

Agglomeration Matters for Trade*

Roberto Ramos
CEMFI

Enrique Moral-Benito
Bank of Spain

[JOB MARKET PAPER]

November 2012

Abstract

We use a unique administrative dataset of Spanish exporters to document the existence of exporters' geographical agglomeration by export-destination. We reveal that firms selling to countries with worse business regulations, dissimilar language and different currency tend to cluster significantly more. We then assess the implications of exporters' geographical agglomeration for firms' behaviour and for the estimated welfare gains from trade. On the one hand, we find that exporters engage in more stable trade relationships with those countries that are the export-destinations of nearby firms. On the other, we introduce agglomeration in a model of international trade à la [Melitz \(2003\)](#). Using our Spanish firm-level data, we find that relative to a model without agglomeration, accounting for this phenomenon increases by 44% the elasticity of welfare with respect to fixed trade-costs.

*We are very grateful to Guillermo Caruana, Rosario Crinó and Claudio Michelacci for their constant guidance and help. We are also thankful to Pol Antràs, Manuel Arellano, Stéphane Bonhomme, David Dorn, Manuel García-Santana, Horacio Larreguy, Marc Melitz, Diego Puga, Rafael Repullo, Rubén Segura-Cayuela, Andrei Shleifer and seminar participants at CEMFI and the European Winter Meeting of the Econometric Society in Konstanz for invaluable comments and useful discussions. We are also grateful to Patry Tello for providing us with the exporters' data. Roberto acknowledges the financial support of the Spanish Ministry of Education under research grant BES-2009-026803.

Contact: CEMFI. Casado del Alisal 5. 28014. Madrid, Spain. roberto.ramos@cemfi.es.

1 Introduction

Firms do not operate in a void. On the contrary, a successful body of literature has shown that firms tend to be geographically concentrated.¹ At least since the early work by Marshall (1920), the concentration of industries in dense areas have attracted much attention from economists, geographers and policy-makers alike. They have built up a large body of research on the sources and effects of these agglomeration economies.²

In this paper, we analyze one type of agglomeration economies that is attracting increasing attention by the literature. This is the agglomeration economies that accrue to export firms. Firms selling abroad incur additional costs in accessing foreign markets and use infrastructure facilities -such as seaports- more intensively than domestic firms. Therefore, the forces driving the agglomeration economies affect differentially these firms and might generate specific concentration patterns.

We use a unique dataset of Spanish firms selling abroad in order to provide evidence on the existence of agglomeration economies that are specific to export firms. We then assess, both empirically and theoretically, the importance of this phenomenon. We provide empirical evidence on the association between exporters' concentration and different micro-level patterns of international trade. From a theoretical perspective, we show that exporters' agglomeration has sizable quantitative effects on the estimation of the welfare gains from trade.

In our analysis, we focus on the different patterns of concentration of exporters across destinations. Therefore, we focus on the diverse location patterns within export firms, rather than differences between export and non-export firms. This allows us to minimize the effect of unobservables that differ between domestic firms and firms selling abroad. We make use of the advantages of our dataset, that provides us with a panel on the universe of Spanish export firms selling above a threshold, with detailed information on location -zip code-, industry and export-destinations. We merge this information with balance sheet data to account for firm characteristics. The source of both datasets is the Bank of Spain.

We uncover the fact that exporters appear spatially concentrated regarding the export-destinations they sell to.³ For this purpose, we adapt the methodology in Duranton and Overman (2005) orig-

¹In, for instance, the US -Ellison and Glaeser (1997)-, France -Maurel and Sédillot (1999) and Devereux et al. (2004)- and the UK -Duranton and Overman (2005)-.

²Some sources of agglomeration economies arise from the benefits of specialization provided by larger markets, sharing indivisible facilities, input suppliers and transport costs, improving the matching quality in the inputs market and speeding up the flow of information and ideas. See Duranton and Puga (2004) for a detailed discussion on the theoretical underpinnings of urban agglomeration economies and Rosenthal and Strange (2004) for empirical evidence.

³Although localization can be defined as agglomeration controlling for that of general manufacturing -for instance, in Duranton and Overman (2005)-, in this paper we use the words agglomeration, localization and concentration interchangeably, as the indices explicitly control for general agglomeration and do not lead to confusion.

inally developed to assess industry agglomeration in the United Kingdom. We show that, for a large number of destinations, exporters are geographically too close to be the result of a random outcome. Crucially, we demonstrate that this is not due to the fact that some countries demand intensively goods from industries that are heavily concentrated.

We observe that the *extent* of agglomeration varies significantly across destinations. For some countries, exporters appear strongly concentrated, whereas for others they are geographically dispersed. We explore the market characteristics that explain these different patterns of concentration. We find that exporters to countries with a worse business environment, dissimilar language and different currency tend to cluster significantly more. We interpret this result as suggestive evidence that harder destinations are more accessible by a cluster of exporters, rather than export firms operating in solitude.

Having established the existence of exporters-specific agglomeration economies, we assess the importance of this phenomenon from an international trade perspective. We construct firm-specific measures of exporters' concentration in order to capture the extent to which exporters are surrounded by other firms serving the same destinations. We then explore how these measures are associated to firms' outcomes and decisions in foreign markets. In particular, we analyse the extent to which exporters' agglomeration is associated to the stability of trade relationships, as well as the entrance and exit from foreign markets. Our results show that exporters surrounded by other exporters sharing the same destinations are more likely to engage in stable trade relationships and exit less from these destinations. Moreover, they are more likely to enter those destinations that are served by firms around the exporter and keep selling once in.

Finally, from a theoretical point of view, we evaluate the quantitative effects that accounting for agglomeration has on estimating the welfare gains from trade. We build a simple extension of a calibrated version of the model by Melitz (2003) by introducing agglomeration -proxied by the number of export firms- as an externality that reduces the fixed export-costs. We solve the model and compute the elasticity of welfare with respect to trade costs, both fixed and variable. We show that this elasticity is increasing in the capacity of agglomeration to drive down the foreign entry costs. We estimate the value of the parameter driving the effect of agglomeration using our Spanish data and find that it can have substantial effects on welfare. Specifically, we find that the fixed and variable trade costs elasticity of welfare increase by 44 and 7 percent, respectively, when we include agglomeration.

The rest of the paper is organized as follows. The next sub section embeds this paper in the existing literature, emphasizing our contributions. Section 2 describes the dataset. Section 3 documents the existence of exporters-specific agglomeration economies and provides an explanation for the factors behind the different degrees of concentration. Section 4 explores the association between agglomeration and the patterns of trade flows at the firm level. Section 5 introduces agglomeration in a standard model of international trade and analyzes the effects on welfare.

Section 6 gives concluding remarks.

1.1 Related Literature

From a broad point of view, we embed this paper in two literatures, for which we establish a link: the literature on measuring the extent of agglomeration economies and the literature of international trade from a micro-level perspective.

As already mentioned, the agglomeration of firms in concentrated areas has attracted the attention of a large body of research over the years. The influential work by Ellison and Glaeser (1997) was the first to derive an index of establishments' geographical concentration from a model of natural advantages, industry spillovers and random chance. Their application to the US manufacturing industry showed that somewhat almost all industries were localized. Taking similar approaches, Maurel and Sédillot (1999) and Devereux et al. (2004) found high degrees of spatial concentration in some industries in France and the UK, respectively. Lately, the paper by Duranton and Overman (2005) improved on previous methods of calculating industry concentration. Their work found significant localization in more than half of the industries in the UK.

We turn the focal point from industry concentration to exporters' concentration, and stress its variation across different foreign markets. As previously argued, some forces driving agglomeration are exporters' specific and hence generate specific patterns of exporters' localization, which vary by destination-country. We test whether the observed proximity of exporters selling to the same market is too high to be the result of a random outcome. To do so, we rely on the methodology developed by Duranton and Overman (2005), used to test for proximity of firms belonging to the same industry. This methodology is very demanding in terms of data. It requires to know the precise location of every firm, as it estimates the distribution of geographical bilateral distances of all firms belonging to the same group -the same industry in the case of Duranton and Overman (2005), the same destination in our case-. This distribution is compared to random distributions, generated by Monte Carlo analysis.

This framework allows us to control for the industry structure of exports to each country. This tackles an essential concern: countries do not demand the same goods, therefore, exporters' concentration may simply reflect the concentration of production of goods that are heavily demanded by certain countries. We show that this is not the case. Hence, we are able to document the existence of exporters-specific agglomeration economies that go beyond industry agglomeration.⁴⁵

We emphasize that understanding the agglomeration patterns of exporters may help to explain

⁴In parallel work, Casey and Schmeiser (2010) use Russian data to study agglomeration of exporters by destination. They do not observe the precise location of the firm, though, hence they allocate them across 89 regions. They also do not control for industry agglomeration hence both forces are not disentangled.

⁵For the sake of completeness, we also show in Appendix B results on the excess concentration of exporters with respect to all production units -both domestic and exporters- operating in the same industry. We find that export firms exhibit significant excess concentration in around half of Spanish manufacturing industries.

some patterns of international trade. There is a large and increasing literature on export spillovers -see, for instance, [Aitken et al. \(1997\)](#), [Greenaway et al. \(2004\)](#) and [Barrios et al. \(2003\)](#)-. Some papers study how the decisions of export firms are influenced by the presence of nearby exporters. For instance, [Koenig \(2009\)](#) and [Koenig et al. \(2010\)](#) use data from French Customs to analyse how the probability of start exporting and exported volumes is affected by the presence of exporters of the same product and/or selling to the same destination in the same employment area.⁶ [Bernard and Jensen \(2004\)](#) also address the issue of spillovers in a panel of US manufacturing firms by including the presence of exporters in the state or industry.

We adopt a similar approach to the previous papers in stressing the association between agglomeration and the patterns of international trade. However, we have a main advantage. We compute agglomeration for every firm as the presence of nearby exporters within a certain radius, regardless of the administrative areas they belong to. That is, we do not need to allocate firms in 'boxes', either employment areas or states. This avoids the problem that bordering areas are treated the same as ones that are further apart. Moreover, it does not involve an arbitrary choice on the level of disaggregation of the administrative area to consider. The work by [Duranton and Overman \(2005\)](#) is precisely aimed to solve these challenges. We come back to this question later on.

A common limitation of the export spillovers literature is that the precise channel through which the spillovers operate is left unexplained. Moreover, endogeneity issues are severe in this context. We limit ourselves to point out the correlation between firm-specific measures of agglomeration and some micro-level patterns of international trade, in order to suggest the importance of accounting for this phenomenon. We see an avenue for further research to provide evidence -perhaps through natural experiments and/or case-studies- on the specific functioning of spillovers, with identification through exogenous shocks. Agglomeration economies may partly explain why firms do not enter foreign markets according to a perfect hierarchy, a fact uncovered by [Eaton et al. \(2011\)](#) and also present in our data.

From a theoretical point of view, we build a simple extension of a calibrated version of the influential model by [Melitz \(2003\)](#) in order to explore the quantitative effects of agglomeration. A broad literature on international trade has shown, both theoretically and empirically, the importance of trade costs to account for the extensive and intensive margins of trade flows.⁷ We assume that fixed trade-costs are reduced whenever new firms enter the foreign market. In this regard, we are close to [Krautheim \(2012\)](#), who introduce agglomeration as an externality that reduces the fixed costs of both producing and exporting. We introduce the externality only in the export market, as we do not have evidence of spillovers working in the domestic market.

⁶Endogeneity issues arising from reverse causality and simultaneity are addressed by lagging the right-hand side variables one period.

⁷See, for instance, [Roberts and Tybout \(1997\)](#), [Eaton and Kortum \(2002\)](#), [Helpman et al. \(2008\)](#), [Chaney \(2008\)](#) and [Morales et al. \(2011\)](#)

There are other papers that model the influence of other exporters on firms' outcomes. In [Casey and Schmeiser \(2010\)](#) agglomeration arises because firms achieve economies of scale in variable costs when serving the same destination. From a learning perspective, [Segura-Cayuela and Vilarrubia \(2008\)](#) and [Wagner and Zahler \(2011\)](#) point out that the success or failure of current exporters reveals information to potential entrants, who benefit from the observed outcomes of pioneers. [Krugman \(1991\)](#) and [Ottaviano et al. \(2002\)](#) explain the emergence of agglomeration economies in an otherwise homogeneous space.

2 Data

Spain is a big developed country located in Southern Europe, very suitable to examine the geographical distribution of export firms. In 2007, it was the 13th largest world exporter, selling virtually to all countries in the World. Its geographical location is privileged and favours international trade. It shares more than 600 km of border with France, the 6th largest world economy, and 1,200 km. with Portugal, a nation belonging to the European Union. It also borders with two small countries, Andorra and Gibraltar (UK), located in the Iberian Peninsula. Mainland Spain is also very close -14 km.- to Morocco, in North Africa, and has a coastline of almost 5,000 km, bordering the Mediterranean Sea and the North Atlantic Ocean. This favour the activity of major seaports, such as Valencia, the 5th busiest container port in Europe. Economic activity is concentrated in several parts of the country, specially the east -Catalonia and Valencia-, the center -Madrid- and the south -Andalusia-.

We exploit a unique administrative database of Spanish exporters for the years 2003 to 2011. This dataset, provided by the Bank of Spain, contains the micro data information used to construct the official Spanish Balance of Payments Statistics. We observe firms making transactions with foreign agents if they are worth more than 12,000 euros.⁸ The dataset accounts for around 97 per cent of aggregate Spanish exports.⁹ For each exporter, we observe the fiscal id, localization -zip code-, industry and value of exports to each country.¹⁰

The dataset has several virtues that make it specially suitable to analyze the location patterns of exporters. First, it is administrative data, therefore very reliable and with a wide coverage. Second, it provides the zip code location of every exporter, then distances between firms can be calculated with a fairly degree of precision. And third, it provides the fiscal id of each firm. Then,

⁸In 2008, this threshold was increased to 50,000 euros.

⁹The official body in charge of recording firm transactions with the rest of the world is the Spanish Customs Office, who collects these data from the information provided by the exporters, regardless of the value of the foreign transactions. The Customs Office treats these data as confidential. The declaration to Balance of Payments is more simplified one, although it well approximates total exports, both in the intensive and extensive margins.

¹⁰The industry of the firm corresponds to the four-digit 2009 National Classification of Economic Activities (CNAE-2009) which corresponds very closely to the Statistical Classification of Economic Activities in the European Community (NACE Rev. 2)

the dataset can be merged with balance sheet information to obtain firm characteristics for a high number of exporters.

We obtain balance sheet information of exporters from a database called Central Balance Sheet Data, also provided by the Bank of Spain. It contains the balance sheets of a representative sample of non-financial companies in Spain -around 800,000 each year- for the period 2002-2010, with information on fiscal id, industry, location, employment, value added and sales. In Appendix B.1 we explain more in detail this database.

Table I displays a radiography of our data for the year 2007. Panel A shows statistics corresponding to the Balance of Payments Data -exporters-, whereas Panel B shows statistics on the Central Balance Sheet Data -all firms-. We restrict our analysis to manufacturing firms. Column (1) of Panel A shows that we observe 18,553 exporters in 2007 distributed in 3,195 zip codes -out of near 11,000 zip codes in Spain- and exporting to 204 destinations. Columns (2) to (7) reports statistics across industries, zip codes and firms of the variables listed in the rows.

Zip codes that host at least one export firm contain on average 5.8 firms selling abroad (median is 2 exporters). The median exporter sells to 2 countries. There are some firms exporting to many countries, driving the average number of markets served to 5. In the Central Balance Sheet Data we observe 73,351 manufacturing firms, located in nearly half of the zip codes in Spain. The median zip code in which there is at least one manufacturing firm hosts 4 firms and 3 industries -means are 13 and 8, respectively-.

3 Localization of Exporters Across Export Destinations

In this section, we provide evidence that exporters tend to be localized by export-destination. In order to account for this fact, we rely on the methodology developed by [Duranton and Overman \(2005\)](#), who study localization of industries in the United Kingdom. This methodology improved on previous methods applied to measure industry geographic concentration, such as [Ellison and Glaeser \(1997\)](#), [Maurel and Sédillot \(1999\)](#) and [Devereux et al. \(2004\)](#). With respect to these works, it has two main advantages. First, it overcomes the 'border effect' problem, that is, the allocation of firms in spatial units -such as counties or states-, by computing the geographical distances between all firms in each industry. Hence, there is no need to make arbitrary choices of spatial units in order assign firms. Second, it is able to provide with a measure of *significant* deviation from randomness. This is done by Monte Carlo analysis.¹¹

¹¹The allocation of firms to spatial units involves several problems. First, it amounts to treat symmetrically plants not belonging to the same spatial unit, regardless of the distance between the spatial units. Second, it involves the arbitrary decision of which spatial unit to take. This is relevant, as different levels of aggregation can lead to very different results. Furthermore, it has been showed that bigger units produce correlations more pronounced. This is called the Modifiable Areal Unit Problem (MAUP) -see [Openshaw and Taylor \(1979\)](#) and [Openshaw \(1984\)](#)-. And third, the previous problem and the fact that spatial units are often not defined due to

We adapt this methodology in order to measure exporters' concentration. The approach consists of estimating the distribution of geographical bilateral distances within firms serving each country and compare this distribution with the one that would be expected if destinations were allocated randomly across exporters. As long as the former significantly deviates from the latter, exporters to a certain country are said to be localized.

