Estudios de Economía
ISSN: 0304-2758
ede@econ.uchile.cl
Universidad de Chile
Chile

Minondo, Asier; Requena, Francisco

Estimating the gravity equation with the actual number of exporting firms

Estudios de Economía, vol. 40, núm. 1, junio, 2013, pp. 5-19
Universidad de Chile
Santiago, Chile

Estimating the gravity equation with the actual number of exporting firms

Estimando la ecuación de gravedad con el número real de empresas exportadoras

Asier Minondo*
Francisco Requena**

Abstract

To estimate correctly the effect of variable trade costs on firms' exports, the gravity equation should control for the number of firms that participate in foreign markets. Due to the absence of these data, previous studies control for this omitted variable using econometric strategies that may also lead to inconsistent estimates. To overcome this problem the present paper estimates a gravity equation using a new database compiled by the OECD and EUROSTAT that reports the number of exporting firms by reporter and partner country. We show that not controlling for the extensive margin of trade introduces very serious biases in the estimated trade cost coefficients.

Key words: *Gravity equation, exporting firms, distance, trade costs, OECD.*

JEL Classification: *F14, F15.*

Resumen

Para estimar correctamente el efecto de los costes de comerciar sobre las exportaciones de las empresas, la ecuación de gravedad debe controlar por el número de empresas que opera en el mercado internacional. Debido a la ausencia de datos, estudios anteriores han controlado esta variable mediante técnicas econométricas que también pueden generar estimaciones sesgadas. Para superar estos problemas este trabajo estima una ecuación de gravedad utilizando una nueva base de datos de la OCDE y EUROSTAT, que incluye el número de empresas exportadoras en cada relación bilateral. Nuestros resultados muestran que no

* Deusto Business School, University of Deusto. Camino de Mundaiz, 50; 20012 San Sebastian - Spain. E-mail: aminondo@deusto.es)

** Universidad de Valencia. Avda. dels Tarongers. 46022 Valencia - Spain. E-mail: francisco.requena@uv.es)

controlar el margen extensivo genera sesgos muy importantes en la estimación de los costes de comercio.

Palabras clave: *Ecuación de gravedad, firmas exportadoras, distancia, costes de comercio, OCDE.*

Clasificación JEL: *F14, F15.*

1. Introduction

The gravity equation is one of the most successful empirical models in international economics. In its simplest form, the gravity equation predicts that bilateral trade flows are positively linked to the economic size of the trading partners and negatively linked with the distance between them. Due to its simplicity and empirical success, in addition to trade in goods, the gravity equation has been applied to understand the determinants of other bilateral flows such as trade in services, foreign direct investment, tourism or migration.

Linking trade barriers with trade flows, the gravity equation is also a very powerful tool to infer the costs that geography, communication or government practices impose on trade. This property is especially relevant given the inaccuracy and incompleteness of many direct costs measures (Anderson and van Wincoop, 2004). Improvements in the accuracy of trade costs' estimations have come hand in hand with advances in the theoretical foundations of the gravity equation (Anderson, 2011). A major step in this process is the model developed in Anderson and van Wincoop (2003). These authors show that bilateral trade flows do not depend on absolute trade costs but on relative trade costs. To capture this fact, their gravity model incorporates two additional variables, the exporter and importer multilateral resistances, which summarize each trading partner's average trade costs. Since bilateral trade costs are positively correlated with multilateral resistances, estimations of trade costs will be biased upwards if the multilateral resistance terms are not included in the gravity equation.

More recently Helpman *et al.* (2008) (hereinafter HMR) have incorporated the insights of the new-new trade models –the presence of fixed costs of exporting and heterogeneity in productivity across firms– into the gravity framework.[1] These two new elements help to explain the existence of zero trade flows between countries, in one or both directions, due to the lack of firms with enough productivity to export to countries with high entry costs. As HMR argue, if the gravity equation does not control for the number of exporting firms, the estimated coefficients can no longer be interpreted as elasticities of a firm's sales abroad with respect to trade costs. Instead, coefficients capture the impact of trade costs on both the number of firms that participate in trade (the extensive margin) and the trade value per firm (the intensive margin), leading to an upward bias in the estimations.

[1] Eaton and Kortum (2002) and Chaney (2008) also introduce heterogeneity into the gravity framework.

