Is the Border Effect an Artefact of Geographic Aggregation?

by

Carlos Llano-Verduras
Asier Minondo
and
Francisco Requena-Silvente

March 2011

Working Papers in Applied Economics

WPAE-2011-08

Departamento de Estructura Económica (Economía Aplicada II), Facultad de Economía, Universitat de València
Avda dels Tarongers, s/n, 46022 – Valencia, Spain
Phone +34 96 382 83 49 / Fax +34 96 382 83 54
http://www.estructuraeconomica.es
e-mail: investigacion.estructura@uv.es

ISSN 2172-3036

Is the Border Effect an Artefact of Geographic Aggregation?

Carlos Llano-Verduras (Universidad Autónoma de Madrid, Spain)
Asier Minondo (Deusto Business School, Spain)
Francisco Requena-Silvente (Universidad de Valencia, Spain)[1]

Abstract

The existence of a large border effect is considered as one of the main puzzles of international macroeconomics. We show that the border effect is, to a large extent, an artefact of geographic concentration. In order to do so we combine international flows with intra-national flows data characterised by a high geographic grid. At this fine grid, intra-national flows are highly localised and dropping sharply with distance. The use of a small geographical unit of reference to measure intra-national bilateral trade flows allows to estimating correctly the negative impact of distance on shipments. When we use sector disaggregated export flows of 50 Spanish provinces in years 2000 and 2005 split into inter-provincial and inter-national flows, we find that the border effect is reduced substantially and even becomes statistically not different from zero in some estimations.

JEL Codes: F14, F15

Keywords: border effect, distance, interregional trade, international trade, Spanish provinces

1. INTRODUCTION

Since McCallum's (1995) seminal paper, a large body of literature has tried to explain why countries trade much more with themselves than with other partners. The notoriety of the border effect is such that it has been argued to be "one of the six major puzzles in international macroeconomics" (Obstfeld and Rogoff, 2000). Researchers have followed four different routes to solve the border puzzle. First, some authors (Rauch, 2001; Combes et al., 2005; Schulze and Wolf, 2009; Nitsch and Wolf, 2009) argue that large border effects should not constitute a puzzle, because information barriers increase gently the costs of doing business not only for partners located in different countries but even for patterns located in different regions within a country. Second, other authors point out that even when trade costs were low, a large elasticity of substitution between domestic and foreign goods may lead to large home-bias effects (Evans, 2003). Chaney (2008) demonstrates that the presence of fixed

[1] We acknowledge financial support from Spanish Ministry of Science and Innovation (ECO2010-21643/ECON; project ECO 2008-04059/ECON), Regional Government of Madrid (project TransporTrade S2007/HUM/497), Department of Education, Universities and Research of the Basque Government, and Generalitat Valenciana (project Prometeo/2009/098).

costs of exporting and productivity heterogeneity across firms also magnify the aggregated trade elasticity with respect to trade barriers. Rossi-Hansberg (2005) and Yi (2009) show how endogenous firm location and vertical specialisation in the presence of intermediate goods may help to magnify initial small trade frictions along national borders. A third group of researchers argue that large border effects arise due to a misspecification of the gravity model used in the econometric analyses (Anderson and van Wincoop, 2003) or to identification problems (Heinemeyer et al, 2008).

Finally, a fourth group of authors have indicated that the mismeasurement of physical distance may explain the large border effects. This concern arose in studies that lacked data on flows across sub-national units and had to input intra-national trade as the difference between production and exports (Wei, 1996; Nitsch, 2000; Head and Mayer, 2000; Helliwell and Verdier, 2001). In these studies the distance travelled by goods in the intra-national flow plays a crucial role. If the internal distance is not estimated accurately, the estimated border effect will not be reliable. In particular, if intra-national distance is overestimated, a larger border effect will be needed to explain the preference for national goods with respect to foreign goods. Head and Mayer (2002) show that most studies that used imputed intra-national flows overestimate intra-national distance, leading to illusory border effects.

The distance mismeasurement problem does not arise, in principle, in those studies, such as McCallum (1995), that combine intra-national flows across sub-national units and across those sub-national units and foreign sub-national units (or countries), as long as the same procedure is followed to calculate intra-national and inter-national distances. However, these studies may also face limitations if they use geographically large sub-national units, such as US states or Canadian provinces. In a recent work, Hillberry and Hummels (2008) shows that the relationship between shipments and distance is highly non-linear: at the beginning, there is a sharp reduction in value with distance; however, once a distance-threshold is achieved the negative effect vanishes. If most of intra-national trade happens at short distances, the choice of the sub-national unit will play a crucial role in the border effect analysis. If the sub-national unit is geographically large, trade between sub-national units may not pick the sharp reduction in value that happens at short distances. Hence, econometric analyses may underestimate the negative contribution of distance on trade, and a larger border effect will be needed to explain the preference for national goods with respect to international goods.

Hillberry and Hummels (2008) analyse whether aggregation may explain why previous studies, such as Wolf (2000) and Hillberry and Hummels (2003), find that US states trade more with themselves than with other US states. They show that the intra-national home bias vanishes once geographically finely disaggregated data, which picks the non-linear relationship between distance and shipments value, are used. Our paper extends Hillberry and Hummels' findings to the international trade level and analyse to what extent large border (frontier) effects can also be an artefact of geographical aggregation. For that purpose we use a unique database that reports the value of Spanish small geographical units (provinces) trade flows to other small geographical units and to other countries in years 2000 and 2005. Previous research by Gil et al. (2005) use trade flows between Spanish large geographical units (regions) to the rest of Spain and to other countries over the period 1995-1998. They found that Spanish regions traded 21.8 times more with the rest of Spain than they did with other foreign countries, a result very similar to that found by McCallum for Canada. We show that the Spanish border effect is reduced substantially, and even disappears in some estimations, once we allow data to show the high geographic concentration of intra-national trade flows. These results are in line with recent research, such as Gorodnichenko and Tesar (2009), which using price rather than quantities has also put into question the excessively large border effects found by previous studies.