Importantly, our aim is to measure exporters' concentration beyond industry concentration. This is a relevant concern. Countries do not demand the same goods. Then, it is possible that exporters to a certain country are localized because they belong to a highly-localized industry that is intensively exported to the country. In this case, exporters' concentration would simply reflect industry concentration. We control for this issue by taking into account the industry structure of Spanish exports to each country in constructing the random distributions. That is, exporters to a country will appear localized if they are significantly more localized than other exporters operating in the industries that are exported to that country. That is, in constructing the random distributions, we do not take into account exporters that operate in industries that are not exported to the country. We explain this issue in detail below.

We proceed in several steps. Our benchmark year is 2007. We consider those destination-countries with at least 10 exporters. These are a total of 164 countries.¹² For each country, we identify those exporters selling to it and compute all the possible bilateral distances between those firms. To do so, we use each firm zip code coordinates and apply the haversine formula to them.¹³

Next, for each country, we estimate the distribution of bilateral distances of all firms selling to it. We use a Gaussian kernel density estimation. We choose the bandwidth so as to minimize the mean integrated squared error. Distances are reflected around zero, following the method proposed by [Silverman \(1986\)](#)- in order to avoid giving positive densities to negative distances.¹⁴

The kernel density estimation for country c at every kilometer d - $\hat{K}_c(d)$ - reads as follows:

$$\hat{K}_c(d) = \frac{2}{n_c(n_c - 1)h} \sum_{i=1}^{n_c-1} \sum_{j=i+1}^{n_c} f\left(\frac{d - d_{i,j}}{h}\right) \quad (1)$$

where n_c is the number of firms exporting to country c , h is the bandwidth and f is the economic significance make difficult the comparison of results across spatial units. See [Duranton and Overman \(2005\)](#) for a discussion

¹²Appendix Table A.2 lists all the countries, number of exporters and value of the baseline localization index, explained later.

¹³The haversine formula calculates the great-circle distance between two points, i.e. the shortest distance over the earth's surface. Contrary to the Euclidean distance, it takes into account the curvature of the earth, as opposed to measuring distance through the earth's interior.

¹⁴The number of bilateral distances for N firms exporting to a country is given by: $\frac{N(N-1)}{2}$. Therefore, it grows exponentially with the number of exporters. As we explain later, the Monte Carlo simulation is computationally intensive. We then follow the simplification suggested by [Ellison et al. \(2010\)](#). This means that we randomly draw without replacement 2000 firms for those countries with more than 2000 exporters -eight countries-. We then calculate the bilateral distances based on this sample.

Gaussian pdf function.

Figure I shows these kernel-estimates for two countries, Syria -Panel A- and Iceland -Panel B-. There are roughly the same number of Spanish firms exporting to each country -161 and 159, respectively-. Note that the density cumulated at short distances in the Syrian distribution doubles that of Iceland. This reflects the fact that there are more exporters to Syria geographically located within short distances than exporters to Iceland. That is, exporters to Syria appear more agglomerated. In fact, 20 per cent of firm-pairs exporting to Syria are within 50 km. whereas this figure is just 11 per cent for exporters to Iceland.

The distribution of bilateral distances of exporters to Syria has a second peak at around 250 km. This reveals the existence of two -or more- clusters of exporters to Syria, separated by 250 km. In this case, these are the exporters located in two Spanish provinces, Barcelona and Castellón, accounting for 57 per cent of exporters to Syria. The distribution of distances of exporters to Iceland is much smoother, showing that firms selling to this country are scattered across the Spanish land. In fact, the peak of the distribution is roughly 400 km. which corresponds to the median distance within manufacturing firms in Spain.

Next, we compare, for each country, this estimated distribution with a counterfactual. This counterfactual represents a set of distances distributions within exporters if this destination were allocated randomly between exporters. We refer to this set of counterfactual distributions as the random distributions. As long as the estimated density of exporters to a country deviates from the random distributions, we say that exporters to a country are significantly localized. This is relevant, because even if locations were chosen randomly, firms would exhibit a certain degree of concentration. This is so because some locations, such as lakes, cannot host manufacturing firms and others, such as populated regions, will attract more. Then, a uniform spatial distribution of manufacturing firms is not the correct benchmark to account for localization, but one that takes into account the issues mentioned above.

The dartboard metaphor introduced by [Ellison and Glaeser \(1997\)](#) is very appropriate. Significant geographic concentration is found whenever the actual spatial distribution of firms deviates from the spatial distribution that would arise if locations were chosen randomly, as if throwing darts to a board. However, these darts are not thrown purely randomly. Say, they cannot land in a lake and they will be more likely to hit a populated region.

In measuring the concentration of exporters, we face an additional concern. It can be the case that a particular foreign country heavily demands goods from an industry that is highly localized. Then, exporters to that country operating in that industry will appear agglomerated. However, it will be not clear whether this localization stems from exporters' localization or from industry localization. To account for this issue, we control for the industry composition of exports to each country. That is, darts can only go to exporters in industries in which at least one exporter is selling to the country. Moreover, the higher the share of industry exporters in total exporters to

the country, the higher the likelihood that darts will hit exporters in this industry. Then, we fully controls for the industry composition of exports to the country, and then we measure exporters' concentration beyond industry concentration.

We next explain how we construct the random distributions, following the idea proposed by [Duranton and Overman \(2005\)](#). Given the random nature of the random distributions, in order to construct confidence intervals, we construct for each destination-country 1,000 random distributions. For each destination-country, we keep only those Spanish exporters that operate in industries in which at least one exporter is actually selling to the country.¹⁵

Then, from this sample of exporters, we draw 1,000 sub-samples of size n_c -the actual number of exporters to the country-.¹⁶ We draw every sub-sample in order to generate the observed industry composition of exporters to the country. That is, if, say, 30 per cent of exporters to a country belongs to the textile industry, and 70 per cent to the automobile industry, every sub-sample is composed of 30 per cent of exporters producing in the textile industry and 70 per cent in the automobile industry.¹⁷ Therefore, our localization measure compares the spatial concentration of exporters to a country with the spatial concentration of exporters in the industries exported to the country, taking into account the observed country-specific industry composition of exporters.¹⁸

We focus on localization at short distances. Our benchmark is localization within 100 km. For each destination, we test whether there are more exporters within 100 km. that the ones that would be expected under randomness.¹⁹

We construct two tests -one of localization and one of dispersion-, both with a significance level of 95 per cent. The former tests whether exporters exhibit significant excess agglomeration at short distances, whereas the latter tests whether they exhibit significant dispersion. Figure II is a zoom in of Figure I displaying both tests. The test of localization is represented by the upper dashed line, -we refer to it as the localization threshold, denoted by $\overline{\overline{K}}_c(d)$ -, whereas the test of

¹⁵Industries are defined at the 2-digit level. Results are robust to considering 3 or 4-digit levels.

¹⁶If there is, for instance, two exporters in a zip code, this zip code is counted twice, and so on. Each sub-sample is drawn without replacement.

¹⁷In other words, we assume that only those sites in which there is an exporter operating in an industry in which at least one firm sells to the country could host a firm that could potentially sells to the country. Note that this assumption is more stringent than simply controlling for the overall distribution of Spanish exporters, regardless of the industries in which they operate. In his study of industry localization, [Duranton and Overman \(2005\)](#) make the usual assumption that any site hosting a manufacturing firm could host a manufacturing firm of either industry.

¹⁸An alternative is to construct the counterfactual considering producers -both exporters and non-exporters in the industries exported to the country. That is, to measure exporters' localization beyond localization in production. If we follow this approach, using the Central Balance Sheet Data, we observe higher degrees of localization.

¹⁹[Duranton and Overman \(2005\)](#) focus on 180 km., which corresponds roughly to the median distance of manufacturing plants in the United Kingdom. [Ellison et al. \(2010\)](#) provide results for the US on several thresholds, ranging from 1,600 to 160 km. We take 100 km. as benchmark because we are interested in agglomeration at short distances, which appears more relevant to explain interaction between exporters. Results do not vary significantly if we consider wider localization thresholds, such as 200 km. or 400 km. -roughly the median distance-. The quantitative index of localization at 100 km. has a correlation of .99 and .91 with the formers, respectively.

dispersion is the lower dashed line -dispersion threshold, $\underline{K}_c(d)$ -.

The localization threshold for a given country is constructed as follows. For each kilometer, we rank our 1,000 random distributions in ascending order. Then, we choose the percentile -the same across all distances- that makes 95 per cent of the random distribution lie below it across all distances. That is, for instance, the upper dashed line for Syria in Figure II represents the 98.4 percentile of the 1,000 simulations at every kilometer. This percentile is hit by 50 -5 per cent- simulations in at least one kilometer, whereas 95 per cent of the simulations lie below that level in all kilometers.²⁰

Then, firms exporting to a country are said to be localized in an specific kilometer if in that kilometer the distance distribution is higher than the localization threshold. In general, we say that exporters to a country are localized if they are localized in at least one kilometer. Graphically, this is so when the actual distance distribution lies above the upper-dashed line. For instance, in Figure II Syria exhibits localization from kilometers 0 to around 50. Iceland, on the other hand, exhibits no localization as the estimated distribution always lays below the localization threshold.

The dispersion threshold is calculated in a similar fashion. It is constructed with the percentile that makes 5 per cent of the simulations lie below it across all distances. Note, however, that densities must sum up to one. Then, localization at some distances may imply dispersion at other distances. Then, exporters to a country are said to be dispersed in a kilometer if the distance distribution in that kilometer is lower than the dispersion threshold and the country is not localized in any kilometer. In general, exporters to a country are said to be dispersed if they they exhibit dispersion in at least one kilometer and they are not localized. Graphically, this is so when the distance distribution lies below the lower-dashed line and never lies above the upper-dashed line. In Figure II, Syria exhibits localization, then it cannot exhibit dispersion. Iceland does not exhibit localization, but its estimated distribution is always above the dispersion thresholds, hence it does not exhibit dispersion either.

Of the 164 countries in our sample, firms exporting to 98 -60 percent- are localized whereas exporters to just 1 country -Portugal- are dispersed. Figure III shows the percentage of countries exhibiting localization at each level of distance. Exporters to 52 per cent of destinations show significant excess agglomeration at very short distances -below 40 kilometers-. This share tends to decline from around 60 km. on. Then, we document the fact that exporters to certain countries are agglomerated more than the the agglomeration of the industries they belong to. Exporters' agglomeration is pervasive across destinations and is concentrated at short distances, declining from medium distances on.

Following [Duranton and Overman \(2005\)](#), we define an index of localization to grasp an idea on *how much* exporters to each country are geographically concentrated, that is, how much do they

²⁰All percentiles that fulfills this criterion lie between the 96.5 and 99.3. In some cases, though, it is not possible to find a percentile making exactly 95 per cent of simulations lie below it. We use linear interpolation when that occurs.

deviate from randomness. This index is defined as the sum across all distances of the difference between the density of the estimated distribution and the localization threshold if this difference is positive and zero otherwise. That is:

$$\Gamma_c \equiv \sum_{d=0}^{100} \max\left(\hat{K}_c(d) - \overline{\overline{K}}_c(d), 0\right) \quad (2)$$

Graphically, it is the area between the estimated distance distribution and the localization threshold, when the former is above the latter. In Figure II, the index for Syria corresponds to the shaded area. Figure IV shows the sum across countries of the difference of the estimated distance distribution and the localization threshold for those countries that exhibit localization at each distance. As we can see, most of the amount of localization takes place at distances below a few kilometers.

Table II shows the countries to which exporters appear most localized. Interestingly, the top three countries according to this index corresponds to West Bank and Gaza, Iraq and Albania. These three countries seem relatively hard to access for Spanish firms. Also, exporters to a small country -Andorra-, with which Spain shares a border, appear highly agglomerated. In the next sub section, we analyse the patterns of exporters' concentration across destinations.

3.1 Factors Behind Agglomeration by Destination

In this section, we explore the factors that explain why exporters to some countries tend to concentrate more than others firms exporting to other countries. We start with a graphical representation. Figure V depicts the spatial distribution of exporters to four countries: Syria, Iceland, New Zealand and Cuba. Every polygon corresponds to a Spanish municipality, then highlighted polygons correspond to municipalities in which there is at least one exporter to the country.

Panels A and B show the spatial distribution of Spanish exporters to Iceland and Syria, respectively. As mentioned before, there are roughly the same number of firms selling to each country -159 and 161, respectively-. However, exporters to Syria appear more spatially concentrated. One possible explanation of this pattern is the difference in business regulations of both countries. Whereas it is relatively easy to deal with the regulation in Iceland, it is much more difficult to do so in Syria. For instance, in terms of the Regulatory Quality Index provided by the World Bank, Iceland ranks the 17th in the World, whereas Syria ranks the 188th. Then, firms would cluster to better deal with an adverse business environment.

Consider now Panels C and D, that is, exporters to New Zealand and Cuba. Again, a similar number of exporters sell to those countries -272 and 312, respectively-. Exporters to each country appear rather scattered across Spain. It is difficult to say visually the exporters to which country are more concentrated. Actually, under the quantitative index of exporters' localization explained before, exporters to Cuba are not significantly concentrated whereas exporters to New Zealand are,

although slightly. Following the previous reasoning, exporters to Cuba should be more spatially concentrated, as the business environment in New Zealand is much better than the one in Cuba -in terms of the previous index, New Zealand and Cuba rank 8th and 200th, respectively-. However, the colonial links between Cuba and Spain, and the fact that both share the same language make exports to Cuba easier, at least in this dimension. This may compensate the worse business environment, decreasing the relative need of exporters being localized, suggesting that dealing with countries culturally more dissimilar is more difficult.

In order to investigate the plausibility of our interpretation of Figure V, we next conduct regression analysis. In principle, the concentration forces can be of very different nature: comparative advantage, export costs et cetera-. In our regression, we control for a wide range of these forces, in order to minimize a possible omitted variable bias.

Marshall (1920) emphasizes that the gains from industry spatial concentration come from a saving in transport costs. These can be for goods, people, or ideas. Firms tend to concentrate in industry clusters in order to be close to customers and suppliers, to access a large labour pool and to speed up the process of sharing innovations and ideas.

The theory of the gains in exporters' concentration regarding destinations is much less developed. Two main forces arise as candidates, which we focus on, comparative advantage and export costs. We include regressors that try to disentangle these forces in order to assess their relative importance. However, we acknowledge that there is not a clear-cut division between them. Actually, they are intertwined, as comparative advantage translates in a decrease in export costs.

Some firms have a comparative advantage in exporting to certain countries. Perhaps the most obvious example is location. Firms close to the border face lower transport costs with respect to inland firms. Therefore, they are more likely to export to border countries. Then, we expect concentration of exporters to countries they share the border with or are very close.

Exporters may also concentrate in order to save export costs. Some countries are more costly or more difficult to access than others for reasons such as language, currency, undeveloped institutions or cultural dissimilarity. Information costs are higher if the language is different. Transactions costs are higher if the importer's institutions do not provide with a friendly business environment or countries do not share the same currency. We hypothesize that the concentration of exporters may save some amount of these costs. An example is learning from foreign bureaucracy, that may be speeded up by several exporters serving the same country being close to each other.

Our dependent variable is the index of localization, Γ_c , explained above. It goes from zero -66 countries that do not exhibit localization- to .23 -West Bank and Gaza-. We normalize it, as well as some covariates, to have unit standard deviation. We control for distance as the main determinant of comparative advantage. We include a dummy variable taking value 1 for countries that share the border with Spain. We also include the log average distance of exporters to the

country’s capital.²¹

We include three regressors that proxy for exports costs. We include a dummy variable taking value 1 if Spanish is spoken in the destination country. We also include an index of the regulatory quality of the import country, proxying for how easy is conducting businesses there. We also include a dummy variable taking value 1 if the destination country does not use the same currency as Spain -the euro-. We finally control for additional characteristics that may affect localization. We control for per capita GDP and the log number of firms exporting to the country.

In order take into account the large number of zeros, we specify the following Tobit model:

$$\begin{aligned} \text{Localization}_c^* &= \beta_0 + \beta_1 \text{Spanish}_c + \beta_2 \text{Institutional Quality}_c + \beta_3 \text{Euro}_c \\ &+ \beta_4 \text{Contiguity}_c + \beta_5 \text{Log Distance to Capital}_c \\ &+ \beta_6 \text{Log Per Capita GDP}_c + \beta_7 \text{Log Number of Exporters}_c \\ &+ \beta_8 \text{Log Population}_c + \epsilon_c \end{aligned} \quad (3)$$

$$\text{Localization}_c = \max(0, \text{Localization}_c^*) \quad (4)$$

Column (1) of Table III shows the results of estimating Equation (3). Conditional on the rest of covariates, exporters to countries with higher export costs appear more agglomerated. Distance variables do not affect localization, either contiguity or distance to the destination’s capital. Column(2) restricts the sample to export-destinations outside the European Union. These countries are more difficult to access by Spanish firms. Language and the regulatory environment become more important in explaining localization. Column (3) replaces the average distance of exporters to the country’s capital by the average distance to the closest port shipping to the country. This distance measure does not have explanatory power in explaining exporters’ localization. In Column (4) we introduce an additional explanatory variable, the log number of immigrants from each importing country living in Spain. This proxy for further links to import countries, and hence may affect localization. Its coefficient enters negatively and significantly. In this specification, language becomes non-significant, as these two variables are positively correlated. Column (5) estimates a probit model for the probability of exhibiting localization. The dependent variable is whether Spanish firms exporting to the country exhibit positive localization. The coefficients report the marginal effects. Interestingly, if the Spanish is spoken in the destination-country, the probability of being localized decreases by 50 per cent. Also, a better regulatory environment reduces the likelihood of exporters to be localized, although the coefficient is quantitatively lower.