The main difficulty in estimating the HMR model is the lack of data on the number of exporting firms by country-pair, for a large sample of countries. To overcome this limitation, HMR estimate this number using an export participation model that, in a second stage, is introduced in the gravity equation. Their empirical analysis confirms that not controlling for the number of exporting firms leads to an upward bias in the estimated coefficients.

However, the strong distributional assumptions needed to estimate the two-stage empirical model in HMR cast some doubts on the consistency of the estimations. In particular, Santos-Silva and Tenreyro (2009) argue that in the two-stage procedure followed by HMR, heteroskedasticity in trade data makes difficult to disentangle the effects of trade costs on the number of firms that participate in trade (the extensive margin), and the value of trade per firm (the intensive value); in this situation it is difficult to get meaningful insights from the estimations.

As an alternative estimation strategy, a number of papers have used the number of exporting firms to disentangle the differences of trade costs on the extensive and intensive margins of trade (Eaton *et al.*, 2004; Bernard *et al.*, 2007; Mayer and Ottaviano, 2007; Crozet and Koenig, 2010; Lawless, 2010). The main limitation of these studies is that they only have data on the number of exporting firms for one country, and therefore they cannot control for the multilateral resistance terms in the gravity equation, leading, as explained above, to biases in the estimated coefficients related to trade costs.

To overcome these problems, for the first time in the literature, we estimate the gravity equation using a new database compiled by the OECD and Eurostat that reports the value of trade and the number of firms that participate in bilateral trade flows for 21 OECD reporting countries and 36 trading partners. As predicted by the HMR model, our estimates confirm that the traditional gravity equation leads to biased estimates if we do not control for the extensive margin of trade. In addition, the bias in the coefficients is much worse than that found by HMR. However, we should also take our conclusions with care. Our database only includes positive trade flows, which might lead to a sample selection problem.

The rest of the paper is organized as follows. Section 2 presents the theoretical model used in the paper. Section 3 derives the empirical equation and explains the database. Section 4 reports and comments the results of the econometric estimations. Section 5 draws the main conclusions of the paper.

2. Theory

This section explains the model that derives the econometric equation that is estimated in the next section. HMR consider a world consisting of J countries, indexed by $j = 1, 2, 3, \ldots, J$. Consumers preferences, which are the same in all countries, take the Constant Elasticity of Substitution or Dixit and Stiglitz (1977) form. In particular utility in country j is defined as:

$$(1) \qquad u_j \left[\int_{l \in B_j} x_j(l)^{\alpha} \, dl \right]^{1/\alpha}, 0 < \alpha < 1$$

where $x_j(l)$ denotes consumption of product l. B_j is the set of goods that can be consumed in country J. Parameter α determines the elasticity of substitution across products, which is defined as $\varepsilon = 1/(1-\alpha)$. The elasticity of substitution across products is the same in all countries.

The price index dual to (1) is:

$$P_j = \left[\int_{l \in B_j} \check{p}_j(l)^{1-\varepsilon} \, dl \right]^{1/(1-\varepsilon)} \tag{2}$$

where $\check{p}_j(l)$ is the price of product l in country j.

Assuming that the expenditure level in country j equals its income (Y_j), given the preferences in (1) the demand for product l in country j is:

$$x_j(l) = \frac{\check{p}_j(l)^{-\varepsilon} Y_j}{P_j^{1-\varepsilon}} \tag{3}$$

Regarding production, HMR assume that each firm produces only one good, which is different from the goods produced by other firms in the home country or in foreign countries. To produce the good a firm in country j uses a bundle of inputs, captured by the variable a, at a cost c. HMR introduce heterogeneity across firms making the bundle of inputs firm-specific. In particular, the model assumes that the productivity level, $1/a$, is different across firms. In contrast, the cost of the input bundle is country-specific (c_j). Firms are assumed to draw their productivity from a cumulative distribution function $G(a)$, with support $[a_L, a_H]$. This distribution function is assumed to be the same in all countries.

If the producer sells in the domestic market the cost of delivering its product is c_{ja}. If the firm aims to sell in a foreign market it will face two additional costs. On the one hand, in order to enter the foreign market, it will have to pay a fixed cost. On the other hand, there will be additional transport costs to deliver the product to the foreign market. Both fixed and variable costs are foreign market-specific. Transport costs are of the iceberg-type, where $\tau_{ij} > 1$ units should be shipped to country i from country j in order to one unit to arrive. It is assumed that transport costs and fixed costs are zero in the domestic market.

