The Geographic Space in International Trade: from Gravity to New Economic Geography

Cafiso, Gianluca

Univeristy of Catania, Economics Department

September 2007

The Geographic Space in International Trade: from Gravity to New Economic Geography

Gianluca Cafiso[*]

Abstract

In this paper we discuss the foundations of two recent trade theories linked by the role that the space-dimension plays in this kind of models. The theories discussed are the Gravity Approach and the New Economic Geography. We dedicate much to the explanation of the micro-foundations of the gravity equation and to the solution of the Border Puzzle achieved in a relevant and innovative paper by Anderson and van Wincoop. Some up-to-date empirical applications, which test or use Gravity and NEG relations, are discussed in order to show how much these two theories are used in empirical trade analysis.

JEL codes: F10
Keywords: Gravity, Trade Costs, Border Puzzle, Economic Geography.

September 2007

[*] University of Catania, Economics Department, 95100 Catania – Italy. E-mail: gcafiso@unict.it

Index

I. Introduction.

The space-dimension had not played any role in Trade Theory for a long time. Indeed the Classical Theory does not account for it at all, as it focuses on endowment and technological differences. Recently, new interest on distance and trade costs has stimulated new research whose results contributed to the formation of the Gravity Approach and the development of New Economic Geography .

In this paper we discuss how the space-dimension is taken into account by Gravity Models, and study the relation between Gravity and New Economic Geography (NEG). This will take us to consider the mechanics of gravity models, their empirical applications, the issue of their micro-foundations, and the foundations of NEG. Among the different alternative theoretical models that could serve for the micro-foundations, we focus on the Increasing Returns of Scale/ Monopolistic Competition (IRS/ MC) model. We choose the IRS/ MC model for two reasons: firstly, because this model appears to be the most used theoretical framework to micro-found gravity models; secondly, because this framework is the same used in NEG and it itself links Gravity to NEG.

We investigate the relation between NEG and Gravity to demonstrate that they develop the same basic intuition in two different directions, and to evaluate the utility of a model where elements of Gravity and NEG are included. Such basic intuition is that the space dimension shapes trade exchanges by means of trade costs. Gravity models can be very different both in terms of theoretical micro-foundations and analytical complexity. Still, conceptually they can be clustered in two groups. The first consists of micro-unfounded gravity equations similar to the original gravity equation (these are considered in chapter II). The second consists of micro-founded gravity equations in the manner of Anderson (1979) (which we consider in chapter III and IV).

In the first four chapters of this paper, we discuss exclusively the gravity approach in a way which is almost chronological. We will review some important papers about gravity estimations in order to show the role played by trade restrictions, how this approach is empirically implemented, and how gravity models have been improved to address empirical puzzles. The last chapter is devoted to NEG. There, we will derive its fundamental relations and we will discuss the relation between NEG and Gravity.

II. The Concept of Gravity and Trade Costs.

A natural incentive for trade between two countries is their nearness. If a country can purchase the same product (in terms of utility maximization) from different sources, we reasonably expect that it purchases the product from the nearest seller because this minimizes proportional-to-distance transportation costs. This is the *Concept of Gravity*: nearness facilitates trade relations because it reduces trade costs.[2] In the first section of this chapter, we discuss how the concept of gravity is embedded into a gravity equation. In the second section, we will argue about trade costs and how they are included into gravity-based relations.[3] We consider only micro-unfounded gravity equations in this chapter.

II.A. The Simplest Gravity Equation.

In its simplest form, the gravity approach is such a natural relation for international trade that Frankel (1998) affirms that it can only be attributed to Isaac Newton (A. 1642 - Ω. 1727). Indeed Tinbergen (1962) simply uses Newton's formula of *Universal Gravitation*, which entails that bilateral trade between two countries is directly proportional to their *size* and inversely proportional to their *distance*. If we suppose that country j's amount of export to i (X_{ij}) is equal to i's amount of import from j (M_{ij}), the gravity equation used by Tinbergen is: $M_{ij} = X_{ij} = \Lambda \frac{Y_i Y_j}{D_{ij}}$, (1)

where Y_i and Y_j are respectively country i and j's GDP, D_{ij} is the distance between country i and j, and Λ is a parameter . Rewritten in logs for estimation, eq.(1) becomes:

$$m_{ij} = \lambda + \alpha * y_i + \beta * y_j + \gamma * d_{ij} + \varepsilon_{ij},$$

where the small letters represent the log-value of the variables in capital letters in eq.(1) and ε_{ij} is the stochastic error. This is easily estimated through the Ordinary Least Squares (OLS) estimator, where the theory-consistent value of the parameters α, β and γ is 1.[4]

