Agglomeration by Export Destination: Evidence from Spain*

Roberto Ramos
Bank of Spain

Enrique Moral-Benito
Bank of Spain

April 2017

Abstract

We use a dataset of Spanish exporters with rich spatial attributes to document the existence of agglomeration economies by export destination. More specifically, we show that, for a large set of export destinations, exporters are geographically too close to be the result of a random outcome. We also analyze the variables that explain the cross-destination heterogeneity in agglomeration. We find that firms selling to countries with worse institutions, a dissimilar language, and a different currency are significantly more agglomerated. These results suggest that the value provided by agglomeration is higher concerning destinations where entry is more difficult.

JEL codes: R12, F14, D22

Keywords: agglomeration economies, export firms, international trade.

*This paper was previously circulated under the title "Agglomeration Matters for Trade." We are very grateful to Guillermo Caruana, Rosario Crinó and Claudio Michelacci for their constant guidance and help. We are also thankful to César Alonso, Pol Antràs, Manuel Arellano, Stéphane Bonhomme, David Dorn, Gino Gancia, Manuel García-Santana, Horacio Larreguy, Carlos Llano, Marc Melitz, Guy Michaels, Hannes Mueller, Diego Puga, Rafael Repullo, Rubén Segura-Cayuela, Andrei Shleifer, at least two anonymous referees, and seminar participants at CEMFI, European Winter Meeting of the Econometric Society (Konstanz), SAEe Vigo, IMT Lucca, Universität Mannheim, IAE, CESifo, 2013 Annual Meeting of the Society for Economic Dynamics (Seoul), 67th European Meeting of the Econometric Society (Gothenburg), 2013 Barcelona Workshop on Regional and Urban Economics, Banco de España, Universidad de Murcia, XVII Conference on International Economics (A Coruña), and Universidad Autónoma de Madrid for comments and useful discussions. We are also grateful to Patry Tello for providing us the exporters dataset. The views expressed in this paper are those of the authors and do not necessarily coincide with the views of the Banco de España or the Eurosystem. Roberto acknowledges the financial support of the Spanish Ministry of Education under research grant BES-2009-026803.

Contact: Bank of Spain. Alcalá 48, 28014. Madrid (Spain). roberto.ramos@bde.es.

1 Introduction

It is now a well-established fact that a large amount of industries are geographically concentrated. This result and the observation that firms are on average more productive in denser areas have attracted much attention from economists and policy-makers, who have built a large body of research on the foundations and effects of agglomeration economies.

In this paper we study a specific form of these agglomeration economies, namely those accruing to export firms. Our main contribution is to uncover the fact that exporters are geographically concentrated by export destination, which is consistent with the existence of agglomeration economies associated to the process of selling to some foreign markets. When selling abroad, exporters incur in additional costs, hence the traditional forces leading to industry agglomeration, namely sharing, matching and learning mechanisms, see Marshall (1920) and Duranton and Puga (2004), apply to this process, and naturally vary by export destination.

We are not the first to emphasize the existence of destination-specific export spillovers. In a regression framework, Koenig (2009), Koenig et al. (2010), and Choquette and Meinen (2014) show that the decision to export to a country is positively affected by the pool of local exporters selling to that destination, and Cassey and Schmeiser (2013) show evidence of clustering by export destination in Russian regions. Our main step forward in this literature is to account for destination-specific agglomeration economies using a firm level dataset with rich spatial attributes that allows us to apply a non-parametric standard test of agglomeration: Duranton and Overman (2005). This method improved previous approaches, such as Ellison and Glaeser (1997), Maurel and Sédillot (1999), and Devereux et al. (2004), for two reasons. First, it treats space as continuous, instead of using an arbitrary set of spatial units. And second, it allows to assess the statistical significance of departures from randomness.[1]

The advantages of this test allows us to refine and expand the results of the extant literature in two fundamental ways. First, we uncover destination-specific exporter agglomeration *beyond* the overall concentration of exporters with respect to domestic firms *and* the industry agglomeration of exports to each country. That is, for each destination the counterfactual is built only from exporters operating in the industries exported to that destination. This addresses a crucial issue: industries are geographically concentrated, and different countries demand goods from different industries. Then, agglomeration by export destination might be the result of countries buying goods

[1] The existing literature on export spillovers studies the agglomeration of exporters in the same administrative unit or economic area. This entails the so-called "border effect" problem, which involves several issues. First, it amounts to treat symmetrically plants not belonging to the same spatial unit, regardless of the distance that separates them. 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 more pronounced correlations. 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 not often defined on the basis of economic significance make the comparison of results across spatial units difficult to interpret.

intensively from agglomerated industries. We show that exporters are significantly concentrated over and above what would be expected by the fact that exporters to individual countries are concentrated by sector and sectors are concentrated geographically. Furthermore, by restricting the benchmark to export firms, we account for different location patterns between domestic and export firms. And second, we compute a quantitative index capturing the *extent* of geographical agglomeration of exporters to each destination. This allows us to investigate the characteristics that explain the cross-destination heterogeneity in agglomeration levels.

Our baseline results show that for more than half of export destinations exporters are significantly concentrated, i.e. they are too close to be the result of a random outcome.[2] We also perform a battery of robustness checks in order to account for other mechanisms that would result in agglomeration by export destination without relying on export-destination spillovers. There are two mechanisms that are worth emphasizing. First, large firms are able to reach a wider set of destinations, hence agglomeration by destination could be the result of large exporters being concentrated with respect to small exporters. And second, it is documented that firms go hierarchically to more and more destinations. Then agglomeration to popular destinations could induce agglomeration to less popular destinations if exporters to the former disproportionately export to the latter. By restricting the counterfactual, we show that our results are not the mechanical consequence of these and other mechanisms that would result in spurious agglomeration.

Our results also show that the level of agglomeration varies meaningfully across destinations. We find that exporters to countries with a dissimilar language, lower institutional quality, and different currency are significantly more agglomerated. We interpret this result as evidence pointing toward agglomeration providing higher value in countries where entry is more difficult.

Overall, our findings are consistent with the existence of externalities in the process of selling to some countries. Although thus far the literature has been to some extent unable to empirically verify the specific driving mechanisms, several possibilities have been rationalized in theoretical models.[3] For example, Segura-Cayuela and Vilarrubia (2008) and Fernandes and Tang (2014) emphasize that firms entering foreign markets reveal information, hence reducing the uncertainty facing potential entrants. Also, Krautheim (2012) and Cassey and Schmeiser (2013) explore the channel of cost reductions brought by an increasing number of exporters in the setting of the Melitz (2003) model. On the empirical front, our results are consistent with Lovely et al. (2005), who show, using the Ellison and Glaeser index, that US exporters headquarter activity is more

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

[3]On the mechanisms driving industry agglomeration, see Klepper (2010), who analyzes the historical clustering of firms in Detroit and Silicon Valley, and Ellison et al. (2010), who test the Marshall (1920) theories of industry agglomeration using coagglomeration patterns.

agglomerated when they sell to countries less integrated in the world economy and with worse credit ratings. Moreover, Koenig (2009), Wagner and Zahler (2015) and Cadot et al. (2013) show that the presence of neighboring export firms and pioneers in foreign markets are significantly associated to a higher probability of foreign entry, suggesting that the flow of information between nearby firms and signals revealed by successful exporters are important in this context. Also, a recent contribution by Paravisini et al. (2015) finds that the distribution of bank lending is skewed toward firms exporting to the same destination, which suggests that agglomeration may be associated to credit markets. We are able to link the observed patterns of concentration across destinations to cultural and institutional differences across importing countries. Beyond this, our paper also remains silent on the specific mechanisms driving the clustering by export destination. Yet, our findings as well as those of the literature give new understandings regarding the behavior of export firms and provide useful policy insights on how to help firms access foreign markets.

The rest of the paper is organized as follows. Section 2 describes the dataset. Section 3 explains the methodology, presents the baseline results and performs a set of robustness checks. Section 4 delves into the determinants of the cross-destination variation in agglomeration levels. Section 5 concludes.

2 Data

We use the firm-level data from which the Bank of Spain constructs the official Spanish Balance of Payments. The dataset contains information on firms making transactions with foreign agents if they are worth more than €12,000 and they are performed through a bank. Therefore, the dataset is likely to exclude only the smallest exporters. In the baseline analysis we rely on the 2007 data and we use previous years to check the stability of the results over time.[4]

The dataset has several advantages in order to study the geographical location of exporters. First, it is made up from administrative records and it has a large coverage. For example, it accounts for 97% of aggregate exports in 2007. Moreover, both transactions within the EU and to third countries are observed. Second, it contains information on total sales of every firm to each export destination. And third, it provides the zip code of every exporter. Hence, we can compute distances between firms and study the agglomeration of export firms by destination. The zip code provided is that of the headquarters, thus our focus in on headquarter agglomeration rather than on establishment agglomeration. We reckon that this feature is likely to play a small role in the results, as we estimate that in our data around 91% of exporters have just one plant.[5] This is

[4]The dataset extends to 2013, but the export threshold was increased in 2008 to €50,000. For this reason and in order to avoid the results being contaminated from the crisis, we use data up to 2007.

[5]This estimation is as follows. According to the 2009 Spanish Survey on Business Strategies, 93.1% of firms with less than 200 employees and 62.5% of firms with more than 200 employees have only one plant; in our data 94.5% of exporters have less than 200 employees (vs. 99.1% of all firms), then we estimate that the percentage of

also the case in the literature. For example, Koenig (2009), Koenig et al. (2010), and Choquette and Meinen (2014) show that export spillovers are not affected by including either single-plant exporters or the headquarter of all exporters. Having said this, focusing on headquarters hints at externalities stemming from information flows, rather than cost-sharing mechanisms, which are more likely to be linked to establishments.

There is one limitation of the dataset that is worth noting: it has no information on the type of goods being traded, hence we rely on the two-digit firm industry to control for the composition of exports. This makes us account for the varieties being exported only to a certain extent (see also Section 3.3.5). Besides, the dataset does not include firm characteristics beyond the industry, fiscal id, and total exports to each country. For this reason, for the robustness check described in Section 3.3.1, we approximate the size of the firm with total firm exports.

In Table I we show descriptive statistics of the exporters in 2007. Our analysis is restricted to manufacturing firms and export destinations with at least 10 exporters. The data include more than 18,000 exporters located in close to 3,200 zip codes, out of a total of around 11,000 zip codes. The median exporter sells to 2 destinations and the median zip code hosts 2 export firms. See also Table A.1 in the appendix for a description of the main variables used in the paper and Table A.2 for the destinations included in the sample.

