Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Continuing Education Center

MSc Economics

Introducing tariff data in the gravity equation

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by

Robert Kunst

Claudia Ciuciu

1325894

Vienna, 08.06.2015

MSc Economics

Affidavit

I, Claudia Ciuciu

hereby declare

that I am the sole author of the present Master's Thesis,

Introducing tariff data in the gravity equation

22 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, 08.06.2015

Signature

Contents

List of Figures

List of Tables

Abstract

This paper studies whether the inclusion of tariffs in the gravity estimations makes a significant change on the other commonly-used trade costs. Including tariffs in the gravity equation is not a common practice, due to the difficulty of obtaining quality aggregated data. Being aware of this pitfall, I have collected the best free available data on applied tariffs and constructed a panel of eighteen countries, along twenty-one years on which I estimated two forms of the gravity-equation, the difference between them being the scaling of the dependent variable. I used two econometric methods to estimate the equations - the least-square dummy variable approach and the Poisson pseudo-maximum likelihood estimator. The results I obtained depend on whether the dependent variable is size-adjusted or not; when just considering exports, including tariffs reduces the bias on the free-trade-agreement coefficient, while leaving the other trade costs coefficients in usual ranges. This finding is not robust to the specification of size-adjusted exports; in this case, omitting tariffs from the gravity equation does not lead to significant changes in the other coefficients.

1 Introduction

The gravity equation has been the workhorse in a substantial amount of empirical trade research. Its success was due to its high explanatory power of trade patterns. The analogy between Newton's law of gravity and trade flows was purely atheoretical, which was always the drawback of this concept and the main reason it was not accepted into the mainstream economics. Consequently, many economists have tried to provide formal models which could yield a gravity-type of relationship between the demand for trade, output and trade costs. One of the most popular of such models was the Anderson and Van Wincoop (2003) model. It requires (like all the other structural models that deliver the gravity-type equation) homothetic preferences for the agents and perfect product specialization(each country produces only one type of good). As for the other models that are not discussed here, they only differ in the supply side: different types of product specialization are considered, e.g. technology differences across countries, increasing return at the firm level (Evenett and Keller (2002)). All these models established that trade flows are a multiplicative function of countries' output, bilateral trade costs and relative trade costs. Concerning the trade cost function, it is empirically assumed to be a loglinear function of observables; distance, dummies for common border and language and FTA membership are commonly used as proxies for trade costs. As pointed out by Anderson and Van Wincoop (2004), theory has no say on the functional form of the trade costs, which lead to arbitrary forms of the functions and the variables considered. This fact represents one of the gravity equation's main critiques.

With the support of a theoretical framework, the popularity of the gravity equation has increased even more. Another direction in which research has developed concerns gravity estimation. As a consequence of its multiplicative form, log-transformations have been performed, making the gravity equation a good candidate for OLS estimators. Estimations have been done on cross-sectional and panel data. There has been a visible preference for panel data, because it allows for time variation of certain trade costs. One novel paper by Silva and Tenreyro (2006) drew attention to the following facts: in the presence of heteroskedasticity, OLS estimators are not consistent, leading to confounding results. They suggest a Poisson pseudo-maximum likelihood estimator for which they argue that it is equivalent to the non-linear estimator which minimizes the sum of least-squares. It relies on the assumption that the conditional mean of trade flows is given by an exponential function and that it must be the true mean. The second fact their paper discusses is the fact that trade observations are zero for various sectors and countries. Again, the log-transform cannot handle this, leaving the zero-observations out, which the Poisson pseudo-maximum-likelihood estimator (from now on, abbreviated as PPML) does not.

What I have always found puzzling in the gravity estimations is the lack of tariff (at country level) variables from the trade cost function. Even though it is argued that they have decreased a lot, there are still significant differences in tariff rates across developed and less developed countries. The reason there were no tariffs is the lack of publicly available data. Tariff data is easier to obtain on a product-level rather than at country-level. Going deeper in how tariffs are structured, average tariffs are quite cumbersome to aggregate: all goods are divided into 21 sections, according to the Harmonized System (which is the standard of classifying traded products), 96 chapters and further headings and subheadings. Each product gets assigned a section, chapter and the rest that follows, and can have a number consisting of 9 digits in the Harmonized System(HS). Tariffs are assigned to goods which are classified under the HS-6 digit. Besides having to aggregate five levels up, the HS has changed its structure several times (1988, 1996, 2002, 2007, 2012) and for a meaningful comparison, one must convert tariffs from several classifications to just one classification. Nonetheless, World Bank's database called WITS (World Integrated Trade Solutions) provides aggregated data on effectively applied tariffs for a large number of countries, for time range 1988-2013. The only issue is that the data is not complete, so in order to make an estimation of the gravity equation using tariffs, I had to resort to some method of interpolation for the missing points. I chose to perform simple averages for the missing yearly tariff rates as I found no compelling reason to resort to more involved interpolation schemes.

Having the tariff data, I constructed a panel for 18 countries, over 1993-2013, for which the details can be found in the section *Tariffs and Data*, on which I estimated a gravity equation using both econometric approaches mentioned above. The following section provides some theoretical background on the gravity equation I used for the estimation. The third section discusses in more detail the estimation methods I used. Section number four elaborates on the tariffs and the data on the other variables, whereas section five contains the estimation and the results. Lastly, section six summarizes the findings.