Due to the extra costs for operating in foreign markets, firms will export as long as profits derived from selling in the foreign market cover the additional costs. To determine the productivity level that establishes this cut-off condition, we first determine the price that a country j firm will charge in country i. As there is monopolistic competition in final products, the profit maximizing price for the country j firm selling product l in country i is:

$$\check{p}_i(l) = \tau_{ij} \check{p}_j(l) = \tau_{ij} \frac{c_j a}{\alpha} \tag{4}$$

where $c_j a / \alpha$ is the mill price, and $1/\alpha$ the standard mark-up.

With the price equation (4), the demand equation (3) and the cost parameters defined above, it is possible to determine the minimum productivity level ($1/a_{ij}$) that a firm in country j should command to sell in foreign country i. This minimum productivity level is determined by the zero profit condition in the foreign market:

(5) $$(1-\alpha)\left(\frac{\tau_{ij}c_j a_{ij}}{\alpha P_i}\right)^{1-\varepsilon} Y_i = c_j f_{ij}$$

where $c_j f_{ij}$ is the fixed cost of serving country i. All country j firms with a productivity level ($1/a$) equal or above the cut-off condition ($1/a_{ij}$) will export to country i. Hence, aggregate imports of country i from country j can be expressed as:

(6) $$M_{ij} = \int_{a_L}^{a_{ij}} (1/a)^{\varepsilon-1}\left(\frac{\tau_{ij}c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i N_j \, dG(a)$$

Define N_{ij}^x as a term indicating the number of country j firms that export to country i, weighted by an index of firm productivity:

(7) $$N_{ij}^x = \begin{cases} \int_{a_L}^{a_{ij}} (1/a)^{\varepsilon-1} N_j \, dG(a) & for \;\; 1/a_L \geq 1/a_{ij} \\ 0, & otherwise \end{cases}$$

Combining equation (6) and equation (7), imports of country i from country j can be written:

(8) $$M_{ij} = \left(\frac{\tau_{ij}c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i N_{ij}^x$$

According to equation (8) imports by country i from country j depend on trade costs, the costs of production in country j, the price index in country i, the demand capacity in country i and the number of firms in country j that export to country i.

3. Econometric Estimation

To get an estimating equation, first we follow HMR in modeling bilateral trade costs as:

(9) $$\tau_{ij}^{1-\varepsilon} \equiv D_{ij}^{\gamma} e^{-u_{ij}}$$

where D_{ij} is the distance between country i and country j, and u_{ij} denotes other unmeasured stochastic trade frictions.

Substituting equation (9) in equation (8), and taking logs we obtain the following estimating equation:

(10) $$m_{ij} = \lambda_i + \lambda_j - \gamma d_{ij} + \eta_{ij}^x + u_{ij}$$

where lowercase variables represent natural logarithms of their respective uppercase variables. $\lambda_i = (\varepsilon - 1)p_i + y_i$ is a fixed effect for the importing country i and $\lambda_j = -(\varepsilon - 1)\ln c_j$ is a fixed effect for the exporting country j. Notice that the omission of the term η_{ij}^x in the equation (10) would imply that the distance coefficient will not only capture the effect of trade costs on firm level trade, but also the effect of trade costs on the number of firms that participate in exports, leading to an upward bias in the estimated coefficient.[2]

The main difficulty in estimating equation (10) lies in the availability of data on the number of firms that participate in exports by country pair for a large sample of countries. As HMR do not have these data, they follow a two-stage procedure to approximate the number of exporting firms. In the first stage they estimate an export participation Probit equation that incorporates variables related with the fixed costs of exporting. They use the estimated coefficients to build a value on the number of firms that participate in exports per country-pair. This estimated value is then introduced in equation (10) to control for firm-heterogeneity. However, Santos-Silva and Tenreyro (2009) and Anderson (2011) show that the HMR two-stage estimation technique demands very strong distributional assumptions that are hardly met in international trade data. Other studies, such as Bernard *et al.* (2007) and Lawless (2010), estimate the gravity equation with the number of exporting firms, but with data for a single country. This limitation does not allow them to control for both sets of multilateral resistances terms (λ_i, λ_j), leading to biases in the estimated coefficients related to trade costs.