²¹This may be relevant for countries that are close to Spain, but probably not for countries that are further away. In a different specification, we include an average distance of exporters to the closest port from which shipments are sent to the country. This variable may better proxy for *how far* exporters are to each country rather than distance to the capital. We do not include it in the main specification, though, because we lack data on exports from Spanish ports for 12 countries, and the distance variables are never significant.

Finally, Column (6) estimates the probit model for countries outside the European Union. The coefficients of language and regulatory environments become somewhat lower but still statistically significant.²²

Summing up, we find evidence that exporters selling to countries with different language and a worse business environment appear to be more spatially concentrated. This would suggest that exporters tend to agglomerate more when the country of destination is more difficult to serve. Sharing costs or information about the know-how in exporting to certain markets might be at the root of this empirical pattern.

4 Exporters' Agglomeration and the Patterns of International Trade

In this section, we explore the implications that agglomeration economies have on the patterns of international trade. In particular, we analyse the effect of exporters' agglomeration on the stability of trade relationships, and the effect on the decisions of entering and exiting market destinations. We analyse these aspects through the lens of the export firm.

For each firm, year and destination, we calculate the share of exporters selling to the destination in a 100 km. radius around the firm. This share provides a firm-country-year specific quantitative index of spatial agglomeration in export markets. We use this measure to analyse the relationship between exporters' geographic concentration and the patterns of international trade.²³

4.1 The Stability of Trade Relationships

In a recent paper, [Békés and Muraközy \(2012\)](#) show that about one third of Hungarian firm-destination export spells are short-lived or temporary. Although the largest portion of aggregate exports is carried out in stable trade relationships, the fact that a large amount of firms enter and exit continuously export markets is relevant to explain the extensive margin of international trade and inform policy.

[Békés and Muraközy \(2012\)](#) show that the likelihood of temporary trade increases with the size, wealth and closeness to the importing country, as well as the productivity and financial stability of the exporting firm. However, one of the factors that may additionally explain the likelihood of

²²Appendix Table A.3 performs robustness checks on the previous specifications. We analyse the robustness of the results to considering different proxies for institutional quality, alternative definitions of localization at 200 and 400 kilometers and dropping those firms with the highest number of exporters, that are sub-sampled in the computation of the localization index. Results are in line with the baseline specification.

²³In this section, we make use of the full panel structure of our database over the years 2003-2011. We restrict ourselves to those firms selling each year more than 50,000 euros, as this is the threshold level imposed in 2008 onwards.

temporary trade is agglomeration economies. Other things equal, firms surrounded by a pool of exporters to a certain destination may find more easily to establish a permanent relationship with that particular destination.

We investigate this hypothesis using the comprehensive dataset of Spanish exporters described above. We compute a firm-specific quantitative variable of exporters' agglomeration. Then, we analyse how these agglomeration economies relate to the likelihood of permanent trade relationships. We acknowledge that causal statements are hard to make in this context, as causality runs in both directions and aggregate shocks might affect both the nature of the trade relationships and the subsequent clustering process. Having said that, we take as benchmark the year 2007, but compute clustering in 2003, in order to somewhat alleviate the effects of this reverse causality concerns.

We take the year 2007 as benchmark. In this year, we observe around 55,000 firm-country observations. We classify each of them as either a temporary or a permanent trade relationship following the approach suggested by [Békés and Muraközy \(2012\)](#). This consists of computing the number of consecutive years including 2007 in which there exists the firm-country trade relationship. If this number is sufficiently large (i.e. four or more years), the trade-relationship is classified as permanent; otherwise, the firm-country trade relationship is labeled as temporary.

More formally, let X_{ic}^{2007} be the exports of firm i to country c in 2007. For each firm-country pair $\{i, c\}$ with $X_{ic}^{2007} > 0$, we define S_{ic}^{2007} as the number of consecutive years, including 2007, in which firm i exported to country c . For instance, if $X_{ic}^{2005} = 0$, $X_{ic}^{2006} > 0$ and $X_{ic}^{2008} = 0$, then $S_{ic}^{2007} = 2$. We finally classify a trade-relationship as permanent if S_{ic}^{2007} is higher than 3 years. In our data, we find that 46 per cent of trade-relationships are temporary, accounting for 7 per cent of aggregate exports.

Note that for constructing the trade-relationship status in 2007, we need data from 2004 to 2010. This is so in order to observe the 3 years before and after our benchmark year and then be able to use the criterion of 4 years of continuous exporting to be considered a firm-country permanent relationship. However, in order to alleviate endogeneity concerns, our measure of agglomeration is computed in the year 2003, i.e. the year before the first observation of the dependent variable.

In order to explore the existence of a relationship between the nature of the trade-relationships and the the agglomeration economies surrounding each exporter, we run linear probability models as follows:

$$\begin{aligned} \text{Permanent}_{ic}^{2007} &= \beta_0 + \beta_1 \text{Agglomeration}_{ic}^{2003} + \beta_2 \text{Firm Controls}_i^{2003} \\ &+ \beta_3 \text{Destination Characteristics}_c^{2007} + \epsilon_{ic}. \end{aligned} \quad (5)$$

$\text{Permanent}_{ic}^{2007}$ is a dummy taking value 1 if the trade-relationship of firm i and country c is permanent and zero otherwise. $\text{Agglomeration}_{ic}^{2003}$ is the firm-specific measure of agglomeration in 2003, that is, the share of exporters in 2003 serving destination c in a radius of 100 kilometers

around the firm. Firm Controls include, depending on the specification, size, productivity, a dummy indicating whether the firm is a permanent exporter from 2004 to 2010, industry dummies and region dummies. Destination Characteristics include country-specific characteristics of the destination country such as population, per capita GDP, and distance with respect to Spain.

Table IV shows the estimates of these regressions. Column (1) excludes from the model the measure of agglomeration, in order to compare our results with those of Békés and Muraközy (2012). In line with their results, we also find that exports to bigger, richer and closer countries are associated with a higher likelihood of being permanent rather than temporary. When we include our firm-specific measure of agglomeration -Column (2)-, we observe that firms closer (in 2003) to other exporters selling to the same destinations are associated with trade relationships more permanent in nature. Quantitatively, a one standard deviation increase in the share of exporters selling to the country is associated with a 5 percentage points increase in the probability of the trade-relationship being permanent. The coefficient on exporters' agglomeration hardly changes when we include firm controls -Column(3)- such as dummies for medium and large firms, and the firm's productivity.²⁴ In Column (4) we restrict the sample to permanent exporters between 2004 and 2010, and find similar effects. In Column (5), we include firm-specific fixed effects to account for individual permanent heterogeneity in exporters. Interestingly enough, the coefficient associated to agglomeration is significantly higher, implying a large quantitative effect -an increase of 1 standard deviation in the agglomeration in 2003 is associated to a 8.6 percentage points increase in the probability of a permanent trade-relationship in 2007. Finally, in Column (6), we replace the share of exporters in a 100 km. radius by the share of exports in a 100 km. radius. The effect is also positive and significant. To sum up, we find evidence in favour of the hypothesis that firms exporting in clusters tend to engage in more permanent trade relationships.²⁵

4.2 Exit of Destinations

In this subsection, we analyse whether exporters operating close to other exporters sharing the same markets are less likely to exit those foreign destinations. To do so, we explore whether our firm-year-destination specific measure of agglomeration is associated to the probability of leaving specific export markets.

We proceed as follows. For all our firm-destination observations in the period 2003-2010, we define a variable taking value 1 if the firm exits this destination one year later and zero otherwise. In our sample, 36 per cent of firm-destination observations are exits the next year. We then estimate linear probability models of the following form:

²⁴Data on firm controls come from Central Balance Sheet Data of the Bank of Spain. We have information on size and productivity for around half of the firms in our original sample. This explains the lower number of observations in the columns including firm controls.

²⁵For the sake of completeness, Appendix Table A.4 reports the marginal effects of probit specifications similar to the previous regressions.

$$\begin{aligned} \text{Out}_{ic}^{t+1} = & \beta_0 + \beta_1 \text{Agglomeration}_{ic}^t + \beta_2 \text{Firm Characteristics}_i^t \\ & + \beta_3 \text{Destination Characteristics}_c^t + \text{Year Dummies}_t + \epsilon_{ic}^t \end{aligned} \quad (6)$$

Out_{ic}^{t+1} is a dummy variable taking value 1 if firm i exits country c in year $t + 1$, conditional on having exported to market c in year t . $\text{Agglomeration}_{ic}^t$ refers to the firm-country specific measures of agglomeration in year t , normalized to have unit standard deviation. Depending on the specification, these measures are two, namely, the share of exporters in year t serving destination c in a radius of 100 kilometers around the firm and the share of exports to country c in a 100 km. radius. Both are normalized to have unit standard deviation. Firm Characteristics includes size, productivity, industry and region of firm i . Destination Characteristics include population, per capita GDP and distance with respect to Spain of country c . Depending on the specification, we also include firm- or country-specific fixed effects. The estimation period is 2003-2010 and standard errors are clustered at the firm level.

Table V presents the results. In Column (1) we regress the dummy of exiting the market in $t + 1$ on the share of nearby exporters that are selling to the destination in t . We find a negative correlation between the two variables. A one standard deviation increase in the share of nearby exporters is associated to a 2 percentage points decrease in the probability of exiting the market next year. Also, find that firms are less likely to exit bigger, richer and closer countries. In Column (2) we add firm controls. The coefficient on agglomeration decreases to 3.5 percentage points and remains statistically significant. In Column (3) we focus on destination-countries outside the European Union. Estimates are very similar. In Column (4) we include firm fixed-effects in order to control for permanent heterogeneity across exporters. The effect of agglomeration is negative and significant, and larger in absolute value. In Column (5) we take a more stringent approach, and we include firm-destination fixed effects. Results do not change qualitatively. Finally, in Column (6) we replace the previous measure of agglomeration by the share of exports to the country in a 100 km. radius. We still find a negative and significant coefficient of agglomeration on the probability of exit, albeit the elasticity is lower. This suggests that the correlation of agglomeration with firm choices and outcomes seem to work in the extensive rather than the intensive margin. In sum, we find evidence through various specifications that exporters close to other exporters sharing the same destinations are less likely to exit these destinations.²⁶

4.3 Selling and Entrance to Destinations

In this subsection, we look at the relationship between the agglomeration economies in export markets and the firm's probability of selling to certain destinations. We take into account all the

²⁶Appendix Table A.5 report the marginal effects of probit specifications similar to equation (6).

firm-destination relationships that existed at least one year over the period 2003 to 2010. We then study how agglomeration affects the probability of entering a given market and keep selling there.

In our data, we observe 125,710 firm-destination relationships that existed at least once in the period 2004 to 2011. Of these, only 7 per cent existed during the whole period. The rest were active for some years but not for others. We construct a balanced database out of the 125,710 observations, expanding for the whole period 2004 to 2011 for all the firm-destinations that existed at least once during this period. We then create a dummy variable -which we called In_{ic}^t - that takes value 1 if firm i exported to country c in year t and zero otherwise.

We then consider the following empirical specification:

$$\begin{aligned} \text{In}_{ic}^t = & \beta_0 + \beta_1 \text{Agglomeration}_{ic}^{t-1} + \beta_2 \text{Firm Characteristics}_i^{t-1} \\ & + \beta_3 \text{Destination Characteristics}_c^{t-1} + \text{Year Dummies}_t \\ & + \text{Fixed Effects}_{ic} + \epsilon_{ic}^t. \end{aligned} \quad (7)$$

$\text{Agglomeration}_{ic}^{t-1}$ refers to measures of exporters' agglomeration, entering with a lag. We normalize them to have unit standard deviations. Firm Characteristics include the log number of employees and log productivity. Destination Characteristics include log population and log per capita GDP. We also include firm-destination fixed effects. Then, we look at the correlation within destinations between changes in the exporters' agglomeration with changes in the status of exporters. That is, we reach identification through the time dimension of our dataset. For this reason, those firms that sell to a country for the whole period provide no information for identification. Note also that we do not include those destinations that were not served by the firm in any year of our sample period. Our control group are those year-destinations on which the firm did not export to the country but this firm exported to this country at least once during 2004 to 2011.

Table VI shows the results of estimating equation (7). In Column (1) we regress the dummy of exporting to a market on the share of exporters that were selling to this market the year before. As mentioned, the inclusion of firm-destination fixed effects makes that identification only go through changes in the export status across years within a given exporter-country pair. Moreover, it controls for permanent characteristics of firms and countries that may explain changes in the export status and in the concentration of exporters. We find a significant correlation of our agglomeration variable and the fact of selling to the market one year later. A one standard deviation increase in the share of close exporters selling to the country is associated with a 5.3 percentage points increase in the probability of exporting to this destination.

In Column (2) we include firm characteristics, such as size and productivity. The coefficient of agglomeration hardly varies. The signs of firm and destination characteristics have the expected signs. Increases in the number of employees and in productivity, as well as increases in the size and wealth of the importing country are associated to a higher probability of entering the

market. In Column (3), we restrict the sample to countries outside the European Union. For this particular sample of destinations the effect of agglomeration quantitatively almost doubles, a one standard deviation increase in agglomeration is associated to a 9.8 percentage points increase in the likelihood of entering a market. In Column (4) we use an alternative measure of agglomeration: the log number of exporters serving the destination in a 100 km. radius. We find a positive and significant coefficient for this agglomeration measure.

Finally, in Columns (5) and (6) we restrict our attention to entrance to specific destinations. We run regressions in line with those of [Koenig \(2009\)](#).²⁷ In order to focus only on entrance to destinations, we drop those firm-destination observations in which the firm keeps or stops exporting. Therefore, the dependent variable takes value 1 if the firm exports in year t and not in $t-1$ and 0 if it does not export to the destination both in t and $t-1$. Then, we restrict our sample to firm-destinations that are zero and then are followed by a decision of exporting. In Column (5) the agglomeration variable is the lag share of exporters selling to the destination. It is positively and significantly correlated with entrance to the foreign market. In Column (6) the agglomeration variable is the log number of firms selling to the country the year before. We find a positive and significant coefficient. A 1 percent increase in the number of exporting selling to the destination is associated to a 5.4 percentage points increase in the probability of starting next year to sell to the market.

Overall, we find a strong correlation between firm-specific measures of agglomeration and the probability of selling to specific destinations. If these destination-specific agglomeration economies increase, the firm is more likely to start selling to this destination during the next year (or less likely to abandon the destination).

5 Agglomeration and the Estimated Welfare Gains from Trade

In this section, we perform a simple extension to a standard model of international trade in order to study the effects of introducing agglomeration in the estimated welfare gains from trade. Our world is that of [Melitz \(2003\)](#) with the distribution of productivity draws calibrated as Pareto. We consider a monopolistic competition model with two fully symmetric countries. We introduce agglomeration as an externality in the fixed costs of exporting. These costs are reduced whenever new firms enter the export market.²⁸

²⁷She looks at the effect of nearby exporters on the probability of entering a market. She models the decision of entering a foreign market in a year taking into account all the possible alternative destinations, regardless of the fact that the firm may actually never reaches them. Instead, we only consider those destinations that were served at least once by the firm.

²⁸See [Casey and Schmeiser \(2010\)](#) and [Krautheim \(2012\)](#) for alternatives in the introduction of agglomeration in a monopolistic competition model à la [Melitz \(2003\)](#). The former focuses on scale economies in variable trade costs

5.1 Setup

Consider a world of two fully symmetric countries, a domestic and a foreign country. The total population of each country is exogenous and is denoted by L . We model the world economy from the point of view of the domestic country.

5.1.1 Demand

Preferences are given by a Constant Elasticity of Substitution utility function over a continuum of goods, denoted by ω :

$$U = \left[\int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \quad (8)$$

with $\sigma > 1$. $q(\omega)$ represents consumption of variety ω . Ω represents the mass of available goods in a country, both of domestic and foreign production. σ is the elasticity of substitution between goods.

A consumer maximizes utility subject to the budget constraint:

$$\int_{\omega \in \Omega} p(\omega)q(\omega)d\omega = w \quad (9)$$

where $p(\omega)$ is the price of variety ω and w is the wage rate, hereafter normalized to one.²⁹

Aggregate demand of variety ω is given by:

$$q(\omega) = p(\omega)^{-\sigma} P^{\sigma-1} L \quad (10)$$

which is individual demand times total population L and

$$P = \left(\int_{\omega \in \Omega} p(\omega)^{-(\sigma-1)} d\omega \right)^{\frac{-1}{\sigma-1}} \quad (11)$$

is the aggregate price index.

5.1.2 Firms

Firms are heterogeneous in productivity, φ , and use only one factor of production, labour, that is supplied inelastically. Production labour used by the firm is composed by a fixed and a variable amount and depends on whether the firm exports to the foreign market or not. If the firm does not export, production labour used by the firm is:

$$l(\varphi) = f_d + \frac{q_d(\varphi)}{\varphi} \quad (12)$$

whereas the latter introduces agglomeration in fixed production-costs, both in the domestic and foreign markets.

