I.S.S.N: 1885-6888

ECONOMIC ANALYSIS WORKING PAPER SERIES

The Border Effect and the Non-Linear Relationship between Trade and Distance

♦

Nuria Gallego and Carlos Llano

Working Paper 6/2014

UNIVERSIDAD AUTONOMA DE MADRID

DEPARTAMENTO DE ANÁLISIS ECONÓMICO:
TEORÍA ECONÓMICA E HISTORIA ECONÓMICA

The Border Effect and the Non-Linear Relationship between Trade and Distance

Nuria Gallego* (nuria.gallego@uam.es)

Carlos Llano* (carlos.llano@uam.es)

* Departamento de Análisis Económico: Teoría Económica e Historia Económica. Facultad de Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid. Campus Cantoblanco. 28049 MADRID.

Abstract

After the seminal paper by McCallum, various authors have estimated the effect of regional and national borders on trade. This paper digs deeper into the matter, estimating how the internal and external border effect is affected by the non-linear relation between trade and distance at different spatial levels, and the econometric procedure used to control for it. Our paper uses a novel dataset that captures intra- and inter-national truck shipments between Spanish regions and regions in eight European countries. To deal with this non-linearity, we use three alternative strategies—segmented distance, piecewise regressions and semi-parametric approaches—that achieve similar results.

Keywords: border effect, gravity equation, interregional trade, non-linearity, European integration.

JEL Codes: F14, F15

1 Introduction

Some years ago McCallum (1995) found that trade between any two Canadian provinces was (on average) 22 times greater than trade between any Canadian province and any U.S. state. After this seminal contribution, many authors have repeated the exercise with other countries[1] and other spatial units. Some have estimated the relevance of international frontiers by comparing the domestic trade volume of one country (region) with its international trade volume (Head and Mayer, 2000; Gil et al., 2005; Minondo, 2007; Chen, 2004), while others have measured the relevance of internal borders, estimating how much more trade a region (province) of a given country conducts with itself than with any other region (province) of the same country (Wolf 1997, 2000; Hillberry and Hummels, 2008; Combes et al., 2005; Garmendia et al., 2012).

A number of factors may explain the effect of regional and national borders on the volume of trade. Chen (2004) classified them into two groups, by their exogenous/endogenous nature. The size of the border could be explained exogenously by tariffs, non-tariff barriers, information differences or transaction-cost differences, or endogenously by a low degree of substitutability between local and foreign products (*home bias* in preferences) or optimal location choices on the part of producers. In this respect, trade frictions would affect trade volumes through two channels. A direct effect would occur as frictions changed relative prices, inducing substitution towards proximate products. The indirect effect would occur through co-location. Firms linked closely in the input-output structure would locate nearby so as to minimize trade costs. The geographic location of firms and the importance of intermediate goods could also promote the appearance of core/periphery structures (Fujita et al., 1999), which enhanced internal flows with respect to external ones. Other authors have suggested additional causes for the border effect, such as the heterogeneity of firms (Evans, 2003; Chaney, 2008), multi-stage production (Kei-Mu, 2010) and the misspecification of

[1] Japan (Okubo, 2004), United States (Wolf, 2000; Hillberry, 2002; Hillberry and Hummels, 2003; 2008; Millimet and Osang, 2007), the European Union (Chen, 2004; Nitsch, 2000, 2002; Evans, 2003), Germany (Shultze and Wolf, 2008), Russia (Djankov and Freund, 2000) and Brazil (Daumal and Zignago, 2008), among others.

econometric models used in estimations (Anderson and van Wincoop, 2003; Silva and Tenreyro, 2006).

In this last strand of the literature, some recent papers describe the border effect as an artifact of "spatial aggregation" (Hillberry and Hummels, 2008; Llano-Verduras et al., 2011) or of a mismeasurement of the distance variable (Head and Mayer, 2000, 2002). Hillberry and Hummels (2008) used a micro-dataset on the truck shipments of U.S. firms in 1997, which offered several spatial levels corresponding to states and zip codes. Investigating the non-linear effect of distance on the extensive and intensive margin of U.S. internal shipments, they found no border effect on internal shipments at certain spatial levels (*Ownzip*). The non-linear relationship between trade and distance was controlled for with a quadratic term for the distance variable. Similarly, Llano-Verduras et al. (2011) revised the estimated effect that national boundaries (*Owncountry*) exert on Spanish domestic and international trade (at the country level) by using flow data at two different spatial scales for the exporting unit: namely, the Spanish regions (Nuts 2) and provinces (Nuts 3). They found that the size of the border effect depended largely on the unit of spatial measurement. This paper—although varying the spatial scale for Spanish units: from regions (Nuts 2) to provinces (Nuts 3)—always scaled the foreign partner at the country level. A complementary study, Garmendia et al. (2012), estimated the effect of the regional borders (*Ownregion*) on truck shipments within Spain at the province level (Nuts 3), taking into account social- and business-network effects.

Although the econometric treatment of non-linearities has been widely considered in fields such as labor economics or growth, it has received little attention in the literature of international trade and gravity equations. One of the exceptions, Mukherjee and Pozo (2011) used a gravity model to analyze the impact of exchange-rate volatility on the volume of bilateral international trade through a semi-parametric regression for a panel of 200 countries. This model considers a non-linear relationship between volatility and trade, avoiding the need to superimpose any linearity restriction on the underlying relationship between exchange-rate volatility and trade. Another interesting example is Mundra (2005), who studied the relationship between U.S. bilateral trade and the stock of immigrants from different countries using a semi-parametric regression, where some variables enter the model linearly and there is no functional form for the proxy of social networks (immigration stock). De Benedictis et al. (2008) investigated the empirical relationship between overall specialization and per capita income using the Balassa

Index of Revealed Comparative Advantages and non-parametric regression models. Finally, Ruiz et al. (2009) studied the non-linear relationship between remittances, institutions and growth, and, like Chami et al. (2005), discussed the advantages of semi-parametric approaches over quadratic terms. Moreover, piecewise regression (spline models), another benchmark approach for dealing with non-linearities, has also been neglected in trade analysis. In fact, we have found no remarkable examples in the field of bilateral flows, and just one on the use of spline techniques when modeling time series of product-specific exports (Martín Rodriguez and Cáceres Hernández, 2010).

An interesting reference for our approach is Henderson and Millimet (2008), who used different parametric and non-parametric methods to discuss the nonlinear relationship between trade and distance. They specifically questioned two main assumptions of the literature: (i) that the relationship between trade and unobserved trade costs is (log) linear and (ii) that the effects of trade costs on trade flows are constant across country pairs. They then estimated gravity models both in levels and logs using two datasets and different parametric and nonparametric methods. Their paper concluded by suggesting two lines for future research that are worth repeating here: first, their exclusion of zero-trade observations; second, their observation that "*while the parametric models outperform their nonparametric counterparts, all of the models perform relatively poorly using cross-country data […]. As a result, there is substantial room for improvement in modeling cross-country trade flows*" (168).

A more classical reference is Eaton and Kortum (2002), where a Ricardian model was tested by means of structural equations with bilateral flows for 19 OECD countries in 1990. In some of their specifications, the distance variable is divided into six intervals, as an alternative to the quadratic form. The length of such intervals is ad-hoc, and its effects over the results are not subject to a robust check against alternative spatial units or division criteria.

At this point, we should specify how our paper differs from the previous literature. With respect to Hillberry and Hummels (2008), our main contribution lies in the discussion of alternative specifications for dealing with the non-linear relationship between trade and distance when the dataset combines domestic and international flows (both at the region-to-region level) any of which may cross up to nine different borders. Whereas

Hillberry and Hummels (2008) were able to disentangle the effect of internal borders (*Ownregion* and *Ownzip*) on internal flows within the U.S. at very fine spatial units, they remained silent on the equivalent puzzle for international deliveries. Similarly, Garmendia et al. (2012) focused on the effect of regional and provincial borders (*Ownregion* and *Ownprovince*) and networks on domestic shipments within Spain, but did not consider the effect of national borders on international flows. Conversely, Llano-Verduras et al. (2011) focused strictly on the effect of national borders (*Owncountry*), leaving aside intra-regional flows and own-region borders.

In this regard, this paper presents a few novelties: (1) It computes the effect of two different types of borders (*Ownregion*, *Owncountry*) simultaneously for inter-regional flows between one country (Spain) and its eight main European partners; it uses region-to-region national and international flows, something that has never been done before in Europe. We obtain robust results for two alternative specifications of the gravity model, based respectively on the fixed effect (Anderson and van Wincoop, 2003) and on the odds-ratio approach (Head and Mayer, 2013; Combes et al., 2005). (2) Like other papers reporting border effects that shrink along with the size of the exporting unit (Hillberry and Hummels, 2008; Llano-Verduras et al., 2011), we obtain this decrease using a lower spatial scale for the importer (foreign regions instead of countries). (3) Finally, it suggests three new alternative strategies for tackling the non-linear relationship between trade and distance that has been discussed by others (Eaton and Kortum, 2002; Henderson and Millimet, 2008; Hillberry and Hummels, 2008). Our paper begins by incorporating a quadratic distance term, so as to capture the fast decrease in trade flows over the shortest distances. It then suggests an alternative strategy, which considers three sub-divisions of the sample by distance travelled. Furthermore, we develop a robust analysis using two innovative approaches—the linear and cubic piecewise regression and a semi-parametric regression—so as to add flexibility without imposing a specific function-form to the non-linear relation between trade and distance. These strategies produce interesting results: (i) a low but persistent *internal border effect* (*Ownregion*), which reaches a factor of 4, robust to several specifications; (ii) a persistent *external border effect* that is also about 4; (iii) finally, a variation in the elasticity of distance when it is segmented by alternative criteria: especially by the well-known power series known as the *Fibonacci sequence* (to the

best of our knowledge, this is the first time that these complementary approaches have been used for this purpose in trade analysis

The rest of the paper is organized as follows. Section 2 describes our method for estimating a region-to-region trade dataset for the Spanish case and offers a descriptive analysis of new trade flows. Section 3 describes the alternative specifications of the gravity equation used in our analysis. Section 4 presents our results, and the final section summarizes our conclusions.

2 The Data

We should state at the outset that there is no official data on region-to-region international trade flows for any country in the EU. Gallego and Llano (2012), however, have laid out a method for estimating region-to-region international flows between Spain and eight European countries[2]. It combines region-to-region freight statistics for Spanish trucking with international price indices (deduced from official trade data[3]) for each region-country variety. They then apply a process of homogenization to ensure a match between the dataset and official international trade statistics at the lowest common level of aggregation (year-region-to-country-road)[4]. This novel dataset for region-to-region international trade flows was connected with equivalent data on (intra-

[2] Although for the sake simplicity we use the label EU, our sample of countries does not fall under any specific administrative category. Moreover, we consider certain countries, like Andorra, as single-region countries.

[3] For most of our EU countries, we use two main sources for the inter-national bilateral flows of goods: (1) Trade statistics on intra-EU trade, which register bilateral flows between pairs of countries, both in volume and in monetary units; for certain countries, like Spain, the trade data identify the exporting or the importing region but never both simultaneously. (2) Transport statistics on intra-national and inter-national freight flows, which in some cases (e.g., road freight) provide information on the type of product transported (quantity) as well as on the regional origin and destination of the flows. Our method aims to build up a region-to-region trade dataset by combining these two sources: (1) region-to-region flows in quantities (road-freight statistics) and (2) specific region-to-country trade prices (from the official trade statistics).

