1=0	SB1=1, PB1=0	PB1=1 (SB1=1 in England)
Countries	Bulgaria Portugal Slovenia	Croatia Czech Rep. Finland Hungary Ireland Norway Romania Slovakia Sweden Switzerland	Austria Denmark France Greece Italy Netherlands Poland Spain	Belgium Germany United Kingdom

We estimate the following regression to assess the role of statistical and political borders:

$$lnS_{nm} = \phi_n^O + \phi_m^D + \sigma D_{nm} + \lambda_1 SB1_{nm} + \lambda_2 SB2_{nm} + \lambda_3 PB1_{nm} + \lambda_4 PB2_{nm} + u_{nm} \quad (13)$$

Our dependent variable will be the normalised trade share between n and m, and we include origin region, ϕ_n^O, and destination region, ϕ_m^D fixed effects as well as the bilateral distance measure, in logs. u_{nm} is the error term. We start by estimating this regression using only within-country flows and pooling all the countries in our sample together.[16] Table 7 column 1 presents the results.

Column 1 shows the coefficients estimated in the pooled regression in equation 13. Our first finding is that trade shares between regions divided by statistical borders are not substantially different from trade shares within regions. The coefficients on the Statistical border dummies are negative in sign but very small. The coefficient on the NUTS1 level border is marginally significant but very close to zero in magnitude, while the coefficient on the NUTS2 dummy cannot be distinguished from zero due to the low precision of the estimate. This is in line with our predictions: statistical borders are drawn just for data aggregation purposes, with no form of self-government behind them. Therefore, they should not be an obstacle to trade.

Our second finding is that regional political borders, on the contrary, seem to reduce trade. Normalised market shares across region pairs divided by a regional political border are much smaller than within-region trade shares. Political borders at both levels matter. A political border separating a region pair at the NUTS2 level (but in the same NUTS1 region) reduces trade to $\exp\{-0.807\} \times 100 = 44.6$ percent of the independence benchmark. In addition, crossing a NUTS1 political border also has a negative effect on trade (coefficient λ_3=-0.461). Therefore, a region pair separated by political borders both at the NUTS1 level and at the NUTS2 level, will have its trade share reduced to $\exp\{-(0.807+0.461)\} \times 100 = 28$ percent of the independence benchmark.[17]

[16]We drop region EL4 in Greece since it is made up by a very large number of small islands, and the border dummy in this case will coincide with the insularity of the region

[17]To obtain the effect of crossing both a NUTS1 border and a NUTS2 borders, we should have to sum both coefficients because regions in a different NUTS1 will also be, by definition, in a different NUTS2. The coefficient λ_3 identifies the additional effect of crossing a NUTS1 border, while the

Table 7: Statistical borders vs Political borders

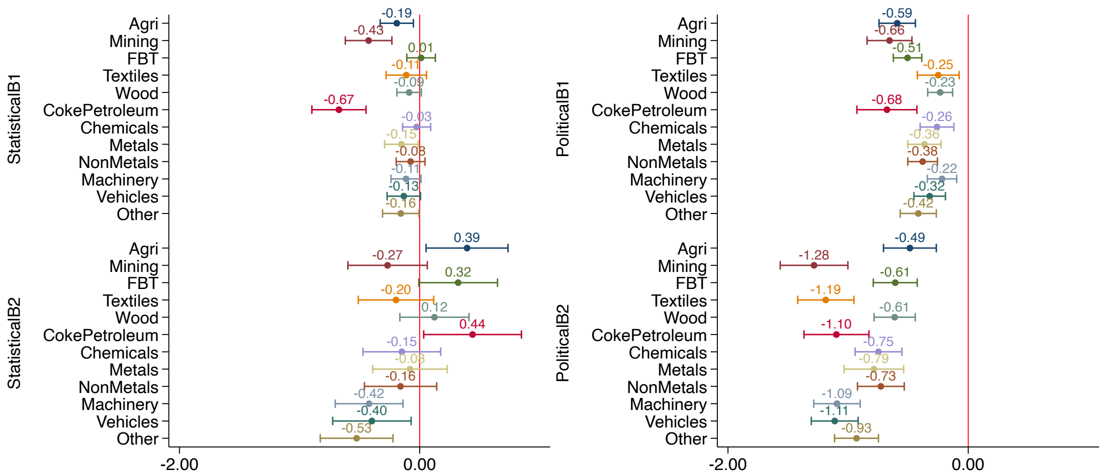
	Pooled (1) Log(S_nm) (1)	Group 1 (2) Log(S_nm) (2)	Group 2 (3) Log(S_nm) (3)	Group 3 (4) Log(S_nm) (4)	Group 4 (5) Log(S_nm) (5)
Statistical Border (NUTS1)	-0.080*	0.187	0.191	-0.086	-0.021
	(0.048)	(0.163)	(0.121)	(0.060)	(0.116)
Statistical Border (NUTS2)	-0.153	-0.344			-0.369
	(0.181)	(0.272)			(0.244)
Political Border (NUTS1)	-0.461***				-0.529***
	(0.081)				(0.086)
Political Border (NUTS2)	-0.807***			-0.737***	-0.810***
	(0.101)			(0.149)	(0.116)
Distance (constant-elasticity)	-1.230***	-1.099***	-1.462***	-1.287***	-1.149***
	(0.030)	(0.154)	(0.064)	(0.050)	(0.042)
Origin FE	Yes	Yes	Yes	Yes	Yes
Dest. FE	Yes	Yes	Yes	Yes	Yes
Observations	5027	61	364	1721	2873
R-squared	0.863	0.951	0.910	0.873	0.781

