Estimating the gravity equation with the actual number of exporting firms

by

Asier Minondo
and
Francisco Requena-Silvente

September 2011

Working Papers in Applied Economics

WPAE-2011-21

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

ISSN 2172-3036

Estimating the gravity equation with the actual number of exporting firms

Asier Minondo (Deusto Business School)

Francisco Requena (Universidad de Valencia)

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 estimations. 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 no controlling for the extensive margin of trade introduces very serious biases in the estimated trade cost coefficients. Moreover, these biases are much larger than predicted by previous studies.

JEL Codes: F14, F15.

Keywords: gravity equation, exporting firms, distance, trade costs.

Acknowledgements: We acknowledge financial support from Spanish Ministry of Science and Innovation (ECO2010-21643/ECON; project ECO 2011-27619/ECON) and Generalitat Valenciana (project Prometeo/2009/098).

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 link to the economic size of the trading partners and negatively link 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 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. 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.

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.

The rest of the paper is organized as follows. Section 2 presents the theoretical model used in the paper. Section 3 derives the estimating 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

HMR consider a world consisting of J countries, indexed by $j=1, 2, 3,\ldots\ldots, J$. 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:

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

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:

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

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

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

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 $c_j a$. 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:

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

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:

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

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:

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

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:

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

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

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

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:

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

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:

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

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*.

As explained in HMR, equation (10) introduces one substantial improvement with respect to previous structural gravity equations, such as Anderson and van Wincoop (2003). In particular, equation (10) includes the term n_{ij}^{x} that controls for the number of firms that export from country *j* to country *i*. If this term is not included in the equation, 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.

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 they do not have these data, HMR 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.

As explained by Santos-Silva and Tenreyro (2009) and Anderson (2011), the HMR two-stage estimation technique demands very strong distributional assumptions that are hardly met in international trade data. If these distributional assumptions are relaxed, it is no longer possible to identify separately the effect of trade costs on the extensive and the intensive margins of trade. 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 the multilateral resistances terms, leading, as explained above, 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.[1] This dataset yields a maximum of 735 observations ((21 exporters*36 importers)-21 importers). However, due to confidentially, the OECD database does not provide data for 19 bilateral exports, leading to a final sample of 716 observations.

[1] The exporting country set is composed by Austria, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Latvia, Lithuania, Luxembourg, Norway, Poland, Portugal, Slovakia, Slovenia, Sweden and the United States. 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 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: 15532; it is followed by Germany (14889), United 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. In fact, as shown in Figure 1, there is a tight relationship between the value of bilateral exports and the number of exporting firms that take part in the export relationship.

<INSERT TABLE 1 HERE>

<INSERT FIGURE 1 HERE>

To estimate the econometric model, in addition to distance, we introduce some dummy variables to control for other bilateral trade costs in equation (10). First, we introduce a dummy variable that takes the value of 1 if the importing and the exporting country have a common official language, and zero otherwise. Second, we introduce a dummy variable that takes the value of 1 if the importing and exporting country share a border, and zero otherwise. Third, we introduce a dummy variable that takes the value of 1 if the importing and exporting country participate in the same regional trade agreement, and zero otherwise. Forth, we introduce a dummy variable that takes the value of 1 if the importing and exporting country use the same currency, and zero otherwise. In our sample, the euro is the only common currency. Finally, we introduce exporter and importer specific fixed effects.

We should point out that estimates in equation (10) could be biased if we do not control for zero trade observations. As HMR explain, in the absence of zero trade observations there is a correlation between the error term and the unobserved trade frictions. In our sample all trade flows are positive, precluding this bias. In any case, as shown by HMR, the bias introduced by selection is small compared to the bias introduced by the absence of control for firm heterogeneity.

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. Second, to control for heteroskedasticity, we estimate the econometric equation with a fixed-effects Poisson model. This estimation does not incorporate either the number of exporting firms. Third, we estimate the gravity equation with the HMR two-stage procedure. Fourth, we estimate the gravity equation with a fixed-effects OLS model which controls for the number of firms. Finally, we estimate the gravity equation with a fixed-effects Poisson model which also controls for the number of exporting firms.

<INSERT TABLE 2 HERE>

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. In Column 2, to control for heteroskedasticity, following Santos-Silva and Tenreyro (2006), we present the results of estimating a Poisson model with fixed effects for exporters and importers. We observe 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.