2 The gravity equation

The idea and the name for the gravity equation appeared in the trade literature in 1962 with a publication of Tinbergen (1962), in which he made a simple analogy between Newton's law of gravitation (the force between two objects is proportional to the product of their mass and inversely proportional to the square of the distance between them) and the trade flows between two countries (where the "mass" of the countries was considered to be the GDP of each country). The main interest of Tinbergen was to be able to explain trade patterns in absence of "discriminatory trade impediments" (Tinbergen, 1962); he used cross-section data and least-square estimation to find that the mentioned right-hand-side variables managed to explain around 80% of the trade flows. Additionally, Tinbergen also studied the influence of trade preferences (Commonwealth, Benelux) on trade flows between countries.

This approach became popular among trade economists who continued to use this simple, statistical model because of its explanatory power and because it seemed a useful tool to asses the impact of free-trade agreements or regional-trade agreements on trade flows. However, because of the lack of a rigorous theoretical framework, it was not generally accepted into mainstream economics.

Anderson (1979) is recognised as the first economist who formulated a general equilibrium model, from which a gravity equation was derived (endowment economy, where each country is specialized in the production of one good and agents have homothetic preferences). The setup of the model, however, was frictionless: no transportation costs, no tariffs, no trade costs at all, which represented its main critique. Bergstrand (1985) allowed for frictions in the model he developed, in additional to putting more structure on the consumer side, but he also assumed an endowment economy.

The two papers represented turning points in the acknowledgment of the gravity equation in trade theory and others models have improved upon those. To close the gap between the Ricardian and Heckscher-Ohlin-Vanek trade theory and the gravity equation, Krugman (1979) and Helpman and Krugman (1985) put more structure on the producer side and market (monopolistic market, with a production function which exhibits increasing returns to scale and one input factor - labor) and this GE approach delivered a gravity-type optimal demand of exports.

The Anderson-van Wincoop model

Until 2003, all gravity-related empirical papers referred to Anderson (1979) and Bergstrand (1985) as the theoretical background. What these two approaches lacked was taking into consideration the endogeneity of the prices. Anderson and Van Wincoop (2003) extended the gravity framework, by using a system of equations to capture the prices.

In what follows, I will provide a description of the Anderson and Van Wincoop (2003) model, as it will be the workhorse of the estimation. The setup of the model is as follows:

All goods are differentiated by places of origin (Armington assumption). Each region is specialized in the production of one good. Supply of each good is fixed. There are homothetic preferences, approximated by a CES utility function and there is a representative consumer for each country. c_{ij} is the consumption by region j of goods from region i. therefore, consumers from region j maximize:

$$\left(\sum_i \beta_i^{(1-\sigma)/\sigma} c_{ij}^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)} \tag{1}$$

subject to the budget constraint

$$\sum_i p_{ij} c_{ij} = y_j \tag{2}$$

Here σ is the elasticity of substitution between goods, β_i is a positive taste parameter (consumers from each country value the goods differently) , y_j is the nominal income in region j, p_{ij} is the price of the good from region i paid by consumers in region j. Prices differ between locations due to trade costs that are not directly observable.

- Let p_i be the prices of the exporter i net of trade costs and let t_{ij} be the trade cost between i and j. The price the consumers pay is $p_{ij} = p_i t_{ij}$. This assumption on the final price relies on what is called in the literature "iceberg" trade costs; part of the value of the good (t_{ij} of its original value) "melts" away during the shipment.

- It is assumed that trade costs are supported by the exporter. For each good shipped from i to j the exporter pays a cost $t_{ij} - 1$ of country i goods.

- Nominal value of exports from i to j is $x_{ij} = p_{ij} c_{ij}$ which is the sum of values of production at origin $p_i c_{ij}$ and the trade cost $(t_{ij} - 1) p_i c_{ij}$ that the exporter passes to the importer.

- Total income of region i is given by $y_i = \sum_j x_{ij}$

The result of the constrained maximization problem, i.e. first order conditions yield the following optimal demand of goods from country i by country j:

$$x_{ij} = \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{1-\sigma} y_j \tag{3}$$

where $P_j = \left[\sum_i (\beta_i p_i t_{ij})^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$ is a CES price index for consumers in country j.

Imposing market clearance:

$$y_i = \sum_j x_{ij} = \sum_j (\beta_i p_i t_{ij}/P_j)^{1-\sigma} y_j = (\beta_i p_i)^{1-\sigma} \sum_j (t_{ij}/P_j)^{1-\sigma} y_j \tag{4}$$

Solving for the scaled prices $\beta_i p_i$ from (4), substituting for them in the optimal demand (3) and in the definition of P_j and using the fact that each country's output is a fraction of the world output $\theta_i = y_i/y^W$ and sum of countries' output gives the world output $y^W = \sum_i y_i$ yields the gravity equation:

$$x_{ij} = \frac{y_i y_j}{y^W}\left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma} \tag{5}$$

where

$$\Pi_i^{1-\sigma} = \sum_j \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{1-\sigma} \frac{y_j}{y^W} \tag{6}$$

is called outward multilateral resistance term (OMR) - as if the sellers in each region shipped to a single world market.It represents the average trade costs supported by seller i to all destinations.

$$P_j^{1-\sigma} = \sum_i \left(\frac{\beta_i p_i t_{ij}}{\Pi_i}\right)^{1-\sigma} \frac{y_i}{y^W} \tag{7}$$

is called inward multilateral resistance term (IMR) - as if the buyers in each region imported from a single world market. It represents the average trade costs supported by the buyer j from all sources of import.