[4] Trucks are the main transportation mode for international Spanish exports to the eight EU countries considered in this paper. The survey that provides the basic information on freight flows (volume) covers Spanish trucks only. In order to avoid bias in the estimation of the external border effect (*Owncountry*), region-to-region international flows by road have been re-scaled to the official data on Spanish exports by "road" to each of the eight EU countries. Thus the levels of the Spanish exports for each "Spanish region-country" match the official value of trade split at the "region-to-country by road" level. Therefore, the region-to-region structure is given by the Spanish truck survey, while trade levels are supported by the official trade data broken down at the lowest common level of disaggregation (region-country-year-road).

and inter-regional) trade flows within Spain, which has been the object of previous analysis (Garmendia et al., 2012). The result is a unique dataset on region-to-region flows for intra-regional, inter-regional and inter-national flows into and out of the regions of Spain (Nuts 2) and the regions of Spain's eight main European partners.

Our distance variable is the *mean actual distance* covered by Spanish trucks between each pair of trading regions, as reported in the survey published by Spain's Ministry of Public Works and Transport (Ministerio de Fomento).[5] This variable has the virtue of capturing the *real distance* travelled by trucks between actual points of departure and destinations. In this sense, it is superior to the variable used by other authors, where intra-national and/or inter-national distance is either an a-priori estimate based on the great circle distance between main cities weighted by population or an ad-hoc estimate by mathematical approximation. By using actual distance, we should be able to account for region-to-region inter-country links that are not attributable to the mere allocation of population. There are specific regional endowments or specificities that weighted distance tends to mask. The Ministry's survey also includes the actual distance travelled by trucks for inter- and intra-regional deliveries. Crucially, this allows us to avoid choosing alternative ad-hoc intra-regional distances, which alter results on border effects (Head and Mayer, 2002). Regional GDPs for the EU regions under consideration are published by Eurostat.

2.1 Descriptive Analysis

Before proceeding to the econometric analysis, we will briefly analyze the novel dataset to show the non-linear relation between trade and distance when the spatial grid is sufficiently fine. With this purpose, we offer a first view of the distribution of trade (always region-to-region) as it depends on distance travelled by trucks, for both domestic and international deliveries. Like in Garmendia et al. (2012); Llano-Verduras et al. (2011) and Hillberry and Hummels (2008), we also use a kernel regression to

[5] We obtained the actual distances used in this paper by first screening out outliers: i.e., distances that are too great for a specific dyad. We then computed the actual distance for each regional dyad (aggregate) for each year, starting from the most disaggregated level (micro-data at the municipality level for the Spanish exporting unit). Finally, we obtained intra- and inter-regional distances by averaging the distances observed in all deliveries from 2004 to 2007 for each specific dyad *i-j*.

generate a nonparametric estimate of the relationship between distance and the intensity of Spanish regional export flows[6].

Figure 1: Kernel Regression: Intra- & Inter-National Trade Relative to GDP (NUTS-2 Region-to-Region) on Distance. Zero Flows Excluded (€). 2004–2007.

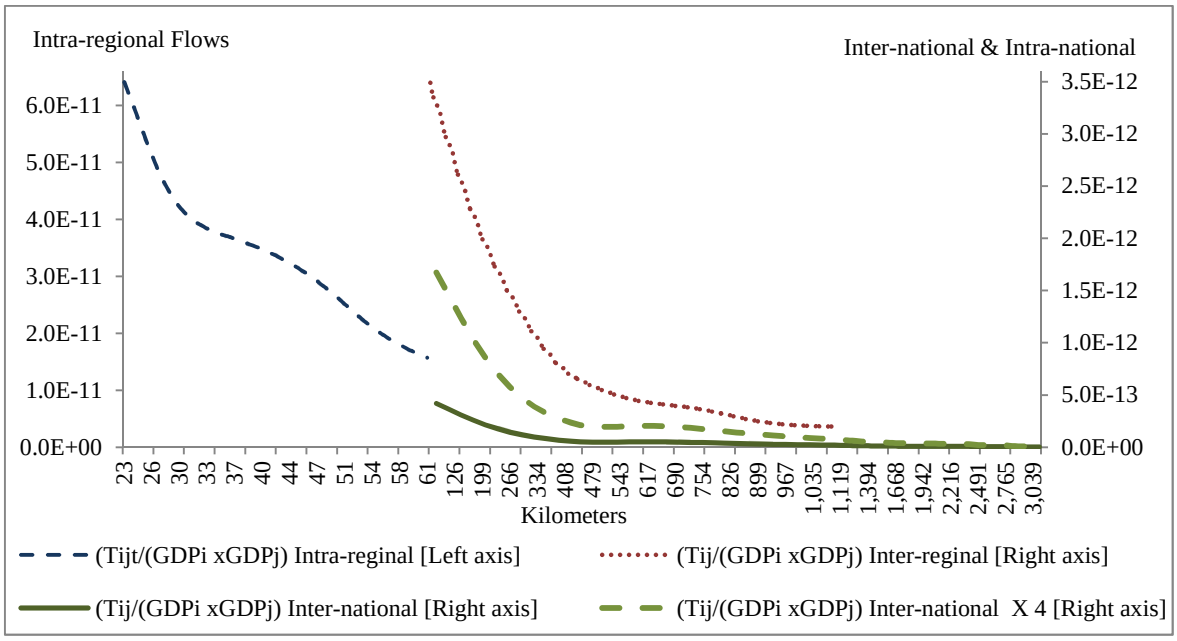


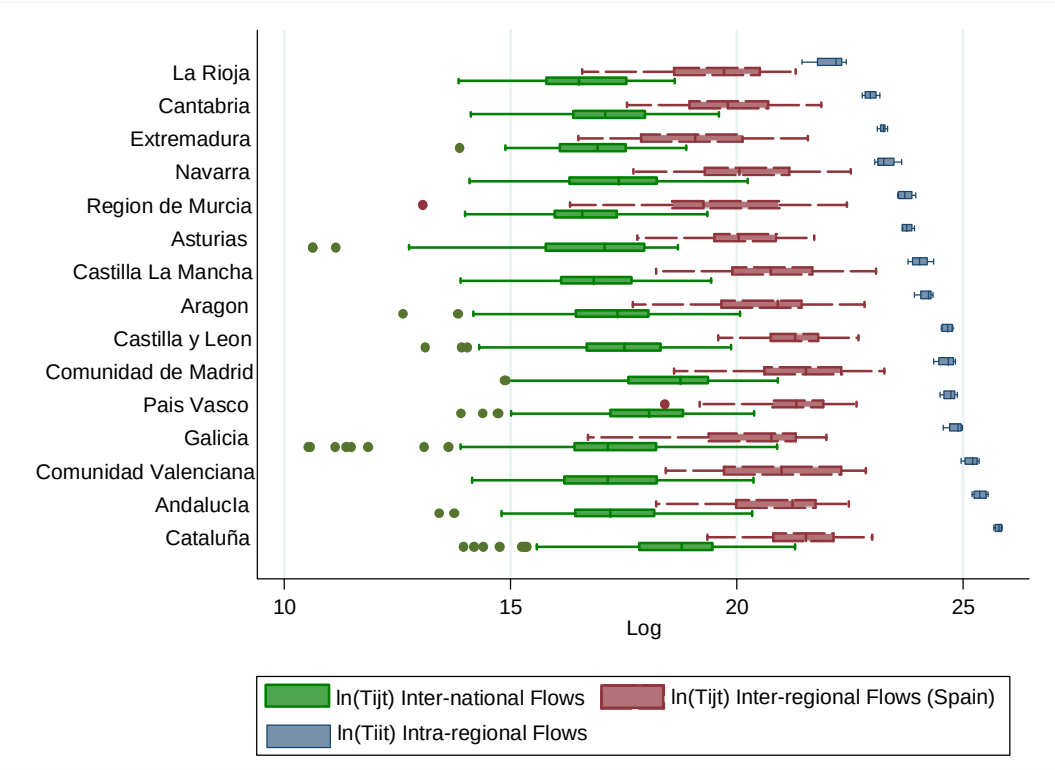
Figure 1 plots the distribution of domestic and international flows (exports) for each region against those regions for the rest of Spain and the eight European countries. Note that trade flows are corrected by the GDP of each exporting/importing region. To illustrate the *multi-level dimension of the non-linear relation between trade and distance*, the figure has a separate plot for the kernel regression of each kind of trade flow: i.e., intra-regional flows within Spain, inter-regional flows within Spain and inter-regional exports from Spanish regions to regions in the eight countries. To bring out the great differences in intensity, the graph displays two different scales: one for intra-regional flows (left axis) and one for the remaining flows (right axis). Moreover, to emphasize the similar shape of each kernel distribution, the *international flows* kernel is plotted twice: with its natural scale and re-scaled at a factor of "x4" (in line with the largest *external border effect* reported in this paper). In distinguishing the great differences in the relative intensity of the flows, we can also see the *regularity* of the

[6] We use the Gaussian kernel estimator in STATA, with n = 100 points and the estimator calculating optimal bandwidth.

non-linear relationship between trade and distance over the shortest distance. By mixing together different types of flows, other papers have emphasized the sharp decrease on the intensity of trade over the shortest distances (e.g., 700 km). Our approach shows how length of flow varies by kind of flow.

From this analysis we can conclude that, regardless of flow type, the bulk of trade takes place over short distances and beyond a certain point the negative effect of distance falls off deeply. Hence the relevance of territorial disaggregation, Llano-Verduras et al. (2011) and Garmendia et al. (2012) have shown that, with insufficient territorial disaggregation of trade, the gravity equation may lead to an overestimation of the border effect and an underestimation of the distance effect. As we will see in the next section, this overall effect can arise not just when regions are used instead of provinces, but also when countries are used instead of regions. Because of this non-linearity, moreover, sharp decreases in trade intensity may or may not coincide with the administrative units where the flows are allocated (and where the borders are!). It would thus be interesting to consider econometric procedures flexible enough to control for that.

Figure 2: Distribution of trade flows by nature and exporting region. 2004–2007.



Note: The box records the second and the third quartiles of the variable, being the line that divides the box between these two quartiles the average. The first and the second whiskers correspond to the first and the fourth quartiles, respectively. The dot values are usually interpreted as outliers, but in this case we consider them as extreme values.

Complementarily, **Figure 2** shows the distribution of each type of flow (in logs and without controlling for GDPs) by exporting region. To interpret this, it is critical to take into account that the figure plots the distribution of observations around the corresponding mean rather than aggregate magnitudes. The aim of this graph is threefold: First, to illustrate the variability of trade in each market (intraregional, interregional within Spain and interregional with the rest of Europe), and show that, although there are some rare small international exports, the intensities for each category are quite stable (structural). Second, as with kernel regressions, to confirm a clear discontinuity in the intensity of trade in the presence of both regional and national borders, note that in almost all regions, there is little overlap between the intensities of interregional flows within Spain and with Europe, and none at all between intra-regional trade and the other two categories. Third, to demonstrate the remarkable variability in the range of trade intensity by region, for the sake of clarity, the exporting regions are ranked by largest flows (intra-regional). Cataluña has the largest intensities for each category, showing outstanding values for intraregional and international flows. Other cases are also worth mentioning: Galicia and Asturias, for example, show wide ranges of flows, with several outliers in the bottom part of their distribution; conversely, Madrid shows a shorter and more compact range (no outliers in the bottom part of the distribution) together with intense flows in all categories.