TABLE I
DESCRIPTIVE STATISTICS:
EXPORTERS IN 2007 (BALANCE OF PAYMENTS)

	Mean	Percentiles		
	(Std. Dev)	25	50	75
	(1)	(2)	(3)	(4)
Panel A: Exporters ($N = 18,715$)				
Total Exports (thousand €)	6,745 (100,757)	54	255	1,435
Destinations ($N = 166$)	5.19 (7.82)	1	2	6
Panel B: Zip Codes ($N = 3,192$)				
Number Exporters	5.89 (10.05)	1	2	6

Notes: This table shows descriptive statistics of Spanish manufacturing exporters in 2007 included in the Balance of Payments micro data. The panel A shows statistics of total exports and number of export destinations per exporter. The panel B shows moments of the distribution of the number of export firms located in the zip codes hosting at least one export firm. N corresponds to the number of distinct observations.

single-plant exporters is approximately 91.4%.

3 Localization of Exporters by Export Destination

In this section we provide evidence that exporters are significantly agglomerated by export destination. To do so, we apply the methodology developed by Duranton and Overman (2005), henceforth DO. Among other advantages discussed in the introduction, this methodology allows us to account for the fact that exports to a country are concentrated by sector and sectors are concentrated geographically. Furthermore, it allows to control for other forces leading to agglomeration by export destination, such as the concentration of large vs small exporters, sequential exporting, and the concentration of exporters in large cities, which we discuss in a robustness section below. Before turning to the baseline results, we describe briefly our application of DO.

3.1 Methodology: Application of Duranton and Overman (2005)

In this section we give a succinct overview on how we use DO to uncover agglomeration by export destination, see Appendix B for a more comprehensive explanation and technical details. In our baseline analysis we use data from 2007 and we consider the 166 destinations with at least 10 exporters. For each country, we kernel-estimate the distribution of bilateral distances of exporters to that country by applying the haversine formula, which computes the shortest distance over the Earth's surface, to the zip-code coordinates. We then compare this distribution with 1,000 counterfactual distributions built from independent random draws of exporters meeting one condition; namely they operate in a two-digit industry in which at least one exporter actually exports to that country. Moreover, each random draw replicates the composition of exports to the country. If, say, 80% of exporters to the country operate in the auto industry and 20% in the apparel industry, each random draw is made up from 80% auto exporters and 20% apparel exporters. Note also that the size of each random draw is the same as the actual number of exporters to the country.

The counterfactual distributions built this way control for two mechanisms that could result in spurious agglomeration by export destination. First, the fact that exporters have special characteristics relative to non-exporters (see for example Bernard et al. (2003)) and therefore they may agglomerate with respect to domestic firms. Indeed, Behrens and Bougna (2015) show that this is the case in 14 to 16% Canadian industries and the literature on export spillovers has documented that pools of local exporters positively affects the decision to enter foreign markets, see for example Koenig (2009). Second, the fact that the industry composition of exports differs across countries. For example, one country may demand heavily goods from an industry that is highly concentrated. Therefore, exporters to this country can be concentrated either because of industry concentration or because of exporter concentration. By making the random draws replicate the destination-specific industry composition of exports, we are able to disentangle the latter from the former.

We then rank, for each kilometer, the 1,000 counterfactual distributions in ascending order and define a localization threshold as the percentile that makes 95% of the counterfactual distributions lie below it across all distances, whereas 5% of the simulations are above it in at least one kilometer. Note that in order to compare the estimated density with the counterfactual distributions we focus on distances below 100 kilometers, which are more relevant in order to explain interactions between exporters. This distance horizon has no substantial effects on the results, see Section 3.3.5.

We define exporters to a country to be significantly localized if the actual distance distribution is above the localization threshold in at least one kilometer. We also define a dispersion threshold as the percentile that makes 5% of the simulations lie below it across all distances. Given that densities must sum up to one, localization at some distances implies dispersion at others. Therefore, we define exporters to a country to be dispersed if the actual distance distribution is below the dispersion threshold in at least one kilometer and the country does not exhibit localization. These definitions follow DO.[6]

Finally, we also construct the country version of the industry quantitative index of localization defined by DO. This country index is computed as the sum across distances of the difference between the density of the actual distance distribution and the localization threshold if the former is above the latter and zero otherwise. This index gives a measure of the *amount* of exporter localization by export destination.

3.2 Baseline Results

In our baseline results we find significant agglomeration for a large number of export destinations: of the 166 countries in our sample, firms exporting to 107 (64%) are significantly agglomerated, whereas only one destination exhibits dispersion. Panel A of Figure I displays the share of destinations in which exporters exhibit significant localization at each level of distance. Close to 60% of destinations exhibit significant agglomeration at distances below 40 kilometers, this share decreasing fast at larger distances. This scale at which agglomeration takes place is similar to the one found by DO regarding industry agglomeration. With respect to the amount of agglomeration at each distance, Panel B plots for each kilometer the sum across destinations of the difference between the distance distribution and the localization threshold when the former is above the latter. As can be observed, the largest amount of agglomeration takes place at very small distances. Again, this result is in line with industry agglomeration patterns.

Table II shows the destinations exhibiting the highest country index of agglomeration and Table A.2 in the appendix reports the value of the index for all destinations. The largest amount of agglomeration is found in rather small destinations, accounting for only 1.4% of all firm-country relationships in 2007. However, larger countries also exhibit significant agglomeration, for example the main EU countries and the US. Overall, 85.8% of firm-country relationships correspond to significantly localized destinations. It is also worth emphasizing that the extent of agglome-

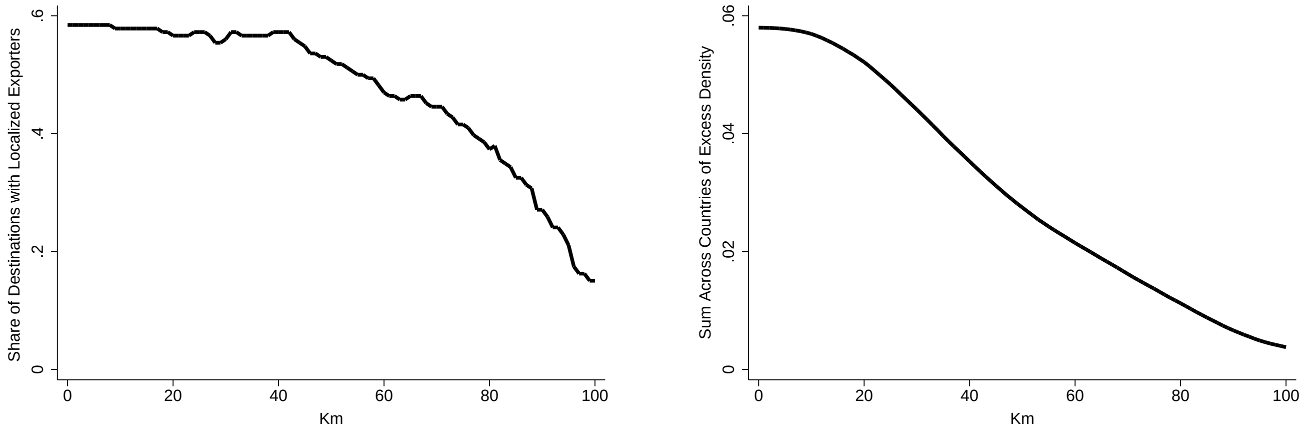
[6]In DO the localization and dispersion thresholds are referred to as global confidence bands.

Figure I

Localization of Exporters by Export Destination

Panel A: Share of Destinations with Significantly Localized Exporters

Panel B: Amount of Localization at each Distance



Notes: This figure shows the pattern of agglomeration by export destination. Panel A displays the share of destinations in which exporters exhibit significant localization at each level of distance, whereas Panel B shows the amount of localization at each distance, i.e. the sum across destinations of the difference between the actual distribution and the localization threshold when the former is above the latter.

ration varies widely across destinations, being the standard deviation of the country index twice as large as the mean. At first sight it is not straightforward to uncover a systematic pattern of cross-destination agglomeration, probably due to the fact that several determinants act in opposing directions. For example, some agglomeration is expected to happen next to the border of neighbouring countries. Indeed, one destination exhibiting very large agglomeration is Andorra, with which Spain shares a border. However, nearby countries such as those belonging to the EU have good business practices, which undermines the value of agglomeration in overcoming trade barriers. Hence, we would expect exporters to these countries to be less agglometated. We acknowledge that given the nature of our data it is very hard to disentangle the specific mechanisms driving the agglomeration by export destination. Yet in Section 4 we discuss several possibilities and show some evidence via regression analysis. Before this, we perform several robustness checks in the next section.

3.3 Robustness Checks

In this section we check the sensitivity of the baseline results to restricting the counterfactual. In so doing, we control for some concentration patterns that would result in agglomeration by export destination without relying on export-destination spillovers. Specifically, we take into account different concentration patterns of large vs small exporters, sequential exporting, the concentration of firms in large cities, the stability of the results over time, and others.

Table II

Most Localized Destinations

Rank	Country	N	Localization	Rank	Country	N	Localization
1	West Bank and Gaza	36	0.29	6	Montenegro	61	0.10
2	Iraq	23	0.18	7	Andorra	887	0.09
3	Suriname	18	0.18	8	Aruba	19	0.09
4	Chad	18	0.16	9	Tanzania	42	0.08
5	Albania	168	0.11	10	Armenia	53	0.08

Notes: This table shows the ten destinations for which exporters are most agglomerated, according to the country index of localization defined in Section 3.1.

3.3.1 Controlling for Destinations Served by Large vs Small Exporters

The heterogeneous firms literature has documented that there is a substantial amount of heterogeneity within exporters, one important dimension of this heterogeneity being size. When studying agglomeration by destination, this is relevant, because it has been documented that large exporters reach a larger set of export destinations, see Helpman et al. (2008) and Eaton et al. (2011). Therefore, our baseline results raise the concern that agglomeration by destination could be the result of large exporters being agglomerated with respect to small exporters. If this were the case, our findings would no exist *within* exporters of comparable size.