Put together, the multilateral resistance terms form a system of equations that can be solved in terms of trade costs t_{ij}. One solution is $\Pi_i = P_i$:

$$P_j^{1-\sigma} = \sum_i P_i^{\sigma-1} \theta_i t_{ij}^{1-\sigma} \tag{8}$$

This makes the optimal demand x_{ij} be equal to:

$$x_{ij} = \frac{y_i y_j}{y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma} \tag{9}$$

The price indexes are referred to as multilateral resistance terms, they are not the usual price indexes for which the CPI data is available (because of the fact that such price indexes consider non-tradable goods as well). What these terms measure is the trade barriers of each country with respect to all its partners. An increase in the trade costs of country i which exports to country j will increase the relative price of good i and divert the imports from i to the other partners of country j.

There is one more aspect to mention: Anderson and van Wincoop suggest that another possibility of interpretation of the gravity equation is to divide equation (9) by the product of each country's output. Then, one would obtain a relationship between size-adjusted exports and trade costs.

$$\frac{x_{ij}}{y_i y_j} = \frac{1}{y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma} \tag{10}$$

Overall, the gravity equation derived in this model shows that trade between countries depends on relative trade barriers (Anderson and Van Wincoop, 2003), more precisely, for a pair of countries, trade is influenced by the bilateral trade barriers relative to the average barriers with the other trading partners.

3 Estimation Methodology

There are two main econometric methods employed throughout the literature which dominate the gravity estimation: linear estimators(OLS) and Poisson pseudo-maximum likelihood estimators. However, there have also been economists which suggested and used non-linear estimators (Anderson and Van Wincoop, 2003; Helpman et al., 2007), given the form of the multilateral resistance terms.

3.1 OLS Fixed-Effects

Due to its multiplicative form, equation (9) can be log-linearized and become a suitable candidate for linear estimators:

$$\log x_{ij} = \log y_i + \log y_j - \log y^W + (1-\sigma)[\log t_{ij} - \log P_i - \log P_j] \tag{11}$$

where x_{ij} represents the exports from country i to j, y_i represents the GDP of the exporter, y_j is the GDP of the importer, t_{ij} are the trade costs, on which I will elaborate below and P_i and P_j are the multilateral resistance terms mentioned above. This is sometimes referred to, in the literature, as the traditional gravity equation, having the logs of GDPs on the RHS.

After some algebra, the LHS of equation (11) becomes

$$\log \frac{x_{ij}}{y_i y_j} = -\log y^W + (1-\sigma)[\log t_{ij} - \log P_i - \log P_j] \tag{12}$$

where the $\frac{x_{ij}}{y_i y_j}$ is the size-adjusted exports.

Until the 1990s, the equation was estimated on cross-section data which did not account for heterogeneity of the countries and results varied across country samples. To control for such heterogeneity, economists have expanded the cross-section data to panel data. This has several advantages, apart from the aforementioned: it allows for the inclusion of fixed effects in the regression, which is a way to capture the countries' time-invariant particular features; this fits well with the gravity specification, given that the multilateral resistance terms are difficult to find data for, but replacing them with fixed effects solves the problem of having to omit them from the estimation (removing omitted variable bias). Another advantage of using panel data, and particularly fixed effects model, is that it allows for the covariates(variables representing the trade costs) to be correlated with the variables representing the fixed effects. Concretely, thinking of multilateral resistance terms and their definition (both of them are functions of trade costs), they are correlated with other trade barriers that are explicitly accounted for (because there is available data). Using OLS fixed effects estimator (or least-square dummy variable estimator), one takes care of the bias arising from mis-specification of trade costs function (which, empirically, is assumed to be linear in observables).

Furthermore, to advocate for the use of fixed-effects OLS estimation, there have been gravity papers (Egger, 2000) which perform a Hausman test for fixed versus random model specification, resulting in favor of the fixed effects model. therefore, the equation to be estimated is

$$\log x_{ij} = \log y_i + \log y_j - \log y^W + \beta_1 \log t_{ij} + \delta_i + \delta_j + \epsilon_{ij} \tag{13}$$

By using fixed effects (one dummy variable for the exporter δ_i, one for the importer δ_j), the multilateral resistance terms are assumed to be time-invariant. However, this does not accurately reflect reality as it is a fair assumption that the MRT change over time. To account for this in the gravity estimation, I added an interaction term between the fixed effect dummies and time.

However, another point that must be mentioned in the case of the OLS estimator is the endogeneity bias, which fixed-effects approach cannot remove (unlike the above omitted variable bias). Trade flows are part of the output (GDP) for every country, so there is cause of concern that exports influence output (y_i, y_j). To remove this bias, what has been done throughout the literature is instrumental variable estimation. The most common instruments are population and factor endowments (Anderson and Van Wincoop, 2004). Nonetheless, the results of the IV estimation for the other coefficients do not change in sign or magnitude if one doesn't account for the endogeneity bias (Evans, 2003).

3.2 Poisson Pseudo-Maximum Likelihood estimator

Silva and Tenreyro (2006) acknowledged that there is one major flaw with the OLS estimators, in the case of gravity equations. The problem arises from the nature of the equation (a multiplicative relationship between the bilateral trade flows and its determinants) and the log transformation. In a deterministic approach, the gravity equation can be written as an exponential function $x_{ij} = \exp(\log y_i + \log y_j - \log y^W + \beta_1 \log t_{ij} + \delta_i + \delta_j)$, where the exponential function is seen as the conditional expectation of the trade flows, given trade determinants. However, given that we cannot measure these variables perfectly, i.e. there is measurement error, trade flows should be estimated (stochastic/probabilistic version) by

$$x_{ij} = \exp(\log y_i + \log y_j - \log y^W + \beta_1 \log t_{ij} + \delta_i + \delta_j) + \epsilon_{ij} \tag{14}$$

where x_{ij} can be 0.