3 The Empirical Model

3.1 Baseline models

As in most of the articles cited previously, the backbone of our investigation is the gravity equation, where the intensity of trade between any two locations (regions or countries) is positively related to their economic size and inversely related to the trade cost (proxy by geographical distance) between them. However, we depart from previous literature by redefining specific border effects to be measured. By *internal border effect*

(Ownregion) we denote the number of times a Spanish region trades more with itself than with any another region in the sample. By *external border effect (Owncountry)* we denote the number of times a Spanish region trades more with another Spanish region than with a foreign region elsewhere in Europe, controlling for a set of factors.

For the sake of brevity, we define two equations that contain the benchmark models:

$$Ln\frac{T_{ijt}^{eu}}{GDP_{it}xGDP_{jt}} = \beta_0 + \beta_1 Ownregion + \beta_2 Owncountry + \beta_3 Internal\_Contig +$$

$$\beta_4 External\_Contig + \beta_5 Ln(dist_{ij}) + \mu_{it} + \mu_{jt} + \mu^u + \mu_t^u + \varepsilon_{ijt} \qquad [1]$$

Where $\frac{T_{ijt}^{eu}}{GDP_{it}GDP_{jt}}$[7] represents bilateral flows originating in Spanish regions and corrected by the GDPs of the trading regions. More specifically, T_{ijt}^{eu} is the flow from region *i* in country *e* to region *j* in country *u* in year *t*. Note that: (a) if *e* = *u* = *Spain* and *i* = *j*, Eq. [1] will capture intra-regional trade flows for a Spanish region *i*; (b) if *e* = *u* = *Spain* and $i \neq j$, Eq. [1] will capture inter-regional trade flows for a pair of regions within Spain *ij*; (c) if $e \neq u$, Eq. [1] will capture inter-regional flows between Spain and another European country in the sample. Since this paper focuses on flows originating in Spanish regions, *e* = *Spain*. Note that if e = Spain, Eq. [1] captures Spanish exports to regions in our sample of eight European countries (cf. list of countries and products in the Annex). Because of the characteristics of our road-flow dataset, we exclude flows where the regional partner is a Spanish island (therefore I=15 for Spain). The variables GDP_{it} and GDP_{jt} are the nominal gross domestic product (GDP) of the exporting and importing region, respectively. The variable $Ln\ (dist_{ij})$ is the logarithm of the distance between region *i* and region *j*.

The variable *Owncountry* is a dummy that takes the value one for inter-regional flows within Spain (*e* = *u* = *Spain*) and zero otherwise. In addition, the variable *Ownregion* takes the value one when the origin and the destination region are the same (intraregional flows *i* = *j*) and zero otherwise. The anti-log of the parameter associated

[7] Note that in the tables of results we refer to this term as T_{ijt}_corr (trade flow corrected for the GDPs' product), instead of $T_{ijt}^{eu}/GDP_{it}GDP_{jt}$, just for saving notation and keeping clear these tables.

with these two variables measures the size of the effect that national and regional borders respectively exert on trade.

To capture the positive effect of adjacency, we introduce two dummy variables: *Internal_Contig* and *External_Contig*. This allows us to consider (simultaneously or independently) the different effects that *adjacency* exerts on trade flows between two contiguous regions in Spain or between a Spanish region and a contiguous foreign one. *Internal_Contig* takes the value one when trading regions i and j are contiguous and both located in Spain and zero otherwise. Similarly, *External_Contig* takes the value one when region i is a Spanish region exporting to a foreign contiguous region j and zero otherwise. These variables conveniently control for higher inter-regional trade flows between contiguous Spanish regions as well as for the higher concentration of trade between border regions of different countries (Spain-Portugal, Spain-France, Spain-Andorra). It is in line with the results of Lafourcade and Paluzie (2011), who have shown that border regions in countries like France and Spain tend on average to capture larger shares of bilateral trade and FDI flows.

The terms μ_{it} and μ_{jt} correspond to the multilateral-resistance fixed effects for each origin and destination region interacted with time, respectively (Anderson and van Wincoop, 2003; Feenstra, 2002). These fixed effects are meant to control for competitive effects exerted by the non-observable price index of partner regions and by other competitors. They are also meant to capture other particular characteristics of the regions in question. To account for the likely heterogeneity between countries and its effect on the estimate of a single border effect, we have also added for each destination country a fixed-effect term (μ^u) as well as a fixed-effect term interacted with time (μ_t^u), with the aim of controlling for country characteristics that vary during the period (national cycle and national political shocks).

Eq. [2] describes an alternative specification with certain refinements in the treatment of distance.

$$Ln\frac{T_{ijt}^{eu}}{GDP_{it}GDP_{jt}} = \beta_0 + \beta_1 Ownregion + \beta_2 Owncountry + \beta_3 Internal\_Contig +$$

$$\beta_4 External\_Contig + \beta_5 dist_{ij} + \beta_6 (dist_{ij})^2 + \mu_{it} + \mu_{jt} + \mu^u + \mu_t^u + \varepsilon_{ijt} \qquad [2]$$

It thus includes, apart from the traditional variable $dist_{ij}$, a new variable$(dist_{ij})^2$. As in Hillberry and Hummels (2008), Llano-Verduras et al. (2011) and Garmendia et al. (2012), the variable $(dist_{ij})^2$ is defined as the square of the distance between trading regions and is expected to capture the non-linear relationship between trade and distance that is observed for kernel regressions in **Figure 1**. Also in line with these papers, we split the interpretation of these two variables (capturing the negative but non-linear effect of distance on trade) into two parts: (i) a negative and direct effect of distance on trade and (ii) a positive effect for the square of the distance, to capture the high concentration of trade over the shortest distance as observed in the kernel regression.

3.2 Alternative specifications

3.2.1. Gravity equation with segmented distance

Next, as an alternative way to deal with the non-linear relationship between trade and distance, we introduce a flexible approach that controls for changes in the slope of our linear estimation for different "segments" of the sample, these segments corresponding to different distances traveled by trucks. Although purely non-parametric techniques such as kernel regression offer certain flexibility, they cannot quantify the border effects under discussion. As we will see in the next section, this new approach generates different results from those of the square of distance. In our view, the variation is due to the differing capacities of the alternative strategies to deal with the non-linear relationship shown in **Figure 1**, which repeats itself at different levels of aggregation, perhaps as flows cross certain thick borders[8]. For each regression using this approach

[8] Two examples of thick borders (i.e., administrative borders coinciding with specific forces that cause considerable agglomeration of trade at a short distance) are: (a) Internal borders defining large metropolitan areas; these may coincide with the space where the forces of economic agglomeration around cities are at work, causing a great volume of intra- and inter-regional flow between contiguous regions. (b) International frontiers, coinciding with disproportionate divisions in terms of legal, cultural, historical and political barriers to trade.

we proceed as follows[9]: (1) we rank the whole sample by increasing distance; (2) we divide the entire range of distance traveled (max-min distance observed in the sample) into "segments" (stretches). For purposes of rigor, we define the "segments" in three alternative ways:

i. **"Naïve":** The first way simply divides the entire range of actual distance traveled into four stretches of equal length (in kilometers). We call it "naïve" because it ignores the expected higher intensity of flows over the shortest distance.

ii. **"Fibonacci":** The second way follows the *Fibonacci sequence*, a "magical" mathematical relation that appears in several natural phenomena (the reproduction of rabbits, the internal structure of sunflowers, etc.). The sequence has been used in architecture and in certain fields of economics and finance but, to the best of our knowledge, never before in trade. One benefit of the sequence is that it produces "segments" of increasing length. Another is that the sequence, although completely exogenous, fits perfectly with the non-linear intensity of trade at the nearest distance, dividing the entire range of distance as follows: first stretch: 8% of distance; second stretch: 8%; third stretch: 17%; fourth stretch: 25%; fifth stretch: 42% (100% in total).

iii. **"Quartile":** The third way assures an equal distribution of the number of observations per segment. It arranges them into quartiles of observation distribution, ranked by distance traveled.

This novel strategy is formally expressed in Eq. [3]:

$$\frac{T_{ijt}^{eu}}{GDP_{it} x GDP_{jt}} =$$

$$\beta_0 + \beta_1 Ownregion + \beta_2 Owncountry + \beta_3 Internal\_Contig + \beta_4 External\_Contig +$$

$$\text{STRETCH}^s * Ln(dist_{ij}) * \theta + \mu_{it} + \mu_{jt} + \mu^u + \mu_t^u + \varepsilon_{ijt} \qquad [3]$$

$STRETCH^s * Ln(dist_{ij})$ denotes the interaction between the log of the distance and a matrix STRETCH, which contains a set of dummy variables identifying each "segment". By including such interactions, we essentially introduce a set of "semi-

[9] Note that segmentation of the sample by range of distance traveled varies for specifications that estimate *internal border effects* (subsample excluding inter-national flows) or focus on *external border effects* (subsample excluding intra-regional flows).

dummy" variables, where $Ln(dist_{ij})$ replaces the value one of a normal dummy for the corresponding stretch. θ is a vector containing the coefficients for each distance stretch. Superscript *s* indicates the three alternative ways of splitting the sample (Naïve, Fibonacci, Quartile). The rest of the variables are the same as those used in previous specifications.

3.2.2. A Piecewise regression approach

As a robust check, three alternative piecewise regressions—namely, two linear and one restrictive cubic spline models—have been estimated. Piecewise models are also known as spline regressions and are described in the literature as efficient ways to approximate true non-linear relationships in data. Their main advantage is that the shape of the estimated function acquires a larger flexibility and is data driven, since no form is imposed a priori. A piecewise linear function is composed of linear segments—straight lines—separated by a number of *knots*. In some econometric packages (i.e., Stata) the number of knots as well as the specific location of each can be set a priori by the researcher, or be automatically assigned by the procedure to find the best fit for the data. In keeping with our previous models, we consider four segments in every one of them. In our case, the three spline models can be described by Eq. [4], where element $f(Ln(dist_{ij}))$ corresponds to the three alternative segment definitions:

$$Ln\frac{T_{ijt}^{eu}}{Y_{it}Y_{jt}} = \beta_0 + \beta_1 Ownregion + \beta_2 Owncountry + f(Ln(dist_{ij})) + \mu_{it} + \mu_{jt} + \mu^u + \mu_t^u + \varepsilon_{ijt} \quad [4]$$

For the first linear spline (M10 in **Table 4**), three equally spaced knots were set. For the second linear spline (M11 in **Table 4**) the three knots were assigned in accord with the sample's quartiles. Finally, we estimated a restricted cubic spline to better capture the strong non-linearity observed in the shortest distance (M12 in **Table 4**), here we also set four segments (knots=3) a priori, although the size of each segment was automatically determined.