It is interesting to compare our results with recent studies that have estimated the gravity equation (Baier and Bergstrand, 2009; Head et al., 2010; Rose and Spiegel, 2011).[2] It is important to point out that these studies use a wider sample of countries than in our study, and two of them use a much longer time-frame. We find that our distance coefficient estimated by OLS has a larger negative effect on trade than the coefficients estimated by these studies: -1.35 in Baier and Bergstrand, -0.91 in Head et al., and -1.33 in Rose and Spiegel; with respect to the Poisson estimation, our coefficient has also a larger negative effect than the one reported by Head et al. (-0.64). Sharing a common border, speaking the same language and having a common currency, which are statistically not significant in the our OLS estimation, are positive and statistically significant in these studies. In the Poisson estimation, as in our study, Head et al (2010) also report a positive and statistically significant coefficient for language; however, in contrast to our study, they also report positive and statistically significant coefficients for adjacency and common border. Finally, regional trade agreement has a positive and statistically significant coefficient in Head et al. (2010), and Rose and Spiegel (2011); however, as in our study, it is not statistically significant in Baier and Bergstrand (2009). These results point out that the countries included in our sample lead to estimations that, in some cases, are not in line with studies that use a wider set of countries.

To analyze whether the use of the actual number of exporting firms improves on the HMR procedure, Column 3 presents the results of estimating the model with the two-stage HMR methodology. We have to stress that the results of this estimation 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 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.[3] 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. As shown in the bottom of Column 3, the number of observations in the HMR estimation, 487, is

[2] For comparison we take Table 3-Column (9) in Baier and Bergstrand (2009), Table 2-Columns (1) and (2) in Head et al. (2010) and Table 1-Column (3) in Rose and Spiegel (2011).

[3] As in HMR, the percentage of population sharing the same religion is the variable that is excluded in the second stage. These authors also use entry costs as the excluded variable. However, due to the lower number of observations for this latter variable, the former was selected as the excluded variable.

considerably lower than in the previous OLS and Poisson estimations: 716.[4] Due to the large drop in the number of variables, to compare properly the results obtained from the HMR model with those obtain previously with OLS and Poisson, in Columns 4 and 5 we re-estimate the OLS and Poisson models with the reduced sample. We observe that with the reduced sample the OLS and Poisson estimates for adjacency and regional trade agreement (RTA) become positive and statistically significant. In contrast, the language coefficient, that was statistically significant in the Poisson model, becomes statistically not significant with the reduced sample. If we compare Column 4 and 5 with Column 3, we observe than the HMR coefficient for distance is slightly lower than the OLS coefficient, but larger than the Poisson coefficient. In contrast, the adjacency coefficient in HMR is slightly larger than in OLS and Poisson; on its hand, the RTA coefficient in HMR is slightly larger than in OLS, but lower than in Poisson. The coefficients for language and common currency remain statistically not significant. Finally, we observe that the coefficient for the estimated number of firms, 3.21, is much larger than the one reported in Helpman et al, (2008): 0.87 (Table IV). 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.

Columns (6) and (7) present the results of estimating the model with the actual number of exporting firms. We observe a remarkable reduction in the negative effect of distance on trade. Compared to the standard OLS gravity estimation (Column 1), the distance coefficient falls by two-thirds. Notwithstanding the limitations of the HMR procedure, this reduction is much larger than that occurring between Column 3 and Column 4. The reduction in the distance coefficient for the Poisson model is also very high: 54%. We can see that the

[4] The reduction in the number of observations is due to the first stage Probit estimation. In this stage, the model excludes those variables that predict the existence of a bilateral export flow perfectly. In the UN-COMTRADE sample, once we remove the observations that do not report data on some independent variables that are included in the first stage, there are six exporter country dummies (Canada, Denmark, France, Germany, Italy and United States) and two importing country dummies (Germany and United Kingdom) that predict the existence of exports perfectly.