To overcome these problems, our paper uses a new database, the OECD-Eurostat Trade by Enterprise Characteristics Database (Araújo and Gonnard, 2011), that reports the number of firms that participate in bilateral exports. To perform the empirical analysis we select the year 2005. The database also offers data for the years 2006 and 2007. We will use these latter data to test the robustness of our results to changes in the period of analysis. For the year 2005, the database identifies 21 exporting countries and 36 importing countries.[3] This dataset yields a maximum of 735 observations ((21 exporters*36 importers)-21 importers); however, due to confidentially, the final sample contains 716 observations.

Table 1 presents some descriptive statistics on the number of exporting firms per bilateral exports in our sample. Among the 21 exporters included in our sample, Italy is the country with the highest average number of exporting firms in bilateral trade relationships: 15,532; it is followed by Germany (14,889), United

2 Strictly speaking η_{ij}^x is a term indicating the number of country j firms that export to country i, weighted by an index of firm productivity.

3 The exporting country set is listed in Table 1. The importing country set is composed by the former countries plus Belgium, Bulgaria, China, Greece, India, Ireland, Japan, Mexico, Malta, Netherlands, Russia, Spain, Switzerland, Turkey and United Kingdom.

TABLE 1
DESCRIPTIVE STATISTICS ON THE NUMBER OF EXPORTING FIRMS PER BILATERAL EXPORT RELATIONSHIP, 2005

Exporter	Average	Standard deviation	Maximum Nº of firms	Maximum partner	Minimum Nº of firms	Minimum partner
Austria	2,380	1,977	8,527	Germany	337	Malta
Canada	2,042	5,816	34,949	United States	100	Malta
Cyprus	48	53	264	Greece	6	Slovakia
Czech Republic	2,023	2,231	9,729	Slovakia	150	Malta
Denmark	1,705	1,576	9,050	Norway	295	Malta
Estonia	302	383	1,519	Finland	11	Luxembourg
Finland	1,046	830	3,708	Russia	85	Malta
France	9,079	7,452	33,216	Switzerland	1,577	Malta
Germany	14,889	7,339	28,053	Austria	3,309	Malta
Hungary	884	965	4,508	United States	54	Malta
Italy	15,532	9,664	45,730	Switzerland	3,944	Estonia
Latvia	230	278	1,021	Russia	9	Luxembourg
Lithuania	522	736	3,253	Russia	20	Malta
Luxembourg	229	317	1,278	Belgium	19	Malta
Norway	1,206	1,615	7,492	Sweden	93	Luxembourg
Poland	2,222	1,873	9,022	Germany	142	Malta
Portugal	1,004	1,083	4,995	Spain	107	Latvia
Slovakia	547	664	3,137	Czech Republic	26	Malta
Slovenia	355	318	1,281	Germany	53	Mexico
Sweden	2,589	3,620	21,511	Norway	238	Malta
United States	12,071	16,545	83,727	Canada	655	Luxembourg

Source: OECD.

States (12071) and France (9079). The lowest average number of exporting firms, 42, is found in the country with the lowest GDP: Cyprus. We observe, as well, that there is a large variation in the number of exporting firms per partner. For example, in the case of Canada the standard deviation is three times larger than the average. The bilateral export relations with the highest number of exporting firms are those with neighboring countries. However, it is striking that for France and Italy the highest number of exporting firms are found in their relationship with Switzerland, rather than with other larger neighboring countries, such as Germany for France (20569 exporting firms vs 33216 with Switzerland), and France for Italy (31572 vs. 45730 with Switzerland). In contrast, in the majority of cases, the lowest number of exporting firms is found in the relationship with small partners, such as Malta or Luxembourg.

We introduce the (log of) bilateral distance to control for bilateral trade costs in equation (10) together with a number of dummies: common language, land border, common membership in a regional trade agreement (RTA), and common currency (in our data the Euro). Finally, we introduce exporter and importer specific fixed effects. We should point that all trade relationships in our database are positive. This is a shortcoming of our database, as it might lead to a sample selection problem. Hence, we should be conservative with the relevance of the results. Notwithstanding this problem, Helpman *et al.* (2008) point out that the bias due to sample selection is much lower than the bias due to the absence of the number of exporting firms.