Even if eq. (1) fits the data well, it takes into account only few variables. The analysis can be enriched by including other explanatory variables in eq. (1) in order to measure their effect on bilateral trade. For instance, qualitative variables have an important effect on trade that is not uniform across country-pairs. In the next section we discuss how their effect on trade is interpreted in terms of *trade costs*.[5]

II.B. Trade Costs.

[2] The concept of gravity is clearly embedded in trade policies. For instance, Regional Trade Agreements are always established among countries near to one another in order to strengthen their trade relations, which are likely to be already high on the mere basis of their closeness.

[3] In this paper we mean for gravity equation: a theoretically unfounded relation; for gravity model: a theoretical framework from which is derived a theoretically founded gravity relation; for gravity approach: a method to describe trade exchanges through gravity relations both theoretically founded and not.

[4] Silva and Tenreyro (2006) affirm that the use of OLS yields inconsistent estimates when the relation is log-linear and heteroskedasticity occurs; this issue is discussed in the *third paper* of this thesis.

[5] If we consider as qualitative variable the presence of a common language between trading partners, we can check how much a common language boosts trade. Indeed, we expect that language diversity decreases bilateral trade because it makes trade more costly.

The category of Trade Costs is very large since it includes all the costs borne to transfer a good to its final user but the marginal cost of producing the good itself. Those are: transportation costs (both freight costs and time costs), policy barriers (tariffs and non-tariff barriers), information costs, contract enforcement costs, legal and regulatory costs, and local distribution costs (wholesale and retail).[6] Measures of trade costs can be direct or indirect. Direct measures (reported in published data) come from two major categories: costs imposed by policy (tariffs, quotas and the like), and costs imposed by the environment (transportation, insurances, translation costs, time costs). Indirect measures are those obtained through inference from quantity (gravity equations) and inference from prices.[7]

Measuring directly trade costs is difficult and often inaccurate since the complexity of a measure-at-the-source approach. Indeed different countries may adopt different definitions of trade costs, and generally their sources are so varied that it is impossible to distinguish each source of trade cost which is added to the other in causing the final import cost. So it may be easier to infer a specific trade cost by estimating a gravity equation in which a trade costs term is included. We can infer trade costs due to quantitative variables (such as the distance between trading partners) as well as qualitative variables (such as the effect of currency unions -Rose and van Wincoop 2001 or Frankel and Rose 2002- language links and ex colonial relationship -Melitz 2002- regional trade agreements -Frankel 1998-, etc.).

Anderson and van Wincoop (2004) propose a bilateral trade costs term (T_{ij}) which is a log-linear function of M observables Z_{ij}^{m}:[8]

$$T_{ij} = \prod_{m=1}^{M} (Z_{ij}^{m})^{\rho m} , \qquad (2)$$

[6] Anderson and van Wincoop (2004) reckon that on the whole trade costs amount to a representative 170% tax equivalent for industrialized countries, of which 55% local distribution costs and 74% international trade costs. The latter is constituted by 21% transportation costs and 44%border related trade barriers. Specifically, 1.70 = [(1.55 x 1.44 x 1.21)-1]. The border related trade barrier can be decomposed in: 8% policy related barrier (based on direct evidence from Tariffs and Non Tariff Barriers), a 7% language barrier, 14% currency barrier, a 6%information costs barrier, and a 3% security barrier.