[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.

regional trade coefficient multiplies almost by two and becomes statistically significant both in OLS and Poisson models. The coefficients for shared language, adjacency and common currency remain statistically not significant. As predicted by the HMR model, the actual number of exporting firms' coefficient has elasticity with respect to imports close to one, especially in the OLS model.[6]

Our results confirm that traditional gravity estimates are not always reliable to infer the elasticity of trade per firm with respect to trade costs. Not controlling for the number of exporting firms, traditional estimates confound the effect of trade costs on the extensive and the intensive margins of trade. 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. In contrast, the traditional gravity equation seems to underestimate the positive effect that regional trade agreements have on the intensive margin of trade. 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 Helpman et al. (2008). In particular, their distance coefficient falls by 35% when controlling for firm heterogeneity;[7] in our sample it falls between 54% and 66%. On its hand, the results presented in Columns (6) and (7) 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 (Eaton et al., 2004; Bernard et al. 2007; Mayer and Ottaviano, 2007; Crozet and Koenig, 2010; Lawless, 2010). These studies also find that the distance coefficient drops severely when estimating the intensive margin of exports.[8] With respect to language, Mayer and Ottaviano (2007) and Lawless (2010) find that the coefficient is no longer statistically significant when estimating the intensive margin of trade; in contrast, Crozet and Koenig (2010) find that to speak French has a positive and statistically significant coefficient. In line with our estimations, these authors find that being adjacent does not have a statistically significant effect on the intensive margin of trade; Mayer and Ottaviano (2007) also find that neither regional trade agreements nor currency unions have a significant effect on the value of exports per firm.

---

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

[7] See Table IV in page 471 in Helpman et al (2008).

[8] Hillberry and Hummels (2008) also obtain a similar result for US intra-national shipments.

Robustness analysis

To analyze the robustness of our results we perform a series of sensitivity analyses. First, in the HMR model, as shown in equation (7), 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 (10), 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 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.

<INSERT TABLE 3 HERE>

Table 3 reports the results of estimating equation (10) with the number of exporting firms weighted by productivity.[9] 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.

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.[10] As shown in Table 4, in all sectors, there is a severe drop in the distance coefficient when introducing the actual number of exporting firms in the estimation. The

[9] 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.
[10] 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.

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 regional trade agreements facilitate the intensive rather than the extensive margin of trade. Common currency and adjacency are statistically not significant in all estimations.

<INSERT TABLE 4 HERE>

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.[11] As shown in Table 5, the main conclusions are not altered. First, we find a severe reduction in the coefficient for distance both in year 2006 and in year 2007; moreover, in the year 2006 distance becomes statistically not significant in the OLS estimation with the number of firms. Second, the language coefficient, which is positive and statistically significant in the Poisson estimation, becomes statistically not significant when introducing the number of firms. Finally, in most of cases, the RTA coefficient becomes larger and statistically significant when introducing the number of exporting firms.

<INSERT TABLE 5 HERE>

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.

[11] 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.

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 between 54% and 66% when introducing the number of exporting firms in the gravity equation. We also find that the traditional gravity estimates overestimate the impact of language, and underestimate the impact of regional trade agreements on the intensive margin of trade.

References

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

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

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

Baier, S.L. and Bergstrand, J.H. (2009). "Estimating the effects of free trade agreements on international trade flows using matching econometrics", *Journal of International Economics*, **77**, 1, 63-76.

Bernard, A.B., Jensen, J.B., Redding, S.J. and 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.

Crozet, M. and Koenig, P. (2010). "Structural gravity equation with extensive and intensive margins of trade", *Canadian Journal of Economics*, **93**, 1, 41-62.

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

Dixit, A.K. and Stiglitz, J. (1977). "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, **67**, 3, 297-308.

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

Eaton, J., Kortum, S.S. Kramarz, F. (2004) "Dissecting Trade: Firms, Industries, and Export Destinations", *American Economic Review, Papers and Proceedings*, **93**, 2, 150-154

Head, K., Mayer, T. and Ries, J. (2010). "The erosion of colonial trade linkages after independence ", *Journal of International Economics*, **81**, 1, 1-14.