Taking logs and estimating the coefficients with OLS is inappropriate because of the following reasons (Silva and Tenreyro, 2006): the values of the trade flows can be 0 (so taking logarithm is not an option); the expected value of the log-linearized error will depend on the regressors and OLS will be inconsistent. therefore, Silva and Tenreyro suggest that a non-linear estimator should be used instead.

A usual non-linear estimator (for the conditional mean being an exponential function $y_i = \exp(x_i\beta)$) would be minimizing the sum of least squares:

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} [y_i - \exp(x_i\beta)]^2 \tag{15}$$

for which the first order conditions are:

$$\sum_{i=1}^{n}[y_i - \exp(x_i\hat{\beta})] \exp(x_i\hat{\beta})x_i = 0 \tag{16}$$

The problems with this type of NLS is that it puts more weight on the observations where $\exp(x_i\hat{\beta})$ is large (the observations with larger variance). If the form of the conditional variance would be known, then a weighted NLS could be used. However, there is little knowledge about the form of the conditional variance. Based on the assumption that the conditional variance is proportional to the conditional mean, a more efficient estimator is given by a pseudo-maximum likelihood estimator. Such an estimator is obtained by maximizing a log-likelihood function associated with a family of distributions (exponential family) that does not necessary contain the true distribution.

The first order conditions for β (using the PML) become

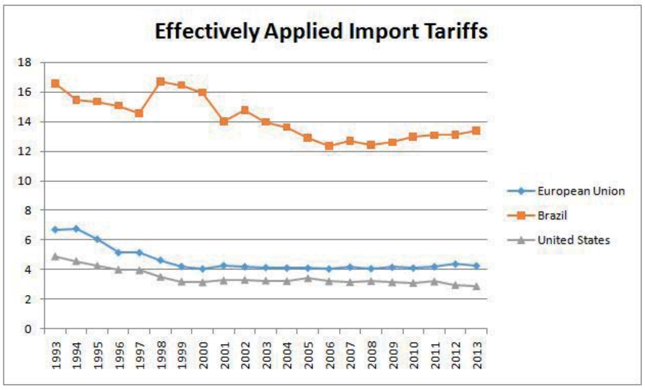
$$\sum_{i=1}^{n}[y_i - \exp(x_i\widetilde{\beta})]x_i = 0 \tag{17}$$

The estimator defined by (17) is numerically equivalent to a Poisson pseudo-maximum likelihood estimator (Silva and Tenreyro, 2006; Winkelmann, 2003). The PPML is a consistent estimator and this property depends on the correct specification of the conditional mean ($E[y_i|x] = \exp(x_i\beta)$). This also implies that the data doesn't have to be Poisson for a consistent estimation.

4 Tariffs and Data

Over the decades, there has been a significant decrease in import tariff rates and an increase in the number of trade agreements, for which the sole purpose is to facilitate trade by lowering trade barriers. Figure 1 reports the bilateral effectively applied tariff rates for a sub-sample of countries (the European Union, United States and Brazil for New Zealand) evolving over a period of twenty years (1993-2013), from which a decreasing trend is noticeable. While currently, most of the tariffs are on average around 5 %, there are still some bilateral tariffs which are above 10%, especially for the developing countries, for trade protection reasons.

Figure 1: Effectively applied import tariffs for New Zealand



Trade barriers are generally represented in the empirical gravity work by distance. In the same category of trade costs, but not being impediments to trade, just facilitators of trade, economists include indicator variables to account for common language and borders and for the existence of a trade agreement. What has been missing from gravity estimation are tariffs, explicit values for import tariffs. The main reason behind this is the unavailability of data. Recently, the World Bankwit (2015) has made publicly available a database that contains tariff information and data for a large number of countries, on a time-span of approximately twenty-five years. The tariffs that it reports can be categorized into three classes:

- effectively applied tariffs (AHS)
- most-favourite-nation tariffs (MFN)
- bound tariffs

According to the definition by the World Banktar (2015), MFN tariff rates are what countries promise to impose on imports from other members of the World Trade Organization(WTO), except for the cases when countries have signed preferential trade agreements (free-trade, customs union). Based on this definition, MFN rates are the highest import tariffs that can be charged.

Bound tariffs are the maximum tariff rates countries can impose onto each other. They are country-pair specific. These tariffs act as threshold: countries can vary their applied tariff rates without preceding negotiations, as long as they stay within the bound rates. If one country applies tariff rates outside the determined bounds on another country's products, the country which was charged differently has the possibility to take legal actions and enforce the bound rates commitment. The country which did not respect its contractual terms can be punished by being charged higher tariffs by the other party. What is desirable is that all commodities that are traded should have bound rates. The reasons for this standardization is that it increases transparency and reduces fluctuations in international traded prices.

Lastly, WITS uses the concept of effectively applied tariff which is defined as the lowest available tariff.

All import tariffs are classified under the Harmonized System(HS) of tariff nomenclature, initiated and maintained by World Customs Organization(WCO). However, the HS provides tariff data on a disaggregated level. WITS offers tariff data for the above mentioned types of tariffs at a sector level and also at aggregate level. They report two types of averages - simple average and trade weighted average (i.e. the average of the tariff rates weighted by the product import shares corresponding to each partner country).