We carry out four tests in order to address this issue. First, we repeat the baseline exercise with subsamples of the largest exporters. Specifically, we consider exporters above the median and above the 75 percentile of total firm exports. Second, we explicitly control for the size of exporters when building the counterfactual. To be precise, we further restrict the counterfactual by conditioning on 10 and 20 bins of total firm exports. That is, we construct counterfactual distributions from exporters in the same industry *and* the same size bin, replicating for each destination the distribution of these variables observed in the data. Hence, this procedure estimates destination-specific agglomeration patterns beyond the concentration of exporters, industries and exporters of comparable size.[7]

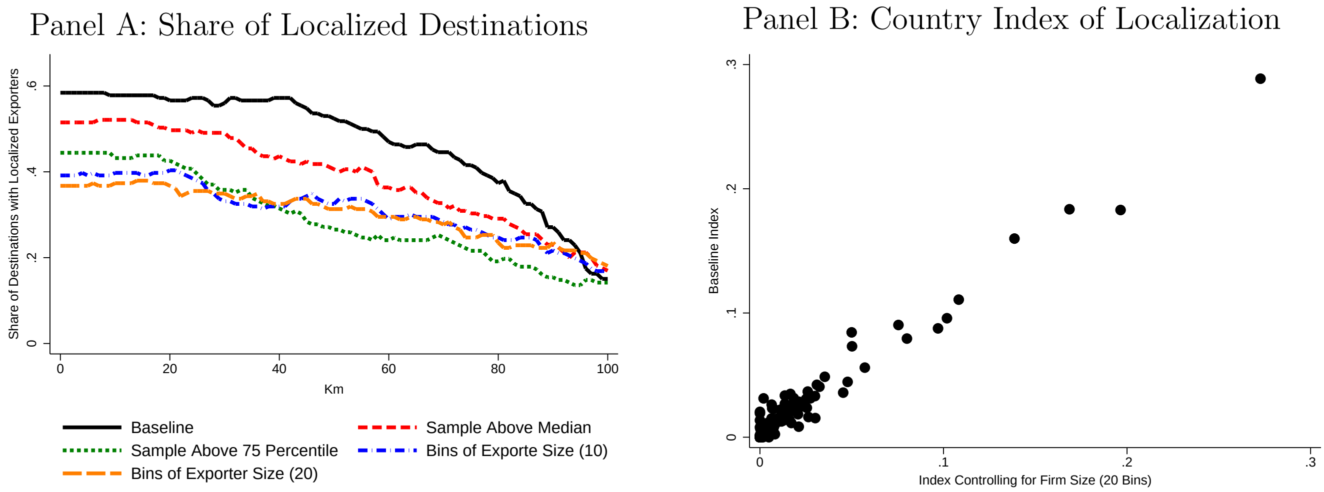
Figure II shows the results of these tests. Focusing on the sample of the largest exporters (those above percentile 75) reduces the number of agglomerated destinations to 49%, from 64% in the baseline. A similar result arises when we restrict the counterfactual to 10 or 20 bins of total firm exports, see Panel A. Hence, some portion of the localization patterns can be explained by agglomeration of firms of similar size, although this portion is small and the bulk of the baseline results are preserved. Indeed, the destinations for which exporters are no longer significantly agglomerated are those with the lowest degree of agglomeration in the baseline, see the scatter plot in Panel B. This suggests that the lower agglomeration levels can be explained (at least partly)

[7]Each size bin contains the same number of exporters. The median amount of total exports is given by €0.25 million and the percentile 75 by €1.4 million.

by the smaller populations from which the random samples of the counterfactual simulations are drawn. For example, the median number of exporters per industry is 626, whereas the median number of exporters per industry and size bin is 37. Hence, it is reassuring that a more stringent counterfactual preserves the general findings of significant agglomeration by export destination, in this case to half of them.

Figure II

Localization Controlling for Exporter Size



Notes: This figure compares the baseline results of agglomeration by export destination with those controlling for the size of exporters. Panel A shows the share of localized destinations at each level of distance. The red dashed and the green short-dashed lines are built from samples of the largest exporters: those above the median and the 75 percentile of total firm exports, respectively. The blue dash-dotted and the orange long-dashed lines are obtained from counterfactuals that control for 10 and 20 bins, respectively, of total firm exports (as well as the industry composition of exports). Panel B displays the correlation between the country index of agglomeration and that from the simulations accounting for 20 bins of exporter size. The value of this correlation is 0.98.

3.3.2 Controlling for Sequential Exporting

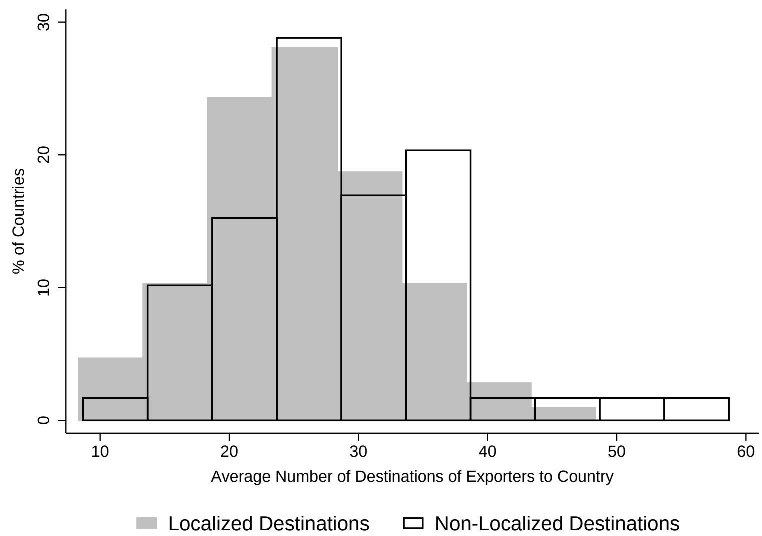
In our empirical specification each destination is treated independently. Given that some firms export to more than one country, this creates a possible dependence of agglomeration patterns across certain markets. Indeed, a large literature has documented that firms go hierarchically to more and more destinations, see for example Albornoz et al. (2012) and Chaney (2014). Therefore, if exporters to one destination are agglomerated and a subset of these exporters disproportionately export to another destination, this destination will also exhibit agglomeration.

We partially addressed this concern when conditioning the counterfactual to firm-size bins, as exporters of the same size are more likely to share some destinations. Here we explore further this issue. First, we inspect if the average number of export destinations of exporters to countries that exhibit significant agglomeration is higher than that of exporters to countries that do not exhibit agglomeration. Finding so would suggest that the large number of agglomerated destinations could be the result of these countries attracting exporters that sell to many (possibly agglomerated) countries. Second, we restrict the counterfactual in the spirit of the previous subsection in order to control for the ability of exporters to sell to more and more difficult destinations.

Regarding the first analysis, we find that the distribution of the average number of export destinations of exporters to agglomerated countries lies *to the left* of the distribution of the average number of export destinations of exporters to countries that do not exhibit agglomeration, see Figure III. For example, the average number of export destinations of exporters selling to countries that exhibit agglomeration is on average 25.1, whereas this statistic is 27.8 in countries that do not exhibit agglomeration. This provides evidence, at least to a first approximation, against the concern that agglomerated destinations are so because exporters selling there also sell to many other countries.

Figure III

Average Number of Destinations of Exporters Selling to Localized vs Non-Localized Destinations



Notes: This figure shows the cross-destination distribution of the average number of destinations reached by exporters. The white bars correspond to countries that exhibit significant agglomeration whereas the gray bars to countries that do not exhibit significant concentration.

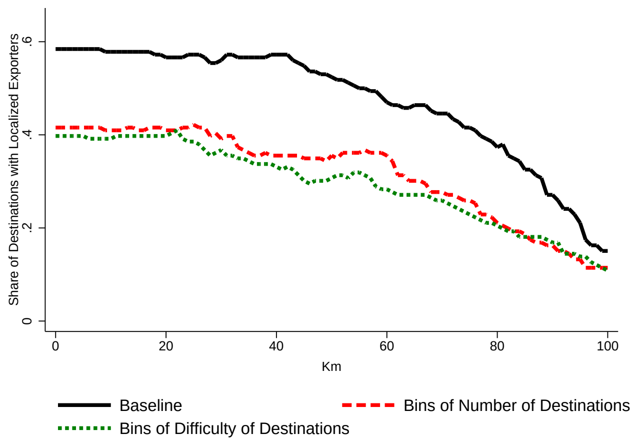
Regarding the second analysis, we carry out two additional robustness checks. First, we condition the counterfactual on two bins: exporters selling to only one destination (which represents roughly 40% of exporters) and exporters selling to more than one country. And second, we condition on how 'difficult' are the destinations each exporter sells to. We proceed as follows, we rank the destinations in our sample from the highest to the lowest number of exporters, that is, from the most popular to the least popular destinations. We then split the sample in five quintiles, the fist quintile containing the most popular destinations and the top quintile the least popular. We then assign each exporter to the quintile of the most difficult destination it sells to. Therefore, firms selling to difficult countries are assigned to the top quintile whereas exporters reaching only the most popular destinations are assigned to the first quintile. Hence, we condition the counterfactual to exporters of the same industry *and* of the same ability to reach difficult destinations.

Figure IV shows that accounting for sequential exporting gives similar results as controlling for firm-size bins. Panel A shows that 40% of destinations exhibit significant agglomeration at small distances, a number that gradually declines across distances. This is a lower share of localized destinations as compared to the baseline, as it was the case in the previous subsection. Reassuringly, we find that the relative agglomeration patterns across countries is preserved. Panel B displays a high correlation between the baseline country index of agglomeration and that accounting for bins of difficulty of destinations. Those countries exhibiting the largest amount of agglomeration in the baseline are still those with the highest levels in the more restricted counterfactual. It is worth noting also that the extent of agglomeration is reduced: the average index of significantly agglomerated destinations is around two thirds that of the baseline. This suggests that sequential exporting can be relevant in determining the extent of agglomeration by export destination, yet the the smaller samples from which the counterfactual simulations are drawn may too partly explain this result.