3.2.3. A semi-parametric regression approach

We have also applied a semi-parametric approach (Pagan and Ullah 1999; Yatchew, 1998) for the same purpose of achieving some flexibility in modeling the non-linearity as well as in estimating our desired parameters (*internal* and *external border effects*). We have followed Robinson (1988),[10] who described a general model of the type in Eq. [5]:

$$y_{ij} = \beta_0 + x_{ij}\beta + f(Ln(dist_{ij})) + \varepsilon_{ijt} \qquad [5]$$

Where y_{ij} is the dependent value expressed in dyadic terms ij (bilateral flows divided by the corresponding income levels), and $x_{ij}\beta$ is a matrix with the corresponding explanatory variables whose parameters are to be estimated (*internal* and *external border effects).* $Ln(dist_{ij})$ is the explanatory variable that enters the equation non-linearly in accord with a non-binding function *f*. This model can be estimated using Robinson's (1988) double residual method, which starts by applying a conditional expectation to both sides of [5]. This leads to:

$$E\left(y_{ij}\middle|Ln(dist_{ij})\right) = \beta_0 + E\left(x_{ij}\middle|Ln(dist_{ij})\right)\beta + f(Ln(dist_{ij})) + \varepsilon_{ijt} \qquad [6]$$

By subtracting [6] from [5], we obtain Eq. [7]

$$y_{ij} - E\left(y_{ij}\middle|Ln(dist_{ij})\right) = \left(x_{ij} - E\left(x_{ij}\middle|Ln(dist_{ij})\right)\right)\beta + \varepsilon_{ijt} \qquad [7]$$

If the conditional expectations are known, parameter vector β can be estimated by means of OLS. If they are unknown, they can be estimated with a non-parametric kernel estimator, as in Robinson (1988).

[10] Verardi and Nicolas (2012) have described this approach and developed the corresponding Stata routine (*semipar*) for implementing it. We are grateful for this contribution.

3.3 A look at the effect of the national border through the odds ratio approach

After the robust check on the treatment of non-linearities, we would now like to focus on the external border effect by country. Our approach takes inspiration from another classic specification in the literature of trade integration (Head and Mayer, 2000; 2013; Poncet, 2003, 2005): the *absolute odds ratio approach* (Combes et al., 2005; Head and Mayer, 2013). This theory-based specification makes use of a convenient feature of CES demand functions, and models the ratio of an inter-regional flow to an intra-regional one, under the assumption that the flow depends only on the ratio of the monadic characteristics of the regions involved. This new specification is formally expressed in Eq. [8]:

$$\mathrm{Ln}\left(\frac{T_{ijt}^{eu}}{T_{iit}^{eu}}\right) =$$

$$\beta_0 + \beta_1 Country\_border + \beta_2 \ln\left(\frac{GDP_{jt}}{GDP_{it}}\right) + \beta_3 \ln\left(\frac{w_{jt}}{w_{it}}\right) + \beta_4 \mathrm{Ln}\left(\frac{\mathrm{STRETCH}^{s} * dist_{ij} * \theta}{dist_{ij}}\right) + \mu_t + \varepsilon_{ijt} \qquad [8]$$

Note that the endogenous variable consists of interregional (national and international) flows divided by intra-regional flows. It is also worth mentioning that the reference group is now intra-regional trade, so the effect should be interpreted in the opposite direction to that of the previous specifications. The constant term of the model measures how much less—on average—a Spanish region trades with another Spanish region than with itself (intra-regional trade). The *Country_border* dummy measures how much less—on average—a Spanish region trades with another foreign region than with itself (intra-regional trade). Thus by subtracting the constant term from the *Country_border* parameter we obtain the external border effect, which should be interpreted as how much greater domestic inter-regional trade is than international inter-regional trade. Note that to emphasize that change we label the new dummy *Country_border* and define it as the inverse of the *Owncountry* dummy. *Country_border* now takes the value 1 when the destination of the flow is abroad; we can consequently expect it to take a negative sign.

Next, in keeping with Head and Mayer (2000), and in order to assure a higher level of comparability with our previous specifications, we impose the assumption of unit elasticity on relative production, passing the ratio of GDPs to the left-hand side of the

equation. By doing so, we avoid the simultaneity problem. Now the gravity equation resembles that in Eq. [9]:

$$\mathrm{Ln}\left(\frac{T_{ijt}^{eu}}{T_{iit}^{eu}}\right) - \mathrm{Ln}\left(\frac{GDP_{jt}}{GDP_{it}}\right) = \beta_0 + \beta_1 Country\_border + \beta_2\,\mathrm{Ln}\left(\frac{w_{jt}}{w_{it}}\right) + \beta_3 \mathrm{Ln}\left(\frac{\mathrm{STRETCH}^{s} * dist_{ij} * \theta}{dist_{ij}}\right) + \mu_t + \varepsilon_{ijt} \qquad [9]$$

4 Results

In this section we analyze the main results for the twenty models estimated in this paper with our novel region-to-region dataset. The first specifications to be considered use *corrected trade* flows $\frac{T_{ijt}^{eu}}{GDP_{it}GDP_{jt}}$ as the endogenous variable as well as all the fixed-effects approaches described above. However, each uses a different treatment of the distance variable. In contrast to previous papers, the effects of external (*Owncountry*) and internal (*Ownregion*) borders are estimated simultaneously in all specifications: that is, with the whole sample considered at the same time. We are thus able to determine whether the two border effects are at work: that is, when certain (international) flows are crossing two borders (one internal, the other external) or more[11]. Note that such results would not be fully comparable with those previously reported, since with just two dummies we would have to control for three types of flows (intra-regional, inter-regional within Spain and inter-regional with other EU countries). However, this approach is close to computing the *internal border effect* (*Ownregion*) within a single country (*Euroland*) with two nested administrative borders, as Hillberry and Hummels (2008) did for the U.S. Our analysis and interpretation of the results will therefore be close to theirs.

The results generated by Eqs. [1] to [3] are reported in **Table 1**. Ordinary Least Square (OLS) estimators are used when the gravity equation is applied to a dataset with no zero-values, in this case we are modeling only the intensity of flows between regions, not the drivers behind the existence or non-existence of said flows.

[11] Note that in some cases (e.g., exports from Spain to Germany) a Spanish truck may cross three or four different national borders. As described in the introduction, this could induce additional "jump" in the intensity of trade.

Table 1. Alternative Estimates for *External Border Effects*.

M1–M2 are based on Eq. [1], M3 on Eq. [2], M4–M6 on Eq. [3].

	M1	M2	M3	M4-Naïve	M5-Fibonacci	M6-Quartile
	OLS	PPML	PPML	PPML	PPML	PPML
VARIABLES	$Ln(T_{ijt}\_corr)$	T_{ijt}_corr	T_{ijt}_corr	T_{ijt}_corr	T_{ijt}_corr	T_{ijt}_corr
$Ln(dist_{ij})$	-1.035***	-0.908***				
	(0.0950)	(0.123)				
$dist_{ij}$			-2.025***			
			(0.335)			
$dist_{ij}^2$			0.541***			
			(0.1000)			
$Ln(dist_{ij}$ stretch1)				-1.179***	-1.104***	-1.195***
				(0.172)	(0.253)	(0.159)
$Ln(dist_{ij}$ stretch2)				-1.129***	-1.151***	-1.116***
				(0.158)	(0.229)	(0.148)
$Ln(dist_{ij}$ stretch3)				-1.082***	-1.128***	-1.082***
				(0.153)	(0.211)	(0.145)
$Ln(dist_{ij}$ stretch4)				-1.044***	-1.048***	-1.046***
				(0.150)	(0.198)	(0.141)
$Ln(dist_{ij}$ stretch5)					-1.002***	
					(0.191)	
Ownregion	1.373***	2.031***	3.513***	1.380***	1.339**	1.289***
	(0.324)	(0.424)	(0.313)	(0.512)	(0.546)	(0.499)
Owncountry	0.682**	0.932***	1.270***	1.259***	1.485***	1.484***
	(0.293)	(0.187)	(0.203)	(0.190)	(0.199)	(0.199)
Internal Contig	0.396***	0.570***	0.997***	0.312	0.199	0.243
	(0.116)	(0.205)	(0.189)	(0.228)	(0.243)	(0.231)
External Contig	0.249	-0.186	0.270	-0.398	-0.471	-0.493
	(0.214)	(0.337)	(0.325)	(0.347)	(0.352)	(0.353)
Constant	-22.66***	-24.22***	-29.29***	-22.90***	-23.36***	-22.97***
	(0.641)	(0.931)	(0.319)	(1.157)	(1.464)	(1.083)
Own Region	4	8	34	4	4	4
Own Country	2	3	4	4	4	4
Null hypothesis†	$Ln(dist_{ij}$ stretch1)-$Ln(dist_{ij}$ stretch2)=0, $\chi2$			5.24**	1.01	14.99***
	$Ln(dist_{ij}$ stretch2)-$Ln(dist_{ij}$ stretch3)=0, $\chi2$			9.36***	0.61	4.97**
	$Ln(dist_{ij}$ stretch3)-$Ln(dist_{ij}$ stretch4)=0 $\chi2$			3.68*	13.91***	5.6**
	$Ln(dist_{ij}$ stretch4)-$Ln(dist_{ij}$ stretch5)=0, $\chi2$				9.28***	
Observations	3,688	6,376	6,376	6,376	6,376	6,364
R-squared	0.812	0.905	0.882	0.906	0.906	0.905

All regressions include an "origin by year", "destination by year", "destination country by year" and "destination country" fixed effect. Tijt_corr = Tijt/GDPixGDPj. The standard errors, which are robust and clustered by the distance variable, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

When zeros are included[12], we use instead the pseudo-maximum likelihood technique (PPML). It was Santos Silva and Tenreyro (2006) who proposed using the PPML approach, which also sorts out Jensen's inequality (note that the endogenous variable is in levels) and produces unbiased estimates of the coefficients by solving the heteroskedasticity problem.

Table 1 reports the results for the first six models. M1 includes the endogenous variable and the distance in logs (OLS without zero flows). M2 include zero flows and use the PPML estimator. Thus the endogenous variable is expressed in levels and distance in logs. In M3 *distance* and the *square of distance* are included in levels. Next, to shed more light on the non-linear relationship between trade and distance, M4–M6 report the corresponding results for three alternative models based on our alternative strategy (Eq. [3]), which segments the sample in three ways by trucking distance. This procedure estimates the elasticity of distance in each interval. Note that in these models the distance variables for each "stretch" are also expressed in logs.

The first three models (M1-M3) generate significant coefficients with the expected signs for all variables except *External_Contig*. This result suggests that the difference in the intensity of trade between a Spanish region and a foreign border region, on the one hand, and between non-adjacent Spanish regions, on the other, is non-significant. However, the coefficient for the *Internal_Contig* variable is positive and significant. In the three models the coefficient for the distance variable is negative and significant, with elasticities that are within normal range. Moreover, the results for distance variables that control for the non-linear relationship between trade and distance in M3 suggest that distance acts as a clear impediment to trade (negative coefficient for $dist_{ij}$), but an impediment that tapers off as distance increases (positive coefficient for $dist_{ij}^2$).

Regarding the *Owncountry* dummy, model M1 reaches a value of 2 while the other two (M2–M3) reach a factor of 3 and 4, respectively. Note that the latter two stand up

[12] The zero values considered in our dataset correspond to region dyads that had non-zero values at least in one year in the period 2004–2007. Zeros corresponding to regions that did not receive any exports from a Spanish region during that period are not considered in our sample.

robustly to alternative specifications (M4-M6), subsamples, estimation procedures and treatments for the non-linear relationship between trade and distance. This persistent value of about 4 for *Owncountry* is very close to the one obtained by Llano-Verduras et al. (2011) with region-to-country [3.3 = exp(1.2)] and province-to-country [4.9 = exp(1.6)] data. Note that these papers use different datasets and definitions of the border dummy, but similar specifications for distance and the same estimation procedures as in M2 and M3 (PPML). By contrast, the results for the *Ownregion* dummy are more puzzling, showing a larger variation with respect to the econometric method used to control for the non-linearity. On the one hand, the factor obtained for *Ownregion* for M1, M4, M5 and M6 is exactly the same (4) and coincides with the one for *Owncountry*. On the other, *Ownregion* slightly rises for M2 (factor of 8) and skyrockets to a factor of 34 in M3. Note that this jump just occurs when the PPML is used with the whole sample (zero flows included) and the quadratic term. Similar results have not been reported in articles where *Ownregion* was estimated alone (Garmendia et al., 2012): that is, with the external border effect and international flows excluded.