[7] The most extensive source of panel data on policy barriers to trade is the Trade Analysis & Information System (TRAINS) of the United Nations Conference on Trade and Development. From a scanning of the panel, tariffs emerge to be low among most developed countries (under 5%), while developing countries continue to have higher tariff barriers (mostly over 10%). Evidence for NTBs shows that these are basically price and quality control measures, whose use is concentrated in few sectors in most economies. TRAINS reports sectoral NTBs coverage ratios for U.S., E.U., Japan and Canada for 1999. NTBs are widely used by developed countries in food products, textiles/ apparel, wood and wood products, and in some other areas of manufacturing. The products involved are quite significant in the trade of developing countries but also somewhat significant in the trade of developed countries with each other. Price comparison measures confirm the high and highly concentrated nature of NTBs in the agriculture sector. European and Japanese agriculture emerge as being more highly protected than U.S. and Canadian one.

[8] There is not concordance if the structure of the trade costs term should be multiplicative, as supposed, or additive as Hummels (2001) suggests.

where $(Z_{ij}^{m})^{\rho m}$ is equal to one plus the ad-valorem equivalent tax of the trade cost. Then we add T_{ij} (right hand side of eq.(2)) to the right hand side of eq.(1), so we obtain an augmented gravity equation in which a trade costs term is included and log-linearity is preserved:

$$M_{ij} = X_{ij} = \Lambda \frac{Y_i Y_j}{D_{ij}} \times \prod_{m=1}^{M} (Z_{ij}^{m})^{\rho m} \text{, in logs:} \quad (3)$$

$$m_{ij} = x_{ij} = \lambda + \alpha * y_i + \beta * y_j + \gamma * d_{ij} + \left[\rho_1 * z_{ij}^1 + ... + \rho_m * z_{ij}^m + ... + \rho_M * z_{ij}^M \right] + \varepsilon.$$

The dummy variable Z_{ij}^{m} is equal to 1 when (in a given pair) both countries belong to the same regional group, 0 otherwise. The estimate of its coefficient ($\hat{\rho}_m$) measures how much trade within each region can be attributed to the specific m^{th} regional effect.[9]

II.C. An Application of the Simplest Gravity Equation: Currency Unions boost Growth.

Frankel and Rose (2002) study if and in which way currency unions boost economic growth. According to economic theory, currency unions have a positive impact on growth because they represent an ultimate credible commitment to non inflationary monetary policies. Frankel and Rose (2002) refute this thesis proving that the effect of currency unions on growth works through trade: a currency union guaranties stability and ease trade relations among its members, the increase in trade makes the country to grow through a positive effect on both demand and supply side.

We discuss Frankel and Rose's paper because it is a good example of the potentialities of the gravity approach when one tries to quantify the effect of qualitative and quantitative variables upon trade. In Frankel and Rose's paper this is achieved through the extension of the Trade Costs term as indicated in eq. (3). They employ a two-step strategy to achieve this result: firstly, they measure the positive relation between currency unions and trade, secondly, they estimate the impact of trade on growth. Furthermore, they prove that there is not a direct effect of currency unions on growth on the mere basis of increased credibility.[10] We now present the main results in Frankel and Rose (2002) explaining how they are achieved in accordance with the Gravity Approach.

---

[9] Common coefficients can be imposed in the cost function. They would involve the constancy of an effect across different trading partners. For instance, the effect of membership in a custom union (or of speaking the same language) on trade may be assumed to be uniform for all its members.

[10] The econometrics used is OLS and Instrumental Variables Estimator (IVE); the IVE is used since endogeneity arises among the regressors. Indeed, whatever proxy used for trade, it is likely to be simultaneous with the growth variable. The data sample is a panel data for almost 8,000 country-pair observations at five years interval from 1970 through 1995.

Currency Unions increase Trade.

The effect of currency unions on trade is quantified through a gravity equation where the dependent variable is the logarithm of bilateral trade. Frankel and Rose's estimation output is reported in Table 3 in the Appendix, each column represents a gravity equation which includes a different number of regressors. Interest is focused on the coefficient of the "Currency Union" dummy variable which is around 1.6 across the different estimations. By considering the value in the second column, the estimated coefficient in level ($e^{1.22} = 3.38$) implies that the presence of a currency union increases trade flows by almost three and half times; in absolute terms, this seems a very high effect and recalls the border puzzle from McCallum (1995).[11]

Currency Unions boost Openness.