4. Econometric Results

To analyze how the introduction of the actual number of exporting firms improves on previous estimations, we present the econometric results in five steps. First, we estimate a gravity equation with a fixed-effects OLS model, without controlling for the number of firms that participate in exports. Table 2, Column 1 presents the results when estimating the gravity equation with a standard OLS model, not controlling for firm heterogeneity but introducing exporter and importer fixed-effects. We observe that distance has a very strong negative effect on trade. In contrast, to speak the same language, to share a land border, and to belong to the same regional integration agreement do not have a statistically significant effect on trade. On its hand, common currency has a negative coefficient, although it is neither statistically significant.

Second, to control for heteroskedasticity, following Santos-Silva and Tenreyro (2006), we estimate the econometric equation with a fixed-effects Poisson model without controlling for the number of firms that participate in exports. The results in Table 2, Column 2, show that distance, compared to the OLS coefficient, has a lower negative effect on trade. In addition to that, the language coefficient becomes statistically significant. On its hand, coefficients on adjacency, regional trade integration and common currency remain statistically not significant.

Third, we estimate the gravity equation with the HMR two-stage procedure. Table 2, Column 3 presents the results. We have to stress that the output should be taken with a lot of care. As all bilateral export relationships are positive in our sample, we cannot estimate the first stage export participation equation of the HMR procedure. To overcome this problem, we estimate the first stage

TABLE 2
RESULTS OF THE ECONOMETRIC ESTIMATIONS

	(1) OLS	(2) Poisson	(3) HMR	(4) OLS with Nº of exporting firms	(5) Poisson with Nº of exporting firms	(6) 2SLS
Distance (log)	–1.66***	–1.23***	–1.60***	–0.56***	–0.57***	–0.29*
	(0.10)	(0.11)	(0.09)	(0.11)	(0.12)	(0.15)
Adjacency	0.19	0.11	0.25**	–0.01	0.05	–0.07
	(0.12)	(0.09)	(0.12)	(0.07)	(0.07)	(0.33)
Language	0.16	0.37**	0.12	0.01	0.03	–0.03
	(0.16)	(0.13)	(0.16)	(0.12)	(0.12)	(0.30)
RTA	0.22	0.24	0.26	0.41*	0.45***	0.34
	(0.27)	(0.17)	(0.26)	(0.17)	(0.14)	(1.07)
Common currency	–0.07	0.01	–0.08	–0.10	–0.02	–0.08
	(0.11)	(0.12)	(0.10)	(0.08)	(0.10)	(0.09)
Estimated Nº of exporting firms			4.47			
			(5.77)			
Actual Nº of exporting firms				0.97***	0.79***	1.28***
				(0.01)	(0.07)	(0.03)
Observations	716	716	716	716	716	596
Adjusted R-square	0.91	0.96		0.94	0.98	0.94

Note: RTA stands for Regional Trade Agreement. All specifications include exporting and importing country fixed effects. Country-pair clustered standard errors in parentheses. *** denotes significance at the 1-percent level, ** significance at the 5-percent level and * significance at the 10-percent level.

using the UN-COMTRADE database, which reports bilateral exports data for more than 200 countries, and includes a sizable set of zero-trade observations. Following the HMR procedure, we use the results of the Probit model to obtain a consistent estimator of the number of firms that participate in each bilateral export relationship.[4] In a second stage, we introduce the estimated number of exporting firms, and estimate the second stage of the model using our original sample that includes 21 exporters and 36 importers.

We can see that HMR estimates are very similar to those obtained with OLS (Column 1). The only exception is that in the HMR estimation adjacency is positive and statistically significant. We observe that the coefficient for the estimated number of firms is much larger than the one reported in HMR (4.47 vs. 0.87). To explain this difference, we should bear in mind that the estimated number of exporting firms is built on the probability of the existence of a bilateral trade relationship predicted by the first stage. As in our sample all bilateral export relationships are positive, the probabilities predicted by the first stage are very high for those observations and, thus, the differences in the estimated number of firms across observations are tiny.[5] Hence, to explain the large differences in bilateral exports across the countries included in our sample, in the second stage the econometric model attributes a large effect to those small differences in the estimated number of exporting firms.