²⁹Due to the symmetry assumption, foreign and domestic wages are the same.

f_d is a fixed cost of producing at home and $q_d(\varphi)$ is the production sold at home. The marginal cost of producing at home is $\frac{1}{\varphi}$, hence more productive firms -higher φ - produce more output with the same labour.

If the firm exports, production labour used by the firm is:

$$l(\varphi) = f_d + \frac{q_d(\varphi)}{\varphi} + f_x(M_x) + \frac{\tau q_x(\varphi)}{\varphi} \quad (13)$$

which is the production labour used in order to sell at home plus the production labour used to sell abroad. There is an iceberg cost τ of selling abroad, meaning that $\tau > 1$ units of the good must be produced for 1 unit of the good to arrive to the foreign market. $q_x(\varphi)$ is the quantity of the good sold in the foreign market.

$f_x(M_x)$ is a fixed cost of serving the foreign market. It accounts for such costs of complying with foreign regulation, labeling, establishing distribution networks, learning about the foreign market, et cetera. Critically, we depart from Melitz (2003) in assuming that these fixed costs depend on the number of exporters, rather than being an exogenous quantity. That is, we incorporate the idea that when more firms serve the foreign market, the fixed costs of exporting are reduced. Therefore, the fixed export costs becomes an endogenous variable. We assume that there is no strategic interaction, and firms take the number of exporters as given.

We assume the next functional form for the fixed export costs:

$$f_x = \bar{f}_x M_x^{-\theta} \quad (14)$$

Note that $f'(M_x) < 0$, therefore the higher the number of firms selling to the foreign market, the lower are the fixed costs. This captures the effects of agglomeration, that facilitates entrance to the foreign market. Parameter θ governs the elasticity of fixed costs with respect to the number of exporters. Then, the higher the θ , the higher the is effect on export costs of an increase in the number of exporters. Then, higher values of θ can be interpreted as more powerful agglomeration forces.³⁰

If θ is zero, then the fixed cost of exporting is \bar{f}_x , independent of the number of exporters, and then we are back to Melitz (2003). Therefore, this extension includes the fixed costs of the model by Melitz (2003) as a particular case. The cost function is always decreasing in the number of exporters, but less so the higher the number of exporters. That is, the cost function is decreasing and convex in the number of exporters.³¹

³⁰This is so when the number of exporters is higher than 1. If $M_x < 1$, then the higher the θ the higher the fixed costs for a given number of exporters. In the calibration exercise, we normalize the number of exporters to 1, by setting the size of the economy sufficiently large, as we explain later. See Section 5.5.

³¹For a sufficiently high number of exporters and/or θ , it can happen that the fixed costs of exporting goes below the fixed costs of selling at home. In the numerical exercise, we assume that that cannot happen, that is, we assume that an export firm always find it profitable to sell at home, as in Melitz (2003).

Conditional on selling to each market, profits of a firm are the sum of profits earned at home and profits earned in the foreign market:

$$\pi(\varphi) = \pi_d(\varphi) + \pi_x(\varphi) \quad (15)$$

where $\pi_d(\varphi)$ and $\pi_x(\varphi)$ stands for profits in the domestic and foreign markets, respectively. Their expressions are:

$$\begin{aligned} \pi_d(\varphi) &= p_d(\varphi)q_d(\varphi) - \frac{q_d(\varphi)}{\varphi} - f_d \\ \pi_x(\varphi) &= p_x(\varphi)q_x(\varphi) - \frac{\tau q_x(\varphi)}{\varphi} - f_x(M_x) \end{aligned} \quad (16)$$

where $p_d(\varphi)$ and $p_x(\varphi)$ are the price of the variety in the domestic and foreign markets, respectively.

Profit maximization implies:

$$p_d(\varphi) = \frac{\sigma}{\sigma-1} \frac{1}{\varphi}, \quad p_x(\varphi) = \frac{\sigma}{\sigma-1} \frac{\tau}{\varphi} \quad (17)$$

Using equations (10), (16), and (17), profits in the domestic and foreign markets can be expressed as:

$$\begin{aligned} \pi_d(\varphi) &= \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \right)^{-(\sigma-1)} \left(\frac{1}{\varphi} \right)^{-(\sigma-1)} P^{\sigma-1} L - f_d \\ \pi_x(\varphi) &= \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \right)^{-(\sigma-1)} \left(\frac{\tau}{\varphi} \right)^{-(\sigma-1)} P^{\sigma-1} L - f_x(M_x) \end{aligned} \quad (18)$$

Firms will sell in each market if they find it profitable to do so. As profits are strictly increasing in productivity, there is a threshold productivity level for selling in the domestic market:

$$\pi_d(\bar{\varphi}) = 0 \Rightarrow$$

$$\bar{\varphi} = \frac{\sigma}{\sigma-1} \sigma^{\frac{1}{\sigma-1}} L^{\frac{-1}{\sigma-1}} P^{-1} f_d^{\frac{1}{\sigma-1}} \quad (19)$$

Similarly, there is a threshold productivity level for selling in the foreign market:

$$\pi_x(\bar{\varphi}^*) = 0 \Rightarrow$$

$$\bar{\varphi}^* = \frac{\sigma}{\sigma-1} \sigma^{\frac{1}{\sigma-1}} L^{\frac{-1}{\sigma-1}} P^{-1} \tau f_x(M_x)^{\frac{1}{\sigma-1}} \quad (20)$$

where $\bar{\varphi}^* > \bar{\varphi}$.

5.2 Firm Entry

We assume that firms in each country must pay a fixed entry cost f_e in terms of labour in order to enter the industry and draw a productivity realization. Entrants draw their productivity from a Pareto distribution with cumulative distribution function $F(\varphi) = 1 - \left(\frac{m}{\varphi}\right)^\gamma$ and probability density function $f(\varphi) = \frac{\gamma m^\gamma}{\varphi^{1+\gamma}}$ where $\gamma - (\sigma - 1) > 0$.

If a firm draws a productivity below $\bar{\varphi}$ it exits immediately without producing. If the productivity is above $\bar{\varphi}$ but below $\bar{\varphi}^*$, then it operates domestically but does not sell to the foreign market. If the productivity is above $\bar{\varphi}^*$, then the firm is an export firm.

As in Melitz (2003), we assume that producing firms exit exogenously the industry if they are hit with a bad shock, that happens with probability δ . We assume that there is an unbounded pool of potential entrants to the industry, then, in equilibrium, expected profits of a firm must equal the entry cost. That is, using the functional forms:

$$E\left[\sum_{t=0}^{\infty} (1 - \delta)^t \pi(\varphi)\right] = f_e$$

$$\int_{\bar{\varphi}}^{\infty} \pi_d(\varphi) f(\varphi) d\varphi + \int_{\bar{\varphi}^*}^{\infty} \pi_x(\varphi) f(\varphi) d\varphi = \delta f_e$$

$$\frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1}\right)^{-(\sigma-1)} \frac{\gamma}{\gamma - (\sigma - 1)} LP^{\sigma-1} \left[\bar{\varphi}^{-(\gamma - (\sigma-1))} + \tau^{-(\sigma-1)} \bar{\varphi}^{*- (\gamma - (\sigma-1))} \right]$$

$$- f_d \bar{\varphi}^{-\gamma} - f_x(M_x) \bar{\varphi}^{*- \gamma} = \frac{\delta f_e}{m^\gamma} \quad (21)$$

5.3 Labour Market Clearing

The labour market clearing condition reads as follows:

$$M \int_{\bar{\varphi}}^{\infty} \left(\frac{q_d(\varphi)}{\varphi} + f_d \right) \frac{f(\varphi)}{1 - F(\bar{\varphi})} + M_x \int_{\bar{\varphi}^*}^{\infty} \left(\frac{\tau q_x(\varphi)}{\varphi} + f_x(M_x) \right) \frac{f(\varphi)}{1 - F(\bar{\varphi})^*} + M_e f_e = L \quad (22)$$

where M_e denotes the mass of potential entrants. The first term corresponds to labour used in production for the domestic market. The second, to production labour used for the foreign market. The third corresponds to the labour used by potential entrants.

In equilibrium, the mass of successful entrants must be the same as the mass of exiting firms:

$$(1 - F(\bar{\varphi})) M_e = \delta M \quad (23)$$

Using this condition and the functional forms, the labour market clearing condition becomes:

$$\begin{aligned} \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} \frac{\gamma}{\gamma-(\sigma-1)} LP^{\sigma-1} \left[M\bar{\varphi}^{(\sigma-1)} + \tau^{-(\sigma-1)} M_x \bar{\varphi}^{*\sigma-1} \right] \\ + Mf_d + f_x(M_x)M_x + M \frac{\delta f_e}{m\gamma} \bar{\varphi}^\gamma = L \end{aligned} \quad (24)$$

5.4 Price Index

Finally, note that the price index, defined in equation (11) can be expressed in terms of productivities and using the functional forms becomes:

$$\begin{aligned} P^{-(\sigma-1)} &= \left(\int_{\omega \in \Omega} p(\omega)^{-(\sigma-1)} d\omega \right) \\ P^{-(\sigma-1)} &= \int_{\bar{\varphi}}^{\infty} p_d(\varphi)^{-(\sigma-1)} M \frac{f(\varphi)}{1-F(\bar{\varphi})} d\varphi + \int_{\bar{\varphi}^*}^{\infty} p_x(\varphi)^{-(\sigma-1)} M_x \frac{f(\varphi)}{1-F(\bar{\varphi}^*)} d\varphi \\ P^{-(\sigma-1)} &= \left(\frac{\sigma}{\sigma-1}\right)^{-(\sigma-1)} \frac{\gamma}{\gamma-(\sigma-1)} \left[M\bar{\varphi}^{\sigma-1} + M_x \tau^{-(\sigma-1)} \bar{\varphi}^{*\sigma-1} \right] \end{aligned} \quad (25)$$

5.5 Equilibrium

A steady state equilibrium is characterized by the productivity cutoffs of the domestic and foreign markets $\{\bar{\varphi}, \bar{\varphi}^*\}$, the mass of producing and export firms $\{M, M_x\}$ and aggregate prices $\{P\}$ such that:

- The cutoff to sell in the domestic market satisfies equation (19).
- The cutoff to sell in the foreign market satisfies equation (20).
- The price index satisfies equation (25).
- The entry condition is satisfied -equation (21)-.
- The labour market clears -equation (24)-.

Due to the introduction of the agglomeration forces in the fixed export costs, the model has no close-form solution.³² Therefore, we solve it numerically. Our aim is to explore the quantitative implications of adding the agglomeration forces. We explore how the elasticity of welfare with respect to trade costs change once we allow for agglomeration, that is, once we vary the parameter θ , that drives the sensitivity of fixed trade costs with respect to the number of exporters.

³²If fixed export costs were independent from the number of export firms, the model, calibrated with the Pareto distribution, can be solved analytically. See for instance [Arkolakis et al. \(2008\)](#).

We solve the model for the five unknowns defined above. We settle the parameters at the values showed in Table VII. In solving the model, we face an issue with the functional form of the fixed export costs -equation (14)-. Note that if the mass of exporters M_x is lower than 1, an increase in θ makes the cost of exporting *higher*, instead of lower, That is:

$$\frac{df_x}{d\theta} = \bar{f}_x M_x^{-\theta} (-1) \log M_x < 0 \iff M_x > 1. \quad (26)$$

This is an undesirable property of the cost function because we interpret the parameter θ as capturing the effect of agglomeration. $\theta = 0$ means that agglomeration plays no role in the model. As θ increases, conditional on the mass of exporters, new exporters have a bigger effect on the export costs.

We address this issue in the following way. As the mass of exporters is a function of the (exogenous) size of the economy L , when solving the model, we adjust the size of the economy in order to generate a mass of exporters M_x equal to 1. This amounts to fixing the mass of exporters to 1 and treat the size of the economy as an equilibrium variable. This approach has the advantage that variations in the parameter driving agglomeration (θ) has no effect on the steady state equilibrium, that is, the equilibrium is independent of θ .

Then, we calculate how welfare changes when trade costs change, and how this effect depends on the parameter driving agglomeration. For this, we calculate the trade costs elasticity of welfare for a small decrease in trade costs and for different values of θ . An increase in θ means that an additional exporter has a higher effect on trade costs.

5.6 Agglomeration and the Estimated Welfare Gains from Trade

Welfare per worker is given by real wages, which corresponds to the inverse of the aggregate price index. We calculate the elasticity of welfare with respect to a decrease in fixed and variable trade costs. Figure VI shows how the fixed costs elasticity of welfare varies with the parameter driving agglomeration. VII shows the percentage change of welfare with respect to a 1 percentage change in variable trade costs, again as a function of θ .

Note that both elasticities are strictly increasing in agglomeration. When we allow for agglomeration ($\theta > 0$), a decrease in trade costs generates new firms entering the foreign market. This further reduces the fixed costs of exporting, which further reduces the productivity cutoff of exporting, generating additional exports and then an additional welfare increase. This effect is higher when the exporters elasticity of trade costs is higher, that is, when θ is higher. Note that the effect can potentially be quantitatively very large. For instance, for values of θ close to 0.7, the fixed costs elasticity of welfare would be six times higher as with a model without agglomeration, and the variable trade cost elasticity of welfare would be roughly 1.5 times higher. In the next subsection, we give a sense of the value θ according to our Spanish data.

5.6.1 Calibration of the Agglomeration Parameter θ

As previously noted, θ corresponds to the elasticity of fixed export costs with respect to the mass of exporters. In order to calibrate it, we make use of the expression driving the probability of exporting, which depends on θ . Note that the probability of exporting conditional of entering the industry, P_x , using equations (14), (19) and (20) and the functional form of the Pareto is:

$$P_x = \frac{1 - F(\bar{\varphi}^*)}{1 - F(\bar{\varphi})} = \left(\frac{\bar{\varphi}}{\bar{\varphi}^*} \right)^\gamma$$

$$P_x = f_d^{\frac{\gamma}{\sigma-1}} \tau^{-\gamma} \bar{f}_x^{\frac{-\gamma}{\sigma-1}} M_x^{\frac{\theta\gamma}{\sigma-1}} \quad (27)$$

Therefore, according to the model, the elasticity of the export probability with respect to the number of exporters is:

$$\frac{d \ln P_x}{d \ln M_x} = \frac{\theta\gamma}{\sigma - 1} \quad (28)$$

This expression relates the effect of a change in the number of exporters to the change in the conditional probability of exporting.³³ We use the coefficient estimate of Column (6) in Table VI, which relates the number of exporters around a firm and this firm's probability of entering a certain market, to approximate the value of the previous elasticity. The coefficient estimate is 0.0543 which we divide by the probability of exporting in our sample -0.175-. This gives us a value of the elasticity of 0.31. Making use of the values of γ and σ in Table VII we obtain a θ of around 0.26.

This approximate value of θ means that the estimated fixed-costs elasticity of welfare is 44 percent higher than the estimated elasticity in a model without agglomeration. The variable-costs elasticity is 7 percent higher. This stresses the fact that the welfare gains from trade are higher when we introduce the agglomeration forces in a standard model of international trade.

6 Concluding Remarks

In this paper we use a unique administrative micro-dataset of Spanish exporters to study the agglomeration economies that accrue to firms selling abroad. We document that exporters are concentrated by export-destination and that this pattern of agglomeration cannot be explained by the concentration of industries producing the goods demanded by each country. Moreover,

³³The number of exporters is an equilibrium variable, therefore, we cannot interpret the above elasticity as a comparative static. We take this shortcut, nevertheless, for the sake of simplicity and because the correlation between the number of exporters and the probability of exporting can be estimated from the data.

we observe that there is high variation in the patterns of exporters' concentration across export-destinations and that part of it can be explained by differences in the business environment and economic and cultural dissimilarities with respect to Spain.

We then assess the relationship between the observed exporters' agglomeration economies and the patterns of international trade. We find a significant association of agglomeration economies and firms' behaviour in international markets. Firms surrounded by a pool of exporters sharing the same destination engage in more stable trade relationships, exits less and enters more in these markets.

Finally, we study the effect of introducing agglomeration economies in a standard model of international trade, in order to assess the consequences on the estimated welfare gains from trade. We observe that when agglomeration is higher, a decrease in trade costs generates higher welfare gains relative to a model in which agglomeration plays no role. We calibrate the parameter driving agglomeration using our Spanish data and find that these excess welfare gains are of around 44 per cent with respect to fixed trade costs and of 7 per cent with respect to variable trade costs.

Overall, we point in the direction that exporters-specific agglomeration economies appear to be a robust finding and that it can have important consequences both in the patterns of international trade and in the way trade economists think on the welfare gains from trade. This paper, then, suggests that agglomeration may be an important issue in understanding exporters' decisions. For instance, in particular, it can help explain why firms do not follow a perfect hierarchy in selling to destinations, as uncovered by [Eaton et al. \(2011\)](#). If an export firm is close to another exporters selling to a hard-to-access country, this export firm may find easier to sell to this difficult destination rather than to an a priori easier one.

Our work, though, remains silent on the specific channels through which these agglomeration economies work. There are several possibilities, such as cost-sharing, peer effects, learning, et cetera. Some of these mechanisms have already been modeled. However, robust empirical evidence on the functioning of the potential channels is still missing. Case studies or natural experiments seem a suitable framework to disentangle specific mechanisms playing a role in generating export spillovers. We see this avenue of further research as promising.