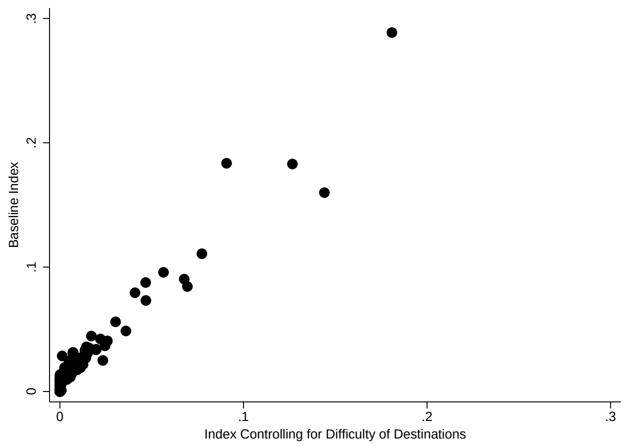
Figure IV

Localization Controlling for Sequential Exporting

Panel A: Share of Localized Destinations



Panel B: Country Index of Localization



Notes: This figure compares the baseline results of agglomeration by export destination with those controlling for sequential exporting. Panel A shows the share of localized destinations at each level of distance. The red dashed line is obtained from a counterfactual that controls for two groups of firms according to the number of countries they sell to: one or more than one. The green dotted line is obtained from a counterfactual of 5 bins according to the most difficult destination each exporter is able to sell to. The degree of difficulty of each destination is assessed according to the total number of exporters. Both counterfactuals account also for the industry composition of exports. Panel B displays the correlation between the country index of agglomeration and that from the simulations accounting for bins of difficulty of destinations. The value of this correlation is 0.97.

3.3.3 Controlling for Destinations Served by Exporters in Large Cities

It may be the case that exporters located in large cities are able to reach a larger set of export destinations, for instance by using better transport facilities, such as airports. Then, if some exporters concentrate in large cities and large cities disproportionately export to difficult destinations, exporters to these destinations will be spatially agglomerated. If this is the case, the destination-specific agglomeration economies would not exist *within* cities of comparable size.

We perform two tests to address this concern. First, we exclude from our sample those firms located in Madrid and Barcelona, which are the largest municipalities in Spain, accounting for about 13% of export firms. This precludes the possibility that agglomeration by destination is driven by exporters located in the largest cities. Second, we control for the size of each city where the exporter is located when building the counterfactual distributions. We construct city-size bins with cutoffs given by 10 thousand, 100 thousand, 250 thousand and 1 million people.[8] We then restrict the counterfactual to exporters in the same industry *and* the same population bin, replicating for each destination the distribution of these two variables observed in the data.

Panel A of Figure V shows that excluding firms located in the largest municipalities has a small effect on the results. At low distances the share of localized destinations does not change, whereas at larger distances agglomeration is higher. Indeed, 60% of destinations exhibit agglomeration until around 70 km, which is a larger scale of agglomeration than is found in the baseline. Accounting for city-size bins reduces the extent of agglomeration at small distances, although to a limited extent. The total number of localized destinations goes down from 64% in the baseline to 62%. Panel B compares the country index of agglomeration with that of the baseline. We find that they are very highly correlated (0.98) and most of the countries lie very close to the 45° line. Indeed, those destinations for which agglomeration is no longer significant are those that exhibit very low levels in the first place. Hence, it seems that the baseline results cannot be explained by firms located in the largest municipalities or in locations of certain size.

3.3.4 Stability of the Results Over Time

In this subsection we check the stability of the results over time. A recent strand of literature documents that a large portion of exporter-destination relationships are short-lived, see Besedes and Prusa (2006), Nitsch (2009) and Békés and Muraközy (2012). This raises the concern that the agglomeration patterns uncovered so far could be fairly volatile over time.

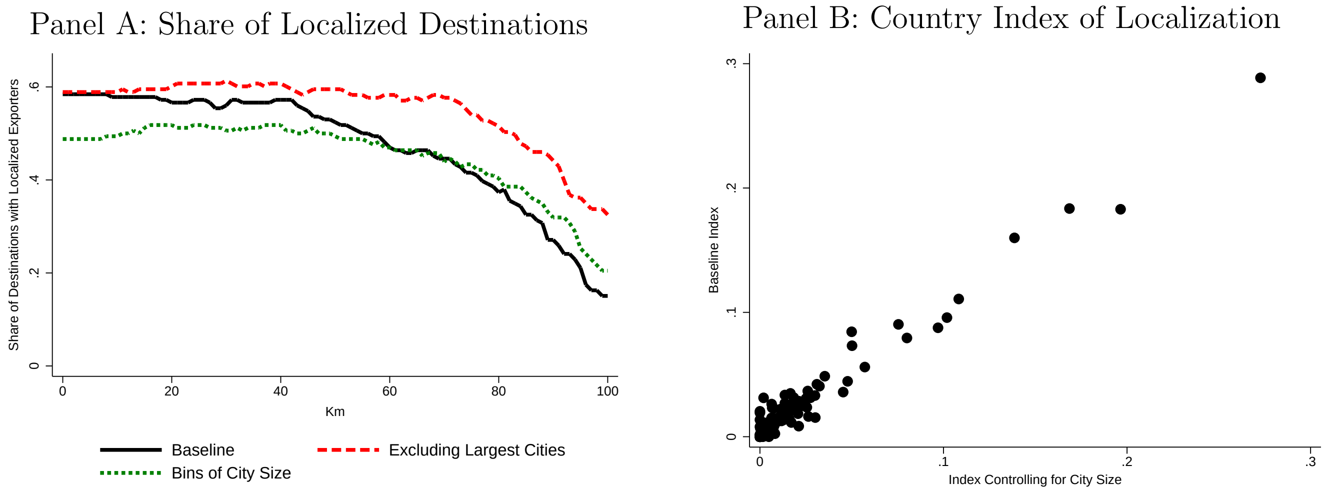
We carry out two tests to address this issue. First, we repeat the baseline analysis for the years 2003 and 2005. And second, we restrict the baseline sample to continuous exporters from 2005 to 2007. Note that this exercise involves restricting the sample to those firm-country relationships that exist during three consecutive years (2005, 2006 and 2007). This criterion implies dropping around 40% of the exporters in the original sample. Note also that the agglomeration patterns by destination are computed only for continuous exporters to that destination, the counterfactual being random draws of continuous exporters to that and other destinations that satisfy the industry criterion. This implies that the number of destinations is reduced to 121.

Figure VI shows the results. The patterns of agglomeration by destination hold broadly across

[8] The distribution of exporters across these bins is: 30%, 32%, 15%, 11% and 13%.

Figure V

Localization Controlling for City Size



Notes: This figure compares the baseline results of agglomeration by export destination with those controlling for the size of cities where exporters are located. Panel A shows the share of localized destinations at each level of distance. The red dashed line corresponds to the sample excluding exporters located in Madrid and Barcelona. The green short-dashed line is built from a counterfactual that controls for 5 bins of city size (as well as the industry composition of exports). Panel B displays the correlation between the baseline country index of agglomeration and that from the simulations controlling for city-size bins. The value of this correlation is 0.98.

years, being the share of localized destinations very similar in 2003, 2005 and 2007 (see Panel A). This is also true for the sample of continuous exporters between 2005 and 2007. The share of localized destinations is 60% vs. 64% in the baseline and the scale of agglomeration at short distances is very similar. This highlights that agglomeration by a significant amount of export destinations is also a feature of the exporters most able to establish permanent trade relationships. In terms of the country index of agglomeration, the results over time are also relatively stable. The correlation between the 2007 index and that of 2005 and 2003 is 0.66 and 0.75, respectively. The Panel B shows the scatterplot between the baseline index and that of the continuous exporters sample. Again, the correlation is significantly high (0.71).[9]

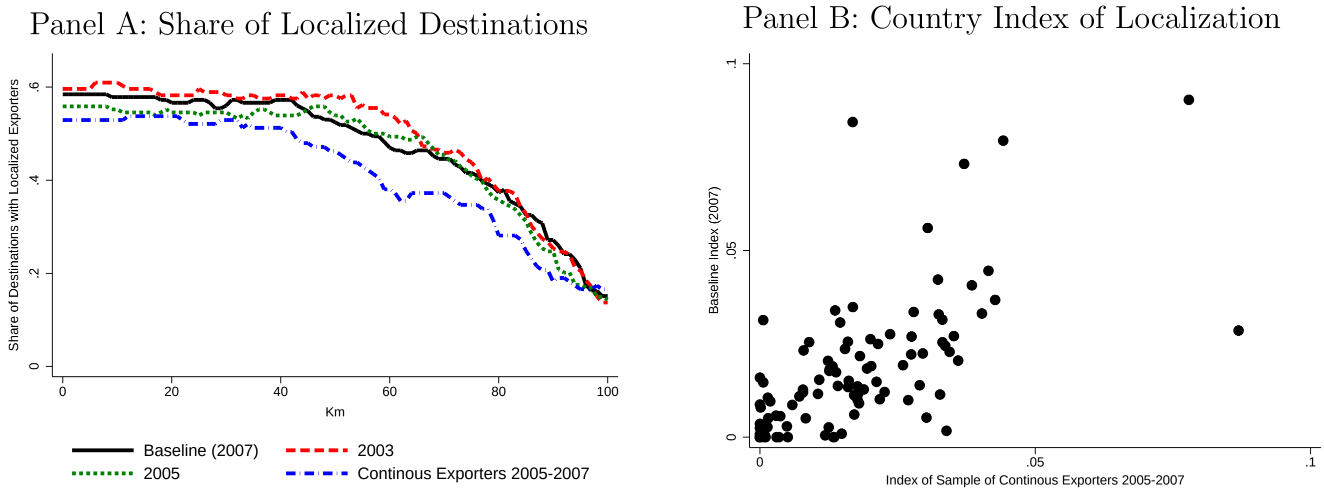
3.3.5 Additional Robustness Checks

In our counterfactual simulations, we controlled for the industry composition of exports up to two-digit industries. This level of aggregation might be too coarse if countries demand specific varieties of a product that are locally produced. This concern could be partly addressed by using detailed enough product codes such as HS-6, yet our data do not contain such information. As an additional robustness check, we controlled in the counterfactual simulations for the industry composition of exports up to four-digit industries, at the cost of losing degrees of freedom when building the counterfactual. We find that the number of destinations with agglomerated exporters is 57%, slightly less than in the baseline but still significantly high. Moreover, the patterns of

[9]Note that in order to keep a meaningful scale this figure excludes two destinations with highly localized exporters in both the baseline and the continuous exporters sample.