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

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

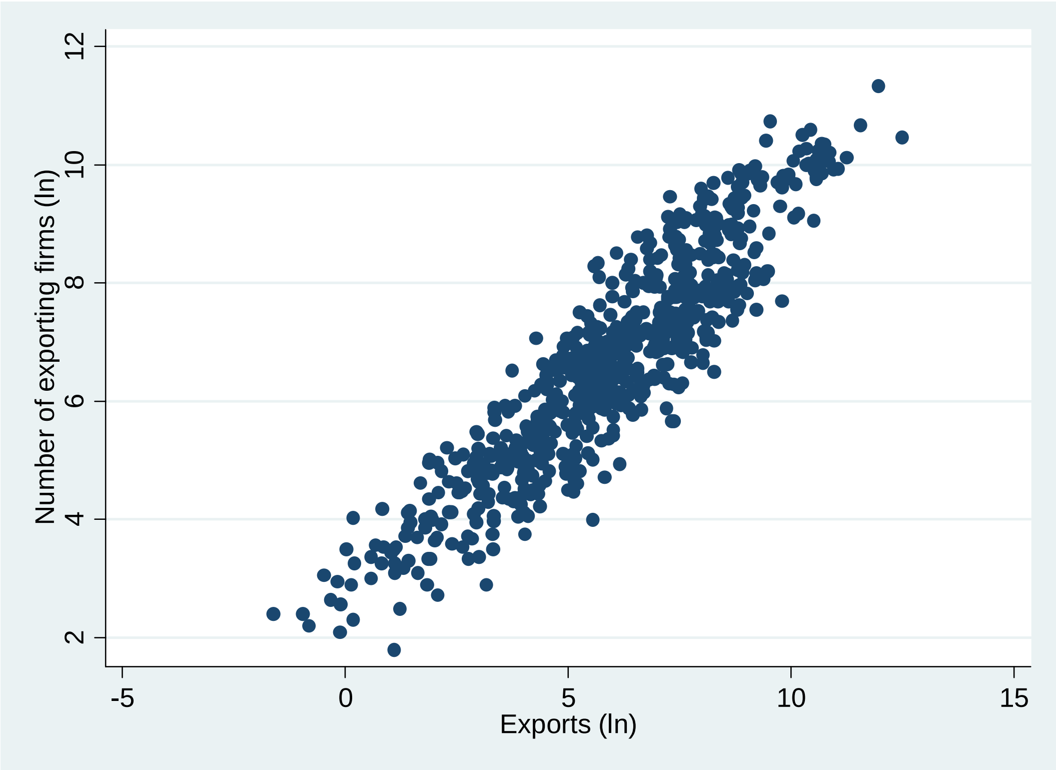
Mayer, T. and Ottaviano, G.I.P. (2007). "The Happy Few: The internationalisation of European Firms. New facts based on firm-level evidence", *Bruegel Blueprint Series Volume III*, Bruegel, Brussels.

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

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

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

Figure 1. Relationship between the number of exporting firms and the value of exports



Source: OECD.

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	2380	1977	8527	Germany	337	Malta
Canada	2042	5816	34949	United States	100	Malta
Cyprus	48	53	264	Greece	6	Slovakia
Czech Republic	2023	2231	9729	Slovakia	150	Malta
Denmark	1705	1576	9050	Norway	295	Malta
Estonia	302	383	1519	Finland	11	Luxembourg
Finland	1046	830	3708	Russia	85	Malta
France	9079	7452	33216	Switzerland	1577	Malta
Germany	14889	7339	28053	Austria	3309	Malta
Hungary	884	965	4508	United States	54	Malta
Italy	15532	9664	45730	Switzerland	3944	Estonia
Latvia	230	278	1021	Russia	9	Luxembourg
Lithuania	522	736	3253	Russia	20	Malta
Luxembourg	229	317	1278	Belgium	19	Malta
Norway	1206	1615	7492	Sweden	93	Luxembourg
Poland	2222	1873	9022	Germany	142	Malta
Portugal	1004	1083	4995	Spain	107	Latvia
Slovakia	547	664	3137	Czech Republic	26	Malta
Slovenia	355	318	1281	Germany	53	Mexico
Sweden	2589	3620	21511	Norway	238	Malta
United States	12071	16545	83727	Canada	655	Luxembourg