Fourth, we estimate the gravity equation with a fixed-effects OLS model which controls for the number of firms. In Column (4) we observe a remarkable reduction in the negative effect of distance on trade with respect to the HMR estimation. We also find a positive and statistically significant effect of regional trade agreements on trade. The coefficients for shared language, adjacency and common currency are statistically not significant. As predicted by the HMR model, the actual number of exporting firms' coefficient has an elasticity with respect to imports close to one.[6] Finally, we estimate the gravity equation with a fixed-effects Poisson model which also controls for the number of exporting firms. The Poisson results in Column (5) are very similar to the OLS ones in Column (4).

Finally, we should address the endogeneity of the number of exporting firms' variable. As the amount of exports and the number of exporting firms are determined simultaneously in equilibrium, the estimations might lead to inconsistent and biased estimates. To solve this problem, we should instrument the number of exporting firms with a variable that is determined outside the model and is highly correlated with the number of exporting firms. In the theoretical section of this paper, we showed that export fixed costs determine the number of exporters; moreover, export fixed costs is a variable exogenous to the model. Export fixed costs combine the costs that the exporter has to face in his country (e.g. the draft of a contract for a foreign delegate) and in the foreign country (e.g.

4 As in HMR, the percentage of population sharing the same religion is the variable that is excluded in the second stage.

5 When we compare the number of exporting firms per bilateral relationship estimated by the HMR model and the actual number of exporting firms the correlation is only 0.26.

6 We also estimate an equation moving the number of exporting firms to the left-hand side of equation (10). Results are not altered.

the legal costs of opening a delegation). HMR argue that the fixed costs faced by the exporter in his home country and the fixed costs faced by the exporter in the destination country can be proxied by the regulation costs of firm entry in each country. The higher the regulation costs for a firm entry in the origin and in the destination country, the higher the fixed costs of exporting. The regulation costs of firm entry are measured by the number of days and the number of legal procedures for an entrepreneur to legally start operating a business.[7] As in HMR, we calculate bilateral fixed costs of exporting as the sum of the number of days and legal procedures in the exporting and importing country above the median.

We confirm empirically that this variable is a suitable instrument for the number of firms. First, there is a negative correlation between the instrument and the number of exporting firms (–0.35), confirming that higher fixed costs in the origin and destination country reduce the number of exporting firms. Second, when we regress the number of exporting firms on the instrument and the rest of explanatory variables included in the regression equation, the coefficient for the instrumental variable is also negative and statistically significant at 1%. Column (6) presents the results of the two-stage least regression. We observe that there is a further reduction in the distance coefficient and the regional trade agreement variable, although positive, becomes statistically not-significant.[8]

Our results confirm that traditional gravity estimates are not always reliable to infer the elasticity of trade per firm with respect to trade costs. According to our estimates, the traditional gravity coefficient overestimates the negative impact of distance on trade because the coefficient captures the large deterring effect of distance on the number of firms that participate in exports. It is important to point out that the size of the bias in our estimations, which uses actual data on the number of exporting firms, is much larger than the bias reported in HMR. In particular, their distance coefficient falls by 35% when controlling for firm heterogeneity;[9] in our sample it falls as much as 83%. On its hand, the results presented in Columns (4), (5) and (6) are in line with studies that have analyzed the impact of trade costs on the intensive and extensive margins of trade using single country data (Bernard *et al.* 2007; Lawless, 2010).

Robustness analysis

To analyze the robustness of our results we perform a series of sensitivity analyses. First, in the HMR model, the number of exporting firms is weighted by an index of firm productivity: the larger the productivity the larger the exports. However, when estimating equation (1), we have use a number of exporting firms which is not weighted by productivity. To gauge the sensitivity of results to the absence of weighting, in this robustness test we introduce a proxy for firms' productivity. To build this proxy we use data on the number of exporting

---

7 The data are obtained from the World Bank's Doing Business database. Available at http://www.doingbusiness.org/

8 There is a drop in the number observations because there are not data on the number of days and the number of legal procedures to legally start a new firm for all countries included in the sample.