Table 1 reports promising results for models M4 to M6, which employ our new controls for non-linearity. The coefficients for $STRETCH*Ln(dist_{ij})$ in each of the segments are negative and highly significant for these three alternative models. More interestingly, in M4 (Naïve) and M6 (Quartile) the negative elasticity for each stretch decreases, which is consistent with a segmentation where the distance variable is shorted in increasing order. In M5 (Fibonacci), however, the negative elasticity of distance increases in the first two segments (from –1.104 to –1.151) and decreases thereafter (from –1.151 to –1.002). The last part of **Table 1** shows the results for the Wald test applied to the Null-hypothesis of equal elasticities between consecutive stretches of distance. The results show that the differences between the stretches are statistically significant, with the exception of the first three consecutive segments in the Fibonacci division. For *Owncountry*, the three alternative procedures for segmenting the sample reach exactly the same positive and significant factor of 4, close to the obtained in M2 and M3. Moreover, the results for *Ownregion*, also point out to a factor of 4, which is significant in all cases. Therefore, we can conclude that *Ownregion* increases when zero flows are included and the PPML is used (M2–M3) but this effect is controlled for when segmented distance is used (M4–M6). Finally, it is interesting to note that the

coefficients for *Internal_Contig* and *External_Contig* become non-significant. This suggests that when the non-linear relationship is controlled for by segmentation of the sample into stretches, the alternative control for the higher intensity of trade over the shortest distance (contiguity) becomes redundant.

Table 2. Characterization for three Alternative Segmentations.

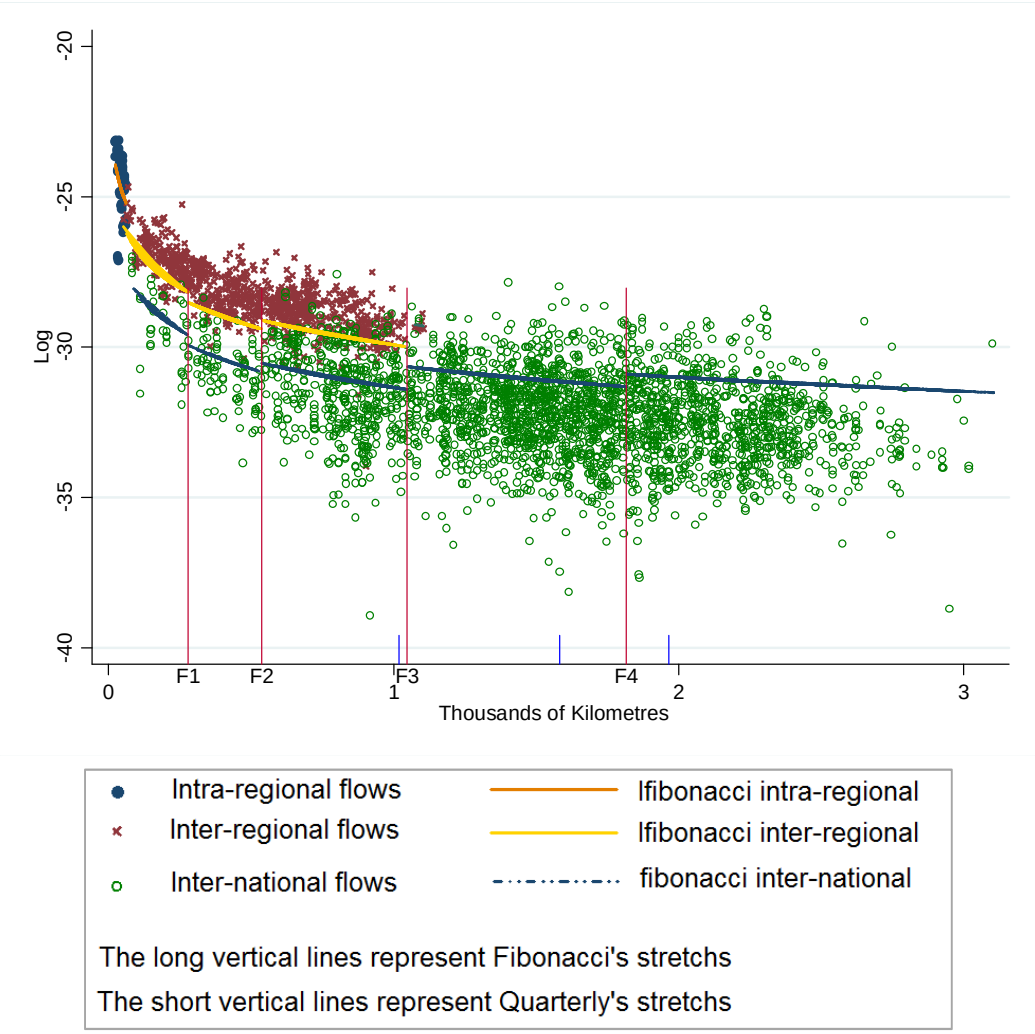
Naive	Km. (thousands)	0.0229 - 0.7921	0.7946 - 1.5606	1.5618 - 2.3285	2.3318 - 3.1		TOTAL
	Rango	25%	25%	25%	25%		100%
	obs. (%)	18%	31%	42%	9%		100%
	zeros (%)	1%	12%	24%	5%		44%
	R2 (Ln($dist_{ij}$))	**90%**	**68%**	**16%**	**23%**		0.905
	R2 ($dist_{ij}$, $dist_{ij}^2$)	**88%**	**68%**	**18%**	**23%**		0.882
Fibonacci	Km. (thousands)	0.0229 - 0.2791	0.2802 - 0.5357	0.5366 - 1.048	1.0488 - 1.8153	1.819333 - 3.1	TOTAL
	Rango	8%	8%	17%	25%	42%	100%
	obs. (%)	5%	6%	15%	42%	33%	100%
	zeros (%)	0%	0%	2%	22%	18%	44%
	R2 (Ln($dist_{ij}$))	**89%**	**75%**	**62%**	**34%**	**18%**	0.905
	R2 ($dist_{ij}$, $dist_{ij}^2$)	**90%**	**75%**	**62%**	**34%**	**18%**	0.882
Quartiles	Km. (thousands)	0.0229 - 1.0187	1.0231 - 1.5823	1.5825 - 1.961	1.9648 - 3.1		TOTAL
	Rango	32%	18%	12%	37%		100%
	obs. (%)	25%	25%	25%	25%		100%
	zeros (%)	2%	12%	15%	14%		44%
	R2 (Ln($dist_{ij}$))	**90%**	**47%**	**34%**	**0.238**		0.905
	R2 ($dist_{ij}$, $dist_{ij}^2$)	**88%**	**46%**	**33%**	**0.238**		0.882

All regressions include an "origin by year", "destination by year", "destination country by year" and "destination country" fixed effect.

To bring home the previous results, **Table 2** summarizes the main features of the three alternative segmentations and provides measures of overall fitness for each segment. Several points are worth mentioning: (i) The three sequences have been defined not by volume or nature of trade but, instead, by distance range and number of observations (zero flows included). Thus the *Fibonacci* and *Quartile* sequences consider segments of different length. (ii) The percentage of zero values for each stretch and each criterion is different. Zero values are highly concentrated in the longest trips (mainly international flows). (iii) For a complementary view, we show the close fit obtained with the models

run separately for each segment of the sample[13]. **Table 2** reports the R^2 for regressions that use ($Ln(dist_{ij})$) or ($dist_{ij}$, $dist_{ij}^2$): i.e., for the counterparts to the specifications used in M2 and M3. Note that, although the three alternative segmentation criteria generate the same R^2 ($Ln(dist_{ij})$) and $R^2(dist_{ij}, dist_{ij}^2)$ for the whole sample (TOTAL column), the quality of the fit is different for each segment and sequence. Throughout the sample (TOTAL), R^2 is always higher with $Ln(dist_{ij})$ (90.5%) than with $dist_{ij}$ and $dist_{ij}^2$ in levels (88.2%). The *Fibonacci* sequence (followed by *Quartile*) shows the best fits when the model is regressed for the last subsamples (largest distances) and generates the highest R^2 ($Ln(dist_{ij})$) and $R^2(dist_{ij}, dist_{ij}^2)$. Conversely, although the *Naïve* sequence performs well for the first two stretches, it fails for the last two.

Figure 3: Trade flows by nature and distance stretches, 2004–2007.



[13] Note that the results in Table 2 consider the whole sample and use the strategy for Eq. [3], where each segment is controlled by a semi-dummy obtained through the interaction of a dummy and the distance variable.

To complement the previous table, **Figure 3** shows the distribution of the dependent variable (in logs) in regards to distance. It uses three different colors for identifying the main categories of trade flows (intra-regional; inter-regional within Spain; inter-regional exports to the eight EU countries). It also includes full-vertical lines in red for identifying the five stretches of the Fibonacci sequence, and short-vertical-lines in blue for the Quartile. The plot shows a clear "jump" in the intensity of intra-national (red-crosses and blue-bullets) and inter-national (green hollow circles) flows. The non-linear relationship is also clear.

4.1 Robust checks with alternative procedures

Before we conclude, we analyze in this section the results obtained for a last set of specifications based on the last two alternative econometric methods described in section 4.2: namely, the spline and semi-parametric regressions. It is now worth considering that these flexible procedures play a competing role against the contiguity dummies, which also tend to control for "jumps" in the relationship between trade and distance. Taking our cue from their non-significant results in certain cases of the previous section, we therefore now exclude adjacency dummies.

Table 3 reports alternative results for *Ownregion* and *Owncountry effects* when they are computed by the methods described in Eqs. [3–5]. The first three models—M7, M8 and M9—correspond to the PPML estimates applied to the whole dataset (with zeros) once the sample has been controlled by the three sets of semi-dummies containing the segmented distance (Naïve, Fibonacci and Quarterly, respectively). Taking into account the non-significant results obtained for the contiguity variables in models M4–M6, we now exclude these variables. The results vary slightly from those reported in **Tables 1**. *Ownregion* remains significant but decreases to a factor of 3, while *Owncountry* reaches a factor of 5. The next models (M10, M11, M12) correspond to three alternative procedures for estimating the spline regression. In M10 the knots of the spline regression are equally spaced over the range of the distance variable, in M11 they are

placed at the quartiles of the distance variable, and in M12 the "natural spline" (i.e., where spline regression creates variables containing a restricted cubic spline) is applied.

Table 3. Alternative Estimates for the *External* and *Internal Border Effect*. M7–M9 are based on Eq. [3], M10–M12 on Eq. [4] and M13–M14 on Eq. [5].