The rest of the paper is organised as follows. Section 2 analyses whether there is a non-linear relationship between distance and the value of trade flows in Spain. Section 3 presents the empirical model and Section 4 the results of the econometric analyses. The final section summarises the main conclusions of our paper.

2. THE GEOGRAPHIC CONCENTRATION OF TRADE FLOWS

Aggregation will lead to an overestimation of the border effect as long as there is a non-linear relationship between distance and the value of trade flows. To document this feature in the case of Spain, we use a new database, *C-Intereg*, which combines fine grid intra-national trade flows data with international exports data (Llano et al., 2010).[2] It offers

[2] The C-intereg database (see www.c-intereg.es) combines the most accurate data on Spanish transport flows of goods by transport modes (road, rail, ship, plane, pipe and electric network) with additional information used to estimate specific export price vectors by province of origin, transport mode and type of product . The methodology also includes a process for correcting problems in the original transport flows database that arise from multi-modal transport flows and international transit flows hidden in the interregional flows. After this first screening process, an initial estimate of interregional trade flows in tons and in current euro is obtained, based on a combination of transport magnitudes and price information. Finally, a process of harmonization is applied to produce flows magnitudes in tons and Euros that are consistent with total output magnitudes from the Spanish Industrial Survey, the National Accounts and the international trade figures from Customs. This final process of

Código de campo cambiado

the value of total trade flows from Spanish provinces (Eurostat NUTS III) to other Spanish provinces and to 17 European countries, disaggregated by 13 manufacturing industries.[3] Distance data is obtained from the Spanish Permanent Survey on Commodity Transport by Road (SPSCTR), both for internal and external trade flows. Our distance measure is the actual average distance travelled by heavy trucks in their deliveries of commodities for the inter-provincial and international flows. This measure of distance has the virtuosity of capturing the distance travelled by trucks between actual origins and destinations.[4] Spain is divided into 50 provinces, a relatively small administrative unit whose average area is similar to an US 3-digit zip code.[5] Due to the characteristic of our distance measure, we exclude from the sample those intra-national flows where a Spanish island-province is a partner (Balearic Islands, Las Palmas and Tenerife).

We use a kernel regression estimator to provide a nonparametric estimate of the relationship between distance and the value of Spanish intra-national and international trade flows for year 2005.[6] As shown in Figure 1, there is a negative relationship between the value of trade flows and distance. However, this negative relationship does not follow a clean pattern. In a first segment, 25 km. to 900 km., the value of shipments drops almost four times, from 65000 thousand Euros to 17000 thousand Euros. This reduction is especially sharp at short distances, when distance increases from 25 km. to 250 km. In a second segment, 900 km. to 1,200 km., paradoxically, we observe an increase in the value of trade with distance. In

---

harmonization allows offering international trade flows with origin (destination) in any of the Spanish provinces and destination (origin) in any country of the world, compatible with the internal trade estimates and the official output figures at the provincial and the sector level. This procedure brings the trade flows to a position close to the paradigm of trade in continuous space.

[3] The European countries are Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia and Sweden. The industries are Food, beverages and tobacco; Textiles; Leather and footwear; Wood; Paper, printing and publishing; Chemicals; Rubber and plastics; Other non-metallic minerals; Basic and fabricated metals; Machinery; Electrical and Optical Equipment; Transport equipment; Other manufacturing.

[4] The estimate of the actual distance travelled by the trucks for the intra-provincial and interprovincial flows is obtained straightforward from the survey. The distance for the international flows between each of the Spanish provinces and the European countries implies a more complex procedure: the SPSCTR also captures bilateral flows between the Spanish provinces and a large list of countries (with specification of the distance travelled but not of the international region of destination/origin). Although in theory, every international flow carried out by Spanish trucks is covered by the survey, the number of observations for every year may be not representative enough for some bilateral flows. In order to increase the quality of our distance measure, and taking into account the stable nature of this variable, we increase the number of observations by pooling together all the international shipments for the period 1995-2006. Thus, a database of about 200.000 observations regarding the average actual distance for the international deliveries was obtained for each dyad of Spanish provinces and European countries in that period.

[5] The average area of a US 3-digit zip code is 13763 square kilometres and the average area of a Spanish province is 10,116 square kilometres (excluding Ceuta and Melilla, two autonomous city-provinces in Africa).

[6] We use the Gaussian kernel estimator in STATA, calculated on n=100 points, and allowing the estimator to calculate the optimal bandwidth.

a third segment, 1,200 km. to 2,700 km. there is again a negative relationship between the value of trade flows and distance; finally, the relationship becomes flat from 2700 onwards.

<INSERT FIGURE 1 HERE >

Next we examine the relationship between the value trade flows and distance for the intra-national flows and international flows, separately, in year 2005. As shown in Figure 2, we now observe a clear non-linear relationship between the value of trade flows and distance for intra-national flows. The value of intra-national trade flows drop from 120000 thousand euros to 28000 thousand Euros when distance increases from 25 km. to 200 km. From this distance onwards the relationship is slightly negative and even flat for some distance ranges.

<INSERT FIGURE 2 HERE >

To analyse the relationship between distance and international trade flows, we divide exports by the product of production at origin and demand at destination (adjusted trade), since international partners are very different in their demand capacity. Figure 3 shows that the non-linearity between the value of adjusted trade flows and distance is much sharper for international shipments. The value of trade flows drops more than 80 times when distance increases from 200 km. to 600 km.; from 600 km. onwards distance does not seem to affect the value of shipments.

<INSERT FIGURE 3 HERE >

Finally, Figure 4 presents the relationship between adjusted trade and distance for intra-national flows when we use highly disaggregated geographical units as origin and destination partners (provinces: PROV-PROV line), and when we use a geographically larger destination partner (Rest of Spain: PROV-ROS line).[7] Again the relationship between adjusted trade and distance is highly non-linear when province is the destination partner

[7] Rest of Spain is the geographical aggregation for the destination sub-national unit used by Gil et al. (2005) in their analysis of the Spanish border effect. As the origin sub-national unit (exporter), they use regions (NUTS - 2). Spanish regions, also called autonomous communities, are aggregations of Spanish provinces. There are 17 regions in Spain, of which seven have only one province. The results are very similar if we use this aggregation for the origin sub-national unit.