9 See Table IV in page 471 in Helpman *et al.* (2008).

firms, and the value of exports per exporting country and size class provided by the OECD-Eurostat database. The database distinguishes four different size classes: firms with 0-9 employees, firms with 10-49 employees, firms with 50-249 employees, and firms with more than 249 employees. We calculate the average exports per firm in each size class, and we divide this figure by the sample mean. We use this ratio as a proxy for productivity. Multiplying the number of exporting firms by the productivity index in each size class, and adding all size classes, we get a productivity-weighted number of exporting firms per each exporting country.

Table 3 reports the results of estimating equation (10) with the number of exporting firms weighted by productivity.[10] We can see that the estimation of the OLS model lead to almost identical results to those obtained with the non-weighted number of firms. In the Poisson model there is a slight increase in the coefficient for distance, and a slight reduction in the coefficient for RTA; moreover, we observe that adjacency becomes statistically significant. Notwithstanding these changes, the results point out that the main conclusions of our analysis are robust to using a productivity-weighted measure of the number of exporting firms.

TABLE 3
ESTIMATIONS WITH THE PRODUCTIVITY-WEIGHTED AVERAGE OF EXPORTING FIRMS

	(1) OLS		(2) Poisson	
Distance (log)	–0.55***	(0.11)	–0.63***	(0.11)
Adjacency	–0.01	(0.08)	0.15*	(1.76)
Language	0.02	(0.11)	–0.08	(0.13)
RTA	0.40*	(0.17)	0.38***	(0.11)
Common currency	–0.11	(0.09)	–0.07	(0.10)
Actual number of exporting firms	0.98***	(0.08)	0.76***	(0.07)
Observations		657		657
Adjusted R-square		0.94		0.98

Note: RTA stands for Regional Trade Agreement. All specifications include exporting and importing country fixed effects. Country-pair clustered standard errors in parentheses. *** denotes significance at the 1-percent level and * significance at the 10-percent level.

In addition to the total figure, the OECD-Eurostat database provides data on the number of exporting firms per country-pair trade relationship for three main economic activities: industry, trade and repair and other sectors. As a second robustness analysis, we study whether results are altered when estimating the gravity equation distinguishing by main economic activity.[11] As shown in Table 4,

10 As Germany and Norway do not provide data on the number of exporting firms and value of exports per size class for all partners, the sample is reduced to 657 observations.

11 As some countries do not report the number of exporting firms per bilateral export relationship for some industries, the number of observations in this sample is lower.

TABLE 4
ECONOMETRIC ESTIMATIONS DISTINGUISH FIRMS' MAIN ECONOMIC ACTIVITY

	Industry				Trade and Repair				Other industries			
	OLS	Poisson	OLS with Nº firms	Poisson with Nº firms	OLS	Poisson	OLS with Nº firms	Poisson with Nº firms	OLS	Poisson	OLS with Nº firms	Poisson with Nº firms
Distance (log)	–1.69***	–1.15***	–0.46***	–0.46***	–2.13***	–1.68***	–0.73***	–0.73***	–2.13***	–0.98***	–1.10***	–0.47*
	(0.10)	(0.10)	(0.09)	(0.11)	(0.12)	(0.14)	(0.17)	(0.16)	(0.14)	(0.19)	(0.17)	(0.20)
Adjacency	0.01	0.12	–0.05	0.03	0.28	0.12	0.02	0.06	0.27	0.22	0.02	0.20
	(0.13)	(0.09)	(0.08)	(0.07)	(0.16)	(0.11)	(0.11)	(0.09)	(0.17)	(0.15)	(0.13)	(0.13)
Language	0.35*	0.34*	0.12	–0.01	0.02	0.22	–0.17	–0.12	0.37*	0.50***	0.18	0.08
	(0.17)	(0.13)	(0.14)	(0.15)	(0.21)	(0.15)	(0.16)	(0.15)	(0.19)	(0.15)	(0.17)	(0.13)
RTA	0.17	0.36*	0.21	0.36**	–0.01	–0.52**	0.46*	0.12	–0.62	0.55*	0.00	0.94***
	(0.27)	(0.17)	(0.18)	(0.13)	(0.35)	(0.23)	(0.23)	(0.21)	(0.36)	(0.27)	(0.27)	(0.26)
Common currency	–0.16	–0.08	–0.14	–0.05	0.13	0.36*	0.01	0.24	–0.16	0.00	–0.16	0.05
	(0.11)	(0.12)	(0.08)	(0.10)	(0.15)	(0.16)	(0.11)	(0.13)	(0.17)	(0.17)	(0.15)	(0.16)
Nº of exporting firms			1.18***	1.01***			1.03***	0.88***			0.82***	0.59***
			(0.08)	(0.08)			(0.10)	(0.07)			(0.10)	(0.08)
Observations	653	653	653	653	653	653	653	653	653	653	653	653
Adjusted R-square	0.91	0.96	0.94	0.98	0.85	0.94	0.89	0.96	0.85	0.97	0.88	0.98