	M7	M8	M9	M10	M11	M12	M13	M14
	Naïve	Fibonacci	Quarterly	Naïve	Quarterly	Cubic		
	PPML	PPML	PPML	SPLINE-OLS	SPLINE-OLS	SPLINE-OLS	SEMI-PAR	SEMI-PAR
VARIABLES	$T_{ijt_}$corr			Ln($T_{ijt_}$corr)				
Ln(D_{ij} stretch1)	-1.318***	-1.132***	-1.282***	-0.891	-1.287***	-1.363***		
	(0.114)	(0.243)	(0.100)	(0.647)	(0.0591)	(0.0842)		
Ln(D_{ij} stretch2)	-1.255***	-1.191***	-1.193***	-1.250***	-0.850***	0.467*		
	(0.103)	(0.212)	(0.0937)	(0.135)	(0.203)	(0.252)		
Ln(D_{ij} stretch3)	-1.204***	-1.166***	-1.157***	-1.233***	-1.509***	-3.966		
	(0.0995)	(0.193)	(0.0920)	(0.0912)	(0.381)	(3.018)		
Ln(D_{ij} stretch4)	-1.162***	-1.080***	-1.119***	-1.089***	-1.315***	8.703		
	(0.0987)	(0.181)	(0.0904)	(0.191)	(0.454)	(16.91)		
Ln(D_{ij} stretch5)		-1.033***						
		(0.175)						
Ownregion	0.943**	1.143**	0.987***	0.997**	0.692***	0.576**	1.499***	1.377***
	(0.368)	(0.456)	(0.355)	(0.403)	(0.226)	(0.238)	(0.395)	(0.423)
Owncountry	1.284***	1.544***	1.541***	0.800***	0.816***	0.820***	0.858***	0.905***
	(0.184)	(0.195)	(0.195)	(0.289)	(0.300)	(0.298)	(0.274)	(0.263)
Constant	-21.98***	-23.12***	-22.41***	-22.94***	-21.17***	-20.77***		
	(0.756)	(1.343)	(0.692)	(2.850)	(0.436)	(0.529)		
Own Region	3	3	3	3	2	2	4	4
Own Country	4	5	5	2	2	2	2	2
Observations	6,376	6,376	6,364	3,688	3,688	3,688	3,688	3,688
R-squared	0.904	0.906	0.905	0.811	0.812	0.811	0.328	0.506
FE by origin x time	YES	YES	YES	YES	YES	YES	NO	YES
FE by destination x time	YES	YES	YES	YES	YES	YES	NO	YES

All the regressions include country-destination fixed effects and country-destination by year fixed effects. Tijt_corr = Tijt/GDPixGDPj. The standard errors, which are robust and clustered by the distance variable, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Although the spline models (M10–M12) resemble our previous approach (M7–M9), there are several differences worth mentioning: (i) the three spline models are based on the OLS estimator and are applied to our restricted sample with no zero flows; (ii)

conversely, our previous approach used the PPML estimator and the complete sample. The consequences are twofold. First, the number of observations considered for the PPML-STRETCH approach is 6,376, whereas that for the SPLINE-OLS is 3,688; segment length in each is therefore different under the quartile criterion. Second, the PPML gives more consideration to the largest-value observations: that is, the ones taking place within Spain. That said, the two approaches rely on similar assumptions and reach coherent results: in both cases the *Ownregion* and *Owncountry* effects are low and significant, with a factor that ranges from 3 to 2 for *Ownregion* and from 5 to 2 for *Owncountry*. Moreover, negative elasticity for the distance variable also varies by stretch: in contrast to our finding when the semi-dummy variables were used for segmenting the distance (M7–M9), the negative elasticities for the first and subsequent stretches of distance in models M10–M12 do not show a clear decreasing pattern (in absolute terms); now, for example, the largest negative elasticity corresponds to the third segment in all cases; moreover, in some of them the coefficient becomes non-significant or even positive (stretch 2, 3 and 4 in M12-Cubic spline).

The last two columns correspond to the results obtained with the semi-parametric regression. In M13 the model is estimated with country fixed effects and with country fixed effects interacted with time, while in M14 time-origin region and time-destination region fixed effects are added. The idea here, as in Benedictis et al. (2008), is to test extent to which our results are affected by the inclusion of a large number of fixed effects. Note that elasticity for distance is excluded, since its effect is captured by the corresponding kernel distribution, and extracted from both sides of the function as expressed in Eqs. [6] and [7]. Now, when this highly flexible approach is applied, *Ownregion* becomes significant and positive again with a factor of 4, while *Owncountry* decreases to a factor of 2.

Finally, to illustrate the performance of these three highly flexible approaches for dealing with the non-linear relationship between trade and distance, we report two informative plots. **Figure 4** shows the scatterplot of the dependent variable (in logs) with respect to distance (levels), along with the trade predicted with the cubic spline regression and the model using the *quadratic* term (D_{ij}^2). Similarly, **Figure 5** shows the scatterplot of the dependent variable against the prediction based on the semi-parametric approach.

Figure 4. Spline regression scatterplot. 2004–2007.
In this regression the zero values and contiguity dummies are not included.

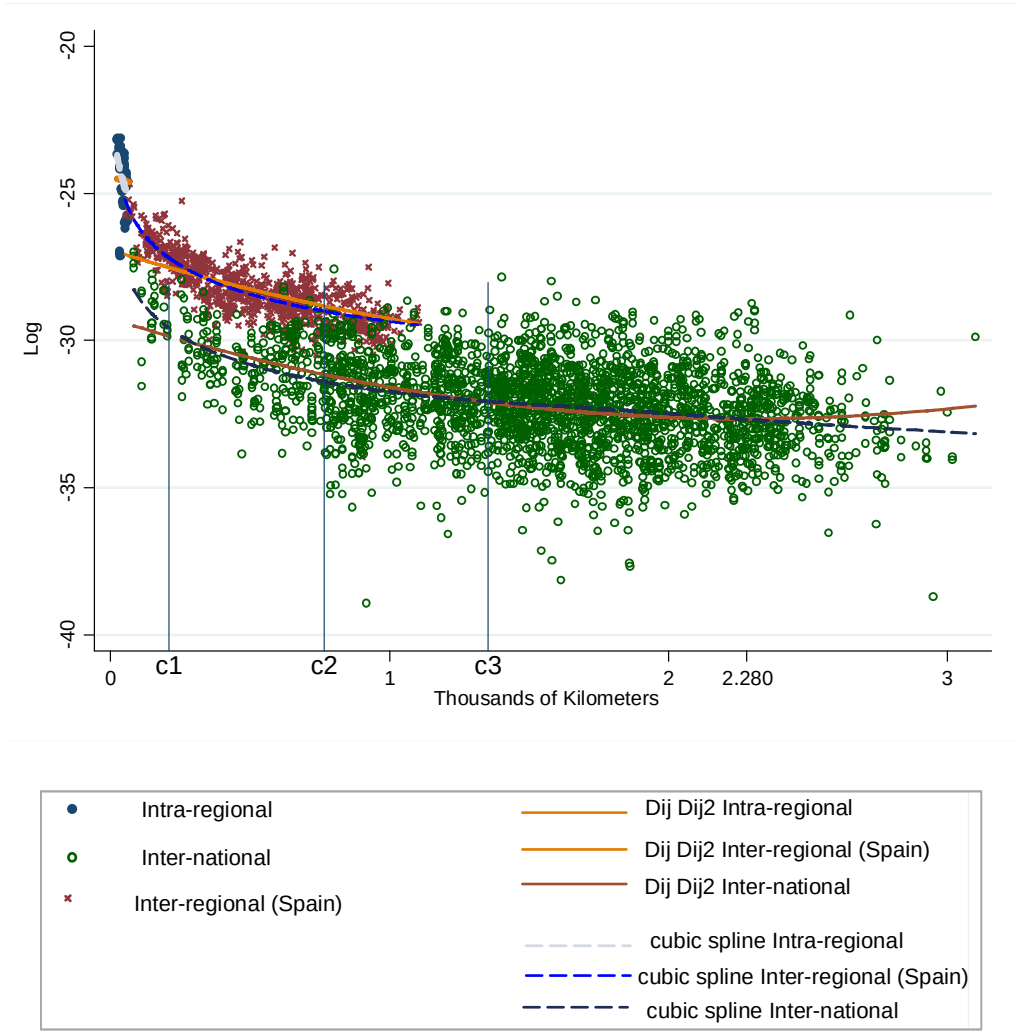


Figure 4 tries to shed some light on the alternative performance of the spline model versus the quadratic form and, more specifically, on the atypical high internal border effect obtained in M3 versus that in any other specification. The blue vertical lines indicate segments automatically set by the cubic spline (C1, C2, C3). A vertical black line indicates the distance 2,280 km, which is the distance at which the quadratic term of the distance variable is reverted: the turning point where the parabola's slope becomes positive[14]. Several results are worth mentioning. First, although the scale of the

[14] Although widely used for dealing with non-linear functions, the quadratic model suffers from a potential limitation: the reversal of the effect's direction. Normally, the quadratic model is used under the assumption that the turning point lies outside the sample (Gould, 1993). In order to compute the point at which the effect changes direction, we use the following expression: $-\beta_5/(2\beta_6)$, where β_5 and β_6 are, respectively, the coefficients for the distance and the square for distance in our Eq. [2]. The 2,280km [–(–2.98523)/(2*0,654596)], corresponds to the estimate plotted in the graph, which was based on OLS and the sub-sample with non-zero flows (equivalent to the spline estimates plotted in the same graph). If

graph does not show this clearly, the spline-model prediction is a better fit than the quadratic-model prediction for the largest flows over the shortest distance (intra- and inter-regional within Spain over the shortest distance). Second, the shape of the predictions based on spline and square-of-distance models for international flows (in green) is very similar. Moreover, if we consider that our sample is of eight EU countries and of relatively short distances (< 3,000 km), the number of flows going beyond the parabola's turning point (2,280 km) is not especially high for non-zero values. However, this could point to a stronger limitation in **Table 2**, where PPML and zero flows are included, since the turning point occurs at 1,873 km.

At this point, it is worthwhile to sum up our results, which might have something to do with the nature of the two border effects considered here. On the one hand, *Ownregion*, far from being explained by external barriers to trade (division or fragmentation), seems most closely related to the economics of agglomeration around metropolitan areas (Diaz-Lanchas, et al., 2013), as well as to the spatial spillover of the strongest regions and their neighbors. It thus seems sensitive mostly to mismeasurement, spatial-unit use (modifiable area unit problem, MAUP) and aggregation bias. *Owncountry*, on the other hand, seems to harder to budge (Wei, 1996). First, region-to-region international flows lead to lower external borders than do region-to-country datasets. However, even when we include zero flows (which tend to increase the external border, since most zero flows correspond to international flows) and control for the non-linear relationship of trade, we obtain a positive and significant factor that ranks between 5 and 4. Finally, according to our results, we find no strong variations in border effects when using alternative treatments for non-linearity (log-log; quadratic terms; and more flexible approaches based on segmented distance and non-parametric approaches) with the exception of M3 for *Ownregion*. Nevertheless, our results show larger variations in the elasticity of distance (by segment) and in the role played by (external and internal) contiguity than in the *border effects* themselves.

the analysis were repeated with a specification equivalent to that in model M3 (Table 2)—i.e., with the PPML and zero flows—then, the turning point would be 1,873 km [–(–2.012/(2*0.537))= 1,873].