Source: OECD

Table 2. Results of the econometric estimations

	(1) OLS	(2) Poisson	(3) HMR	(4) OLS Reduced sample	(5) Poisson Reduced sample	(6) OLS with nº of exporting firms	(7) Poisson with nº of exporting firms
Distance (log)	-1.66*** (0.10)	-1.23*** (0.11)	-1.63*** (0.12)	-1.72*** (0.12)	-1.42*** (0.10)	-0.56*** (0.11)	-0.57*** (0.12)
Adjacency	0.19 (0.12)	0.11 (0.09)	0.47*** (0.15)	0.39* (0.16)	0.37** (0.11)	-0.01 (0.07)	0.05 (0.07)
Language	0.16 (0.16)	0.37** (0.13)	-0.22 (0.31)	-0.16 (0.32)	0.08 (0.08)	0.01 (0.12)	0.03 (0.12)
RTA	0.22 (0.27)	0.24 (0.17)	1.91*** (0.24)	1.87*** (0.26)	2.68*** (0.21)	0.41* (0.17)	0.45*** (0.14)
Common currency	-0.07 (0.11)	0.01 (0.12)	0.09 (0.14)	0.13 (0.15)	-0.09 (0.13)	-0.10 (0.08)	-0.02 (0.10)
Estimated nº of exporting firms			3.21 (5.77)				
Actual nº of exporting firms						0.97*** (0.01)	0.79*** (0.07)
Observations	716	716	487	487	487	716	716
Adjusted R-square	0.91	0.96		0.88	0.91	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, ** significance at the 5-percent level and * significance at the 10-percent level.

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.

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*** (0.10)	-1.15*** (0.10)	-0.46*** (0.09)	-0.46*** (0.11)	-2.13*** (0.12)	-1.68*** (0.14)	-0.73*** (0.17)	-0.73*** (0.16)	-2.13*** (0.14)	-0.98*** (0.19)	-1.10*** (0.17)	-0.47* (0.20)
Adjacency	0.01 (0.13)	0.12 (0.09)	-0.05 (0.08)	0.03 (0.07)	0.28 (0.16)	0.12 (0.11)	0.02 (0.11)	0.06 (0.09)	0.27 (0.17)	0.22 (0.15)	0.02 (0.13)	0.20 (0.13)
Language	0.35* (0.17)	0.34* (0.13)	0.12 (0.14)	-0.01 (0.15)	0.02 (0.21)	0.22 (0.15)	-0.17 (0.16)	-0.12 (0.15)	0.37* (0.19)	0.50*** (0.15)	0.18 (0.17)	0.08 (0.13)
RTA	0.17 (0.27)	0.36* (0.17)	0.21 (0.18)	0.36** (0.13)	-0.01 (0.35)	-0.52** (0.23)	0.46* (0.23)	0.12 (0.21)	-0.62 (0.36)	0.55* (0.27)	0.00 (0.27)	0.94*** (0.26)
Common currency	-0.16 (0.11)	-0.08 (0.12)	-0.14 (0.08)	-0.05 (0.10)	0.13 (0.15)	0.36* (0.16)	0.01 (0.11)	0.24 (0.13)	-0.16 (0.17)	0.00 (0.17)	-0.16 (0.15)	0.05 (0.16)
Nº of exporting firms			1.18*** (0.08)	1.01*** (0.08)			1.03*** (0.10)	0.88*** (0.07)			0.82*** (0.10)	0.59*** (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.

Table 5. Econometric estimations for the year 2006 and the year 2007

	Year 2006				Year 2007			
	OLS	Poisson	OLS with nº firms	Poisson with nº firms	OLS	Poisson	OLS with nº firms	Poisson with nº firms
Distance (log)	-1.44***	-1.26***	-0.14	-0.53***	-1.54***	-1.27***	-0.36***	-0.50***
	(0.19)	(0.07)	(0.22)	(0.08)	(0.12)	(0.08)	(0.08)	(0.10)
Adjacency	0.44	0.08	0.08	0.04	0.29	0.06	0.00	0.04
	(0.23)	(0.07)	(0.10)	(0.05)	(0.17)	(0.07)	(0.08)	(0.05)
Language	-0.24	0.44***	-0.21	0.16*	-0.07	0.45***	-0.05	0.15
	(0.39)	(0.11)	(0.21)	(0.08)	(0.26)	(0.17)	(0.14)	(0.09)
RTA	0.80*	0.28*	0.77***	0.40***	0.56	0.17	0.56**	0.30*
	(0.39)	(0.14)	(0.19)	(0.10)	(0.32)	(0.15)	(0.18)	(0.12)
Common currency	-0.04	0.07	-0.07	0.05	-0.13	0.12	-0.08	0.09
	(0.11)	(0.10)	(0.08)	(0.07)	(0.11)	(0.11)	(0.08)	(0.08)
No. of exporting firms			1.24***	0.88***			1.12***	0.90***
			(0.13)	(0.06)			(0.06)	(0.08)
Observations	681	681	681	681	635	635	635	635
Adjusted R-square	0.89	0.97	0.94	0.98	0.91	0.97	0.95	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.