Figure VI

Localization Across Different Years



Notes: This figure shows agglomeration by export destination across different years and for the sample of continuous exporters in 2007 (i.e. exporter-destination relationships that exist during 2005, 2006 and 2007). Panel A shows the share of localized destinations at each distance. Panel B displays the correlation between the baseline country index of agglomeration and that obtained from the continuous exporters sample. The value of this correlation is 0.71. Note that this panel excludes two destinations with highly localized exporters in both the baseline and the continuous exporters sample, in order to keep a meaningful scale.

agglomeration resemble that of the baseline. The share of localized destinations is close to 50% until about 40 km, decreasing fast from that distance on, see Panel A of Figure VII.

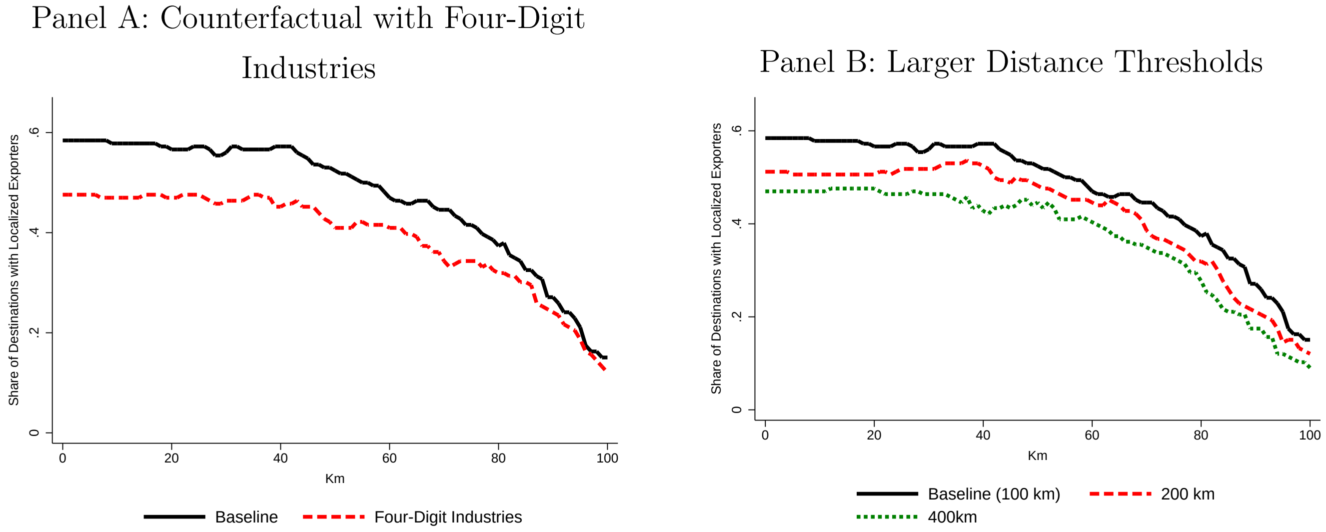
Finally, we also checked if the agglomeration patterns we document vary if we extend the distance horizon at which the distance distribution and the counterfactual are compared. Note that DO focus on 180 km, which corresponds approximately to the median distance of manufacturing plants in the United Kingdom, and Ellison et al. (2010) provide results for the US on thresholds ranging from 100 to 1,000 miles. Panel B of Figure VII shows that extending the distance horizon to 200 and 400 km from the 100 km baseline deliver similar agglomeration patterns. Moreover, the country indices of agglomeration have a large correlation with the baseline, of 0.99 and 0.93, respectively.

4 Factors Behind Agglomeration by Destination

In this section we aim at uncovering the variables that explain the cross-destination variation in agglomeration levels documented above. The process of selling overseas involves both fixed and variable costs, such as learning about the foreign market, establishing a distribution network, tailoring the products to foreign tastes and regulations, clearing the goods through customs, etc. Therefore, the industry agglomeration sources emphasized by the literature, related to sharing, matching, and learning mechanisms (see for instance Duranton and Puga (2004)) may too lead, at least to some extent, to agglomeration by export destination. Indeed, there exist potential gains from pooling the costs of selling abroad and from extracting the information revealed by nearby

Figure VII

Additional Robustness Checks



Notes: This figure shows additional robustness checks on the baseline results. Panel A shows agglomeration by export destination at each distance when the counterfactual controls for four-digit industries vs two-digit industries in the baseline. Panel B extends the distance horizon at which the density and the counterfactual distributions are compared to 200 and 400 km vs. 100 km in the baseline.

exporters as shown, for instance, by Segura-Cayuela and Vilarrubia (2008) and Fernandes and Tang (2014). Agglomeration may also be the result of geographical proximity to the importing country, which would lead to export firms being concentrated close to the border.

In order to shed some light on the determinants of agglomeration by export destination, we regress the country index of agglomeration on a set of country characteristics borrowed from the gravity literature, which connects trade flows with size and trade barriers, see Helpman et al. (2008). If agglomeration results in lower transport costs, this literature provides a natural guideline to assess the cross-destination heterogeneity of exporter agglomeration. Then, we include proxies of geographical distance, cultural dissimilarity, transaction costs and institutions among other determinants. In order take into account our left-censored dependent variable, we specify the following Tobit model:

$$\begin{aligned}\text{Localization}^*_c &= \beta_0 + \beta_1 \text{Spanish}_c + \beta_2 \text{Institutional Quality}_c + \beta_3 \text{Euro}_c \\ &\quad + \beta_4 \text{Contiguity}_c + \beta_5 \text{Log Distance to Capital}_c \\ &\quad + \beta_6 \text{Log Per Capita GDP}_c + \beta_7 \text{Log Number of Exporters}_c \\ &\quad + \beta_8 \text{Log Population}_c + \epsilon_c \qquad (1)\end{aligned}$$

$$\text{Localization}_c = \max\left(0, \text{Localization}^*_c\right) \qquad (2)$$

Our variables capturing trade barriers are the following. Cultural similarity is proxied by a dummy variable taking value one if Spanish is spoken in the destination country, and zero otherwise. The institutional environment in the importing country is captured by the principal factor of the six dimensions comprising the Worldwide Governance Indicators, namely rule of law, political stability, control of corruption, government effectiveness, regulatory quality, and voice and accountability (see Appendix C for details). We also include a dummy taking value one if the euro is the currency of the destination. Finally, we add the log average distance between the exporters and the destination's capital and a dummy taking value one if the destination shares a border with Spain. In a different specification we include the average distance of exporters to the closest port from which shipments are sent to the country.[10] Last, we add a set of controls including per capita GDP, population, and the log number of firms exporting to the country. The latter is itself a function of export costs. We include it in order to facilitate the reading of the results: we compare the extent of exporter agglomeration across destinations reached by the same number of firms, and assess how this varies according to characteristics such as language, institutions and distance.[11] Note finally that we rescale the dependent variable by its standard deviation in order to ease the interpretation of the coefficients.

Table III shows the results. In column (1) we find that conditional on the rest of the covariates exporters to countries with a different language, a different currency and a worse business environment are significantly more geographically agglomerated. The most significant variable is language: if Spanish is spoken in the importing country, agglomeration decreases by 0.89 standard deviations.[12] In column (2) we restrict the sample to the destinations not belonging to the European Union, where entry is more difficult and hence the value provided by agglomeration is potentially larger. In this specification language and the institutional environment become more important in explaining the concentration of exporters. Speaking Spanish reduces the extent of agglomeration by 1.3 standard deviations and a one standard deviation increase in the business environment is associated to a 0.66 standard deviations decrease in agglomeration, both results pointing to a positive relationship between agglomeration and the difficulty in conducting businesses abroad.[13] Our findings in columns (1) and (2) also suggest that distance plays no role in

[10]We do not include it in the baseline because we lack data on exports from Spanish ports for 13 countries and the distance variables are never significant. Also, including the simple distance between the most populated cities yields very similar results.

[11]Moreover, Helpman et al. (2008) shows the importance of accounting for the extensive margin of trade in the gravity equation framework. We also checked that excluding this variable does not affect the overall results.

[12]Note that our baseline regression is performed on 150 countries because we lack data on 16 small countries. GDP data are missing in 14 countries and institutional quality in 6 observations.

[13]We also inspected the role of some specific elements of the institutional environment by replacing the institutional factor in column (1) by each Governance Indicator (one by one). We found that a better rule of law, less corruption, and more political stability are significantly associated to lower agglomeration, whereas the rest yield the expected signs (better institutional quality associated to lower agglomeration) but the estimates are not significant. Moreover, we also found that specific measures of investors protection, such as the number of procedures required to enforce a contract, are also negatively and significantly associated to agglomeration. We also tried another proxies of import costs. For example, we found that the higher the number of days and the number of

explaining the concentration of exporters. In column (3) we replace the exporters average distance to the country's capital with the average distance to the closest port shipping to the country. This measure does not have either explanatory power in accounting for exporter concentration.

In column (4) we quantify the role of immigrants in explaining agglomeration across countries. Specifically, we test whether the local concentration of immigrants can explain some of the patterns that we document. A line of research has shown that immigrants help overcome trade barriers, for example by providing specific knowledge about their home countries. For instance, Herander and Saavedra (2005) find an effect of local immigrant groups on export volumes in the US. In order to delve into this issue, we construct an origin-specific index of immigrant dispersion, defined as the median distance between immigrants from each country (a higher distance meaning more dispersion). Our analysis is restricted to the only 28 countries of which we have information on the population distribution across municipalities, therefore we raise a flag of caution on interpreting the results. With this caveat in mind, column (4) shows that there is a significant relationship between the dispersion of immigrants and the agglomeration of firms selling to their home countries. Conditional on the rest of controls, a 10% increase in the dispersion of immigrants is associated to a 0.53 standard deviation decrease in the degree of agglomeration. Therefore, agglomeration is higher in countries whose immigrants exhibit some concentration.

In column (5) we add region fixed effects. We include 10 region dummies: Western Europe, Eastern Europe, Western & Central Asia, South-East Asia, Northern Africa, Central & Southern Africa, North America, Central America & Caribbean, South America, and Oceania. We find that the coefficient on language barely changes with respect to the baseline, although it is imprecisely estimated, whereas that of institutional quality is somewhat lower, though it is still statistically significant. Interestingly, controlling for regions increases the estimated effect of the currency: belonging to the euro area reduces the extent of agglomeration by 0.90 standard deviations. These results suggest that the mechanisms connecting agglomeration with trade barriers hold also within broad geographic and economic areas.