Using WITS database, I have collected data on the effectively applied tariffs, since those are the ones who are actually paid in practice. The tariff data is yearly reported, but for some countries in the sample, data on all years was not available . In this situation, for the missing yearly observation, I used a simple average between the previous and subsequent year to replace the missing value for the tariff. However, this interpolation leaves room for bias in the aggregation.

Data for the exports, GDP, GDP deflator are taken from IMF Direction of Trade Statistics database; the list of the countries considered for the analysis can be found in the appendix. The distance considered in the regressions is computed as the distance between the capital cities of each country-pair. The information about a country's participation in preferential trade agreement is taken from World Trade Organization website WTO (2015).

5 Estimation and Results

The trade costs that I assumed for this analysis enter the gravity equation in the following form:

$$t_{ij} = dist_{ij}^{\gamma_1} tariff_{ij}^{\gamma_2} exp(\gamma_3 lang_{ij} + \gamma_4 border_{ij} + \gamma_5 FTA_{ij}) \tag{18}$$

Hence, the gravity regression is specified in logs and levels:

$$\begin{aligned} \log exports_{ij,t} =& \alpha + \beta_1 \log GDP_{i,t} + \beta_2 \log GDP_{j,t} + \gamma_1 \log dist_{ij} + \gamma_2 \log tariff_{ij,t} \\ &+ \gamma_3 lang_{ij} + \gamma_4 border_{ij} + \gamma_5 FTA_{ij} + \delta_i + \delta_j + \epsilon_{ij,t} \end{aligned} \tag{19}$$

where the δ_i and δ_j represent the country-specific fixed effects (which are an empirical approximation to the multilateral resistance terms).

A reformulation of the above equation, where the exports are size-adjusted:

$$\begin{aligned} \log \frac{exports_{ij,t}}{GDP_{i,t} GDP_{j,t}} =& \alpha + \gamma_1 \log dist_{ij} + \gamma_2 \log tariff_{ij,t} \\ &+ \gamma_3 lang_{ij} + \gamma_4 border_{ij} + \gamma_5 FTA_{ij} + \delta_i + \delta_j + \epsilon_{ij,t} \end{aligned} \tag{20}$$

For the PPML estimation, the gravity equation, both with the dependent variable being exports and size-adjusted exports is given by

$$\begin{aligned} exports_{ij,t} =& \alpha + \beta_1 \log GDP_{i,t} + \beta_2 \log GDP_{j,t} + \gamma_1 \log dist_{ij} + \gamma_2 \log tariff_{ij,t} \\ &+ \gamma_3 lang_{ij} + \gamma_4 border_{ij} + \gamma_5 FTA_{ij} + \delta_i + \delta_j + \epsilon_{ij,t} \end{aligned} \tag{21}$$

Size-adjusted version:

$$\begin{aligned} \frac{exports_{ij,t}}{GDP_{i,t} GDP_{j,t}} =& \alpha + \gamma_1 \log dist_{ij} + \gamma_2 \log tariff_{ij,t} \\ &+ \gamma_3 lang_{ij} + \gamma_4 border_{ij} + \gamma_5 FTA_{ij} + \delta_i + \delta_j + \epsilon_{ij,t} \end{aligned} \tag{22}$$

Using size-adjusted exports would give less weight to the observation with a large product of GDPs, assuming that the variance is proportional to the square of this product (Silva and Tenreyro, 2006). This means whether to choose exports or size-adjusted exports depends on what pattern of heteroskedasticity the data has. I will estimate the gravity equations using both types of exports as dependent variable and look at how the results change.

Table 1 presents the results of estimating (19), considering data for effectively applied tariff rates. The results are presented in contrast with the estimated gravity without the variable tariff. Firstly, a least-square dummy variable estimator is used, with two different approaches with respect to fixed-effects estimation.

Dependent Variable: Log of Exports				
Variable name	Country fixed effects		Country time-varying fixed effect	
	with tariffs	without tariffs	with tariffs	without tariffs
	(1)	(2)	(3)	(4)
Log of GDP Exporter	.113*	.271***	-.086	−.102
	(.023)	(.000)	(.275)	(.157)
Log of GDP Importer	.356***	.433***	.053	.667
	(.000)	(.000)	(.156)	(.071)
Log of tariffs	−6.622***		−6.981**	
	(.000)		(.001)	
Log of distance	−.720***	−.744***	−.759***	−.802***
	(.000)	(.000)	(.000)	(.000)
Common language	.653***	.641***	.657***	.646***
	(.000)	(.000)	(.000)	(.000)
Common border	.077	.102	.064	.0808
	(.801)	(.761)	(.838)	(.808)
FTA dummy	.789***	.949***	.658***	.7731***
	(.000)	(.000)	(.000)	(.000)
Obervations	6315	6315	6315	6315
R^2	.8817	.8744	.897	.888

p-values are in parenthesis
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 1: LSDV estimation of the gravity equation

The first two columns are estimating equation (19) using a dummy variable for the exporting country and another one for the partner country(the fixed effect). Column one includes the import tariff as an additional trade barrier, whereas the second column repeats the estimation without tariff data. The log-linear form of the gravity equation imposes that the elasticities of exports with respect to the GDPs should be one. Empirically, however, both of the coefficients are considerably different from one. This result is not in accordance with the gravity concept.