(PROV-PROV line); on the contrary, when Rest of Spain is used as destination partner the relationship is flat for the whole distance-range (PROV-ROS line).

<INSERT FIGURE 4 HERE >

The comparison of the PROV-PROV line and the PROV-ROS line helps to understand how the use of highly aggregated geographical sub-units may lead to an overestimation of the border effect. The use of aggregated units does not allow capturing the highly non-linear relationship between distance and the value of intra-national shipments, leading to an underestimation of the negative impact of distance on trade at short distances.[8] In this paper we want to examine to what extent such underestimation may lead, in turn, to create an "artificial" border effect. In particular, if distance (and other controls for trade impediments) does not explain accurately interregional trade within a country, then the Spanish frontier will be capturing it, exaggerating the "true" frontier effect. The next sections examine this possibility.

3. EMPIRICAL MODEL

We estimate the border effect based on the structural gravity equation developed by Anderson and van Wincoop (2003).[9] The estimating equation takes the following form:

$$\begin{aligned} X_{ijkt} = {} & \beta_0 + \beta_1 Border + \beta_2 \ln Y_{ikt} + \beta_3 \ln Y_{jt} + \beta_4 \ln dist_{ij} + \\ & \beta_5 Contig\_Int\_Partner_{ij} + \beta_6 Contig\_Nat\_Partner_{ij} + P_{i,k} + P_{j,k} \end{aligned} \qquad (1)$$

where X_{ijkt} are exports from province *i* to a trading partner *j* of industry *k* in year *t*. *Border* is a dummy variable that takes the value of 1 if the shipment is intra-national and zero if the shipment is international; Y_{ikt} is production of industry *k* in province *i* in year *t* (source: INE); Y_{jt} denotes trading partner's GDP *j* in year *t* (source: WDI-online). $dist_{ij}$ is the actual average

[8] In Figure 3 we observe as well that the value of shipments drops sharply with distance for international flows. In this case the use of geographically aggregated national units would underestimate the border effect as the distance travelled by goods in international shipments would be overestimated. In any case, this problem seems to be of a lower magnitude, because differences in travelled distance would only be relevant for countries that are very close to Spain (France and Portugal).

[9] Helpman, Melitz and Rubinstein (2008) develop a structural gravity model that incorporates heterogeneous firms. They argue that previous studies' estimates are biased because they do not take into account the large number of zeros in bilateral trade. As explained in the text, our estimation technique controls for this problem.

distance travelled by heavy trucks in their deliveries of commodities from province *i* to trading partner *j*; it is important to stress that our distance measure helps to address the aggregation problem both for international and intra-national shipments. *Contig_Int_Partner* is a dummy variable that takes value of one if the Spanish exporting province is adjacent to an importing international trading partner (a foreign country), zero otherwise, and *Contig_Nat_Partner* is a dummy variable that takes value of one if the Spanish exporting province is adjacent to an importing national partner (a province), zero otherwise.[10] We expect these variables to be positive if our sample does not capture the whole non-linear relationship between trade and distance that happens at short distances. To control for multilateral resistances, we include industry-specific exporter and industry-specific importer fixed effects (P_i^k, P_j^k).

In equation (1) distance is introduced in logs, which helps to linearise the relationship between trade and distance. To capture non-linearity between the two variables we also estimate an equation where distance and the square of distance are introduced in absolute terms:

$$\begin{aligned} X_{ijkt} = \beta_0 + \beta_1 Border + \beta_2 \ln Y_{ikt} + \beta_3 \ln Y_{jt} + \beta_5 dist_{ij} + \beta_6 dist_{ij}^{\ 2} + \\ \beta_6 Contig\_Int\_Partner_{ij} + \beta_7 Contig\_Nat\_Partner_{ij} + P_{i,k} + P_{j,k} \end{aligned} \qquad (2)$$

The equations (1) and (2) are estimated using data from years 2000 and 2005. To account for zero trade flows in our database, we estimate equations (1) and (2) using the Poisson pseudo-maximum likelihood technique, first proposed by Santos Silva and Tenreyro (2006).[11] In the robustness section we compare our preferred results with those obtained using alternative econometric techniques (OLS and Tobit regressions).

The definition of a Spanish trading partner plays a key role in our exercise. While the number of international trading partners will be constant and equal to 17 countries in all the estimations, the number and economic size of the Spanish trading partners will change. In the first sample we aggregate all Spanish trading partners in one, the rest of Spain, so each exporting province has 18 trading partners: Rest of Spain plus 17 European countries. We call this sample "SPAIN". Our second sample considers the smallest geographic unit that we have

[10] When a province sells to the same province the dummy variable takes the value of 1.
[11] We estimate the Poisson model using the ppml command for Stata developed by Santos Silva and Tenreyro (http://privatewww.essex.ac.uk/~jmcss/LGW.html). This command overcomes the convergence problems that may arise with other algorithms when, as in our case, the dependent variable has many zeros and the model includes many dummy variables.

as Spanish trading partner, 47 provinces. In this sample, that we call "PROVINCE", an exporting province has 64 trading partners (47 provinces and 17 European countries). Notice that the value of the variable *distance* will change when we choose different definitions of Spanish trading partner. In particular, the value of distance from one province to Rest of Spain is calculated as the simple average of the distance travelled by trucks from that province to other Spanish provinces. We expect a narrow definition of Spanish trading partners to control adequately for the fact that intra-national shipments are highly localised. Trading with bigger partners (i.e. with the rest of Spain as a whole instead of with each province) will tend to upward bias the border effect since the border effect will be capturing part of the non-linear relationship between distance and shipments.