Usually, Regional Trade Agreements are assessed positively if they cause *trade creation* and not *trade diversion*. Indeed, trade diverted from a cheaper (non member of the block) towards a more expensive trade partner (member of the block) causes RTAs not to be beneficial. Frankel and Rose (2002) test this hypothesis through the dummy variable "Currency Union/ Non Union" (which is equal to 1 when just one of the two countries in the pair belongs to the union), their aim is to test if member countries' trade with non member countries worsens after the creation of the Currency Union, namely if trade diversion occurs. The coefficient of the dummy is 0.37 which means that Currency Unions do not cause trade diversion but increase trade between members and non-members of the 44%. Then, Currency Unions boost openness in general.

Trade increases National Income.

The final step in Frankel and Rose's analysis is to test the correlation between openness and GDP per capita growth. This involves a serious problem of simultaneity between the dependent variable and the Trade variable, then the Instrumental Variable Estimator is necessary. Frankel and Rose instrument the Trade variable with the gravity equation estimated at the first step. Indeed, variables such as Distance, Population, Common Border and Common Language are plausibly exogenous and highly correlated with trade so being good instrumental variables.[12] The equation which they estimate is:

---

[11] In McCallum (1995) trade flows among the Canadian provinces are 12 to 22 times more than trade between Canadian provinces and U.S. states; we will discuss the border puzzle afterwards. However, Frankel and Rose suggest that one reason for such a higher trade among the Canadian provinces is that they are part of a common federation which uses the same currency, while U.S.-Canada trade requires currency exchanges.

[12] They proceed sequentially. At first they estimate bilateral trade using the exogenous regressors provided by the gravity model, then they aggregate the exponential of fitted data across a country's trade partners to create a prediction of its overall trade. In the second stage, they use this predicted trade as an instrument for actual trade in the Output equation (4) to estimate the influence of trade on output.

$$\ln(Y/Pop)_{90,i} = \alpha_0 + \alpha_1 \log(Pop)_i + \alpha_2 \log(Aerea)_i + \alpha_3([X+M]/Y)_{90,i} + \alpha_4(I/Y)_i + \alpha_5 n_i + \alpha_6 School1_i + \alpha_7 Shool2_i + \alpha_8 \ln(Y/Pop)_{70,i} + u_i \quad (4)$$

where the dependent variable is the logarithm of country *i's* GDP per capita at the end of 1990 (measured in PPP), while the regressors are respectively: "Pop", country *i* 's population; "Area", a measure of country *i* 's size; "([X+M]/ Y)", openness to trade where *X* is aggregate exports and *M* is aggregate imports; "(I/ Y)", normalised investment; "*n*", population growth rate; "*School1* and *School2*", estimates of human capital investment based respectively on primary and secondary schooling enrolment rates; "$\left(Y/Pop\right)_{70,i}$", country *i* ' s GDP per capita at the end of 1970. The alphas denote coefficients and *u* is the disturb term, the coefficient of interest is α_3. We report Frankel and Rose's estimations of eq.(4) when different regressors are included in Table 4 in the Appendix. OLS estimates show that α_3 is positive, statistically significant and economically large whether or not controls are included (0.33 with, 0.79 without). A coefficient of 0.33 indicates that holding constant for 1970 income, income in 1990 was 0.3% higher for every 1% increase in the trade-GDP ratio.[13]

As shown by Frankel and Rose (2002), the gravity approach has great potentialities in terms of policy analysis even when the gravity equation considered is of the simplest form and interpretation. In the following sections we will show how the insights from gravity analysis are much more when economic theory is behind the gravity equation.