The coefficient on the tariffs is meant to represent the trade-costs elasticity of trade flows; from the theoretically derived equation, $1 - \sigma = -6.622$ which means an elasticity of substitution of 7.6 between goods. As Anderson and Van Wincoop (2003) point out , the model assumes one elasticity for all goods, but, in reality, the elasticity depends on the industry: some industries may even have goods close to being perfect substitutes (Feenstra, 1994). Therefore, an average elasticity of 7.6 seems to fit the empirical literature. The classic interpretation of this elasticity is as follows: if the import tariff decreases by 1 % then the exports will rise with approximately 7.6 %.

The other usual trade barriers have appropriate signs, except for the common border dummy, which is not significantly different than 0. Exports are negatively correlated with distance and positively correlated with sharing a common border and being part of a trade agreement. Common language is a usual proxy for cultural/historical proximity. These “non-traditional” determinants of economic exchange turn out to be important factors in trade patterns (Head and Mayer, 2013).

The coefficient on the FTA dummy indicates that a pair of countries signing a trade agreement will experience an increase in their bilateral trade of approximately 120 % ($e^{.789} - 1 = 1.20$). However, the coefficient on the FTA dummy has been the center of debate in the empirical gravity literature in the late 2000s. The main issue refers to the fact that the coefficient of the FTA is usually biased: countries who share similar economic characteristics may self-select themselves into an FTA. These characteristics may be represented by covariates already included in the RHS of the gravity equation or other unobservable (that are still correlated with the before-mentioned covariates). However, the purpose of this work is not to deal with the FTA endogeneity. Baier and Bergstrand (2007) have provided a treatment effect approach to control for the endogeneity. Their results indicate a range for the average treatment effect of an FTA, from 0.46 to 0.68.

Column two reports the same country-fixed effects OLS regression, but without the tariff data. Not accounting for tariff data increases the coefficients on the GDP, so country output plays a greater role in influencing trade flows, but countries are less willing to trade when tariffs exist, based solely on their outputs. Hence, trade is less elastic to countries outputs when tariffs enter the gravity estimation.

The presence of tariffs though does not affect the coefficients on distance and language in a significant manner, but another result is that the coefficient on the FTA dummy has a much higher magnitude. Hence, just by comparing these two columns, introducing explicit tariff data reduces the bias on the FTA dummy.

Column three and four estimate the gravity equation (with and without tariffs) allowing for a time effect in addition to the country-specific features. In other words, this specification allows for the multilateral resistance terms to change over time. The results are as follows: the coefficients on the GDPs become insignificant, which means that, in this specification, the economic size of the countries does not play a role in demand for trade, which is not in accordance with theory or empirics. Yet, comparing the first two column with the third and fourth (with the exception of GDP), the results stay in a similar range.

Table 2 estimates equation (20)-dependent variable is the size-adjusted exports, using the fixed effect approach, just as in table 1. Looking at the first column, the coefficient on tariffs is not statistically different from 0 and when introducing tariffs in the regression, the other coefficients change very little. From this specification, one can conclude that omitting tariffs from the gravity equation doesn't lead to more bias in the other coefficients; the other trade costs and the fixed effects are capturing the effects of tariffs on trade.

Column three and four estimate the same equation (20), but allowing for a time effect of the dummies which are a proxy for the multilateral resistance terms. In this case, the coefficient on tariffs is statistically significant, and implies an elasticity of trade of approximately 7 %, which is in the range of other estimates in the literature. The coefficients on the common border and language do not vary significantly, but here, again, the FTA coefficients is smaller in the case when tariffs are included then in the equation. Hence, in this case including the tariffs has a direct effect on the FTA dummy, by reducing its bias.

The results on the other trade cost coefficients look similar when the dependent variable is given by the exports or by the size-adjusted exports. This means that the country-pair observations for the size-adjusted exports do not have large variance to make the results sensitive to the type of exports considered.

Dependent Variable: Log of Size-adjusted Exports				
Variable name	Country fixed effects		Country time-varying fixed effect	
	with tariffs	without tariffs	with tariffs	without tariffs
	(1)	(2)	(3)	(4)
Log of tariffs	.473		-5.976^{**}	
	(.661)		(.006)	
Log of distance	$-.772^{***}$	$-.774^{***}$	$-.743^{***}$	$-.780^{***}$
	(.000)	(.000)	(.000)	(.000)
Common language	$.643^{***}$	$.642^{***}$	$.652^{***}$	$.643^{***}$
	(.000)	(.000)	(.000)	(.000)
Common border	.089	.091	.075	.0889
	(.779)	(.776)	(.812)	(.789)
FTA dummy	$.845^{***}$	$.857^{***}$	$.740^{***}$	$.838^{***}$
	(.000)	(.000)	(.000)	(.000)
Obervations	6315	6315	6315	6315
R^2	.6015	.6014	.6814	.678

p-values are in parenthesis
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 2: LSDV estimation of the gravity equation

Table 3 presents the results of estimating equation in the form of (18) - having the exports(LHS) in levels. I have estimated two specifications: country fixed-effects and country-pair fixed effects using the PPML estimator.

Looking at the first two columns from table 2, the coefficients on GDP, distance and common border have the appropriate signs and are similar in magnitude. This can be interpreted as follows: accounting for the tariff variable as an explicit trade cost in the gravity equation has no significant impact on the other coefficients, it simply shows the elasticity of trade to trade costs, it does not create or reduce bias on the other trade barriers. There is, nonetheless, a visible change in the coefficient for the FTA dummy. Given the endogeneity bias I mentioned above, it seems that including tariffs in the gravity equation works in the direction of reducing the bias on the FTA coefficient. What seems to be a problem is that the coefficient on the common language dummy has a negative sign, which is counter-intuitive and not in accordance with the literature.