4. RESULTS OF THE ECONOMETRIC ANALYSIS

A MAIN RESULTS

Table 1 reports the results of estimating Eq. (1) and (2) using the two samples based on the definition of Spanish trading partner. The first two columns use the sample SPAIN and the second two columns the sample PROVINCES. When we aggregate intra-national shipments as the rest of Spain, the coefficient associated to the border effect is around 3.7; that is, a Spanish province trades around 40 times more with the rest of Spain than with a European country. This border estimate is larger than the one obtained by Gil et al. (2005). As explained in the sensitivity-analyses section, the discrepancy can be explained by the fact that we use different periods, countries and estimation techniques. In the first estimation, where we are assuming a linear (log-log) relationship between trade and distance, the distance coefficient is negative and statistically significant. In the second estimation, we introduce distance and the square of distance in absolute terms to capture the non-linear relationship between trade and distance. The distance coefficient is negative and statistically different from zero, but the coefficient for the square of distance is not statistically different from zero. It seems that the sample SPAIN is capturing the large non-linearity between trade and distance for international shipments (Figure 3); however, as the relationship for intra-national flows is flat (Figure 4; PROV-ROS line), the overall non-linearity is not very robust. With respect to other independent variables, in both specifications the coefficients for economic size of trading partners (exporter's production and importer's GDP) and contiguity are positive.

<INSERT TABLE 1 HERE >

Next, we present the results of estimating Eq. (1) and (2) using the sample PROVINCES, that is, other provinces as trading partners. We observe that the border effect is reduced substantially: the coefficient drops from 3.8 to 1.9 when distance is introduced in logs, and from 3.7 to 1.6 when we introduce distance and the square of distance in absolute terms. According to the last coefficient, a Spanish province trades around 5 times more with another Spanish province than with a European country once we control for economic size and other trade impediments.

As argued in this paper, the severe reduction of the border effect from 40 to 5 is explained by the use of a more geographically disaggregated sample (PROVINCES), which allows capturing the high concentration of shipments at low distances. When distance is introduced in logs we observe that the distance coefficient has a much larger negative impact on trade (from -0.486 in Column 1 to -0.810 in Column 3). On the other hand, as shown in Column (4), now both distance and square of distance coefficients have the expected sign and are statistically significant, confirming the high non-linear relationship between trade and distance. As distance has a much larger negative impact on trade, now a lower border effect is needed to explain why Spanish provinces show a preference to trade with other provinces.

We also observe that the coefficient on contiguity between provinces is positive and statistically significant. As exports to contiguous province encompass the shortest distance shipments, the positive coefficient may point out that our sample still does not capture the whole non-linearity between trade and distance. However, the coefficient on contiguity between countries is negative and statistically significant. This apparently counterintuitive result can be explained by the fact that the gravity equation includes the GDP of each Spanish province but only the GDP of France and Portugal. We expect that a Spanish province that shares a border with a French or Portuguese province will trade more with each other. However, if the GDP of the border provinces in France and Portugal is below the country average, the trade flow predicted by our gravity equation will be larger than the actual trade flow.[12] Our distance coefficient, which is calculated from the actual distance travelled by trucks, does not fully attenuate this bias, leading to a negative country-adjacency coefficient. Finally, as in the previous estimations, the origin province output and the destination partner demand coefficients are positive.

To sum up, we observe that the border effect is reduced enormously when the definition of Spanish trading partner changes from a highly aggregated geographical unit (rest of Spain as a whole) to a much smaller unit (provinces). Hence, our results indicate that the border effect is, to a large extent, an artefact of the geographic aggregation of intra-national trade flows.

Table 2 reports the industry-specific border coefficients for 13 manufacturing sectors. In both samples, SPAIN (column 1) and PROVINCES (column 2), there are differences in the border coefficients across industries. If we focus on the sample SPAIN (Column 1), we observe that most industries have a border coefficient in the 3.3-3.7 range. There is one industry with a much larger border coefficient (leather & shoes) and two industries with lower than average coefficients (electric & electronic goods and transport equipment). As we found before, when we use the sample PROVINCES (column 2), the border coefficients tend to become smaller as we use a smaller geographic unit to define Spanish trading partners. Moreover, for three industries the border coefficient is not statistically significant (textile & apparel, paper & print and electric & electronic equipment).

<INSERT TABLE 2 HERE>

B ROBUSTNESS ANALYSIS

As mentioned above, the border coefficient we obtain in Table 1 (Columns 1 and 2) is larger than the one obtained by Gil et al. (2005) using the same sub-national aggregated destination partner: the Rest of Spain. It is important to analyse where those differences come from. In order to do so, we reproduce as close as possible the exercise performed by Gil et al. (2005) using our database. These authors estimate the Spanish border effect using a random effects model and data on aggregate shipments from Spanish regions to the rest of Spain and to 27 OECD countries in the period 1995-1998. Their distance measure is based on geodesic

[12] According to Eurostat, four of the five French NUTS-3 regions that are adjacent to Spain had in year 2005 a GDP below the country average; in the case of Portugal nine of ten adjacent NUTS-3 regions had a GDP below the country average.

information. We can match all the characteristics of their sample, except for the number of countries, which is reduced to 25 OECD countries in our sample and the sample period (2000 and 2005). Table 3 - Column (1) reproduces the results obtained by Gil et al. (2005). Their border effect for Spain is 3.08. Column (2) reproduces Gil et al.'s exercise using our database. We can see that our border effect, 2.75, is now smaller than the one obtained by Gil et al. (2005). Hence, we can conclude that the differences in the Spanish border effect between previous estimates and those reported in Table 1 - Columns (1) and (2), are explained by differences in the sample and the estimation technique.