Finally, column (6) replaces the baseline country index of agglomeration with that obtained from the counterfactual that accounts for 20 bins of exporter size, see Section 3.3.1. This is a pertinent analysis because restricting the counterfactual in some cases reduced the number of significantly agglomerated destinations. However, given the high correlation between the country indices of agglomeration, the results tend to confirm the baseline findings. In fact, the point estimates are even larger in absolute value regarding language, institutional quality and currency. Adding the other country indices constructed in Section 3.3 confirms the baseline estimates.

Overall, the previous results suggest that there exists a relationship between trade barriers to enter a country and the degree of spatial agglomeration of exporters selling to it. One limitation

documents required to import goods, the higher is the level of agglomeration, although these relationships were not statistically significant.

of this approach is that the precise mechanism driving these patterns cannot be uncovered, and we cannot rule out that other omitted factors may contaminate this relationship, hence we do not pursue causality. However, they show an insightful correlation between proxies of export costs and the extent of geographical concentration that is systematic and robust. Moreover, the results suggest that agglomeration can be more effective regarding those destinations from which information is more valuable and, in this regard, they inform the theoretical and empirical literature on export spillovers and learning from neighboring firms cited in the introduction.

5 Concluding Remarks

In this paper we document the existence of agglomeration economies that accrue to firms selling to certain foreign markets. In the pursuit of shedding light on the interpretation of our results, we show that these patterns of geographical concentration are not driven by the spatial location of large vs small exporters, hierarchical exporting, or exporters located in large cities. Moreover, we find that these location patterns are quite stable over time. Regarding the determinants of agglomeration, we show that the cross-destination variability in agglomeration levels can be partly explained by language, currency, and institutional quality, being agglomeration higher the higher the export costs.

These findings are consistent with the existence of externalities in selling to certain foreign countries, having implications for international trade. For example, agglomeration could help reduce destination-specific fixed costs, which would rationalize why firms do not follow a strict hierarchy on export destinations, a fact uncovered by Eaton et al. (2011). Also, some policy implications can be derived. The pattern of concentration by export destination suggests that easing the flow of information from exporters to potential entrants can pay off. Moreover, the fact that concentration is higher concerning more difficult destinations suggests that the benefits of these policies can be specially helpful in countries where entry is harder. Also, helping companies penetrate new markets can lead nearby firms to follow them. Interestingly, given how we defined the counterfactual, these benefits are not bounded to firms of the same industry, but also to firms belonging to different industries.

The nature of our data prevents us from digging deeper into the specific channels through which agglomeration economies might work. More detailed data would allow to disentangle some sources of export spillovers, such as those related to information (via headquarters) from those linked to costs (via establishments). Also, a larger time span and more categories of goods would allow a geographical analysis of new products exported and new markets accessed. In general, we think that there is room in the literature to test empirically which are the most important channels through which agglomeration economies in international trade operate. Case studies or natural experiments seem a suitable framework to disentangle specific mechanisms playing a role

Table III

Factors Behind Exporters' Agglomeration by Destination

	Baseline	Non-EU	Ports	Immigrants	Region FE	Exporters' Size
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: Country Index of Exporters' Localization						
Spanish	-0.8901***	-1.2840***	-1.1375***	-0.9416**	-0.9035	-1.2434**
	(0.3358)	(0.3750)	(0.3200)	(0.3305)	(0.8289)	(0.4834)
Institutional Quality	-0.4756*	-0.6561**	-0.5395**	-0.2008	-0.3968*	-0.7163**
	(0.2437)	(0.3035)	(0.2577)	(0.1622)	(0.2251)	(0.3092)
Euro	-0.6491**		-0.2722	0.3424	-0.9010***	-0.7463*
	(0.3052)		(0.2548)	(0.2001)	(0.3264)	(0.4362)
Contiguity	0.7024		-0.7600	0.6388***	0.6165	1.4376
	(1.0417)		(0.8099)	(0.2175)	(0.9288)	(1.2682)
Log Distance to Capital	-0.1684	0.0043		0.2156	-0.2861	-0.2592
	(0.1914)	(0.2008)		(0.1614)	(0.3398)	(0.2521)
Log Distance to Port			0.0833			
			(0.2651)			
Log Per Capita GDP	0.2849	0.3498	0.3386*	0.7116***	0.1605	0.3276
	(0.1963)	(0.2370)	(0.1914)	(0.2309)	(0.1655)	(0.2079)
Log Number of Exporters	0.1371	0.2096	0.1474	-0.6587**	-0.0083	0.1746
	(0.1701)	(0.2110)	(0.1398)	(0.2321)	(0.2122)	(0.1919)
Log Population	-0.0148	-0.0060	0.0628	0.2162	0.0268	-0.0476
	(0.1111)	(0.1360)	(0.0941)	(0.1322)	(0.1419)	(0.1334)
Dispersion of Immigrants				-2.3675***		
				(0.7900)		
Region Fixed Effects	No	No	No	No	Yes	No
Observations	150	123	141	28	150	150
Pseudo R-squared	0.04	0.04	0.04	0.62	0.10	0.05
Log Likelihood	-201.60	-165.90	-184.60	-8.67	-187.60	-181.40

Notes: This table shows the regression of the country index of exporters' localization (i.e. a variable capturing to what extent exporters to each export destination are significantly agglomerated) against measures of export costs, comparative advantage and several covariates. The specification is a tobit model described in equation (1). Column (1) presents the baseline regression. Column (2) restricts the sample to countries not in the European Union. Column (3) replaces the variable distance with the average distance to the closest port shipping to the country. Column (4) introduces a measure of the concentration of immigrants from each country, proxied as the median distance between them. Column (5) introduces seven region fixed effects: Western Europe, Eastern Europe, Western & Central Asia, South-East Asia, Northern Africa, Central & Southern Africa, Central America & Caribbean, North America, South America, and Oceania. Finally, in column (6) the dependent variable is the country index built from a counterfactual that controls for firm-size bins (Section 3.3.1). Robust standard errors are in parenthesis. Significance levels: *: 10%; **: 5%; ***: 1%.

in generating export spillovers. We see this avenue of further research as promising.

References

Albornoz, F., H. F. Calvo Pardo, G. Corcos, and E. Ornelas (2012): "Sequential Exporting," *Journal of International Economics*, 88, 17–31.

Behrens, K. and T. Bougna (2015): "An Antatomy of the Geographical Concentration of Canadian Manufacturing Industries," *Regional Science and Urban Economics*, 51, 47–69.

Békés, G. and B. Muraközy (2012): "Temporary Trade and Heterogeneous Firms," *Journal of International Economics*, 87, 232–246.

Bernard, A. B., J. Eaton, B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *The American Economic Review*, 93, 1268–1290, bernard03.

Besedes, T. and T. J. Prusa (2006): "Ins, outs, and the Duration of Trade," *Canadian Journal of Economics*, 39, 266–295.

Cadot, O., L. Iavacone, M. D. Pierola, and F. Rauch (2013): "Success and Failure of African Exporters," *Journal of Development Economics*, 101, 284–296.

Cassey, A. J. and K. N. L. Schmeiser (2013): "The Agglomeration of Exporters by Destination," *The Annals of Regional Science*, 51, 495–513.

Chaney, T. (2014): "The Network Structure of International Trade," *The American Economic Review*, 104, 3600–3634.

Choquette, E. and P. Meinen (2014): "Export Spillovers: Opening the Black Box," *The World Economy*.

Devereux, M. P., R. Griffith, and H. Simpson (2004): "The Geographic Distribution of Production Activity in the UK," *Regional Science and Urban Economics*, 35, 533–564.

Duranton, G. and H. G. Overman (2005): "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 72, 1077–1106.

Duranton, G. and D. Puga (2004): "Micro-Foundations of Urban Agglomeration Economies." *Handbook of Regional and Urban Economics*, 4, 2063–2117.

Eaton, J., S. Kortum, and F. Kramarz (2011): "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 79, 1453–1498.

Ellison, G. and E. L. Glaeser (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105, 889–927.

ELLISON, G., E. L. GLAESER, AND W. R. KERR (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100, 1195–1213.

FERNANDES, A. P. AND H. TANG (2014): "Learning to Export from Neighbors," *Journal of International Economics*, 94, 67–84.

HELPMAN, E., M. J. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, CXXIII, 441–487.

HERANDER, M. G. AND L. A. SAAVEDRA (2005): "Exports and the Structure of Immigrant-Based Networks: The Role of Geographic Proximity," *The Review of Economics and Statistics*, 87, 323–335.

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2010): "The Worldwide Governance Indicators Methodology and Analytical Issues," The World Bank. Policy Research Working Paper.

KAUFMANN, D., A. KRAY, AND M. MASTRUZZI (2009): "Governance Matters VIII: Aggregate and Individual Governance Indicators 1996-2008," World Bank Policy Research Working Paper 4978.

KLEPPER, S. (2010): "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics*, 67, 15–32.

KOENIG, P. (2009): "Agglomeration and the Export Decision of French Firms," *Journal of Urban Economics*, 66, 186–195.

KOENIG, P., F. MAYNERIS, AND S. PONCET (2010): "Local Export Spillovers in France," *European Economic Review*, 54, 622–641.

KRAUTHEIM, S. (2012): "Heterogeneous Firms, Exporter Networks and the Effect of Distance on International Trade," *Journal of International Economics*, 87, 27–35.

LOVELY, M. E., S. S. ROSENTHAL, AND S. SHARMA (2005): "Information, Agglomeration, and the Headquarters of U.S. Exporters," *Regional Science and Urban Economics*, 35, 167–191.

MARSHALL, A. (1920): *Principle of Economics*, London: McMillan.

MAUREL, F. AND B. SÉDILLOT (1999): "A Measures of the Geographic Concentration of French Manufacturing Industries," *Regional Science and Urban Economics*, 29, 575–604.

MAYER, T. AND S. ZIGNAGO (2011): "Notes on CEPII's Distances Measures: the *GeoDist* Database," CEPII Working Paper 25.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Nitsch, V. (2009): "Die Another Day: Duration in German Import Trade," *Review of World Economics*, 1, 133–154.

Openshaw, S. (1984): *The Modifiable Areal Unit Problem*, Geobooks. Norwich, England.

Openshaw, S. and P. Taylor (1979): "A Million or so Correlation Coefficients: Three Experiments on the Modifiable Area Unit Problem," *Statistical Applications in the Spatial Sciences*, 127–144, n. Wrigley (ed). London. Pion.

Paravisini, D., V. Rappoport, and P. Schnabl (2015): "Specialization in Bank Lending: Evidence from Exporting Firms," Working Paper.