The results from column three and four (which have one dummy - fixed effect) for a country-pair seem to be somewhat consistent: the coefficients on GDP and distance are similar in both specifications (with and without tariffs), which has the same interpretation as above - having tariffs doesn't have any significant impact. For the common language coefficient, not having tariff as an explanatory variable leads to a positive coefficient, but quite high in magnitude. In the specification with tariffs, the coefficient is negative (and even higher in absolute value than the one from column four). The coefficient on the FTA dummy, however, follows the same pattern as before: when explicitly introducing tariff data into the gravity equation, the FTA bias decreases.

Table 4 presents the results of estimating equation (22) using PPML estimator. There are the same two types of fixed effects specifications as in the previous tables. The first two columns take into consideration a dummy for each country

Dependent Variable: Exports				
Variable name	Country FE		Country time-varying fixed effects	
	with tariffs	without tariffs	with tariffs	without tariffs
	(1)	(2)	(3)	(4)
Log of GDP Exporter	.322***	.339***	.337***	.334***
	(.000)	(.000)	(.000)	(.000)
Log of GDP Importer	.399***	.443***	.403***	.439***
	(.000)	(.000)	(.000)	(.000)
Log of tariffs	−4.105***		−3.269**	
	(.000)		(.000)	
Log of distance	−.509***	−.507***	−.9056***	−1.003
	(.000)	(.000)	(.000)	(.000)
Common language	−.144***	−.1738***	−3.767***	2.645***
	(.000)	(.000)	(.000)	(.000)
Common border	.997***	1.006***	3.568***	−2.039***
	(.000)	(.000)	(.000)	(.000)
FTA dummy	.4670***	.6339***	.5311***	.7367***
	(.000)	(.000)	(.000)	(.000)
N	6340	6426	6321	6405
R^2	.925	.914	.946	.934

p-values are in parenthesis
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 3: PPML estimation of the gravity equation

in the pair; from the first column, the coefficient on the tariff is not significantly different than 0, just like in the OLS estimation (table 2). All the other coefficients on the trade costs are similar in the case with tariffs and the case without them. Again, just like in table 2, omitting tariff data does not change the usual results of the gravity equation. However, concerning the estimation methodology, the PPML provides different results for the FTA dummy - coefficients are smaller than in table 2, which is preferable, from a view-point of current research on the bias of the FTA coefficient.

Columns three and four have in addition to the first two columns, a time interaction dummy, again, to represent the concept that multilateral resistance terms may vary in time. Here, in contrast with the first column, the coefficient on tariffs is significant and it shows an elasticity of exports to tariffs of 8.275 %. The other coefficients change slightly, but they are in the usual range of the literature. In this case, it's difficult to say whether having an explicit variable for tariffs improves the overall results or not. In comparison to the OLS estimates, PPML estimates a lower coefficient on the FTA dummy, but this seems to be a matter of the estimation procedure and not of having tariffs or not.

Summing up, when using OLS fixed-effects estimators, introducing tariff data plays a role only on two coefficients: from table 1, the coefficients on GDP and FTA decrease when including tariffs. This would mean that when explicit tariffs are considered, the size of the countries plays a smaller role in determining the export size. If the importer faces a rise in the tariffs, then exports will react less to the GDP of the importer. The other change appears in the FTA coefficient,

Dependent Variable: Size-adjusted Exports				
Variable name	Country FE		Country time-varying fixed effects	
	with tariffs	without tariffs	with tariffs	without tariffs
	(1)	(2)	(3)	(4)
Log of tariffs	-0.328		-7.275^{**}	
	(.061)		(.000)	
Log of distance	-1.014^{***}	-1.015^{***}	$-.9263^{***}$	-1.013
	(.000)	(.000)	(.000)	(.000)
Common language	$.356^{***}$	$.351^{***}$	$.415^{***}$	$.350^{***}$
	(.000)	(.000)	(.000)	(.000)
Common border	$.324^{***}$	$.327^{***}$	$.340^{***}$	$.3309^{***}$
	(.000)	(.000)	(.000)	(.000)
FTA dummy	$.349^{***}$	$.363^{***}$	$.2931^{***}$	$.3702^{***}$
	(.000)	(.000)	(.000)	(.000)
N	6340	6426	6340	6426
R^2	.701	.700	.8323	.8318

p-values are in parenthesis
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 4: PPML estimation of the gravity equation

which decreases from .949 to .789. As mentioned above, it is well-known that unless special treatment, a simple estimation will render the coefficient on the FTA biased. With this data-set, just including tariffs in the gravity equation, one can see a reduction in the bias. Clearly, the FTA and the tariffs are related: countries which join an FTA have smaller tariffs, besides other reductions in trade barriers (like importing quota, non-tariff barriers). Hence, by introducing tariff data, the proportion of the FTA corresponding to the tariffs is separated from the other benefits of an FTA. But this does not eliminate the whole bias, there is still the issue of the self-selection bias that needs to be addressed. However, this is outside the scope of this work. As cited above, Baier and Bergstrand (2007) have dealt with this issue by using a treatment approach and later on using matching econometrics (Baier and Bergstrand, 2009).