<INSERT TABLE 3 HERE>

Second, we analyse the robustness of our results when alternative estimation methods are used. We use the Tobit model as another method for the treatment of zeros in our database. Following Eichengreen and Irwin (1998) and Chen (2004), we express the dependent variable as Ln(1+X_{ijkt}).[13] In addition to that, we drop the zero trade values from the sample and estimate the equation using OLS. As shown in Table 4, in both models we observe that the border effect experiences a very large reduction when we shift from an aggregated intra-national partner to a more geographically disaggregated intra-national partner. Moreover, in the Tobit model the border coefficient is statistically not different from zero, denoting that a Spanish province trade as much with other Spanish provinces as with other European countries, once we control for other trade impediments. In the OLS model, when distance is introduced in logs, the border coefficient is slightly larger than in the PPML model. In contrast, when distance and the square of distance are introduced in absolute terms the border effect drops to 0.8, denoting that a Spanish province trades only twice more with another province than with a European country.

<INSERT TABLE 4 HERE>

Third, we test whether our results are robust to changes in the exporting geographic unit. In the new sample, instead of provinces, we select regions as the exporting geographic unit; in this sample, the number of geographic units is reduced from 47 to 15. As shown in Table 5, when we use the sample SPAIN (columns 1-2), the border coefficient is 3.5 when

[13] We use the McDonald and Moffitt (1980) procedure to recover the elasticity at sample means.

distance is introduced in logs and 3.3 in the non-linear distance specification, suggesting that regions trade around 27-33 times more with the rest of Spain than with another country. These border effects are lower than the ones reported in Table 1. When we use the sample REGIONS (columns 3-4), we find a large reduction in the border effect: coefficients drops from 3.5 to 2.2 in log-linear specification, and from 3.3 to 1.2 in the non-linear specification. Therefore, the border effect also falls when we use a more aggregated geographical unit.

<INSERT TABLE 5 HERE>

4. CONCLUSIONS

During the last decades a large number of studies have tried to explain why countries trade more with themselves than with other countries. We show that the border effect in Spain is, to a large extent, an artefact of geographical aggregation. Based on a geographic fine grid data, we find that intra-national trade flows are highly localised, with value dropping sharply with distance. If intra-national trade flows are aggregated this localised pattern is lost, which leads to an overestimation of the distance travelled by goods sold domestically and to the creation of an artificial border effect. Our econometric estimations confirm this fact. While a sizable border effect exists for Spain when intra-national partners are aggregated, the border effect decreases substantially, and even disappears in some estimations, once we allow data to show the localised nature of intra-national trade flows.

REFERENCES

Anderson, J.E. and E. Van Wincoop (2003), 'Gravity with Gravitas: A Solution to the Border Puzzle', *American Economic Review*, 93, 1, 170-192.

Chaney, T. (2008), 'Distorted Gravity: The Intensive and Extensive Margins of International Trade', *American Economic Review*, 98, 4, 1707-1721.

Chen, N. (2004), 'Intra-National Versus International Trade in the European Union: Why Do National Borders Matter?', *Journal of International Economics*, 63, 1, 93-118.

Combes, P. P., M. Lafourcade and T. Mayer (2005), 'The trade-creating effects of business and social networks: evidence from France', *Journal of International Economics*, 66, 1–29.

Eichengreen, B. and Irwin, D. (1998), 'The Role of History in Bilateral Trade Flows', in Frankel, J. (ed.), *The Regionalizatrion of the World Economy*, University of Chicago Press, Chicago.

Evans, C.L. (2003), 'The Economic Significance of National Border Effects', *American Economic Review*, 9**3**, 4, 1291-1312.

Gil-Pareja, S., R. Llorca-Vivero, J.A. Martínez Serrano and J. Oliver-Alonso (2005), 'The Border Effect in Spain', *The World Economy*, **28**, 11, 1617-1631.

Gorodnichenko, Y. and L. Tesar (2009), 'Border effect or country effect? Seattle may not be sor far from Vancouver after all', *American Economic Journal: Macroeconomics*, **1**, 1, 219-241.

Head, K. and T. Mayer (2000), 'Non-Europe: the causes and magnitudes of market fragmentation in the EU', *Weltwirtschaftliches Archiv*, 136, 2, 285-314.

Head, K. and T. Mayer (2002). 'Illusory Border Effects: Distance mismeasurement inflates estimates of home bias in trade', *CEPII Working Paper 2002-01*.

Heinemeyer, H., M-H. Chulze and N. Wolf (2008), 'Endogenous Borders? The Effects of New Borders on Trade in Central Europe 1885-1933', CESifo Working Paper No. 2246

Helliwell, J.F. and G. Verdier (2001), 'Measuring internal trade distances: a new method applied to estimate provincial border effects in Canada', *Canadian Journal of Economics*, 34, 4, 1024-1041.

Helpman, E., M. Melitz, and Y. Rubinstein (2008), 'Estimating Trade Flows: Trading Partners and Trading Volumes', *Quarterly Journal of Economics*, 123, 2, 441-487.

Hillberry, R. and D. Hummels (2003), 'Intra-National Home Bias: Some Explanations', *The Review of Economics and Statistics*, 85, 4, 1089-1092

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

Llano, C. (2004), 'The interregional trade in the context of a multirregional input-output model for Spain', *Estudios de Economía Aplicada*, 22, 3, 539-576.

Llano, C., A. Esteban, , A. Pulido and J. Pérez (2010), 'Opening the Interregional Trade Black Box: The C-intereg Database for the Spanish Economy (1995-2005)'. *International Regional Science Review*. doi:10.1177/0160017610370701.

McCallum, J. (1995), 'National Borders Matter: Canada-U.S. Regional Trade Patterns', *American Economic Review*, 85, 3, 615-623.

McDonald, J.F. and R.A. Moffitt (1980), 'The Uses of Tobit Analysis', *The Review of Economics and Statistics*, 62, 2, 318-321.

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

Nitsch, V. and Wolf, N. (2009), 'Tear down this wall: on the persistence of borders in trade', *CEPR Working Paper 7545*.

Obstfeld, M. and K. Rogoff (2000), 'The Six Major Puzzles in International Macroeconomics. Is there a Common Cause?', *NBER Working Paper 7777*, National Bureau of Economic Research, Cambridge, MA.