In 2007, the Spanish central government agency in charge of promoting exports (ICEX) was funded with nearly 250 million euros. This adds to the funds that each Spanish region allocates to foster exports. Our results hints the existence of some externalities that make firms benefit from being close to successful exporters. Then, it seems that policies aimed a fostering the working of these externalities can provide benefits. For instance, the organization of meetings in which successful exporters share their experience in foreign markets can reduce information costs and speed up the learning process in dealing with exporting to new countries. Also, helping companies penetrate new markets can lead nearby firms to follow through.

References

- Aitken, B., G. H. Hanson, and A.E. Harrison, “Spillovers, Foreign Investment, and Export Behavior,” *Journal of International Economics*, 1997, 43, 103–132.
- Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow, and Andrés Rodríguez-Clare, “Endogeneous Variety and the Gains from Trade,” *American Economic Review: Papers & Proceedings*, May 2008, 98 (2), 444–450.
- Barrios, S., H. Görg, and E. Strobl, “Explaining Firms’ Export Behavior: R&D, Spillovers and the Destination Market,” *Oxford Bulletin of Economics and Statistics*, 2003, 65, 475–496.
- Békés, Gábor and Balázs Muraközy, “Temporary Trade and Heterogeneous Firms,” *Journal of International Economics*, 2012, 87, 232–246.
- Bernard, Andrew B. and J. Bradford Jensen, “Why Some Firms Export,” *The Review of Economics and Statistics*, 2004, 86 (2), 561–569.
- Casey, Andrew J. and Katherine N. L. Schmeiser, “The Agglomeration of Exporters by Destination,” November 2010. Working Paper.
- Chaney, Thomas, “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, September 2008, 98 (4), 1707–1721.
- Devereux, M. P., R. Griffith, and H. Simpson, “The Geographic Distribution of Production Activity in the UK,” *Regional Science and Urban Economics*, 2004, 35 (5), 533–564.
- Duranton, Gilles and Diego Puga, “Micro-Foundations of Urban Agglomeration Economies.,” *Handbook of Regional and Urban Economics*, 2004, 4, 2063–2117.
- and Henry G. Overman, “Testing for Localization Using Micro-Geographic Data,” *Review of Economic Studies*, 2005, 72 (4), 1077–1106.
- Eaton, Jonathan and Samuel Kortum, “Technology, Geography, and Trade,” *Econometrica*, September 2002, 70 (5), 1741–1779.
- , – , and Francis Kramarz, “An Anatomy of International Trade: Evidence from French Firms,” *Econometrica*, September 2011, 79 (5), 1453–1498.
- Ellison, Glenn and Edward L. Glaeser, “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 1997, 105 (5), 889–927.
- , – , and William R. Kerr, “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns,” *American Economic Review*, June 2010, 100 (3), 1195–1213.

- Greenaway, D., N. Sousa, and K. Wakelin**, “Do Domestic Firms Learn to Export from Multinationals?,” *European Journal of Political Economy*, 2004, *20*, 1027–1043.
- Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein**, “Estimating Trade Flows: Trading Partners and Trading Volumes,” *The Quarterly Journal of Economics*, May 2008, *CXXIII* (2), 441–487.
- Kaufmann, Daniel, Aart Kray, and Massimo Mastruzzi**, “Governance Matters VIII: Aggregate and Individual Governance Indicators 1996-2008,” June 2009. World Bank Policy Research Working Paper 4978.
- Koenig, Pamina**, “Agglomeration and the Export Decision of French Firms,” *Journal of Urban Economics*, 2009, *66*, 186–195.
- , **Florian Mayneris, and Sandra Poncet**, “Local Export Spillovers in France,” *European Economic Review*, 2010, *54*, 622–641.
- Krautheim, Sebastian**, “Heterogeneous Firms, Exporter Networks and the Effect of Distance on International Trade,” *Journal of International Economics*, 2012, *87*, 27–35.
- Krugman, Paul**, “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 1991, *99*, 483–499.
- Marshall, Alfred**, *Principle of Economics*, London: McMillan, 1920.
- Maurel, F. and B. Sédillot**, “A Measures of the Geographic Concentration of French Manufacturing Industries,” *Regional Science and Urban Economics*, 1999, *29* (5), 575–604.
- Mayer, Thierry and Soledad Zignago**, “Notes on CEPII’s Distances Measures: the *GeoDist* Database,” December 2011. CEPII Working Paper 25.
- Melitz, Marc J.**, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, November 2003, *71* (6), 1695–1725.
- Morales, Eduardo, Gloria Sheu, and Andrés Zahler**, “Gravity and Extended Gravity: Estimating a Structural Model of Export Entry,” January 2011. Working Paper.
- Openshaw, S.**, *The Modifiable Areal Unit Problem*, Geobooks. Norwich, England, 1984.
- and **P. Taylor**, “A Million or so Correlation Coefficients: Three Experiments on the Modifiable Area Unit Problem,” *Statistical Applications in the Spatial Sciences*, 1979, pp. 127–144. N. Wrigley (ed). London. Pion.
- Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jaques-Francois Thisse**, “Agglomeration and Trade Revisited,” *International Economic Review*, 2002, *43* (2), 409–435.

Roberts, Mark J. and James R. Tybout, “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs,” *The American Economic Review*, September 1997, 87 (4), 545–564.

Rosenthal, Stuart S. and William Strange, “Evidence on the Nature and Sources of Agglomeration Economies,” *Handbook of Regional and Urban Economics*, 2004, 4.

Segura-Cayuela, Rubén and Josep M. Vilarrubia, “Uncertainty and Entry into Export Markets,” 2008. Bank of Spain. Documento de Trabajo 0811.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.

Wagner, Rodrigo and Andrés Zahler, “New Exports from Emerging Markets: Do Followers Benefit from Pioneers?,” January 2011. Working Paper.

FIGURE I

DISTANCES DISTRIBUTION OF FIRMS EXPORTING TO SYRIA AND ICELAND
Panel A: Syrian Arab Republic
Panel B: Iceland

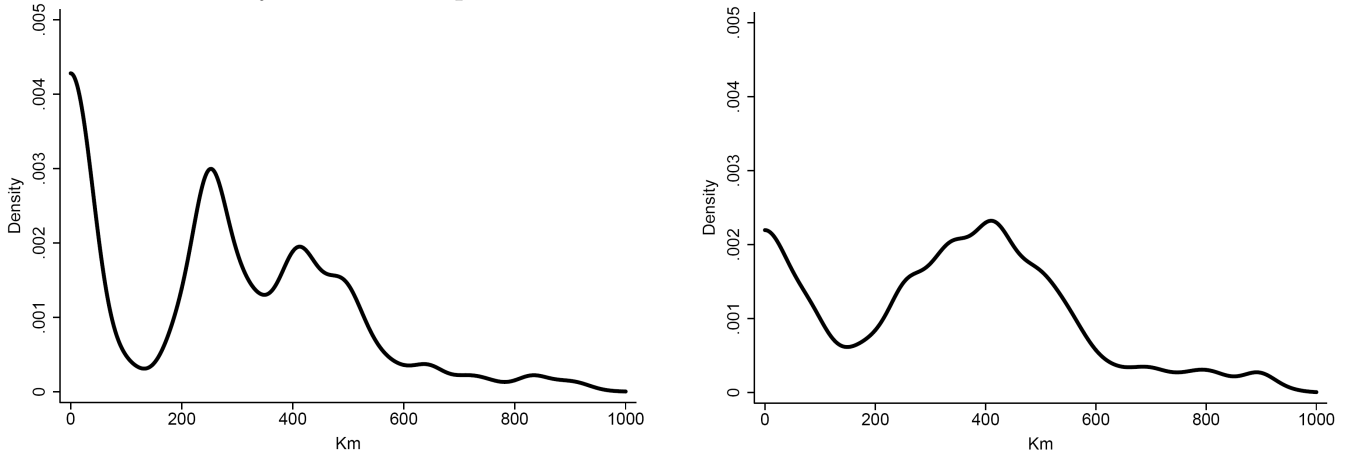


FIGURE II

LOCALIZATION AND DISPERSION THRESHOLDS OF SYRIA AND ICELAND
Panel A: Syrian Arab Republic
Panel B: Iceland

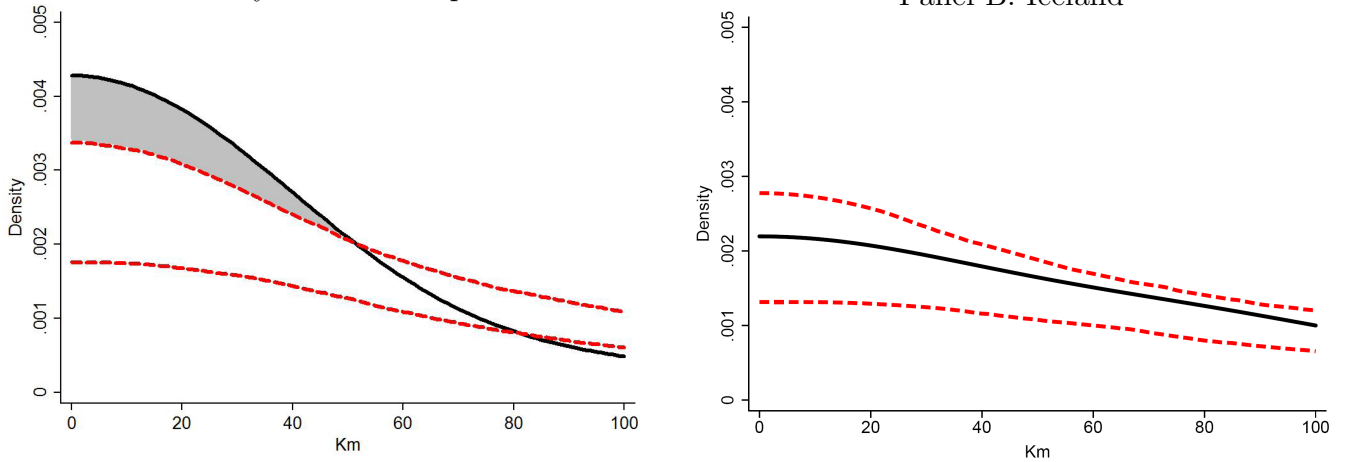


FIGURE III

SHARE OF DESTINATION-COUNTRIES IN WHICH FIRMS EXPORTING TO ARE
GEOGRAPHICALLY LOCALIZED

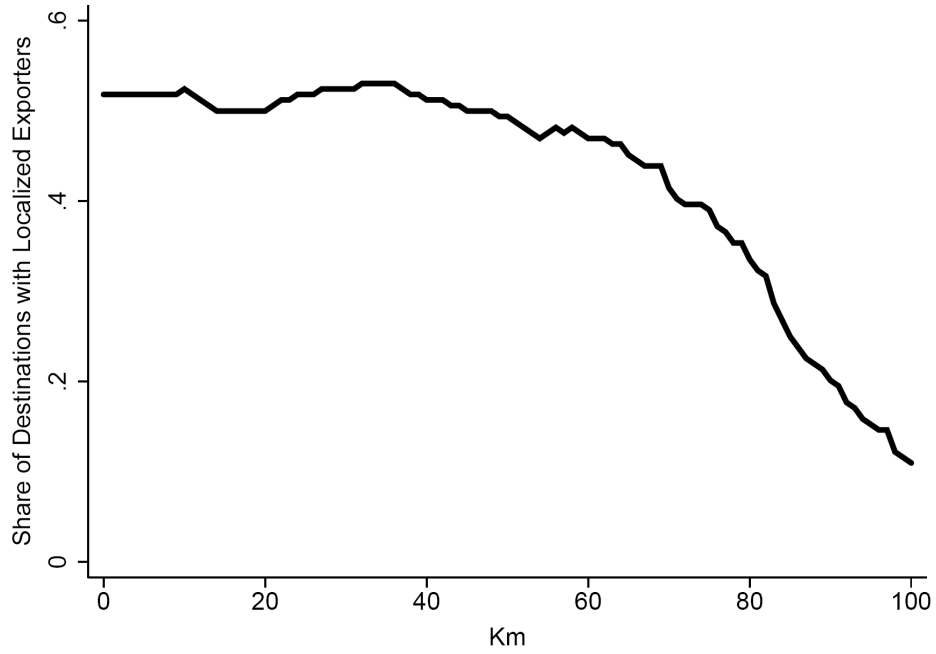


FIGURE IV

SUM OF LOCALIZATION INDEX ACROSS COUNTRIES FOR EACH DISTANCE

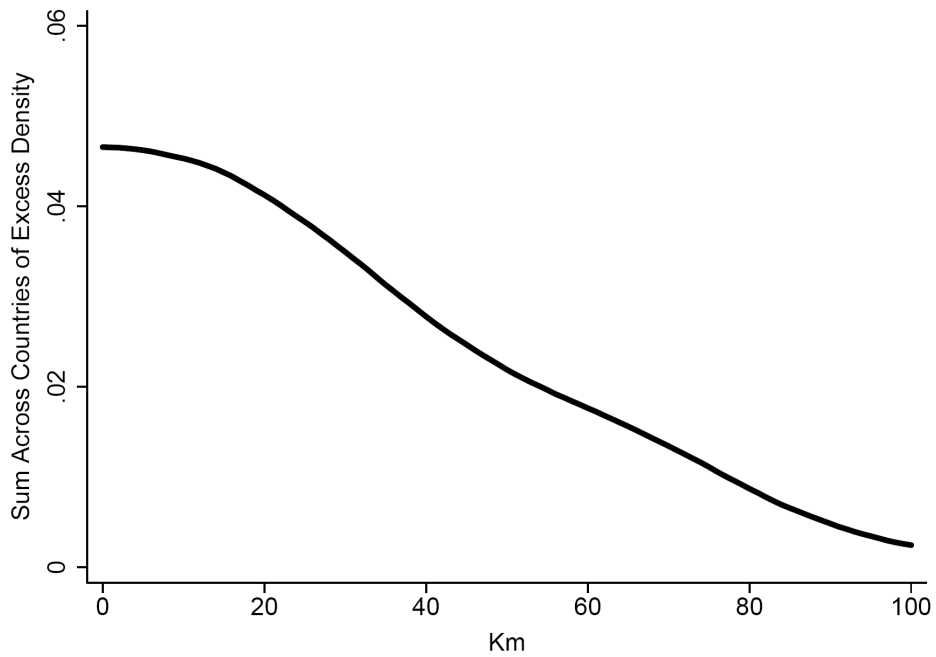
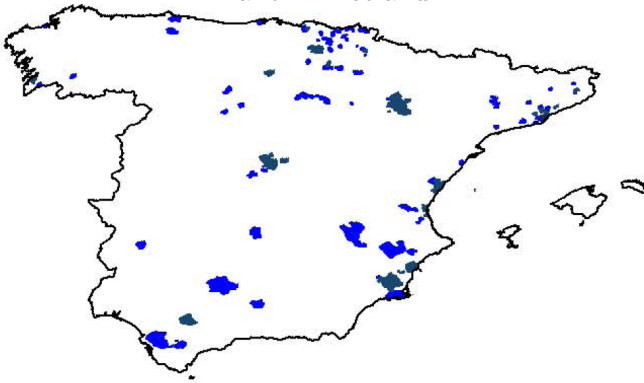