4.2 Alternative specification and country-specific analysis

We would now like to discuss the national border effect in greater detail, taking each country separately. As described in section 3.3, this analysis is based on the odds-ratio specification, which uses intra-regional flows as the reference group. Note that intra-regional flows are not included in the sample, but are used as the denominator on the left-hand side of the equation. **Table 4** reports our results with six alternative specifications for the *border effect* of each importing country: those for M15, M16 and M17 are based on Eq. [8], while those for M18, M19 and M20 correspond to Eq. [9]. In all cases, the treatment of non-linearity in distance is based on our three alternative segmented-distance variables (Naïve, Fibonacci, and Quarterly). The "external border" effect is obtained for each country by the anti-log of the coefficient of the corresponding *Country_border* dummy. It must be noted that these *Country_border* dummies (one for each foreign country) are defined in opposition to *Owncountry*. By contrast, the "internal border" is deduced from the intercept term, as described above in Section 3.3. Note that the effects are now expressed negatively, indicating, for example, how much less a Spanish region exports to a non-adjacent foreign region than to a non-adjacent Spanish region, ceteris paribus.

The results are ranked by increasing order of *Country_border* coefficients in M15 (the ranking is pretty homogeneous for the six specifications, with the exception of the last position, for models based on Eq. [8] rather than on Eq. [9]). The lowest effects for *border* with M15 used as a benchmark are obtained for Portugal (4), France (5), Belgium (5), Germany (6) and the Netherlands (8), followed by Italy (10) and the UK (16). Andorra border effect was dropped to avoid a perfect multicollinearity effect, since the constant term takes part of the regression.

If we consider *border* as a measure of integration between Spanish regions and the regions of the eight European partners—with size, bilateral distance and relative wages previously controlled for—it is remarkable to find the highest levels of integration not only with the regions of the nearest countries (Portugal and France) but also with the regions in Belgium and Germany. Segmented distance performs similarly here and in the spline regressions, where in some cases the first segments show a lower negative elasticity than the next.

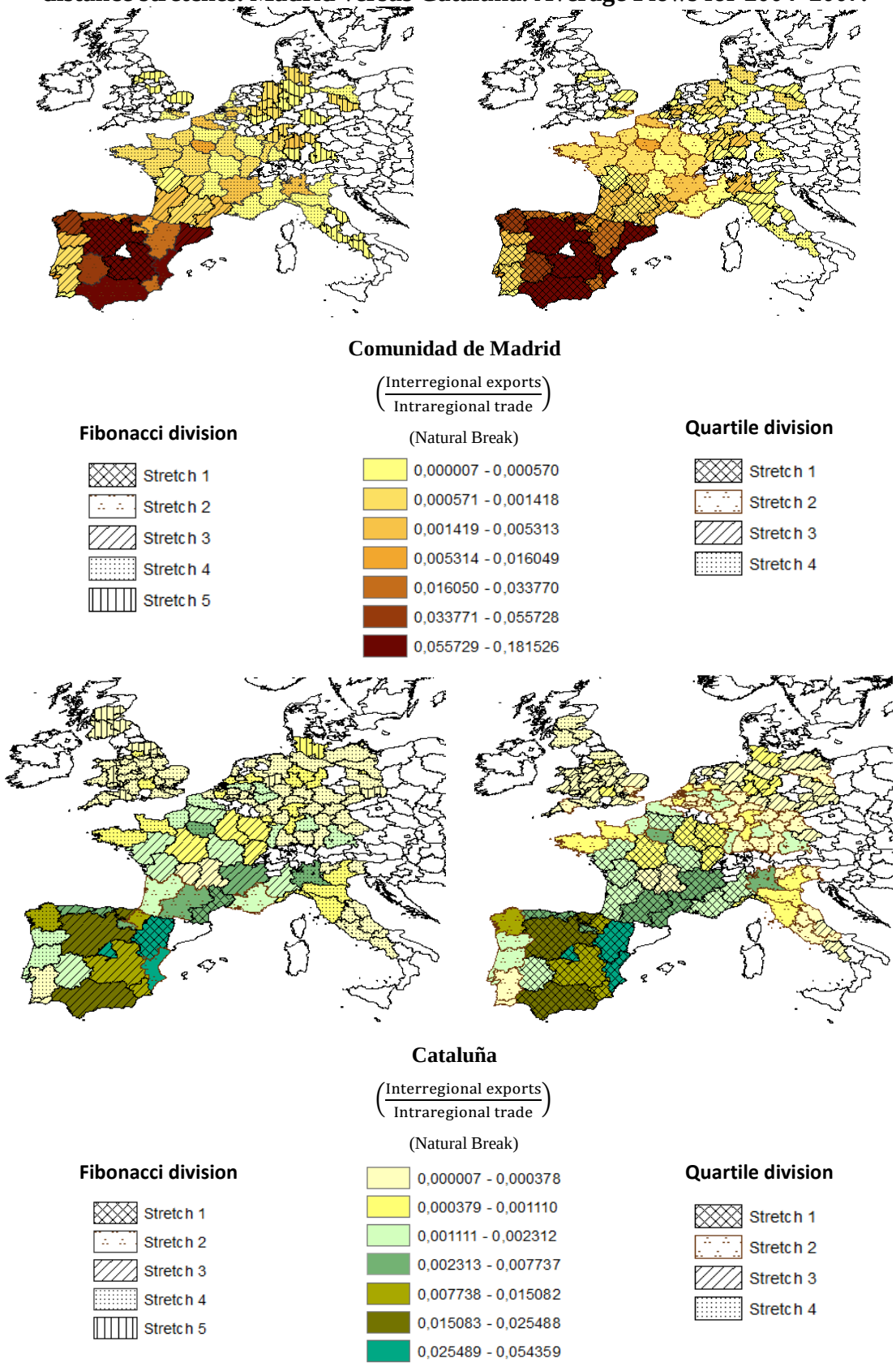
Table 4. External Border Effects by Country. Region-to-Region Spanish exports, 2004–2007. M15–M7 are based on Eq. [8] and M18–M20 are based on Eq. [9].

	M15	M16	M17	M18	M19	M20
	Naïve	Fibonacci	Quarterly	Naïve	Fibonacci	Quarterly
	PPML	PPML	PPML	PPML	PPML	PPML
VARIABLES	Tijt/Tiit			$e^{Ln(Tijt/Tiit) - Ln(GDPj/GDPi)}$		
Ln (GDP_j/GDP_i)	0.691***	0.694***	0.685***			
	(0.0600)	(0.0605)	(0.0598)			
Ln(wj/wi)i	-0.0977*	-0.103**	-0.0940*	-0.404***	-0.402***	-0.405***
	(0.0523)	(0.0527)	(0.0524)	(0.0263)	(0.0248)	(0.0263)
Ln($dist_{ij}$ stretch1/$dist_{ii}$)	-0.981***	-0.850***	-1.000***	-0.997***	-1.051***	-0.998***
	(0.0535)	(0.103)	(0.0510)	(0.0792)	(0.160)	(0.0795)
Ln($dist_{ij}$ stretch2/$dist_{ii}$)	-1.014***	-0.931***	-0.993***	-1.054***	-0.978***	-1.087***
	(0.0507)	(0.0722)	(0.0685)	(0.0621)	(0.114)	(0.0629)
Ln($dist_{ij}$ stretch3/$dist_{ii}$)	-0.940***	-0.938***	-0.938***	-0.934***	-1.096***	-0.961***
	(0.0540)	(0.0617)	(0.0608)	(0.0642)	(0.0947)	(0.0684)
Ln($dist_{ij}$ stretch4/$dist_{ii}$)	-0.906***	-0.914***	-0.918***	-0.873***	-0.998***	-0.889***
	(0.0922)	(0.0651)	(0.0570)	(0.127)	(0.0880)	(0.0650)
Ln($dist_{ij}$ stretch5/$dist_{ii}$)		-0.793***			-0.840***	
		(0.0649)			(0.0876)	
Border_PT	-2.338***	-2.348***	-2.356***	-2.639***	-2.594***	-2.644***
	(0.130)	(0.132)	(0.128)	(0.213)	(0.225)	(0.214)
Border_FR	-2.562***	-2.587***	-2.623***	-2.418***	-2.392***	-2.425***
	(0.124)	(0.127)	(0.143)	(0.0885)	(0.0902)	(0.0896)
Border_BE	-2.635***	-2.783***	-2.810***	-2.521***	-2.595***	-2.606***
	(0.181)	(0.176)	(0.210)	(0.199)	(0.182)	(0.196)
Border_DE	-2.880***	-3.073***	-3.099***	-2.631***	-2.797***	-2.830***
	(0.152)	(0.169)	(0.184)	(0.135)	(0.144)	(0.153)
Border_NL	-3.044***	-3.241***	-3.263***	-2.772***	-2.931***	-2.960***
	(0.215)	(0.227)	(0.242)	(0.284)	(0.274)	(0.274)
Border_IT	-3.336***	-3.496***	-3.536***	-3.749***	-3.854***	-3.919***
	(0.171)	(0.183)	(0.202)	(0.189)	(0.190)	(0.203)
Border_UK	-3.784***	-3.954***	-3.967***	-3.592***	-3.743***	-3.775***
	(0.188)	(0.191)	(0.212)	(0.168)	(0.170)	(0.175)
Constant	-1.022***	-1.194***	-0.984***	-0.961***	-0.901***	-0.958***
	(0.120)	(0.172)	(0.121)	(0.194)	(0.294)	(0.194)
Internal Border	3	3	3	3	2	3
Portugal	4	3	4	5	5	5
France	5	4	5	4	4	4
Belgium	5	5	6	5	5	5
Germany	6	7	8	5	7	7
Netherlands	8	8	10	6	8	7
Italy	10	10	13	16	19	19
United Kingdom	16	16	20	14	17	17
Observations	5,983	5,983	5,980	5,983	5,983	5,980
R-squared	0.691	0.699	0.690	0.510	0.515	0.510

All regressions include a time fixed effect. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Digging deeper into this analysis, **Figure 5** plots the spatial concentration of exports delivered from two key Spanish regions, Cataluña and Madrid, divided by their corresponding intra-regional flows. In the four maps, the *palette* corresponds to the number of flows, with seven color intensities automatically determined by the ArcGis's "natural break" option. We use this so that the data can speak for themselves. The first two are for Madrid, the others for Cataluña. We then use different frames to identify the regions included in the corresponding *Fibonacci* and *Quartile* stretches. It is worth mentioning that the color intensity shows a clear discontinuity in the relevance of trade flows between Spanish and European markets, even for a highly open border region such as Cataluña. It is also interesting to note which regions are included in each stretch for the two cases considered here: in the case of Madrid, the first stretch of the *Fibonacci* division (completely exogenous to our dataset) perfectly matches the two contiguous regions (Castilla y León and Castilla–La Mancha), while the second stretch captures the rest of the Spanish regions with the sole exceptions of Cataluña and Galicia. By contrast, the first stretch of the *Quartile* division is now broader, including all the Spanish regions as well the Portuguese and the nearest French ones. However, if we consider the regions classified in every stretch for Cataluña, we see that the stretches for national and international markets do not exactly correspond. We should thus emphasize that the composition of each stretch will naturally depend on the specific location of each Spanish exporting region.

Figure 5. Interregional exports (divided by the intra-regional trade) and main distance stretches. Madrid versus Cataluña. Average Flows for 2004–2007.



(*) The variable represented in these graphs consider the average across the whole period of the interregional export divided by the intraregional flows $\sum_t \frac{\overline{T}_{ijt}}{T_{ii,t}}$.

5 Conclusions

In this article we aim to shed new light on the *non-linear relationship between trade and distance* and its effect on the regional and national border effects of a country. With this purpose, we have made use of a novel dataset for inter-regional trade flows by Spanish trucking, including intra-national and inter-national flows between Spanish regions (NUTS 2) and the regions of Spain's eight main European partners and considering actual distance for the shipments.