Rauch, J.E. (2001), 'Business and Social Networks in International Trade', *Journal of Economic Literature*, 39, 4, 1177-1203.

Rossi-Hansberg, E. (2005), 'A Spatial Theory of Trade'. *American Economic Review*, 95(5): 1464–1491.

Santos Silva, J. and S. Tenreyro (2006), 'The log of gravity', *Review of Economics and Statistics*, 88, 4, 641-658.

Shultze, M.S. and N. Wolf (2008), 'On the origins of border effects: insights from the Habsburg Empire', *Journal of Economic Geography*, 9, 1, 117-136.

Wei, S. (1996), 'Intra-National Versus International Trade: How Stubborn are Nations in Global Integration?', *NBER Working Paper 5531*.

Wolf, H.C. (2000), 'Intranational home bias in trade', *Review of Economics and Statistics*, 82, 4, 555-563.

Yi, K.M. (2009), 'Can Multi-Stage Production Explain the Home Bias in Trade?', *American Economic Review*, 100, 1, 364-93.

FIGURE 1.

Kernel regression:value of intra-national and inter-national trade flows on distance, year 2005

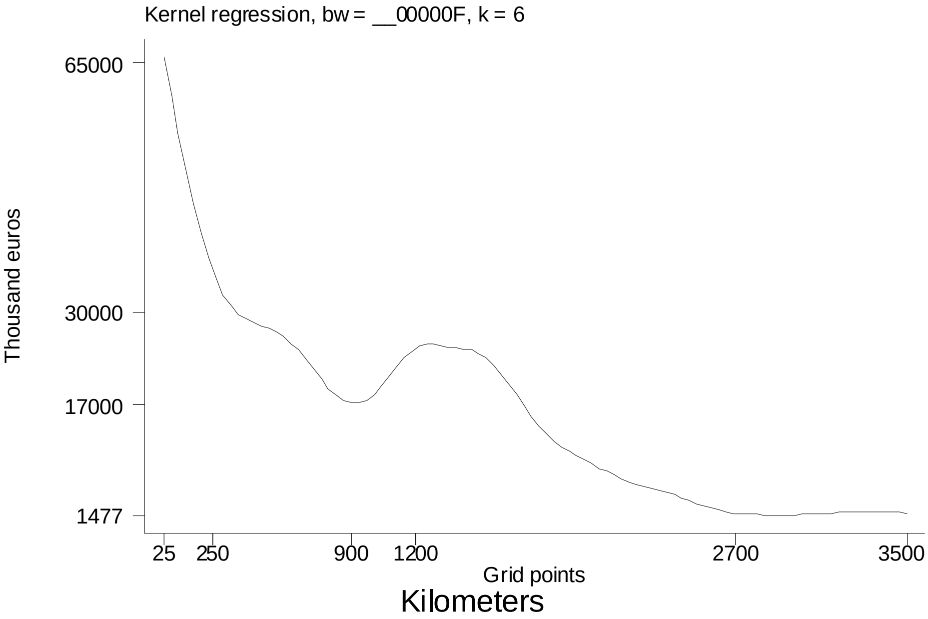

FIGURE 2.

Kernel regression: value of intra-national trade flows on distance, year 2005

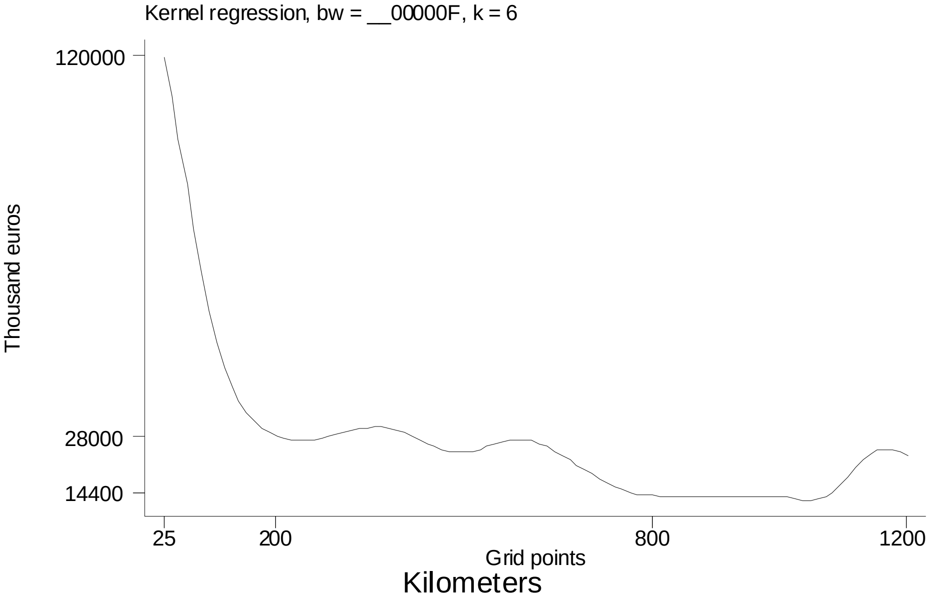

FIGURE 3.

Kernel regression: value of adjusted international trade flows on distance, year 2005