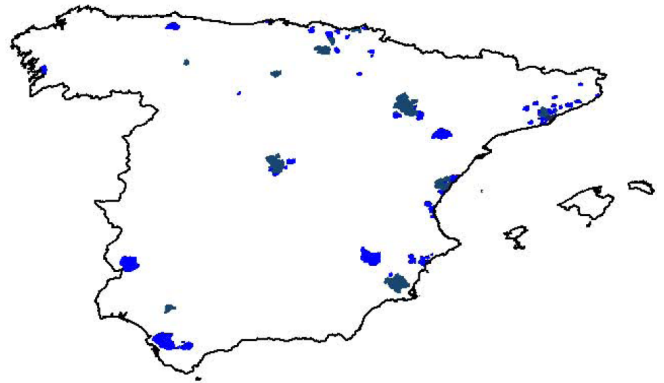
FIGURE V

SPATIAL DISTRIBUTION OF EXPORTERS TO FOUR COUNTRIES

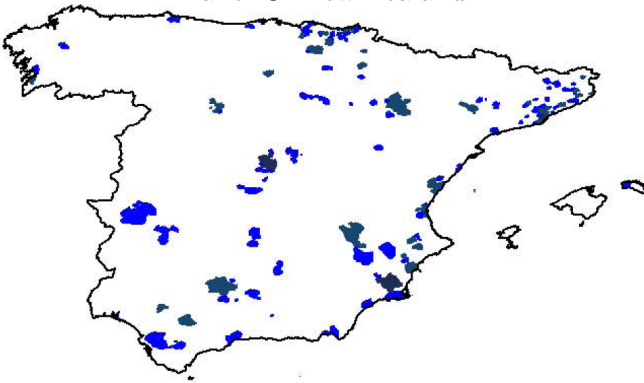
Panel A: Iceland



Panel B: Syria



Panel C: New Zealand



Panel D: Cuba

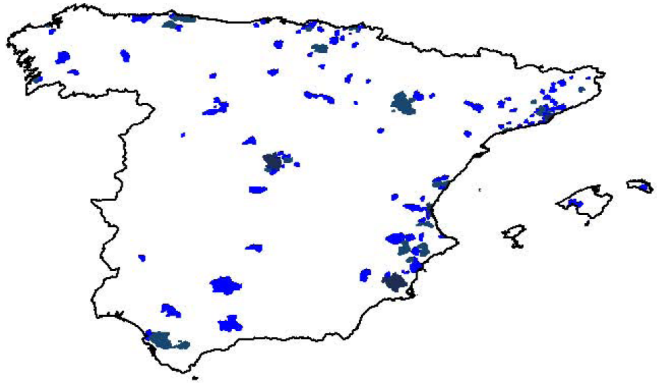


FIGURE VI
FIXED TRADE COSTS ELASTICITY OF WELFARE

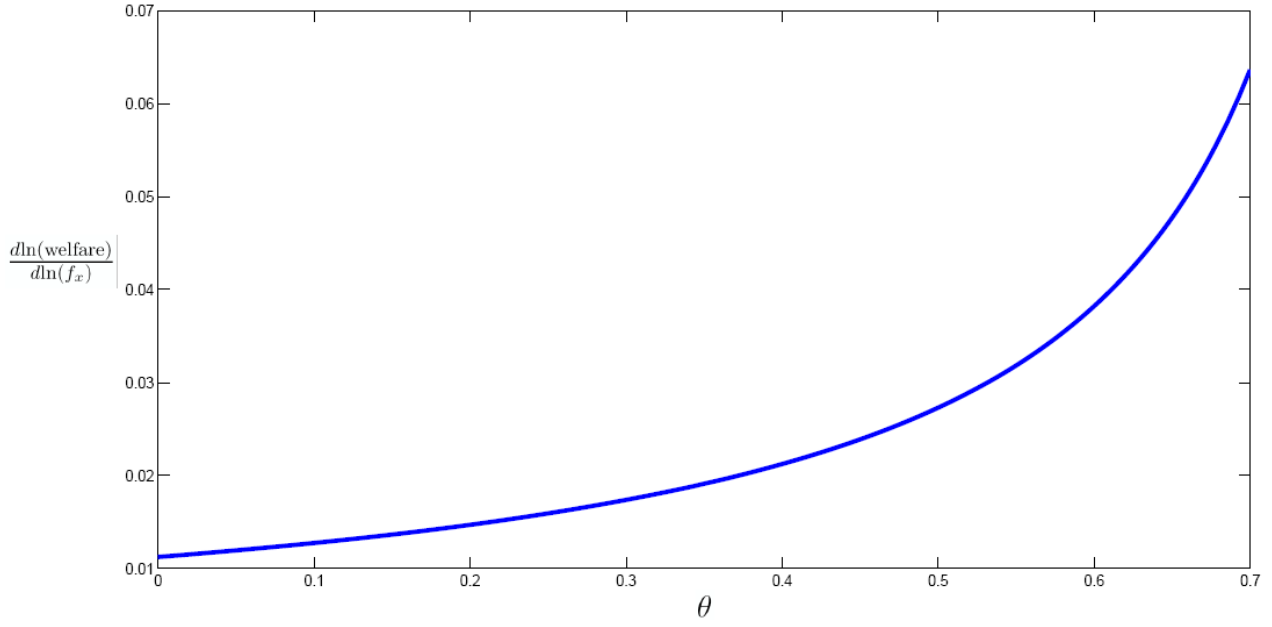


FIGURE VII
VARIABLE TRADE COSTS ELASTICITY OF WELFARE

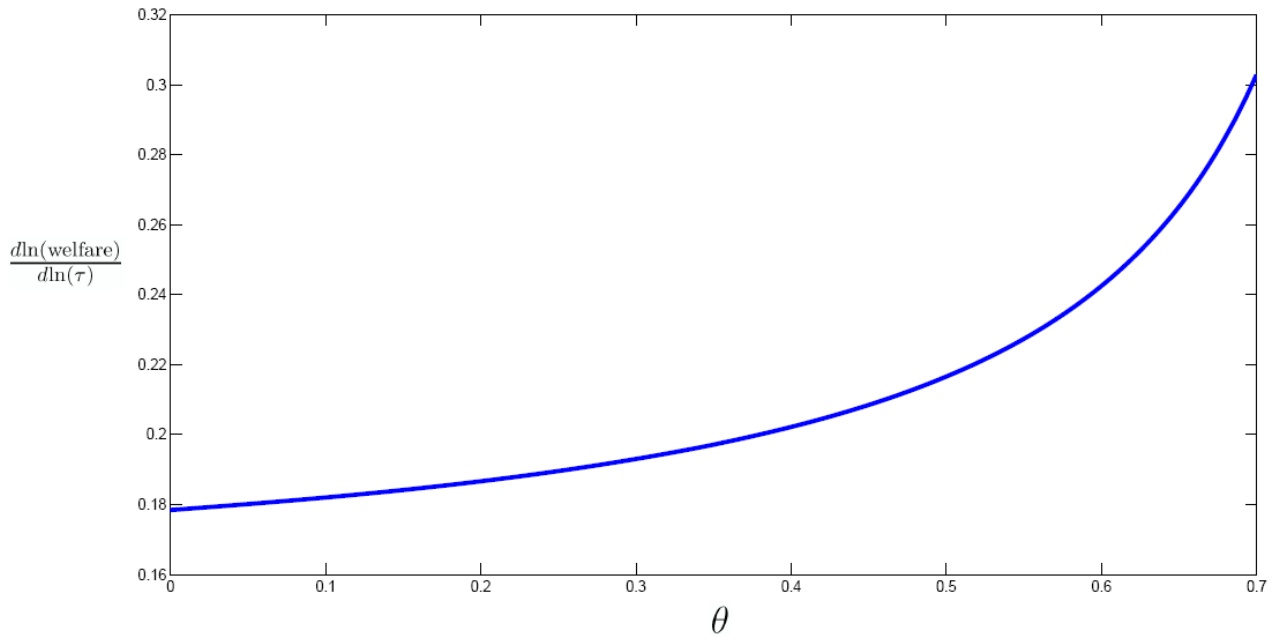


TABLE I
BALANCE OF PAYMENTS AND CENTRAL BALANCE SHEET DATA: 2007

	Total	Across Industries		Across Zips		Across Firms	
		Mean (Std. Dev)	Median	Mean (Std. Dev)	Median	Mean (Std. Dev)	Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>Panel A: Exporters (Balance of Payments)</u>							
Number	18553	90.95 (105.05)	55	5.81 (9.84)	2		
Zip Codes	3195	68.82 (65.08)	48				
Export Countries	204	69.59 (30.8)	68	14.96 (19.35)	7	5.38 (7.99)	2
Industries	204			4.39 (6.34)	2		
<u>Panel B: All Firms (Central Balance Sheet Data)</u>							
Number	73351	359.56 (685.03)	144	13.40 (27.03)	4		
Zip Codes	5472	211.60 (291.58)	111				
Industries	204			7.89 (11.14)	3		

Table I shows descriptive statistics of Exporters -Panel A- and All Firms -Panel B- for the year 2007 from Balance of Payments and Central Balance Sheet Data, respectively. Rows 'Number' shows the number of exporters or firms. Row Zips shows the zip codes where firms locate. Row 'Export Countries' shows the destination countries exporters sell to. Row 'Industries' shows the industries which firms belong to. Column (1) corresponds to the total number of each row. Columns (2) and (3) give statistics across industries; columns (4) and (5) across zip codes; and columns (6) and (7) across firms. For instance, column (4) in Panel A row 'Zip Codes' indicates that zip codes in which there is an exporter firm contain, on average, 5.81 exporters.

TABLE II
TOP 10 DESTINATIONS FOR WHICH EXPORTERS ARE MOST LOCALIZED

Rank	Country	N	Γ_c	Rank	Country	N	Γ_c
1	West Bank and Gaza	34	0.21	6	Syrian Arab Republic	161	0.09
2	Iraq	24	0.17	7	Montenegro	59	0.08
3	Albania	164	0.14	8	Macedonia, FYR	66	0.08
4	Suriname	17	0.11	9	Tanzania	42	0.07
5	Andorra	873	0.09	10	Nigeria	101	0.07

Table II shows the ten countries for which exporters appear most localized, according to the localization index (Γ_c) defined in Section 3.

TABLE III
FACTORS BEHIND EXPORTERS' AGGLOMERATION BY DESTINATION

	Baseline	Non-EU	Ports	Immigrants	Probit	Probit Non-EU
	(1)	(2)	(3)	(4)	(5)	(6)
Spanish	-1.0043*** (0.3711)	-1.4603*** (0.4342)	-1.3189*** (0.3738)	-0.5575 (0.3879)	-0.5027*** (0.1630)	-0.5658*** (0.1603)
Rule of Law	-0.4024* (0.2329)	-0.6196** (0.2965)	-0.4921* (0.2491)	-0.4727* (0.2730)	-0.1397* (0.0810)	-0.2222** (0.0916)
Euro	-0.6283** (0.3074)		-0.2387 (0.2413)	-0.6179** (0.2605)	-0.3324 (0.2349)	
Contiguity	0.9881 (1.0820)		-0.6319 (0.6464)	0.8991 (0.8507)	-0.2935 (0.4586)	
Log Distance to Capital	-0.1171 (0.1699)	0.0542 (0.1800)		-0.3835* (0.2158)	-0.0091 (0.0824)	0.0897 (0.0873)
Log Distance to Port			-0.0563 (0.1906)			
Log Per Capita GDP	0.2686 (0.1792)	0.3927* (0.2295)	0.3149* (0.1841)	0.4452 (0.2769)	0.1254** (0.0619)	0.1596** (0.0712)
Log Number of Exporters	0.0864 (0.1511)	0.0940 (0.1981)	0.0796 (0.1305)	-0.0531 (0.2174)	0.1730** (0.0700)	0.1837** (0.0799)
Log Population	0.0529 (0.1071)	0.1038 (0.1371)	0.0907 (0.0871)	0.2061 (0.1434)	0.0218 (0.0443)	0.0320 (0.0497)
Log Number of Immigrants				-0.1310* (0.0668)		
Constant	-2.2247 (1.9367)	-5.6027** (2.7831)	-3.8838* (2.2419)	-2.0918 (2.2481)		
Observations	150	123	141	110	150	123
Pseudo R-squared	0.05	0.06	0.06	0.10	0.29	0.27
Log Likelihood	-178.70	-144.00	-164.80	-119.00	-71.29	-61.73

Table III shows the regression of the index of exporters' localization against measures of export costs, comparative advantage and several covariates. The specification is a tobit model described in equation (3). Column (1) presents the baseline regression. Column (2) estimates the model only for countries not in the European Union. Column(3) replaces the variable Log Distance by the average distance to the closest port shipping to the country. Column (4) introduces the number of immigrants as an explanatory variable. Column (5) estimates a probit model of the probability of a country being localized. Column (6) estimates the probit model for countries outside the European Union. Robust standard errors are in parenthesis. Significance levels: *: 10%; **: 5%; ***: 1%. See Appendix Table A.1 and Section 3.1 for further details.

TABLE IV
TEMPORARY TRADE AND EXPORTERS' AGGLOMERATION:
LINEAR PROBABILITY MODEL

	No Agglom- -eration	Agglom- -eration	Firm Controls	Permanent Exporters	Fixed Effects	Agglom. Exports
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dep. Variable: Permanent Trade-Relationship in 2007</u>						
Share Exporters 100km ₂₀₀₃		0.0467*** (0.0037)	0.0467*** (0.0112)	0.0544*** (0.0143)	0.0858*** (0.0085)	
Share Exports 100km ₂₀₀₃						0.0235*** (0.0056)
Medium ₂₀₀₃			0.1881*** (0.0088)	0.1037*** (0.0108)		0.1881*** (0.0088)
Large ₂₀₀₃			0.3330*** (0.0136)	0.2382*** (0.0149)		0.3330*** (0.0136)
Log Value Added per Worker ₂₀₀₃			0.0907*** (0.0078)	0.0648*** (0.0093)		0.0907*** (0.0078)
Log Population ₂₀₀₇	0.0469*** (0.0016)	0.0314*** (0.0019)				
Log per Capita GDP ₂₀₀₇	0.0479*** (0.0025)	0.0335*** (0.0027)				
Log Distance	-0.0439*** (0.0030)	-0.0143*** (0.0038)				
Industry Dummies	YES	YES	YES	YES	NO	YES
Region Dummies	NO	YES	YES	YES	NO	YES
Country Dummies	NO	NO	YES	YES	YES	YES
Firm Fixed Effects	NO	NO	NO	NO	YES	NO
Observations	55,780	55,780	25,866	17,655	56,061	25,866
R-squared	0.05	0.05	0.14	0.15	0.12	0.14
Number of Firms	12702	12702	6324	2141	12,715	6324

Table IV shows the regression of the nature of trade-relationships in 2007 -permanent or temporary- on exporters' agglomeration and several controls. Column (1) excludes the measure of exporter' agglomeration in order to compare the results with [Békés and Muraközy \(2012\)](#). Column (2) includes the share of exporters selling to the destination in 2003 in a 100 kilometers radius within the firm. Column (3) adds firm controls. Column (4) restricts the sample to permanent exporters between 2004 and 2010. Column (5) estimates a firm fixed-effects model. Column (6) uses the share of exports in a 100 kilometers radius as the variable proxying for agglomeration. Robust standard errors are in parenthesis, clustered at the firm level. Significance levels: *: 10%; **: 5%; ***: 1%. See the Appendix Table A.1 and Section 4.1 for further details.

TABLE V
EXIT OF MARKETS AND EXPORTERS' AGGLOMERATION:
LINEAR PROBABILITY MODEL

	Share Exporters	Firm Controls	Non-EU	Firm FE	Firm- Country FE	Share Exports
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dep. Variable: Firm Exits Country in $t + 1$</u>						
Share Exporters 100 km _{<i>t</i>}	-0.0226*** (0.0024)	-0.0349*** (0.0059)	-0.0361*** (0.0087)	-0.0570*** (0.0056)	-0.0332*** (0.0128)	
Share Exports 100 km _{<i>t</i>}						-0.0067*** (0.0023)
Log Employment _{<i>t</i>}		-0.0685*** (0.0017)	-0.0600*** (0.0026)	-0.0819*** (0.0082)	-0.1326*** (0.0071)	-0.0694*** (0.0017)
Log Value Added per Worker _{<i>t</i>}		-0.0687*** (0.0037)	-0.0651*** (0.0059)	-0.0435*** (0.0057)	-0.0635*** (0.0045)	-0.0692*** (0.0039)
Log Population _{<i>t</i>}	-0.0251*** (0.0011)	-0.1935*** (0.0430)	-0.1915*** (0.0606)	-0.2455*** (0.0428)	-0.1451*** (0.0551)	-0.2563*** (0.0500)
Log Per Capita GDP _{<i>t</i>}	-0.0188*** (0.0016)	-0.0973*** (0.0284)	-0.0777* (0.0431)	-0.1298*** (0.0279)	-0.0495 (0.0352)	-0.1309*** (0.0328)
Log Distance	0.0212*** (0.0023)					
Industry Dummies	YES	YES	YES	NO	NO	YES
Region Dummies	YES	YES	YES	NO	NO	YES
Country Dummies	NO	YES	YES	YES	NO	YES
Year Dummies	YES	YES	YES	YES	YES	YES
Firm Fixed Effects	NO	NO	NO	YES	NO	NO
Firm-Country Fixed Effects	NO	NO	NO	NO	YES	NO
Observations	355,074	169,670	58,910	169,670	169,670	146,946
R-squared	0.06	0.12	0.10	0.10	0.12	0.13
Number of Firms	21,567	14,485	6,818	14,485	14,485	13,477
Firm-Country Pairs					72,037	

Table V shows the regression of the probability on exiting a market-destination in year $t + 1$ on exporters' agglomeration and several controls in year t -equation (6)-. Column (1) computes agglomeration as the share of exporters selling to the destination in a 100 km. radius around the firm. Column (2) adds firm controls. Column (3) restricts the sample to country-destinations outside the European Union. Column (4) and (5) estimate a firm and firm-country fixed-effects models, respectively. Column (6) uses the share of exports in a 100 kilometers radius as the variable proxying for agglomeration. Robust standard errors are in parenthesis, clustered at the firm level. Significance levels: *: 10%; **: 5%; ***: 1%. See the Appendix Table A.1 and Section 4.2 for further details.