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We can now compare these numbers to our estimates of the effect of national borders obtained in section 3. The effect of national borders on trade depends on whether region pairs share the same language and the same currency. At one extreme, a national border separating a region pair that shares both language and currency reduces bilateral trade to $\exp\{1.491\} \times 100 = 22.52$ percent of the independence benchmark. At the other extreme, a national border separating a region pair that shares neither language nor currency reduces bilateral trade to $\exp\{2.744\} \times 100 = 6.43$ percent of the independence benchmark. The effect of regional political border is smaller, on average, than the effect of national borders. Most countries only have one regional border, that has smaller trade-reducing effects. However, in countries with two levels of political borders, our estimates almost overlap with the lowest bound of the effects of national borders. Indeed, regional political borders at both NUTS1 and NUTS2 levels reduce trade between two regions to 28 percent of the independence benchmark. Thus, regional political borders are in some cases as damaging for domestic trade as

coefficient λ_4 identifies the effect on a trade flow that crosses a NUTS2, but stays within a NUTS1 region.

Figure 17: Statistical and political borders by industry



Notes: Figure shows the coefficients from estimating equation 13 in each industry. Panel 1 shows the coefficients on the Statistical border dummies, where $StatisticalB1$ is the coefficient on the NUTS1 statistical border, while $StatisticalB2$ is the coefficient on the NUTS2 statistical border. Panel 2 shows the coefficients on the political border dummies, where $PoliticalB1$ is the coefficient on the NUTS1 political border, while $PoliticalB2$ is the coefficient on the NUTS2 political border.

a national border that divides a region pair with the same language and currency. Columns 2 to 5 show the result of estimating regression 13 in each group of countries independently. The results confirm our findings in the pooled sample. Across all groups, we cannot distinguish the coefficient on statistical borders from zero, while the coefficient on political borders is large, negative and significant. The magnitude varies slightly across groups but the negative effect of political borders on domestic trade is always clear.

Finally, we perform the same exercise for each industry. Figure 17 shows the results of estimating equation 13 by industry. The first panel reports the coefficients on statistical borders (at NUTS1 and NUTS2 levels), while the second panel reports the coefficients on political borders (at NUTS1 and NUTS2 levels). The results confirm our aggregate findings: statistical borders do not have a substantial effect on trade, while political borders clearly deter trade. Panel 1 shows that in 75 percent of all industries statistical borders are not significantly different from zero. Panel 2 shows the opposite for political borders. In all industries, the coefficient on the political border dummies is negative, significant and much more precisely estimated. In terms of magnitude, the political borders at the NUTS2 level are more heterogeneous by industry than at the NUTS1 level. A few potential explanations come to mind as to why political borders deter trade across regions. First, regional and local governments may create regulation that facilitates trade flows within regional borders but imposes barriers on cross-border participation. Second, regional governments with a higher degree of self-government may also have stronger preferences for local goods and/or shape the preferences of local consumers. Differences in degrees of self-government and regional autonomy across countries could be used to explore this hypothesis. While studying the source of these domestic barriers to trade is outside the scope of this paper, we believe this is an exciting avenue for future research.