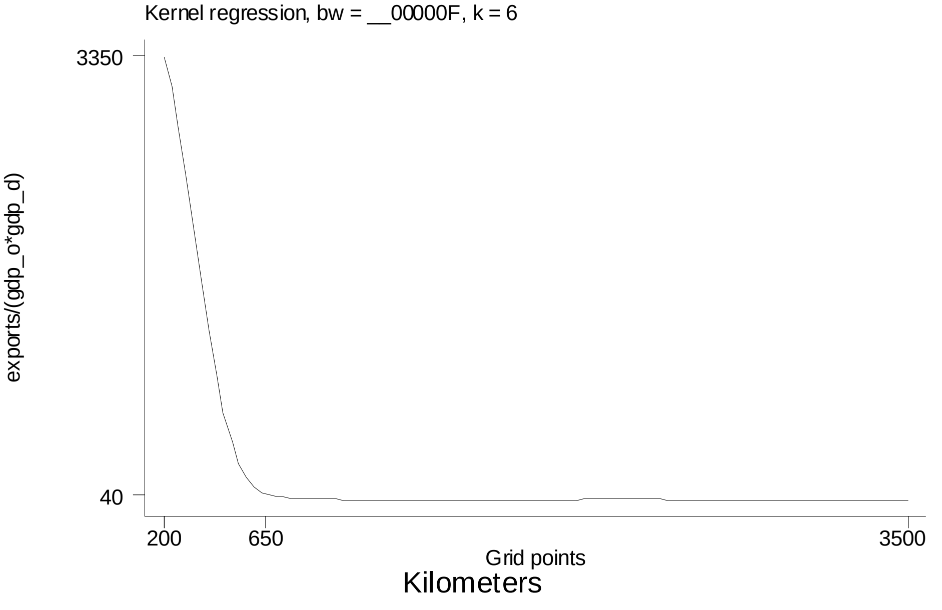

FIGURE 4.

Kernel regression: value of adjusted intra-national trade flows on distance, year 2005

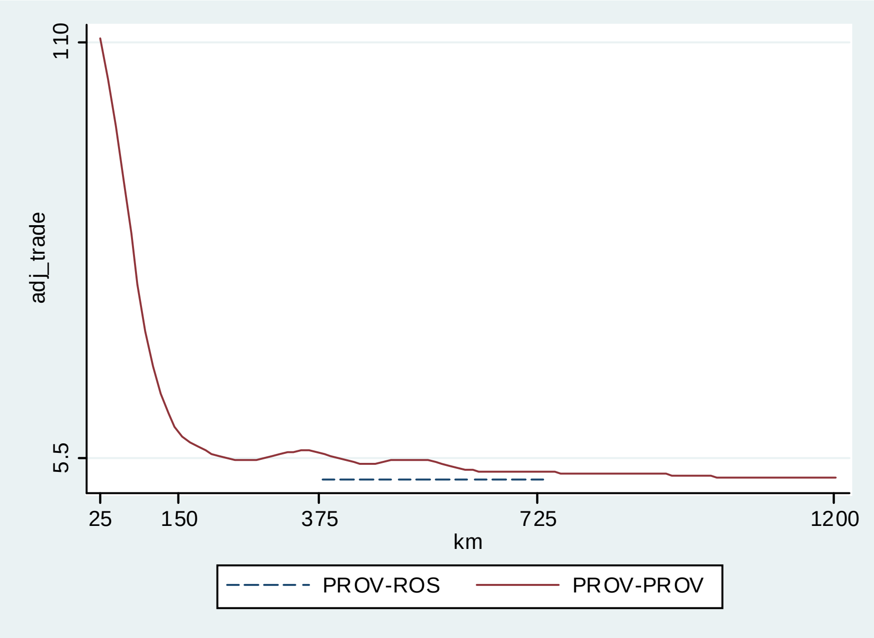

TABLE 1.
Poisson Pseudo-Maximum Likelihood (PPML) estimation of equations (1) and (2)

	Sample SPAIN				Sample PROVINCES			
	(1)		(2)		(3)		(4)	
Origin								
Provinces	47		47		47		47	
Destination								
Provinces	-		-		47		47	
Rest of Spain	1		1		-		-	
Countries	17		17		17		17	
Sectors	13		13		13		13	
Years	2		2		2		2	
N	21996		21996		78208		78208	
Border	3.837	***	3.683	***	1.929	***	1.632	***
	(31.58)		(31.63)		(13.17)		(8.02)	
Gross output exp.	0.968	***	0.948	***	0.878	***	0.919	***
	(44.25)		(59.09)		(46.68)		(50.42)	
GDP importer	0.680	***	0.564	***	0.673	***	0.713	***
	(13.28)		(8.56)		(20.22)		(29.20)	
Log Distance	-0.486	***			-0.810	***		
	(-6.120)				(-12.23)			
Distance			-0.447	**			-2.326	***
			(-1.97)				(-7.43)	
Distance^2			-0.00005				0.00042	***
			(-0.70)				(5.84)	
Cont. Int. Part	0.149		0.280	*	-0.234	*	-0.427	***
	(1.14)		(1.86)		(-1.84)		(-2.72)	
Cont. Nat. Part					0.156	**	0.841	***
					(2.35)		(10.07)	

Notes:
Absolute distance measured in thousand kilometres. All specifications include industry-specific exporter, industry-specific importer and year dummies. Province-clustered robust t-statistics in parenthesis with *** denoting significance at the 1-percent level, ** significance at the 5-percent level and * significance at 10-percent level.

TABLE 2

Industry-specific border effect.

	Sample SPAIN (1)		Sample PROVINCES (2)	
Origin				
Provinces	47		47	
Destination				
Provinces	-		47	
Rest of Spain	1		-	
Countries	17		17	
Sectors	13		13	
Years	2		2	
Observations	21996		78208	
Border Food & drinks	3.675	***	1.621	***
	(24.82)		(7.72)	
Border Textile & apparel	3.559	***	-0.536	
	(18.63)		(-0.64)	
Border Leather & shoes	7.612	***	3.195	***
	(29.36)		(3.91)	
Border Wood products	3.660	***	1.020	**
	(21.92)		(1.99)	
Border Paper & print	3.652	***	0.556	
	(21.37)		(1.38)	
Border Chemical products	3.316	***	0.813	***
	(19.27)		(2.82)	
Border Plastic & rubber	3.541	***	1.606	***
	(23.85)		(3.22)	
Border Nonmetal mineral	3.596	***	1.472	***
	(22.78)		(4.22)	
Border Metal products	3.508	***	1.706	***
	(23.42)		(5.38)	
Border Mechanical machinery	3.607	***	0.763	**
	(25.80)		(1.97)	
Border Electric & electronic goods	2.539	***	0.448	
	(17.20)		(1.17)	
Border Transport equipment	3.030	***	1.620	***
	(15.57)		(3.54)	
Border Other manufactures	3.386	***	0.820	*
	(24.08)		(1.76)	
Gross Industry Output Exporter	0.975	***	0.897	***
	(87.18)		(58.25)	
GDP Importer	0.536	***	0.709	***
	(7.90)		(29.37)	
Distance	-0.562	**	-2.330	***
	(-2.34)		(-7.38)	
Distance ^ 2	-0.00002		0.00042	***
	(-0.31)		(5.82)	
Contig. Int. Partner	0.240		-0.434	***
	(1.59)		(-2.73)	
Contig. Nat. Partner			0.841	***
			(10.04)	

Notes:

Absolute distance measured in thousand kilometres All specifications include industry-specific exporter, industry-specific importer and year dummies. t-statistics in parenthesis with *** denoting significance at the 1-percent level, ** significance at the 5-percent level and * significance at 10-percent level.