Note: All specifications include year specific exporting and importing country fixed effects. Country-pair clustered standard errors in parentheses. *** denotes significance at the 1-percent level, ** significance at the 5-percent level, and * significance at the 10-percent level.

in all sectors, there is a severe drop in the distance coefficient when introducing the actual number of exporting firms in the estimation. The language coefficient, that was positive and statistically significant in the traditional estimation for industry and other sectors, becomes statistically not significant when controlling for the number of exporting firms. This result points out that language has a negative effect on the extensive rather than the intensive margin of trade. In contrast, the RTA coefficient tends to increase its value, and to become statistically significant when controlling for the number of exporting firms. This result points out that RTAs facilitate the intensive rather than the extensive margin of trade. Common currency and adjacency are statistically not significant in all estimations. Finally, to analyze the robustness of the results to changes in the time-period, we estimate the model for the year 2006 and for the year 2007.[12] The main conclusions are not altered.

5. Conclusions

To estimate correctly the effect of variable trade costs on firms' exports, the gravity equation should control for the number of firms that participate in foreign markets. Due to the absence of data, previous studies estimate this number using an export participation model and, then, introduce it in a gravity equation. However, due to the strong distributional assumptions demanded by this procedure, some authors cast doubts on the validity of these estimates. To overcome these problems, this paper estimates a gravity equation using the actual number of exporting firms for a sample of 21 exporters and 36 importers.

Our results show that some traditional gravity estimates are severely biased if they do not control for firm heterogeneity. We also find that the size of the bias is much larger than reported in previous multi-country studies. In particular, we find that the distance coefficient falls as much as 83% when introducing the number of exporting firms in the gravity equation. We also find that the traditional gravity estimates overestimate the impact of language on the intensive margin of trade.

Acknowledgements

We thank an anonymous referee for very helpful comments. We acknowledge financial support from Spanish Ministry of Science and Innovation (ECO2010-21643/ECON; project ECO 2011-27619/ECON, cofinanced with FEDER), the Basque Government Department of Education, Language policy and Culture and Generalitat Valenciana (project Prometeo/2009/098).

12 There are some changes in the sample with respect to the year 2005. In particular, in 2006 and 2007, the US does not report data on the number of exporting firms. In addition to that, in 2007 Poland does not report either the number of exporting firms.

References

Anderson, J.E. (2011). The Gravity Model. *Annual Review of Economics,* 3, 133-160.

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

Anderson, J.E.; Van Wincoop, E. (2004). Trade Costs. *Journal of Economic Literature*, 42 (September), 691-751.

Araújo, S.; Gonnard, E. (2011). Selling to Foreign Markets: a Portrait of OECD Exporters. *OECD Statisticals Brief*, Nº 16.

Bernard, A.B.; Jensen, J.B.; Redding, S.J.; Schott, P.K. (2007). Firms in International Trade. *Journal of Economic Perspectives*, 21 (3), 105-130.

Chaney, T. (2008). Distorted Gravity: the Intensive and Extensive Margins of International Trade. *American Economic Review*, 98 (4), 1707-1721.

Disdier, A.; Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics*, 90 (1), 37-48.

Eaton, J.; Kortum, S.S. (2002). Technology, Geography, and Trade. Econometrica, 70 (5), 1741-79.

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

Lawless, M. (2010). Deconstructing gravity: trade costs and the extensive and intensive margins of trade. *Canadian Journal of Economics*, 43 (4), 1149-1172.

Rose, A.K.; Spiegel, M.M. (2011). The Olympic Effect. *Economic Journal*, 121 (June), 652-677.

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

Santos-Silva, J.; Tenreyro, S. (2009). Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubisntein Model Empirically. *Centre for Economic Performance Discussion Paper Nº 935*.