5 Concluding remarks

This paper has provided an integrated view of intranational and international trade in Europe using the new dataset we constructed in Santamaría et al. (2020). The picture that emerges is clear: (i) European regional trade has a strong home and country bias, (ii) geographic distance and national borders are important determinants of regional trade, but they cannot explain the strong home bias and (iii) this home bias is quite heterogeneous across regions and seems to be driven by political borders at the regional level.

Our findings open up several interesting questions. Why is it that political borders and geographical distance still remain such a strong impediment to trade in the context of Europe? How does the behaviour of governments shape regional trade flows, contributing to the large home bias in trade? Which factors explain the heterogeneous home bias and border effects that we see across countries? Providing a sound answer to these questions might have important policy implications.

The key tool that we have used to explore trade interactions is the matrix of bilateral trade. This matrix provides a snapshot of all trade flows within and across European regions and countries. Unfortunately, many important economic indicators such as migration flows, foreign direct investment or bank lending relationships, are not yet available at the region-pair level in such a unified way. Has Europe achieved a higher degree of integration in these areas? It would also be useful to construct similar matrices for other social and cultural interactions such as travel and tourism, cultural exchanges, sports competitions, joint research projects, and so on. These matrices would help us form an accurate picture of how European citizens interact with each other.

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.

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.

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.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics 97*(1), 76–85.

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., V. Vicard, and S. Zignago (2019). The cost of non-europe, revisited. *Economic Policy 34*(98), 145–199.

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

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' Economique 46*(1), 154–179.

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

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

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

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

A Additional Figures and Tables

Table 8: Border effects for country pairs

Country	Border (Mean)	Border (SD)	Highest (1)	Highest (2)	Highest (3)	Lowest (1)	Lowest (2)	Lowest (3)
AT	-2.93	0.88	-5.24 (FI)	-4.68 (IE)	-3.84 (NO)	-1.45 (DE)	-1.55 (SI)	-1.96 (SK)
BE	-2.71	1.03	-5.30 (FI)	-4.57 (IE)	-4.32 (NO)	-1.35 (FR)	-1.53 (NL)	-1.68 (CZ)
BG	-3.18	1.25	-7.84 (NO)	-4.57 (HR)	-3.93 (PT)	-1.21 (IE)	-1.67 (UK)	-2.28 (EL)
CH	-3.50	0.77	-5.18 (IE)	-4.90 (SI)	-4.88 (HR)	-2.14 (DE)	-2.77 (SK)	-2.77 (BE)
CZ	-2.58	0.76	-4.07 (IE)	-3.82 (FI)	-3.65 (HR)	-0.84 (SK)	-1.48 (DE)	-1.68 (BE)
DE	-2.53	1.07	-5.17 (FI)	-4.99 (IE)	-3.92 (NO)	-1.43 (SK)	-1.45 (AT)	-1.48 (CZ)
DK	-3.18	0.68	-4.61 (FI)	-4.44 (UK)	-4.42 (IE)	-2.28 (BG)	-2.29 (PL)	-2.33 (SE)
EL	-2.90	1.15	-5.19 (HR)	-4.05 (SE)	-3.83 (ES)	-1.80 (NO)	-2.28 (BG)	-2.36 (UK)
ES	-3.17	1.10	-5.60 (IE)	-5.18 (FI)	-4.82 (NO)	-1.43 (PT)	-1.70 (FR)	-2.01 (BE)
FI	-4.47	1.39	-6.50 (PT)	-6.39 (IE)	-6.39 (UK)	-2.60 (BG)	-2.99 (SE)	-3.82 (CZ)
FR	-3.10	1.12	-5.36 (IE)	-5.06 (FI)	-4.89 (NO)	-1.35 (BE)	-1.70 (ES)	-1.93 (SI)
HR	-4.11	0.91	-5.87 (IE)	-5.59 (PT)	-5.27 (FI)	-1.91 (SI)	-3.02 (AT)	-3.13 (HU)
HU	-2.94	0.97	-5.09 (IE)	-4.69 (NO)	-4.29 (FI)	-1.50 (SI)	-1.59 (DE)	-1.72 (SK)
IE	-4.57	1.47	-6.39 (FI)	-6.19 (PT)	-5.87 (HR)	-1.21 (BG)	-3.31 (UK)	-4.02 (NL)
IT	-2.98	0.84	-4.93 (IE)	-4.68 (FI)	-4.48 (NO)	-1.88 (SI)	-2.04 (SK)	-2.13 (PL)
NL	-2.84	0.76	-4.89 (FI)	-4.02 (IE)	-3.93 (NO)	-1.53 (BE)	-1.72 (DE)	-2.08 (PL)
NO	-4.07	1.17	-7.84 (BG)	-4.97 (IE)	-4.89 (FR)	-1.80 (EL)	-2.13 (SE)	-2.72 (DK)
PL	-2.73	0.69	-4.13 (HR)	-4.07 (IE)	-3.94 (FI)	-1.58 (DE)	-1.76 (BE)	-2.08 (NL)
PT	-3.38	1.24	-6.50 (FI)	-6.19 (IE)	-5.59 (HR)	-1.43 (ES)	-2.11 (SK)	-2.20 (FR)
RO	-3.25	0.66	-4.70 (FI)	-4.32 (HR)	-4.27 (NO)	-2.28 (BE)	-2.43 (HU)	-2.48 (UK)
SE	-3.65	0.79	-5.39 (IE)	-4.97 (UK)	-4.64 (FR)	-2.13 (NO)	-2.33 (DK)	-2.86 (PL)
SI	-2.86	1.11	-5.01 (IE)	-4.90 (CH)	-4.30 (NO)	-1.31 (SK)	-1.50 (HU)	-1.54 (DE)
SK	-2.60	1.00	-5.64 (IE)	-3.91 (FI)	-3.51 (HR)	-0.84 (CZ)	-1.31 (SI)	-1.43 (DE)
UK	-3.37	1.03	-6.39 (FI)	-4.97 (SE)	-4.78 (NO)	-1.67 (BG)	-2.29 (PL)	-2.36 (EL)