Segura-Cayuela, R. and J. M. Vilarrubia (2008): "Uncertainty and Entry into Export Markets," Bank of Spain. Documento de Trabajo 0811.

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

Wagner, R. and A. Zahler (2015): "New Exports from Emerging Markets: Do Followers Benefit from Pioneers?" *Journal of Development Economics*, 114, 203–223.

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 the 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.
Government Effectiveness	Kaufmann et al. (2009)	Quality of public services, the civil service, policy formulation and implementation, and credibility of the government's commitment to such policies.
Voice and Accountability	Kaufmann et al. (2009)	Extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
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 the country's capital.
Distance to Ports	Puertos del Estado	Distance between zip code and closest port from which shipments 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.

Table A.1
Data Definitions and Sources (Continued)

VARIABLE	SOURCE	DEFINITION
Location of Immigrants	Instituto Nacional de Estadística	Number of immigrants in each municipality by country of origin.
European Union		1 if country belongs to the European Union.
Doing Business Index	World Bank	Ranking of economies that assess business regulations and their enforcement.
Contract Enforcement	World Bank	Number of procedures required to enforce a contract.
Time to Import	World Bank	Number of calendar days necessary to comply with all the procedures required to import goods.

Notes: This table shows definitions and sources of the main variables used throughout the paper.

Table A.2
List of Destinations

Country	N	Localization	Country	N	Localization	Country	N	Localization
Afghanistan	13	0.0000	Gambia, The	20	0.0000	Netherlands	3010	0.0190
Albania	168	0.1108	Georgia	101	0.0025	New Caledonia	34	0.0000
Algeria	826	0.0027	Germany	6427	0.0174	New Zealand	278	0.0036
Andorra	887	0.0903	Ghana	72	0.0000	Nicaragua	63	0.0000
Angola	89	0.0000	Gibraltar	90	0.0000	Niger	44	0.0000
Antigua and Barbuda	20	0.0000	Greece	1768	0.0178	Nigeria	99	0.0560
Argentina	818	0.0100	Guam	10	0.0000	Norway	733	0.0004
Armenia	53	0.0794	Guatemala	226	0.0000	Oman	103	0.0017
Aruba	19	0.0876	Guinea	19	0.0000	Pakistan	189	0.0407
Australia	795	0.0128	Haiti	13	0.0000	Panama	492	0.0000
Austria	1620	0.0148	Honduras	107	0.0000	Paraguay	67	0.0193
Azerbaijan	28	0.0024	Hong Kong SAR, China	708	0.0315	Peru	435	0.0001
Bahamas, The	48	0.0000	Hungary	778	0.0276	Philippines	191	0.0096
Bahrain	154	0.0329	Iceland	160	0.0001	Poland	1627	0.0182
Bangladesh	67	0.0285	India	643	0.0232	Portugal	6861	0.0000
Barbados	32	0.0000	Indonesia	228	0.0136	Qatar	212	0.0445
Belarus	89	0.0000	Iran, Islamic Rep.	463	0.0109	Romania	902	0.0205
Belgium	3500	0.0217	Iraq	23	0.1835	Russian Federation	915	0.0189
Belize	79	0.0089	Ireland	1308	0.0000	San Marino	19	0.0000
Benin	26	0.0000	Israel	732	0.0307	Saudi Arabia	854	0.0134
Bermuda	16	0.0000	Italy	5166	0.0116	Senegal	86	0.0000
Bolivia	96	0.0000	Jamaica	48	0.0085	Serbia	178	0.0185
Bosnia and Herzegovina	103	0.0422	Japan	829	0.0222	Seychelles	31	0.0000
Brazil	1048	0.0269	Jordan	358	0.0154	Sierra Leone	15	0.0313
British Virgin Islands	124	0.0087	Kazakhstan	102	0.0000	Singapore	429	0.0335
Bulgaria	511	0.0151	Kenya	81	0.0000	Slovak Republic	447	0.0099
Burkina Faso	30	0.0000	Korea, Dem. People's Rep.	38	0.0000	Slovenia	412	0.0224
Cabo Verde	22	0.0000	Korea, Rep.	561	0.0271	Solomon Islands	11	0.0000
Cameroon	47	0.0000	Kuwait	335	0.0331	South Africa	729	0.0262
Canada	848	0.0057	Kyrgyz Republic	188	0.0114	Sri Lanka	50	0.0000
Cayman Islands	20	0.0000	Latvia	679	0.0255	Sudan	43	0.0163
Chad	18	0.1599	Lebanon	494	0.0139	Suriname	18	0.1829
Chile	872	0.0001	Libya	152	0.0127	Swaziland	21	0.0359
China	1007	0.0102	Liechtenstein	56	0.0159	Sweden	1264	0.0051
Colombia	642	0.0117	Lithuania	642	0.0137	Switzerland	2095	0.0112
Congo, Dem. Rep.	19	0.0000	Luxembourg	282	0.0080	Syrian Arab Republic	166	0.0339
Congo, Rep.	15	0.0000	Macao SAR, China	34	0.0010	Taiwan, China	458	0.0229
Costa Rica	274	0.0000	Macedonia, FYR	66	0.0732	Tanzania	42	0.0844
Croatia	328	0.0147	Madagascar	23	0.0000	Thailand	343	0.0204
Cuba	314	0.0000	Malaysia	281	0.0026	Togo	17	0.0000
Cyprus	708	0.0236	Mali	22	0.0000	Trinidad and Tobago	97	0.0000
Czech Republic	1080	0.0245	Malta	247	0.0121	Tunisia	743	0.0061
Cte d'Ivoire	76	0.0009	Marshall Islands	10	0.0011	Turkey	1385	0.0367
Denmark	1286	0.0184	Mauritania	56	0.0000	Uganda	13	0.0000
Dominica	80	0.0000	Mauritius	65	0.0000	Ukraine	311	0.0052
Dominican Republic	456	0.0000	Mexico	1964	0.0051	United Arab Emirates	962	0.0086
Ecuador	337	0.0008	Micronesia, Fed. Sts.	30	0.0487	United Kingdom	5521	0.0120
Egypt, Arab Rep.	588	0.0254	Moldova	45	0.0000	United States	3882	0.0106
El Salvador	128	0.0000	Monaco	99	0.0348	Uruguay	247	0.0091
Equatorial Guinea	60	0.0000	Montenegro	61	0.0958	Uzbekistan	10	0.0000
Estonia	380	0.0254	Morocco	1703	0.0000	Venezuela, RB	675	0.0056
Ethiopia	28	0.0000	Mozambique	19	0.0000	Vietnam	128	0.0005
Finland	897	0.0029	Namibia	38	0.0154	Virgin Islands (U.S.)	22	0.0068
France	8946	0.0250	Nauru	15	0.0000	West Bank and Gaza	36	0.2886
French Polynesia	20	0.0000	Nepal	12	0.0000	Yemen, Rep.	86	0.0313
Gabon	41	0.0000						

Notes: N denotes the number of exporters and Localization is the country index of localization, which measures the amount of geographical concentration exhibiting exporters to each destination.

B Details on the Application of Duranton and Overman (2005)

In this section we explain in detail the application of DO to uncover agglomeration by export destination. We proceed as follows. For each export destination we compute the unique bilateral distances between exporters by applying the haversine formula to the zip code coordinates. Next, we estimate the distribution of bilateral distances of each country via kernel estimation. As in DO, we use a Gaussian kernel, choosing the bandwidth so as to minimize the mean integrated squared error. Distances are reflected around zero, using the method proposed by Silverman (1986) in order to avoid giving positive densities to negative distances. Note also that firms within the same zip code are computed as being separated by 0 kilometers.

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

$$\hat{K}_c(d) = \frac{2}{n_c(n_c-1)h} \sum_{i=1}^{n_c-1} \sum_{j=i+1}^{n_c} f\left(\frac{d-d_{i,j}}{h}\right) \tag{3}$$

where n_c is the number of export firms to country c, h is the bandwidth and f is the Gaussian probability density function.[14] The Panel A of Figure VIII shows the spatial distribution of exporters to India in 2007. The existence of several clusters of exporters is apparent. The Panel B plots the histogram and the kernel estimation of the distance distribution. The high density at very small distances stems from the large number of exporters that are located within very close zip codes. The second peak in the distribution at around 400 kilometers marks the distance that separates the clusters.

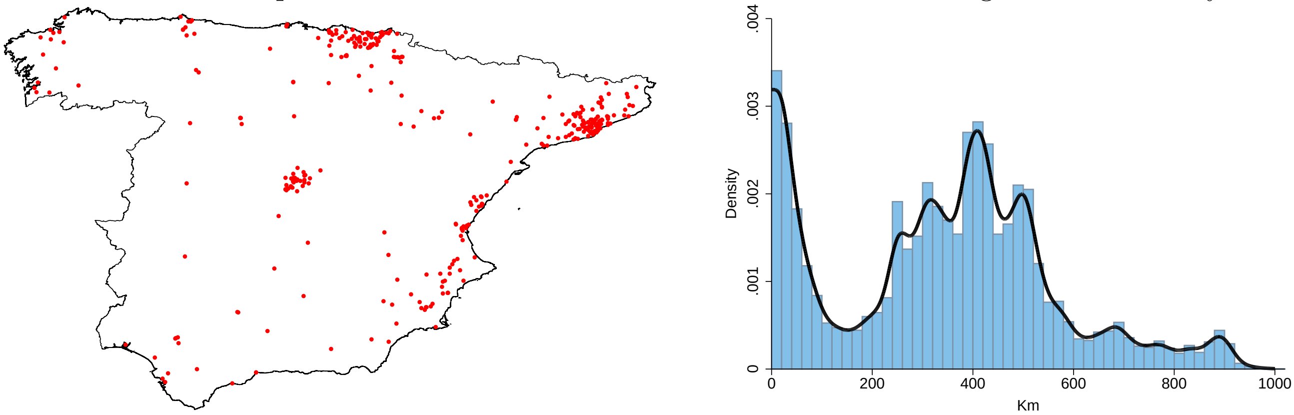
In order to test for significant agglomeration, the observed spatial distribution is compared with the counterfactual. As stated in the main text, the counterfactual controls for both the spatial distribution of exporters, which may be agglomerated with respect to domestic firms, as well as the industry composition of exports to each country. We proceed as follows. For each pair of country and two-digit industry, we draw 1,000 random samples from exporters in that industry; each draw of size the actual number of exporters to that country operating in that industry. Then, for each country we aggregate each draw across the different industries to collect 1,000 random samples of size n_c (the actual number of exporters to the country) with an industry composition that replicates the one observed in the data.