TABLE VI
SELLING TO MARKETS AND EXPORTERS' AGGLOMERATION:
FIXED EFFECTS MODEL

	Share Exporters	Firm Controls	Non-EU	Number Exporters	Entrance	Entrance N. Exporters
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: Firm Exported to Country in Year t						
Share Exporters 100km $_{t-1}$	0.0527*** (0.0158)	0.0529*** (0.0165)	0.0980*** (0.0082)		0.0352*** (0.0123)	
Log Number Exporters 100km $_{t-1}$.				0.0712*** (0.0039)		0.0543*** (0.0046)
Log Employment $_{t-1}$		0.1198*** (0.0030)	0.1159*** (0.0047)	0.1196*** (0.0030)	0.0825*** (0.0035)	0.0826*** (0.0035)
Log ValueAdded per Worker $_{t-1}$		0.0563*** (0.0023)	0.0529*** (0.0035)	0.0561*** (0.0023)	0.0372*** (0.0028)	0.0373*** (0.0028)
Log Population $_{t-1}$	0.5050*** (0.0223)	0.5662*** (0.0298)	0.3677*** (0.0330)	0.5130*** (0.0269)	0.4043*** (0.0314)	0.3609*** (0.0310)
Log Per Capita GDP $_{t-1}$	0.4233*** (0.0170)	0.4799*** (0.0215)	0.3083*** (0.0246)	0.4179*** (0.0184)	0.3541*** (0.0215)	0.3041*** (0.0211)
Year Dummies	YES	YES	YES	YES	YES	YES
Firm-Country Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	996,272	435,034	192,651	429,807	266,236	262,046
R-squared	0.04	0.04	0.02	0.04	0.03	0.03
Firm-Country Pairs	125,710	99,334	43,435	98,747	80,898	80,265

Table VI shows the regression of the firm selling to a market in each year on exporters' agglomeration and several controls -equation (7). All regressions include firm-country specific fixed effects. Column (1) measures exporter' agglomeration as the share of exporters selling to the country in a 100 km. radius. Column (2) adds firm controls. Column (3) restricts the sample Non European Union destinations. Column (4) replaces the measure of agglomeration by the log number of exporters to the country within a 100 km. radius. Columns (5) and (6) use as dependent variable a dummy taking value 1 if the firm starts to export to a destination and zero if it does not enter. These specifications do not consider the observations in which the firm keeps exporting or stops selling to the market. That is, they restrict the focus to entrance in export-destinations. Column (5) uses the share of exporters within 100 km. as the measure of agglomeration. Column (6) uses the log number of exporters within 100 km. Robust standard errors are in parenthesis, clustered at the firm level. Significance levels: *: 10%; **: 5%; ***: 1%. See the Appendix Table A.1 and Section 4.3 for further details.

TABLE VII
PARAMETER VALUES OF THE MODEL

Parameter		Value
\bar{f}_x	Fixed export cost	2
τ	Variable trade Cost	1.5
f_d	Fixed production cost	1
σ	Elasticity of substitution	3.8
γ	Curvature of Pareto Distribution	3.4
m	Minimum of Pareto Distribution	1
f_e	Entry Cost	1

Table VII shows the parameter values used for solving numerically the model developed in Section 5

Appendices

A Appendix Tables

TABLE A.1
DATA DEFINITIONS AND SOURCES

VARIABLE	SOURCE	DEFINITION
Zip Code Coordinates	Geonames	Self-Explanatory
Distance Between Zip Codes		Apply haversine formula to zip code coordinates
Spanish	Mayer and Zignago (2011)	1 if a Spanish is spoken by at least 9% of the population
Rule of Law	Kaufmann et al. (2009)	Quality of contract enforcement, property rights, the police, the courts, and likelihood of crime and violence
Control of Corruption	Kaufmann et al. (2009)	Extent to which public power is exercised for private gain, including corruption, as well as "capture" of the state by elites and private interests.
Regulatory Quality	Kaufmann et al. (2009)	Ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
Political Stability	Kaufmann et al. (2009)	Likelihood that the government will be destabilized or overthrown by unconstitutional or violent means
Euro		1 if country's currency is the euro
Contiguity	Mayer and Zignago (2011)	1 for contiguity with respect to Spain
Distance to Country's Capital		Average distance of exporters to country's capital
Distance to Ports	Puertos del Estado	Distance between zip code and closest port from which positive exports are sent to the country. We assume that from the 3 main Portuguese ports -Aveiro, Leixoes, Lisbon-, all countries are served.
Per capita GDP	World Bank	Log Real per Capita GDP in constant 2000 US dollars
Population	World Bank	Country's population
Number of Immigrants	Instituto Nacional de Estadística	Number of immigrants living in Spain by country of origin.

TABLE A.1
DATA DEFINITIONS AND SOURCES (CONTINUED)

VARIABLE	SOURCE	DEFINITION
European Union		1 if country belongs to the European Union
Medium	Central Balance Sheet Data	1 if firm has between 50 and 249 employees.
Large	Central Balance Sheet Data	1 if firm has more than 249 employees.
Employment	Central Balance Sheet Data	Firm's number of employees.
Value Added	Central Balance Sheet Data	Gross value added at factor cost.
Distance	Mayer and Zignago (2011)	Simple distance between Spain and import country (most populated cities).
Region	Balance of Payments	Autonomous community in which the firm is located.

Table A.1 shows definitions and sources of all variables used throughout the paper.

TABLE A.2
LIST OF DESTINATION-COUNTRIES

Country	N	Γ_c	Country	N	Γ_c	Country	N	Γ_c
Afghanistan	14	0.00	Gabon	41	0.00	Netherlands	2980	0.01
Albania	164	0.09	Gambia, The	19	0.00	New Caledonia	33	0.00
Algeria	811	0.00	Georgia	98	0.00	New Zealand	272	0.00
Andorra	873	0.09	Germany	6345	0.01	Nicaragua	64	0.00
Angola	81	0.00	Ghana	67	0.00	Niger	45	0.00
Antigua and Barbuda	20	0.00	Gibraltar	85	0.00	Nigeria	101	0.05
Argentina	804	0.01	Greece	1736	0.01	Norway	722	0.00
Armenia	53	0.06	Guam	10	0.00	Oman	99	0.00
Aruba	19	0.04	Guatemala	224	0.00	Pakistan	186	0.03
Australia	782	0.01	Guinea	17	0.00	Panama	481	0.00
Austria	1600	0.01	Haiti	11	0.00	Paraguay	63	0.02
Azerbaijan	27	0.01	Honduras	106	0.00	Peru	427	0.00
Bahamas, The	50	0.00	Hong Kong SAR, China	700	0.03	Philippines	187	0.01
Bahrain	152	0.03	Hungary	756	0.03	Poland	1592	0.01
Bangladesh	71	0.06	Iceland	159	0.00	Portugal	6811	0.00
Barbados	31	0.00	India	644	0.02	Qatar	212	0.04
Belarus	87	0.00	Indonesia	221	0.01	Romania	891	0.01
Belgium	3468	0.01	Iran, Islamic Rep.	450	0.01	Russian Federation	894	0.01
Belize	75	0.00	Iraq	24	0.17	San Marino	19	0.00
Benin	27	0.00	Ireland	1277	0.00	Saudi Arabia	835	0.01
Bermuda	15	0.00	Israel	735	0.03	Senegal	83	0.00
Bolivia	96	0.00	Italy	5125	0.00	Serbia	173	0.01
Bosnia and Herzegovina	99	0.03	Jamaica	48	0.01	Seychelles	30	0.00
Brazil	1027	0.02	Japan	814	0.02	Sierra Leone	14	0.03
Bulgaria	497	0.01	Jordan	349	0.01	Singapore	426	0.03
Burkina Faso	30	0.00	Kazakhstan	97	0.00	Slovak Republic	433	0.01
Cameroon	47	0.00	Kenya	84	0.00	Slovenia	403	0.01
Canada	832	0.00	Korea, Dem. Rep.	39	0.00	Solomon Islands	10	0.00
Cape Verde	19	0.00	Korea, Rep.	562	0.02	South Africa	722	0.03
Cayman Islands	20	0.00	Kuwait	339	0.03	Sri Lanka	52	0.00
Central African Republic	10	0.00	Kyrgyz Republic	186	0.01	Sudan	41	0.03
Chad	19	0.06	Latvia	666	0.02	Suriname	17	0.16
Chile	855	0.00	Lebanon	495	0.01	Swaziland	21	0.02
China	993	0.01	Libya	150	0.01	Sweden	1263	0.00
Colombia	629	0.01	Liechtenstein	50	0.01	Switzerland	2060	0.01
Congo, Dem. Rep.	20	0.00	Lithuania	628	0.01	Syrian Arab Republic	161	0.03
Congo, Rep.	14	0.00	Luxembourg	281	0.00	Taiwan	449	0.02
Costa Rica	275	0.00	Macao SAR, China	33	0.00	Tanzania	42	0.09
Cote d'Ivoire	75	0.00	Macedonia, FYR	66	0.06	Thailand	335	0.02
Croatia	335	0.01	Madagascar	21	0.00	Togo	18	0.00
Cuba	312	0.00	Malaysia	275	0.01	Trinidad and Tobago	94	0.00
Cyprus	708	0.02	Mali	22	0.00	Tunisia	726	0.00
Czech Republic	1078	0.02	Malta	243	0.00	Turkey	1350	0.04
Denmark	1257	0.01	Marshall Islands	10	0.00	Uganda	13	0.00
Dominica	75	0.00	Mauritania	57	0.00	Ukraine	312	0.01
Dominican Republic	446	0.00	Mauritius	68	0.00	United Arab Emirates	962	0.01
Ecuador	334	0.00	Mexico	1945	0.00	United Kingdom	5440	0.00
Egypt, Arab Rep.	583	0.02	Micronesia, Fed. Sts.	30	0.03	United States	3836	0.00
El Salvador	127	0.00	Moldova	43	0.00	Uruguay	243	0.00
Equatorial Guinea	60	0.00	Monaco	92	0.02	Venezuela, RB	661	0.00
Estonia	371	0.02	Montenegro	59	0.09	Vietnam	128	0.00
Ethiopia	27	0.00	Morocco	1690	0.00	Virgin Islands (U.S.)	19	0.00
Finland	889	0.00	Mozambique	18	0.00	West Bank and Gaza	34	0.23
France	8820	0.02	Namibia	36	0.00	Yemen, Rep.	83	0.02
French Polynesia	20	0.00	Nepal	13	0.00			

Table A.2 shows the list of country-destinations with at least 10 Spanish exporters. N denotes the number of firms exporting to the country. Γ_c corresponds to the localization index that measures how geographically close are firms exporting to the destination. See Section 3 for details on this variable.

TABLE A.3

FACTORS BEHIND EXPORTERS' AGGLOMERATION BY DESTINATION: ROBUSTNESS

	Corruption	Regulatory Quality	Political Stability	200 km.	400 km.	N _j 2000
	(1)	(2)	(3)	(4)	(5)	(6)
Spanish	-0.8653*** (0.3192)	-0.7949** (0.3440)	-0.8618*** (0.3073)	-1.0756*** (0.3788)	-1.3154*** (0.3640)	-1.4538*** (0.4461)
Control of Corruption	-0.3491** (0.1745)					
Regulatory Quality		-0.1940 (0.1976)				
Political Stability			-0.4384** (0.2123)			
Rule of Law				-0.4493* (0.2350)	-0.6754*** (0.2464)	-0.5364** (0.2636)
Euro	-0.6333** (0.3104)	-0.7366** (0.2971)	-0.6165** (0.2935)	-0.6129** (0.3096)	-0.9129*** (0.3372)	-0.1583 (0.2705)
Contiguity	0.9653 (1.0379)	0.8916 (1.0064)	0.9016 (1.0342)	1.2257 (1.2181)	0.8880 (1.1407)	4.1589*** (0.6898)
Log Distance to Capital	-0.1764 (0.1824)	-0.2050 (0.1883)	-0.1228 (0.1622)	-0.1454 (0.1637)	-0.1331 (0.1634)	0.0228 (0.1668)
Log Per Capita GDP	0.3092* (0.1795)	0.2137 (0.1666)	0.2238 (0.1516)	0.2974* (0.1754)	0.2515 (0.1681)	0.2856 (0.1809)
Log Number of Exporters	0.0077 (0.1718)	0.0134 (0.1730)	0.0912 (0.1438)	0.0790 (0.1425)	0.3296** (0.1478)	0.1578 (0.1513)
Log Population	0.0869 (0.1137)	0.0917 (0.1176)	-0.0100 (0.0959)	0.0574 (0.1007)	-0.0246 (0.0983)	0.0402 (0.1141)
Constant	-2.1920 (1.8573)	-1.2766 (1.7820)	-0.8899 (1.5512)	-2.2501 (1.9311)	-1.8760 (1.9094)	-3.6769* (1.9837)
Observations	148	148	150	150	150	141
Pseudo R-squared	0.05	0.04	0.06	0.06	0.10	0.08
Log Likelihood	-177.10	-178.40	-176.90	-174.30	-173.70	-164.60

Table A.3 shows results on some robustness checks to the Tobit model described in Equation (3). Column (1), (2) and (3) use as covariates different proxies for institutional quality of the import country. Column (4) and (5) use as dependent variable an index of exporters' localization within 200 and 400 km., respectively. Column (6) excludes from the sample those countries with number of exporters higher than 2,000, which were sub-sampled when computing the localization index. Robust standard errors are in parenthesis. Significance levels: *: 10%; **: 5%; ***: 1%. See Appendix Table A.1 and Section 3.1 for further details.

TABLE A.4
TEMPORARY TRADE AND EXPORTERS' AGGLOMERATION:
PROBIT

	No Agglom- -eration	Agglom- -eration	Firm Controls	Permanent Exporters	Agglom. Exports
	(1)	(2)	(3)	(4)	(5)
<u>Dep. Variable: Permanent Trade-Relationship in 2007</u>					
Share Exporters 100km ₂₀₀₃		0.0492*** (0.0040)	0.0533*** (0.0126)	0.0748*** (0.0192)	
Share Exports 100 km ₂₀₀₃					0.0269*** (0.0063)
Medium ₂₀₀₃			0.2018*** (0.0095)	0.1137*** (0.0116)	0.2018*** (0.0095)
Large ₂₀₀₃			0.3391*** (0.0149)	0.2469*** (0.0159)	0.3391*** (0.0149)
Log ValueAdded per Worker ₂₀₀₃			0.1032*** (0.0089)	0.0743*** (0.0104)	0.1032*** (0.0089)
Log Population ₂₀₀₇	0.0486*** (0.0017)	0.0324*** (0.0021)			
Log Per Capita GDP ₂₀₀₇	0.0496*** (0.0026)	0.0345*** (0.0029)			
Log Distance	-0.0452*** (0.0031)	-0.0142*** (0.0040)			
Industry Dummies	YES	YES	YES	YES	YES
Region Dummies	NO	YES	YES	YES	YES
Country Dummies	NO	NO	YES	YES	YES
Observations	55,780	55,780	25,834	17,632	25,834
Pseudo R-squared	0.0354	0.04	0.11	0.12	0.11
Log Likelihood	-37142	-36969	-15882	-10343	-15882

Table A.4 shows the probit regression of the nature of trade-relationships in 2007 -permanent or temporary- on exporters' agglomeration and several controls -equation (5)-. Column (1) excludes the measure of exporter' agglomeration. Column (2) includes the share of exporters selling to the destination in 2003 in a 100 kilometers radius within the firm. Column (3) adds firm controls. Column (4) restricts the sample to permanent exporters between 2004 and 2010. Column (4) uses the share of exports in a 100 kilometers radius as the variable proxying for agglomeration. Robust standard errors are in parenthesis, clustered at the firm level. Significance levels: *: 10%; **: 5%; ***: 1%. See Appendix Table A.1 and Section 4.1 for further details.

TABLE A.5
EXIT OF MARKETS AND EXPORTERS' AGGLOMERATION:
PROBIT

	Share Exporters	Firm Controls	Non-EU	Share Exports
	(1)	(2)	(3)	(4)
<u>Dep. Variable: Firm Exits Country in $t + 1$</u>				
Share Exporters 100 km _{<i>t</i>}	-0.0263*** (0.0031)	-0.0425*** (0.0063)	-0.0401*** (0.0096)	
Share Exports 100 km _{<i>t</i>}				-0.0080*** (0.0027)
Log Employment _{<i>t</i>}		-0.0755*** (0.0019)	-0.0645*** (0.0030)	-0.0768*** (0.0020)
Log ValueAdded per Worker _{<i>t</i>}		-0.0746*** (0.0039)	-0.0698*** (0.0063)	-0.0756*** (0.0042)
Log Population _{<i>t</i>}	-0.0248*** (0.0013)	-0.2183*** (0.0442)	-0.2122*** (0.0639)	-0.2960*** (0.0522)
Log Per Capita GDP _{<i>t</i>}	-0.0186*** (0.0016)	-0.1169*** (0.0298)	-0.0884* (0.0463)	-0.1608*** (0.0349)
Log Distance	0.0203*** (0.0025)			
Industry Dummies	YES	YES	YES	YES
Region Dummies	YES	YES	YES	YES
Country Dummies	NO	YES	YES	YES
Year Dummies	YES	YES	YES	YES
Observations	355,074	169,609	58,853	146,889
Pseudo R-squared	0.05	0.10	0.08	0.10
Log Likelihood	-219491	-98339	-36017	-85435

Table A.5 shows the marginal effects of the probit regression of the likelihood of exiting a country next year on exporters' agglomeration and several controls -equation (6). Column (1) measures agglomeration as the share of exporters in year t selling to the destination in a 100 km. radius within the firm. Column (2) includes firm controls. Column (3) restricts the sample to Non European Union countries. Column (4) uses the share of exports in a 100 kilometers radius as the variable proxying for agglomeration. Robust standard errors are in parenthesis, clustered at the firm level. Significance levels: *: 10%; **: 5%; ***: 1%. See Appendix Table A.1 and Section 4.2 for further details.

B The Excess Industry Localization of Exporters

In this Appendix section, we show descriptive results on the excess localization of exporters with respect to all firms -domestic and exporting- in a given industry. We start by describing the dataset of Spanish manufacturing firms. We then compare the concentration of exports with respect to total sales across industries, computing the so-called [Ellison and Glaeser \(1997\)](#) (henceforth EG) index of agglomeration. Although our focus is on the geographical agglomeration of firms, we compute an agglomeration measure of exports and sales for the sake of completeness. In the third subsection, we compute the excess geographical agglomeration of exporters following the methodology by [Duranton and Overman \(2005\)](#).