Table 9: Gravity: PPML Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	S_{nm}	S_{nm}	S_{nm}	S_{nm}	S_{nm}	S_{nm}
Border Effect	-1.808***	-2.002***				
	(0.123)	(0.108)				
Border / common language / common currency dummy			-1.724***	-1.725***		
			(0.214)	(0.182)		
Border / common language / different currency dummy			-1.855***	-1.833***		
			(0.146)	(0.151)		
Border / different language / common currency dummy			-1.719***	-1.995***		
			(0.148)	(0.147)		
Border / different language / different currency dummy			-1.848***	-2.096***		
			(0.145)	(0.127)		
Distance (constant-elasticity)	-1.412***		-1.410***		-1.473***	
	(0.0644)		(0.0655)		(0.0708)	
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Distance (variable-elasticity)	No	Yes	No	Yes	No	Yes
Border dummies for each country pair	No	No	No	No	Yes	Yes
Observations	46505	46505	46505	46505	46505	46505
R^2	0.975	0.977	0.975	0.977	0.975	0.977

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.

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

Table 10: Gravity: PPML Regressions

	(1) S_{nm}	(2) S_{nm}	(3) S_{nm}	(4) S_{nm}	(5) S_{nm}	(6) S_{nm}
Border Effect	-2.187***	-1.871***				
	(0.119)	(0.120)				
Border / common language / common currency dummy			-1.713***	-1.505***		
			(0.142)	(0.128)		
Border / common language / different currency dummy			-1.840***	-1.704***		
			(0.145)	(0.141)		
Border / different language / common currency dummy			-2.136***	-1.829***		
			(0.141)	(0.142)		
Border / different language / different currency dummy			-2.317***	-2.002***		
			(0.135)	(0.134)		
Home Bias	1.475***	2.128***	1.508***	2.143***	1.486***	2.122***
	(0.414)	(0.522)	(0.418)	(0.526)	(0.480)	(0.555)
Distance (constant-elasticity)	-0.783***		-0.762***		-0.776***	
	(0.141)		(0.145)		(0.183)	
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes	Yes
Distance (variable-elasticity)	No	Yes	No	Yes	No	Yes
Border dummies for each country pair	No	No	No	No	Yes	Yes
Observations	46505	46505	46505	46505	46505	46505
R^2	0.988	0.991	0.988	0.991	0.988	0.991

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.