TABLE 3.
Robustness analysis I. Comparison with Gil et al. (2005)

	Gil et al. (2005) estimation	Estimation with our database
Border	3.08 (10.27)***	2.75 (8.09)***
GDP exporter	1.08 (36.71)***	1.23 (17.67)***
GDP importer	1.08 (36.80)***	1.07 (24.62)***
Distance	-1.28 (-21.69)***	-1.53 (-20.42)***
Time period	1995-1998	2000-2005
Estimation technique	Random effects model	Random effects model
Exporter	Spanish region	Spanish region
Importer	Rest of Spain + 27 OECD countries	Rest of Spain + 25 OECD countries
Data	Aggregated trade	Aggregated trade

Notes:
Gil el al. (2005) coefficients correspond to their Table 1-Column 1 estimation. All estimations include year-specific dummies. t-statistics in parenthesis with *** denoting significance at the 1-percent level.

TABLE 4.
Robustness analysis II. Estimation with Tobit and OLS models

	TOBIT				OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Spain	Spain	Provinces	Provinces	Spain	Spain	Provinces	Provinces
Origin								
Provinces	47	47	47	47	47	47	47	47
Destination								
Provinces	-	-	47	47	-	-	47	47
Rest of Spain	1	1	-	-	1	1	-	-
Countries	17	17	17	17	17	17	17	17
Sectors	13	13	13	13	13	13	13	13
Years	2	2	2	2	2	2	2	2
N	21996	21996	78208	78208	16276	16276	33071	33071
Border	3.514***	3.879***	0.098	-0.125	3.800***	3.632***	2.211***	0.830**
	(11.52)	(22.36)	(0.58)	(-0.51)	(15.68)	(14.96)	(8.19)	(2.35)
Gross output exp.	0.280***	0.279***	0.147***	0.145***	0.341***	0.342***	0.657***	0.660***
	(3.94)	(19.95)	(3.79)	(3.80)	(3.50)	(3.50)	(8.58)	(8.76)
GDP importer	1.170***	1.141***	0.783***	0.786***	0.976***	0.969***	0.978***	0.947***
	(21.02)	(29.25)	(24.16)	(23.20)	(18.09)	(16.56)	(19.29)	(16.99)
Log Distance	-1.429***		-1.130***		-1.254***		-1.488***	
	(-9.10)		(-14.34)		(-11.62)		(-12.23)	
Distance		-1.488***		-2.516***		-2.239***		-4.368***
		(-9.35)		(-11.81)		(-9.15)		(3.54)
Distance^2		0.00002***		0.00045***		0.00004***		0.00074***
		(4.70)		(9.24)		(6.66)		(10.70)
Cont. Int. Part	0.091	0.806***	-0.381***	-0.378**	0.289	0.509**	-0.208	-0.763***
	(0.30)	(5.77)	(-2.79)	(-2.41)	(1.37)	(2.20)	(-0.88)	(-3.19)
Cont. Nat. Part			0.664***	1.8774***			-0.206	0.597***
			(6.51)	(14.15)			(-1.36)	(4.47)
Pseudo R2	0.301	0.301	0.301	0.186	0.734	0.734	0.518	0.522

Notes:
In Tobit estimations are at sample mean elasticities. Absolute distance measured in thousand kilometres. All specifications include industry-specific exporter, industry-specific importer and year dummies. t-statistics in parenthesis with *** denoting significance at the 1-percent level, ** significance at the 5-percent level and * significance at 10-percent level.

TABLE 5.
Robustness analysis III. Regions as origin

	Sample SPAIN		Sample REGIONS	
	(1)	(2)	(3)	(4)
Origin				
Regions	15	15	15	15
Destination				
Regions	-	-	15	15
Rest of Spain	1	1	-	-
Countries	17	17	17	17
Sectors	13	13	13	13
Years	2	2	2	2
N	7020	7020	12480	12480
Border	3.512 ***	3.349 ***	2.165 ***	1.240 ***
	(13.16)	(13.38)	(11.80)	(5.19)
Gross output exp.	1.054 ***	1.034 ***	0.951 ***	0.928 ***
	(36.79)	(54.31)	(38.48)	(46.61
GDP importer	0.588 ***	0.483 ***	0.670 ***	0.691 ***
	(7.14)	(5.69)	(10.67)	(17.21)
Log Distance	-0.864 ***		-0.839 ***	
	(-5.66)		(-8.18)	
Distance	***	-1.078 ***		-3.479 ***
		(-3.94)		(-10.98)
Distance^2		0.00005		0.0006 ***
		(1.03)		(8.15)
Cont. Int. Part	-0.528 **	-0.413 *	-0.443 ***	-1.150 ***
	(-2.51)	(-1.93)	(-3.09)	(-7.26)
Cont. Nat. Part			0.343 ***	0.296 ***
			(2.77)	(2.65)

Notes:
Absolute distance measured in thousand kilometres. All specifications include industry-specific exporter, industry-specific importer and year dummies. Regions-clustered robust t-statistics in parenthesis with *** denoting significance at the 1-percent level, ** significance at the 5-percent level and * significance at 10-percent level.