B.1 Data

We use a firm-level dataset containing operating and financial information of a representative sample of non-financial companies in Spain -around 800,000 firms each year-.³⁴ This database is named Central Balance Sheet Data (*Central de Balances*) and is provided by the Bank of Spain.

The database is comprised of two complimentary datasets. The first one is based on a standardized voluntary survey handled to companies at the time of requesting compulsory accounting information. Each year around 9,000 companies fill this survey. The information gathered is very detailed, from equity structure to leveraging, but the sample size is low and there is an over representation of big firms.

The second dataset contains the balance sheets of a much larger number of companies. It originates from the firms' legal obligation to deposit their balance sheets on their corresponding Mercantile Registries. Therefore, coverage is much wider.

The Bank of Spain Central Balance Sheet Office is in charge of collecting and cleaning these datasets. All of the variables contained in the second database are included in the first one, that is, they are standardized. For each firm, we observe its fiscal id, zip code, industry, revenue and employment. We restrict to those industries with at least 10 firms. We are left with around 70,000 manufacturing firms distributed in 204 industries.

B.2 Industry Localization of Exporters: [Ellison and Glaeser \(1997\)](#)

We compute the concentration of industry exports and compare it to the concentration of total industry sales for all manufacturing industries. We refer to localization of exports and sales as the localization of the firms that produce the goods generating these exports and sales, regardless of the location of the final consumer. We choose as the spatial unit of analysis the zip code, in order to make use of the detailed location that our data provide, and in order to adopt a conservative

³⁴According to the Spanish National Statistics Institute (INE), in 2010 they were around 1.2 million firms in Spain.

approach as bigger units tend to produce correlations more pronounced -the Modifiable Areal Unit Problem-.

EG is an index derived from a model of location choice, in which either natural advantages or industry spillovers lead firms to cluster together. It has three main virtues. First, it is comparable across industries. It is able to define a threshold -zero- above which an industry can be regarded as concentrated. It also can rank industries in terms of concentration. Second, it controls for the concentration of overall manufacturing. That is, concentration is measured above and beyond concentration of manufacturing. And third, it takes into account the size distribution of the industry. Industries in which the main activity is carried out by a few firms will exhibit concentration, but not for either natural advantages or spillovers, but because activity is concentrated in a few operating units.

The EG index for industry d , γ_d , reads as follows:

$$\begin{aligned} \gamma_d &\equiv \frac{G - \left(1 - \sum_{z=1}^Z x_z^2\right)H}{\left(1 - \sum_{z=1}^Z x_z^2\right)(1 - H)} \\ &\equiv \frac{\sum_{z=1}^Z (s_z - x_z)^2 - \left(1 - \sum_{z=1}^Z x_z^2\right)^2 \sum_{j=1}^N z_j^2}{\left(1 - \sum_{z=1}^Z x_z^2\right)\left(1 - \sum_{j=1}^N z_j^2\right)}. \end{aligned} \quad (29)$$

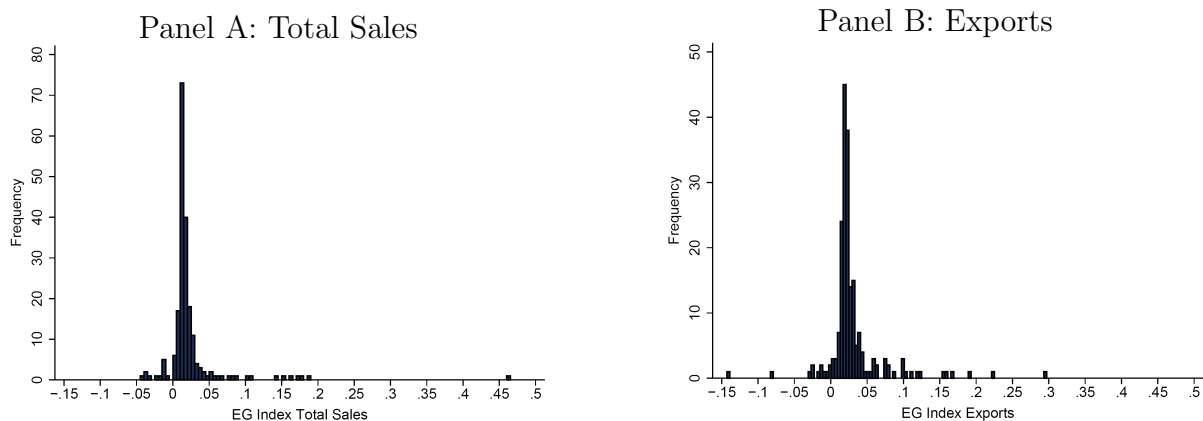
s_z is the share in total industry sales of sales of firms in zip code z . x_z is the share in total manufacturing sales of sales of firms in zip code z . Therefore, $G = \sum_{z=1}^Z (s_z - x_z)^2$ sums across the total number of zip codes, Z , the squared deviations of each zip code's share in total industry sales and each zip code's share in total manufacturing sales. Hence, it provides a measure of an industry's geographic concentration taking into account the concentration of overall manufacturing. N is the total number of firms in the industry and z_j is the share of industry sales of firm j . Then, $H = \sum_{j=1}^N z_j^2$ is the Herfindahl index of the industry's firm size distribution. As shown by [Ellison and Glaeser \(1997\)](#), in a context of firms choosing their location randomly, regardless of natural advantages and industry spillovers, the expectation of γ_d is zero. Therefore, positive values of the index are indicative of concentration.

We compute this index for our 204 manufacturing industries. We do so for total sales and exports. When computing the index for total sales, we take the total number of manufacturing firms as in the Central Balance Sheet Data. When computing the index for exports, we take the total number of export firms as in the Balance of Payments micro data.

Figure VIII shows the distribution of the index for both calculations: total sales -Panel A- and exports -Panel B-. We find that 192 out of 204 industries exhibit a certain degree of concentration in terms of sales and exports. Following [Ellison and Glaeser \(1997\)](#) in defining highly concentrated industries as those with an index value above .02, we find that 56 are highly concentrated in terms of total sales and 114 in terms of exports.

FIGURE VIII

ELLISON AND GLAESER INDEX



In Figure IX, we estimate, via a kernel function, and compare the densities of both distributions.

³⁵ The concentration of exports seems to be on the right to the concentration of total sales.

FIGURE IX

ELLISON AND GLAESER INDEX: TOTAL SALES VS. EXPORTS

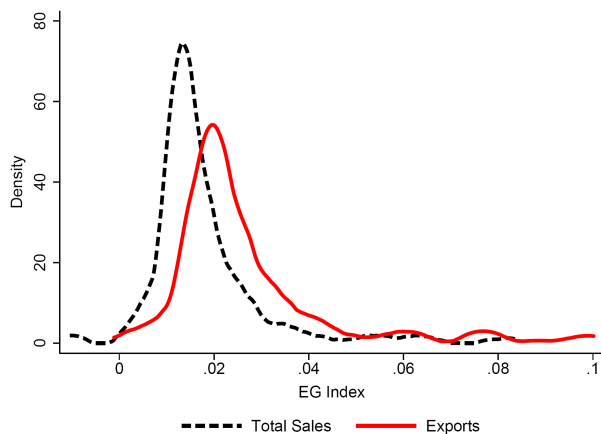


Table B.1 displays a set of moments of both distributions. A comparison of columns (1) and (2) confirms the findings of Figure IX: the distribution of the EG index of exports lies to the right of that of total sales. Column (3) shows that in the ample majority of industries the concentration of exports is higher than the concentration of total sales, although for some industries the opposite is true.

Summing up, it seems that across industries the concentration of exports is higher than that of total sales. These results must be taken cautiously, though, since our focus is on the geographical location of exporters as compared to all firms, rather than the amount of exports with respect to total sales. Moreover, the previous results suggest that the concentration of exports is higher than that of total sales; however, it is hard to argue that the differences are of significant magnitude. Finally, as mentioned in the main text, the EG index suffers from the 'border effect' problem, that

³⁵For presentational purposes, the 5 per cent left and right tails of every distribution are not displayed.

TABLE B.1

DISTRIBUTION OF ELLISON AND GLAESER INDEX: TOTAL SALES VS. EXPORTS

	Total Sales	Exports	Difference
	(1)	(2)	(3)
Mean	0.0236	0.0302	0.0066
Std. Dev.	0.0439	0.0405	0.0516
Percentile 5	-0.0105	-0.0026	-0.0322
Percentile 25	0.0117	0.0172	0.0013
Percentile 50	0.0147	0.0220	0.0065
Percentile 75	0.0209	0.0309	0.0140
Percentile 95	0.0844	0.1002	0.0605
Number of Industries	204	204	204

Table B.1 shows moments of the distribution of the Ellison and Glaeser Index for total sales -column (1)-, exports -column (2)- and the distribution of the difference of exports and total sales.

is, firms belonging to different zip codes are treated the same, regardless of the distance within those zip codes.

Then, in the next subsection, we analyze the agglomeration patterns of exporters with respect to all firms in every industry following the methodology developed by [Duranton and Overman \(2005\)](#). We slightly adapt this methodology to account for localization of some firms -exporters- within industries, instead of industries in general, which motivated the original application.

B.3 Industry Localization of Exporters: [Duranton and Overman \(2005\)](#)

Our way of proceeding is very similar to the exporters' localization indices across country destinations developed in Section 3. For each industry, we estimate the density of bilateral distances of all the exporters operating in that industry. Distances between firms are calculated applying the haversine formula to their zip code coordinates. The density is estimated by means of a Gaussian kernel, as in Equation (1). We estimate each density between 0 and 400 kilometers, following the median criterion of [Duranton and Overman \(2005\)](#). As before, the bandwidth is chosen so as to minimize the mean integrated squared error and data is reflected around zero.

Then, we test for each industry whether the distances within exporters significantly deviate from randomness. We assume that every zip code in which there is a manufacturing firm of one industry can host an export firm of that same industry. This amounts to take into account the spatial distribution of every manufacturing industry in order to test whether exporters exhibit an excess localization beyond that of their industries.

We take for each industry the population of firms in Central Balance Sheet Data and draw

1,000 random samples of size the number of exporters in the same industry in Balance of Payments Data. Then, for each industry and each sample, we calculate the bilateral distances and estimate a density between 0 and 400 kilometers. We are then left with 1,000 random distributions for each industry.

From these 1,000 random distributions, we construct two thresholds, a localization threshold and a dispersion threshold, each with significance level of 95 per cent. These thresholds are constructed in the same way as in Section 3, noting, however, that the random distributions in this Section have a different meaning.

Then, we define exporters in an industry to exhibit localization in a precise distance -excess localization from that of the industry- if the estimated density of exporters' distances is above the localization threshold. Exporters in an industry are said to be localized if this industry exhibit localization in at least one distance.

Similarly as Section 3, exporters in an industry are said to exhibit dispersion in a precise distance if the estimated density is below the dispersion threshold and exporters in that industry are not localized. Exporters in an industry are said to be dispersed if those exporters exhibit dispersion in at least one kilometer and they do not exhibit localization.

Out of 200 industries with enough number of firms, exporters in 103 industries -52 per cent- are localized whereas in 20 -10 per cent- they are dispersed. Figure X shows the proportion of industries in which exporters are localized and dispersed at each level of distance. In 31 per cent of industries, exporters are localized at short distances. Since around 50 km. localization declines fast, reaching just 5 per cent of firms at medium distances. Dispersion, on the other hand, slightly increases with distance.

FIGURE X

SHARE OF INDUSTRIES WITH EXPORTERS EXHIBITING GLOBAL LOCALIZATION OR DISPERSION

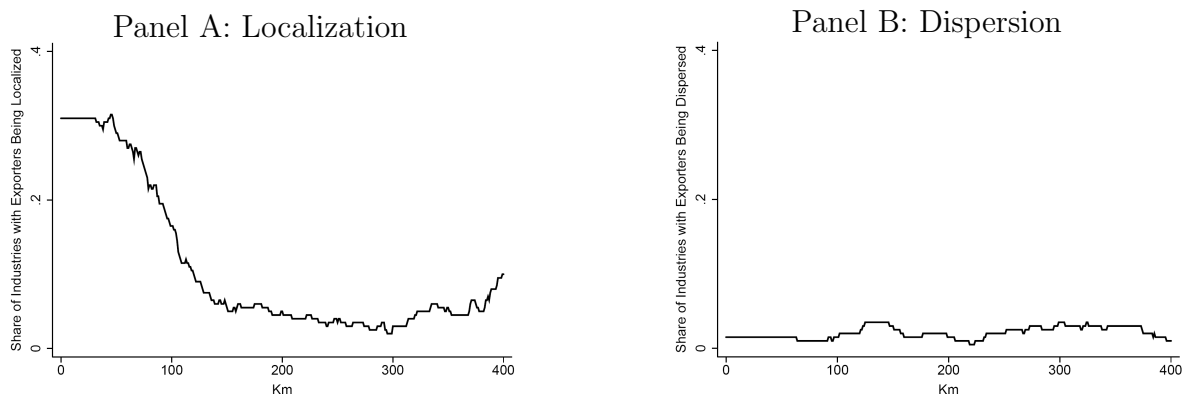
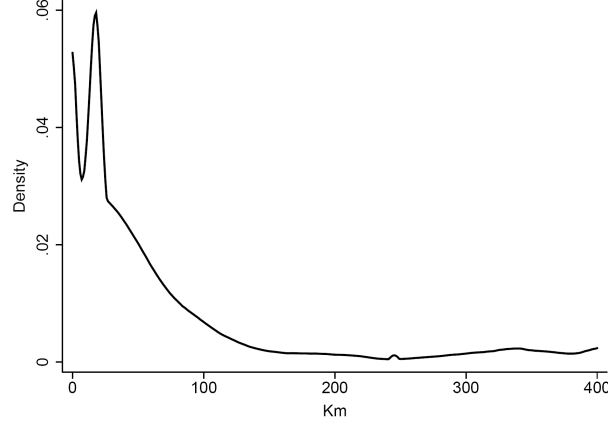


Figure XI plots, for each kilometer, the sum across those industries in which exporters exhibit localization of the difference between the estimated density and the localization threshold. As expected, almost all of the amount of localization takes place at distances within 50 km. The peak

we observe at around 20 km. responds to one industry -ceramic tiles- in which exporters are highly concentrated in several towns within that distance.

FIGURE XI

TOTAL LOCALIZATION AT EACH DISTANCE



Finally, we construct indices of exporters' excess localization and dispersion. The index of localization is computed as follows. For each industry, we compute at each distance the difference between the estimated density and the the localization threshold. We then sum across all distances the positive differences. That is, the localization index of industry i , Γ_i , is given by:

$$\Gamma_i \equiv \sum_{d=0}^{400} \max(\hat{K}_i(d) - \bar{\bar{K}}_i(d), 0) \quad (30)$$

where $\hat{K}_i(d)$ is the estimated density of industry i at distance d and $\bar{\bar{K}}_i(d)$ is the localization threshold.

For computing the index of dispersion we subtract at each distance the estimated density from the dispersion threshold. Then we sum across distances the positive differences. If the index of localization is zero, then the index of dispersion is this sum. If the industry is localized, then the index of dispersion is zero. That is:

$$\Psi_i \equiv \begin{cases} \sum_{d=0}^{400} \max(\underline{\underline{K}}_i(d) - \hat{K}_i(d), 0) & \text{if } \Gamma_i = 0 \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

where Ψ_i is the index of dispersion of industry i and $\underline{\underline{K}}_i(d)$ is the dispersion threshold.

Table B.2 shows the 10 industries in which exporters show the highest degree of excess concentration and dispersion, according to these indices. The industry of top excess localization of exporters -ceramic tiles- is also one of the most concentrated industries. The correlation between the index of excess concentration of exporters and that of industry concentration is not very high -.16-, though.

TABLE B.2

TOP TEN INDUSTRIES IN WHICH EXPORTERS SHOW THE HIGHEST EXCESS
LOCALIZATION AND DISPERSION

Panel A: Localization

CNAE09	Industry	Index
2331	Manufacture of ceramic tiles and flags	0.40
1021	Processing and preserving of fish, crustaceans and molluscs	0.22
1043	Manufacturing of olive oil	0.17
2451	Casting of iron	0.14
2370	Cutting, shaping and finishing of stone	0.10
2442	Aluminium production	0.08
1022	Manufacturing of fish canned	0.07
1084	Manufacture of condiments and seasonings	0.06
2573	Manufacture of tools	0.06
1392	Manufacture of made-up textile articles, except apparel	0.06

Panel B: Dispersion

CNAE09	Industry	Index
1622	Manufacture of assembled parquet floors	0.04
3320	Installation of industrial machinery and equipment	0.03
1439	Manufacture of other knitted and crocheted apparel	0.03
2931	Manufacture of electrical and electronic equipment for motor vehicles	0.02
3012	Building of pleasure and sporting boats	0.02
2110	Manufacture of basic pharmaceutical products	0.01
2529	Manufacture of other tanks, reservoirs and containers of metal	0.01
1394	Manufacture of cordage, rope, twine and netting	0.01
1610	Saw milling and planing of wood	0.01
1511	Tanning and dressing of leather; dressing and dyeing of fur	0.00

Table B.2 shows the ten industries in which exporters show the highest excess localization -Panel A- and dispersion -Panel B-.