The reduction in the FTA bias is robust across most specifications. There is a clear decrease in the FTA coefficients in tables 1 and 3. In tables 2 and 4, where the dependent variable is the size-adjusted export, results go along the same line: in the first specification(when only using country-fixed effects), introducing tariffs plays no role on the FTA dummy (or any others). In this scenario, one can argue that the fixed effects for both countries absorb all the effect of the tariffs, which means that there is not enough variation in the tariffs in the sample considered. However, this contradicts the next specification (column three and four) of the table mentioned: firstly, the coefficient on the tariffs becomes highly-significant and secondly, there is some modification of the FTA (not as large in magnitude as in the other tables). A possible explanation can be that the underlying omitted variables(the multilateral resistance terms) do vary with time and the previous specification could not capture this. Because these multilateral resistance terms are function of trade costs (and therefore, tariffs), which vary in time, not modeling this aspect leads to wrong estimated coefficients. Nonetheless, this only happens when the dependent variable is the size-adjusted exports. Therefore, even though, theoretically, one should get the similar results in the size-adjusted versus non-adjusted, empirically this does not hold.

Comparing OLS fixed-effects estimator with PPML, the fact that the coefficients on the importer's and exporter's GDP are statistically significant is in favor of the PPML, but the coefficients on the common language, however, have a wrong sign (table 3).

6 Conclusion

The gravity equation has proved to be a powerful tool in estimating trade flows. It managed to overcome its critique of not being theoretically founded and improvement has been made on the econometrics part as well. Further research should focus on grounding the trade cost function in theory. Up to now, a loglinear functional form of observables has been adopted to approximate it and this form has prevailed in most studies. One obvious trade cost that was not explicitly considered in the estimation was a quantitative tariff variable and the reason was the lack of data. Recently, the World Bank has made available a database with country-level effectively applied tariffs which allowed me to estimate the gravity equation (where the dependent variable is expressed both in logs and levels) with tariffs and see whether omitting them from the estimation can lead to bias. As terms of comparison, I have used commonly-accepted ranges of coefficients published in the literature.

My main finding is that including tariff data reduces the bias on the FTA coefficient, while not affecting the other coefficients of the trade costs. This result is robust across all specifications I considered, except for the case when the dependent variable is the size-adjusted exports, with country-specific fixed effects. In this setup, the FTA coefficient shows no change when omitting the tariffs from the regression. therefore, whether omitting tariffs from the gravity estimations or not seems to be sensitive to the data on the trade flows. One can perform these regressions on simulated data to have a confirmation. Also, it depends on which specification one believes: if the multilateral resistance terms are time-varying, then one does not need to worry about the results mentioned above; in this case (columns three and four of all the tables), having tariffs in the gravity equation is an advantage for the bias-reduction on the FTA coefficient. I believe that further work to argue this claim can be done by improving the tariff-data quality.

References

(2015): “Tariffs: World Integrated Trade Solutions-World Bank,” http://wits.worldbank.org/wits/wits/witshelp/Content/Data_Retrieval/P/Intro/C2.Types_of_Tariffs.htm.

(2015): “World Integrated Trade Solutions-World Bank,” http://wits.worldbank.org/default.aspx.

Anderson, J. E. (1979): “A theoretical foundation for the gravity equation,” *The American Economic Review*, 69, 106–116.

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

——— (2004): “Trade costs,” Working Paper 10480, National Bureau of Economic Research.

Baier, S. L. and J. H. Bergstrand (2007): “Do free trade agreements actually increase members’ international trade?” *Journal of international Economics*, 71, 72–95.

——— (2009): “Estimating the effects of free trade agreements on international trade flows using matching econometrics,” *Journal of International Economics*, 77, 63–76.

Bergstrand, J. H. (1985): “The gravity equation in international trade: some microeconomic foundations and empirical evidence,” *The Review of Economics and Statistics*, 67, 474–481.

Egger, P. (2000): “A note on the proper econometric specification of the gravity equation,” *Economics Letters*, 66, 25–31.

Evans, C. L. (2003): “The economic significance of national border effects,” *The American Economic Review*, 93, 1291–1312.

Evenett, S. J. and W. Keller (2002): “On theories explaining the success of the gravity equation,” *Journal of Political Economy*, 110, 282–316.

Feenstra, R. C. (1994): “New product varieties and the measurement of international prices,” *The American Economic Review*, 84, 157–177.

Head, K. and T. Mayer (2013): “Gravity equations: Workhorse, toolkit, and cookbook,” Discussion Paper DP9322, Center for Economic Policy and Research.

Helpman, E. and P. R. Krugman (1985): *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*, MIT press.

Helpman, E., M. Melitz, and Y. Rubinstein (2007): “Estimating trade flows: Trading partners and trading volumes,” Tech. rep., National Bureau of Economic Research.

Krugman, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of international Economics*, 9, 469–479.

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

Tinbergen, J. (1962): *Shaping the world economy*, New York: Twentieth Century Fund.

Winkelmann, R. (2003): *Econometric analysis of count data*, Springer Science & Business Media.

WTO (2015): "List of Regional Trade Agreements," http://rtais.wto.org/UI/PublicAllRTAList.aspx.

A Appendix

List of the countries considered in the panel:

Region	Country
Asia	China
Asia	Japan
South America	Argentina
South America	Brazil
North America	Canada
North America	Mexico
North America	United States
Europe	France
Europe	Germany
Europe	Iceland
Europe	Norway
Europe	Portugal
Europe	Spain
Europe	United Kingdom
Pacific	New Zealand
Pacific	Australia