In line with previous papers, we have considered three classic ways for dealing with non-linearity (log-log OLS; log-PPML; quadratic terms and PPML). In addition, we have developed a new strategy to deal with this *non-linearity*. Namely, we segment the sample, considering alternative stretches of the distance variable. Moreover, we have applied two additional estimation methods (piecewise regression and semi-parametric approaches) to estimate the desired parameters while managing the non-linearity in the most flexible way. The results obtained with these alternative strategies are quite robust: the internal border effect seems to be a robust factor of 4, reaching lower values (2) for some specifications. The effect of the national border (*Owncountry*) reaches a significant factor that oscillates between 4 and 7. We consider that these results support the call of Henderson and Millimet (2008) for further research on the appropriate gravity equation functional form, using parametric and non-parametric procedures. We have made the attempt here by using inter-national and inter-regional flows simultaneously.

References

Atkin D. and Donaldson D. (2012) "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs", Working paper.

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

Anderson, M., and S. Simth (1999) "Do National Borders Really matter? Canada-US Regional Trade Reconsidered" *Review of International Economics* **7** (2): 219-227.

Baum, C. F., Caglayan, M. and Ozkan, N. (2004) "Nonlinear effects of exchange-rate volatility on the volume of bilateral exports", *Journal of Applied Econometrics*, **19**, 1–23.

Belke A. Spies J. (2008) "Enlarging the EMU to the east: what effects on trade?" *Empirica*, **35**,369–389

Brakman S., Garretsen H. van Marrewijk C. (2009) "The New Introduction to Geographical Economics", *Cambridge University Press*.

Brulhart M. (2011) "The spatial effects of trade openness: a survey", *Review of the World Economy*, 147:59–83.

Brülhart, M., Crozet, M., & Koenig, P. (2004) "Enlargement and the EU periphery: The impact of changing market potential", *World Economy*, 27(6), 853–875.

Carrere, C. (2006) "Revisiting the Effects of Regional Trade Agreements on Trade Flows with Proper Specification of the Gravity Model", *European economic review*, 50, 2, 223.

Chami, R., Fullenkamp, C. and Jahjah, S. (2005) "Are immigrant remittance flows a source of capital for development?" IMF Staff Paper, 52, 1.

Chaney, T. (2008) "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review*, 98, 4, 1707-21.

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., Lafourcade, M. and Mayer, T. (2005) "The trade-creating effects of business and social networks: evidence from France", *Journal of International Economics*, 66: 1–29.

Cornwell, C., Schmidt P., and Wyhowski D.J. (1992) "Simultaneous Equations and Panel Data", *Journal of Econometrics* 51,151–81.

Cosar A.K. and Fajgelbaum P. D. (2013) “Internal Geography, International Trade, And Regional Specialization”, NBER Working Paper 19697.

Daumal M, Zignago S (2008) “Border effects of Brazilian states”, Working Papers 2008-11, CEPII Research Center

De Benedictis L., Gallegati M. and Tamberi M., (2008) “Semiparametric analysis of the specialization-income relationship”, *Applied Economics Letters*, 15, 301–306.

Deardorff AV (1998) “Determinants of bilateral trade: does gravity work in a neoclassic world?”, J. Frankel JA (ed) *The Regionalization of the World Economy*. University of Chicago Press, Chicago

Díaz-Lanchas J., Llano C., Zofío J. L., (2013) “Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy”, Working Paper Series. Departamento de Análisis Económico: Teoría Económica e Historia Económica. Universidad Autónoma de Madrid.

Djankov S, Freund C (2000) “Disintegration and trade flows: evidence from the Former Soviet Union”, Policy Research Working Paper Series 2378. *The World Bank*

Duranton G.; Morrow P. M.; Turner M. A. (2012) “Roads and Trade: Evidence from the US” CEPR Discussion Papers 9393, C.E.P.R. Discussion Papers.

Eaton J, Kortum S. (2002) “Technology, geography, and trade”, *Econometrica* 70: 1741–1799.

Egger P. (2002) “An econometric view on the estimation of gravity models and the calculation of trade potentials”, *World Economy*, 25,297–312

Egger P. (2004) “On the problem of endogenous unobserved effects of the gravity models”, *Journal of Economic, Integration* 19, 182–191.

Egger P. (2005) “Alternative Techniques for Estimation of Cross-Section Gravity Models”, *Review of International Economics*, 13(5)**,** 881–891.

Evans, C.L. (2003) “The Economic Significance of National Border Effects”, *American Economic Review*, 93, 4, 1291-1312.

Feenstra R. (2002) “Border effect and the gravity equation: consistent methods for estimation”, *Scottish Journal of Political Economy*, 49(5), 1021–1035

Feenstra R. (2004) “Advanced International Trade”, *Princeton University Press*.

Fujita, M., Krugman, P., Venables, A.J., (1999) “The Spatial Economy: Cities, Regions and International Trade”, The MIT Press, Cambridge, MA, London, England.

Fujita and Mori (1996) “The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect”, *Journal of Development Economics* 49, pp. 93-120.

Gallego N. and Llano C. (2012) “Towards a region-to-region international trade database for the EU: methodology of estimation and first results for the Spanish case”, Working paper presented in the NARSC 2011. Miami.

Garmendia, A., Llano-Verduras, C. and Requena-Silventre, F. (2012) “Network and the disappearance of the intranational home bias,” *Economic Letters* 116, 178-182.

Ghemawat, P., Llano-Verduras. C., Requena-Silventre, F., (2010) “Competitiveness and interregional as well as and international trade: The case of Catalonia”, *International Journal of Industrial Organization,* 28, 415–422

Gil-Pareja, S., Llorca-Vivero, R., Martínez Serrano J.A. (2006) “The Border Effect in Spain: The Basque Country Case”, *Regional Studies*, 40.4, 335–345.

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

Gould, W. (1993) “Linear splines and piecewise linear functions”, STATA Technical Bulletin. STB-15. September.

Hanson, G. H. (1996) “Localization Economies, Vertical Organization, and Trade.” *American Economic Review* 86, 5, 1266-1278..

Hanson, G. H. (2001) “US-Mexico integration and regional economies: Evidence from border-city pairs”, *Journal of Urban Economics*, 50, 2, 259–287.

Hausman J, Taylor W (1981) “Panel data and unobservable individual effects”, *Econometrica* 49(6), 1377–1398

Head, K. and Mayer, T. (2000) “Non-Europe: the causes and magnitudes of market fragmentation in the EU”, *WeltwirtschaftlichesArchiv*, 136, 2, 285-314.

Head, K. and Mayer, T. (2002) “Illusory Border Effects: Distance mismeasurement inflates estimates of home bias in trade”, CEPII Working Paper 2002-01.

Head, K. and Mayer, T. (2013) *Gravity Equations: Workhorse, Toolkit, and Cookbook* Handbook of International Economics Vol. 4, eds. Gopinath, Helpman, and Rogoff. Forthcoming.

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

Helble (2007). "Border Effect Estimates for France and Germany", *Review of World Economics*, Vol. 143 (3)

Helliwell JF (1998). "How much do national borders matter?", *Brookings Institution Press, Washington, DC.*

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

Helliwell, JF (1996). "Do national borders matter for Quebec's trade?", *Canadian Journal of Economics*, 29(3), 507–522.

Helliwell, JF (1997) "National Borders, Trade and Migration", *Pacific Economic Review* 3 (3): 507-522

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

Hillberry RH, Hummels D (2003) "Intranational home bias: some explanations", *Review of Economics and Statistics*, 85(4), 1089–1092.

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

Hillberry, RH (1998) "Regional Trade and "the Medicine Line": The External border effect in US Commodity Flow Data", *Journal of Borderland Studies* 13: 1-17.

Hillberry, RH (2002) "Aggregation bias compositional change, and the border effect", *Canadian Journal of Economics*, 35(3), 517–530.

Henderson D.J. and Millimet D.L (2008) "Is Gravity Linear?" *Journal of Applied Econometrics*, 23: 137–172.

Jackson, R., Walter R. Schwarm, Yasuhide Okuyama, Samia Islam (2006) "A method for constructing commodity by industry flow matrices", *Annals of Regional Science*, 40, 909–920.

Kei-Mu, Y. (2010) "Can Multistage Production Explain the Home Bias in Trade?" *American Economic Review* 2010, 100:1, 364–393.

Lafourcade, M. and Paluzie, E. (2011) "European Integration, Foreign Direct Investment (FDI), and the Geography of French Trade", *Regional Studies*, 45: 4, 419 — 439.

Llano-Verduras C.; Minondo A., Requena-Silvente F. (2011) "Is the Border Effect an Artefact of Geographical Aggregation?" *The World Economy*, 34, 10, 1771–1787.

Martín Rodríguez G., Cáceres Hernández J.J. (2010) “Splines and the proportion of the seasonal period as a season index” *Economic Modelling* 27, 83–88.

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

Millimet D, and Osang Th (2007) “Do state borders matter for US. intranational trade? The role of history and internal migration”, *Canadian Journal of Economics*, 40(1), 93–126.

Minondo A. (2007) “The disappearance of the border barrier in some European Union countries’ bilateral trade”, *Applied Economics*, 39, 119-124.

Mitze T. (2010) “Estimating Gravity Models of International Trade with Correlated Time-Fixed Regressors: To IV or not IV?”, MPRA Paper No. 23540. http://mpra.ub.uni-muenchen.de/23540/

Mukherjee D. and Pozo S. (2011) “Exchange-rate volatility and trade: a semiparametric approach”, *Applied Economics*, 43, 1617–1627.

Mundra K. (2005) “Immigration and International Trade: A Semiparametric Empirical Investigation”, *J. Int. Trade & Economic Development*, 1, 65 – 91.

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

Nitsch, V., (2002) “Border effects and border regions: lessons from the German unification” Banakgesellschaft, Berlin.

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

Okubo, T., (2004) “The border effect in the Japanese market: a gravity model analysis", *Japan International Economics*, 18:1–11

Pagan, A. & Ullah, A. (1999) “Nonparametric Econometrics” (Cambridge: Cambridge University Press).

Poncet, S. (2003) “Measuring Chinese domestic and international integration”, China Economic Review 14 (2003) 1 –21.

Poncet, S. (2005) “A Fragmented China: Measure And Determinants Of Chinese Domestic Market Disintegration”, *Review Of International Economics* 13.3 (2005): 409-430. Business Source Complete. Web. 16 Jan. 2014.

Robinson P.M. (1988) “Root-N-consistent semiparametric regression”, *Econometrica* 56: 931-954.

Ruiz I. Shukralla E.and Vargas-Silva C. (2009) “Remittances, Institutions and Growth: A Semiparametric Study”, *International Economic Journal*, Vol. 23, No. 1, 111–119, March 2009

Serlenga L., Shin Y. (2007) “Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors”, *Journal Applied Econometrics* 22: 361–381.

Shultze, M.S. and N. Wolf (2008) “On the origins of border effects: insights from the Habsburg Empire”, *Journal of Economic Geography*, 1-20.

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

Verardi, V. and Nicolas D. (2012) “Robinson’s square root of N consistent semiparametric regression estimator in Stata”, *Stata Journal*, 12,4, 726-735.

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

Wolf, H.C. (1997) “Patterns of Intra- and Inter-State Trade”, *NBER Working Paper 5939*. National Bureau of Economic Research, Cambridge, Mass.

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

Yatchew A. (1998) “Nonparametric regression techniques in economics”, *Journal of Economic Literature* 57: 135-143.