[14] Given that our aim is to uncover agglomeration of export firms, we treat each firm as one observation and therefore we do not weigh the distances when estimating the distribution. An alternative would be consider the agglomeration of export values and hence weigh distances by exports to the country. There are three reasons that advise us against this strategy. First, the construction of the counterfactual involves firms that do not export to the country, hence they lack a weight. Second, if total firm exports were used as weights, we would disregard the specialization of firms to certain markets, given that weights would not have variation across export destinations. And third, total exports is a highly skewed variable, hence a few firms could distort the results.

Figure VIII

Distribution of Distances of Firms Exporting to India

Panel A: Spatial Distribution | Panel B: Histogram and Density



Notes: This figure plots the spatial distribution of exporters to India in 2007 (panel A) and the histogram of the unique bilateral distances between them as well as the kernel estimate of the probability density function (panel B).

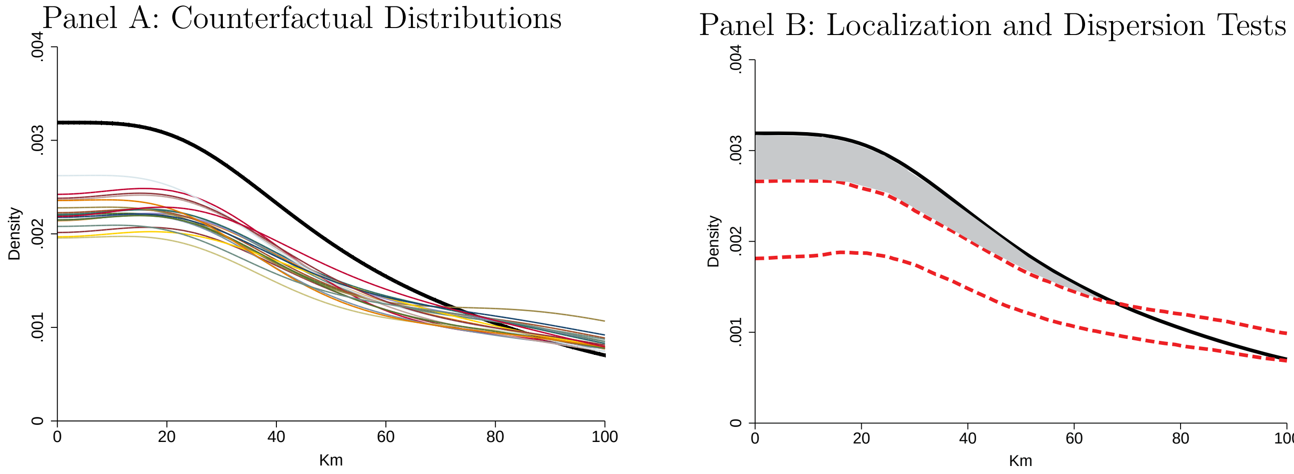
In our baseline analysis we carry out the test of significant agglomeration at distances below 100 km. As in DO we construct two tests, one of localization and one of dispersion, both with a significance level of 95%. We do the following. For each kilometer, we rank our 1,000 counterfactual distributions in ascending order and then pick the percentile that makes 95% of the counterfactual distributions lie below it across all distances. When it is not possible to find a percentile making exactly 95% of the simulations be below it, we use linear interpolation. Note that in our data all the percentiles fullfilling this criterion range between 96.4 and 99.5. This percentile is referred to as the localization threshold, whereas a dispersion threshold is defined in a similar way, i.e. the percentile that makes 5% of the counterfactual distributions lie below it across all distances. Note that in DO the localization and dispersion thresholds are referred to as global confidence bands. The Panel A of Figure IX plots the density of the distance distribution below 100 km and a small sample of the counterfactual distributions. The Panel B displays the localization and dispersion thresholds (the upper and lower dashed-lines, respectively).

We define exporters to a destination to be significantly localized if the distance distribution is above the localization threshold in at least one kilometer. Similarly, exporters to a destination are defined to be dispersed if the distance distribution is below the dispersion threshold in at least one kilometer and the country does not exhibit localization. Note that the last condition stems from the fact that densities sum up to one, hence localization at some distances implies dispersion at others. Following this criteria, the panel B of Figure IX shows that exporters to India are significantly localized. Note finally that localization and dispersion can be assessed at each distance, by comparing the distance distribution and the thresholds at each kilometer. In the example, localization takes place at distances below 70 km.

Finally, we define a country index of agglomeration as the sum across distances of the difference

Figure IX

Localization and Dispersion Thresholds of India



Panel A of Figure IX shows the estimated density of the distances of exporters to India in 2007 as well as a sample of 20 distributions of distances of potential exporters to India in this year -referred to as random distributions-. A potential exporter is defined as an export firm that operates in an industry in which at least one exporter is selling to India. Panel B displays the tests of localization and dispersion, represented by the upper and lower dashed lines, respectively. 95 and 5 per cent of the 1,000 simulated random distributions lie below the upper and lower dashed lines, respectively. See Section 3 for details.

between the distance distribution and the localization threshold if the former is above the latter and zero otherwise. This index accounts for the amount of exporter agglomeration by destination and it is the counterpart of the industry index of agglomeration defined by DO. In panel B of Figure IX it is depicted as the shaded area between the distance density and the upper dashed-line.

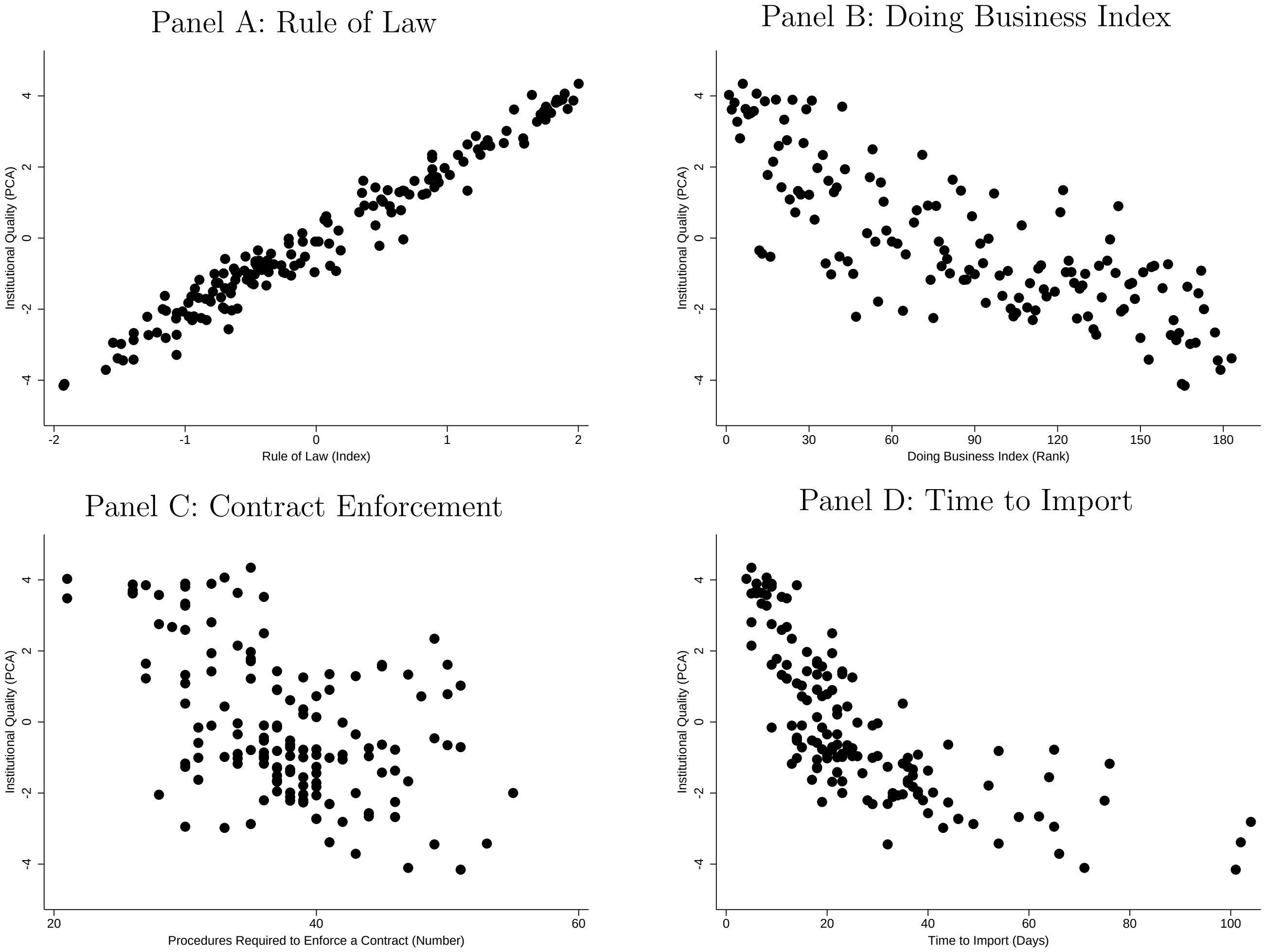
C Details on the Principal Component Analysis

In Section 4 we create an index of institutional quality by applying a principal component analysis (PCA) on the World Bank Worldwide Governance Indicators (WGI). These measures account for six dimensions of governance, namely voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption, see Kaufmann et al. (2010). Given the high correlation between these measures, a PCA is a useful tool to summarize all the information and construct a synthetic index that accounts for the overall institutional quality in each country. The PCA finds a set uncorrelated linear combinations of the measures that accounts for most of the variance. In our case, the first of such combinations accounts for 88% of the variance and constitutes the index of institutional quality that we include in Table III.

Figure X plots the correlation of the institutional quality index with the rule of law (WGI), the Doing Business Index, investor protection, and time to import goods. The high correlations suggest that our index of institutional quality provides a good approximation of the business environment that firms face when exporting to every foreign country.

Figure X

Correlation of the Synthetic Index of Institutional Quality (PCA) with Other Measures



This figure shows the correlation of our synthetic index of institutional quality, computed from a PCA on the Worldwide Governance Indicators, on other measures proxying the institutional environment of every country.