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

Table 11: Home Bias: Determinants - by Industry

	Agri	Mining	FBT	Textiles	Wood	Coke/Pet
	(1)	(2)	(3)	(4)	(5)	(6)
	Home	Home	Home	Home	Home	Home
Distance	-0.271	-0.187	0.199	-0.762*	-0.175	-0.00967
	(0.196)	(0.233)	(0.203)	(0.448)	(0.251)	(0.212)
Log(European Remoteness)	0.184	0.489	1.471***	2.331**	1.351**	0.551
	(0.454)	(0.509)	(0.468)	(0.920)	(0.579)	(0.499)
Island Region	1.784***	0.581	1.388***	1.869*	2.310***	0.554
	(0.289)	(0.510)	(0.353)	(1.023)	(0.440)	(0.363)
Mountain Region	0.336***	0.0968	0.242**	0.300	0.214*	0.0826
	(0.103)	(0.100)	(0.0964)	(0.187)	(0.111)	(0.101)
Major Port Region	-0.251**	0.0190	-0.0792	0.0156	-0.0859	-0.250**
	(0.0974)	(0.132)	(0.105)	(0.252)	(0.129)	(0.111)
Motorway Density	-2.032	1.010	-4.170***	-11.90***	-5.245***	-4.121**
	(1.608)	(1.494)	(1.420)	(3.126)	(1.842)	(1.617)
Log(Population)	-0.606***	-0.786***	-0.553***	-0.889***	-0.549***	-0.752***
	(0.0662)	(0.0814)	(0.0750)	(0.130)	(0.0736)	(0.0659)
Share of Emp. (Manuf.)	-4.761**	-7.716***	-3.778**	-11.24***	-10.27***	-4.877**
	(1.942)	(2.152)	(1.881)	(3.769)	(2.387)	(2.067)
Share of Emp. (Public)	4.018	-2.725	-5.189	-6.696	1.462	2.999
	(5.205)	(5.219)	(5.714)	(9.221)	(6.176)	(4.713)
Sh. Secondary or tertiary educ	-2.094**	-1.605	-0.259	-2.029	-3.105***	-0.580
	(1.011)	(1.029)	(0.994)	(1.985)	(1.126)	(1.106)
Share Migrant Pop.	-0.639	-2.321**	-0.187	-3.550**	-0.118	0.0618
	(0.824)	(0.952)	(0.781)	(1.698)	(0.919)	(0.960)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	265	265	265	254	265	265
R^2	0.838	0.827	0.843	0.625	0.866	0.817

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.

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

Table 12: Home Bias: Determinants - by Industry (cont.)

	Chem	Non-Metal	Metal	Mach.	Vehicles	Other
	(1)	(2)	(3)	(4)	(5)	(6)
	Home	Home	Home	Home	Home	Home
Distance	0.478*	0.0678	0.509**	-0.287	-0.300	-0.0701
	(0.264)	(0.214)	(0.243)	(0.201)	(0.361)	(0.317)
Log(European Remoteness)	2.130***	1.176**	1.904***	2.029***	3.167***	2.991***
	(0.554)	(0.464)	(0.522)	(0.508)	(0.754)	(0.859)
Island Region	2.548***	0.264	2.379***	1.238***	0.946*	2.164***
	(0.665)	(0.377)	(0.753)	(0.264)	(0.560)	(0.691)
Mountain Region	0.226	0.129	0.0782	0.260***	0.180	0.260**
	(0.140)	(0.0928)	(0.141)	(0.0918)	(0.139)	(0.109)
Major Port Region	-0.197	-0.0357	-0.0174	-0.157	0.0162	-0.0142
	(0.150)	(0.105)	(0.134)	(0.147)	(0.210)	(0.216)
Motorway Density	-5.406***	1.304	-5.518***	-8.651***	-10.75***	-7.042***
	(2.009)	(1.664)	(1.849)	(1.732)	(3.232)	(2.164)
Log(Population)	-0.901***	-0.838***	-0.791***	-0.785***	-0.887***	-0.599***
	(0.102)	(0.0715)	(0.0990)	(0.0689)	(0.111)	(0.111)
Share of Emp. (Manuf.)	-6.680***	-5.895***	-11.43***	-11.76***	-13.94***	-5.466**
	(2.253)	(1.923)	(2.171)	(2.127)	(2.820)	(2.679)
Share of Emp. (Public)	-9.292*	-2.286	-4.960	-6.169	0.836	-9.645
	(5.606)	(4.283)	(5.250)	(4.832)	(6.817)	(6.522)
Sh. Secondary or tertiary educ	0.467	-1.795*	-0.675	-1.950**	-1.659	-1.912
	(1.064)	(0.966)	(1.021)	(0.927)	(1.364)	(1.364)
Share Migrant Pop.	2.693**	-1.220	2.207**	-2.229**	-0.767	-0.312
	(1.283)	(0.963)	(1.066)	(0.947)	(1.608)	(1.055)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	264	265	264	263	261	258
R^2	0.836	0.852	0.821	0.854	0.809	0.787

Standard errors in parentheses. Distance is measured as the (log) of average kilometers travelled.

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