III. Micro-founded Gravity Models and the Identification issue.

Eq. (1) and eq. (3) are not linked to any trade theory.[14] Consequently, they do not explain but just quantify trade flows. Furthermore, we assumed them as true macroeconomic relations without a formal derivation through a utility maximization procedure; both are not micro-founded. These drawbacks cause equations such as (1) and (3) to be old-fashioned (micro-foundation issue) and useful only to describe bilateral trade but not to identify what causes it (identification issue). Hence, the Identification issue of the gravity equation is a concept different from its Micro-Foundation. Although, this difference is not relevant in practical terms because when we micro-found the gravity approach we derive a gravity relation from a theory which includes a motivation

[13] The IVE output is displayed in Table 4. The effect is respectively 0.33 with controls and 1.22 without. As regards Currency Unions, Frankel and Rose's results make clear in a very simple way that: i) currency unions stimulate trade, ii) it matters with whom one enters a currency union (results for this are not reported here), iii) the propulsive effect on trade relies on size, proximity and other linkages. The estimates show that every one percent increase in trade (relative to GDP) raises income per capita by roughly 1/ 3 of a per cent over 20-year period.

[14] For instance, they do no account for comparative advantages due to different productivities (Ricardian model) or different factor endowments (Heckscher-Ohlin model). Moreover, gravity equations describe bilateral trade between two countries, while neoclassical trade models study a country's relations with the rest of the world.

for trade. As a result, the gravity equation is contemporaneously micro-founded and identified. For this reason, hereinafter when we use the term Identification, we mean both the proper identification aspect (what causes the observed pattern of trade) and the micro-foundation one (derivation of a gravity relation through a utility maximization procedure).

The attempts to identify gravity equations have been many and reasonably successful. Probably the first micro-founded gravity relation is due to Anderson (1979), with further developments by Bergstrand (1985, 1989). In the following section we discuss the derivation of a gravity relation in a typical IRS/ MC setting. In the last section of this chapter we discuss a paper which tackles the identification problem from an empirical perspective.

III.A. Micro-Foundations.

Theory-based equations for bilateral trade are obtained in a wide class of models, in which the allocation of trade across countries may be analyzed separately from the allocation of production and consumption within countries. These models are said to be *trade separable*. Trade separability is built on the assumption of separable preferences and technology. The class of trade separable models yields gravity relations without any further assumption about what specific model accounts for the observed output structure and output allocation. Bilateral trade is determined in conditional general equilibrium whereby product markets for each good produced in each country clear conditional on the allocations. Three additional assumptions are made to derive gravity relations: CES preferences, an identical aggregator for each variety distinguished by country of origin, and ad-valorem equivalents of trade costs not depending upon the quantity of trade.[15]

We discuss the micro-foundation of the gravity approach on the basis of the IRS/ MC model by considering demand-side and supply-side micro-foundations separately (Harrington 2001).[16]

Supply–side micro-foundations concern the conditions characterizing the market structure and firms' output decisions on the base of the maximization of their profit function. Those are:

i) *Perfect Specialization*. Every country specializes in the production of some varieties of a good. Varieties are partially substitutable and trade is of the intra-industry kind (Armington's assumption, 1969).

[15] The assumption of products differentiated symmetrically by country of origin has became associated with Armington (1969). Anderson (1979) uses it adopting a linear expenditure system, in which the preferences for a variety are assumed to be homothetic and uniform across importing countries.

[16] Even though the IRS/ MC model suits the gravity approach well, this does not mean that it is the only one. Indeed, Deardorff (1998) shows how to derive gravity relations from a Heckscher-Ohlin framework (where we recall that constant returns of scale and perfect competition hold). His aim is to show that the empirical success of the gravity approach does not necessarily support the IRS/ MC model more than a Heckscher-Ohlin based pattern of trade. In the next section we discuss Evenett and Keller (2002) where the identification problem is tackled from an empirical perspective.

ii) *Increasing Returns of Scale*. In the New Trade Theory this is a result of Helpman and Krugman's (1985) IRS/ MC model who formalize it for the zero transport costs case. Firms produce the selected variety at an average cost which is decreasing.

iii) *Monopolistic Competition*. Firms enjoy market power on the produced variety since consumers are assumed to have Dixit-Stiglitz (1977) *love of variety* preferences. Then, firms gain a mark-up over the marginal cost of production that allows them to achieve monopolistic profits, if and only if, a no-entry condition is imposed.

Demand-side micro-foundation derives from the maximization of consumers' utility function, which indicates the optimum amount of import that a country should purchase. We assume that country *i's* consumers have homothetic preferences of the Dixit-Stiglitz kind (so we can focus on the representative agent) and demand varieties of the differentiated good.[17] The gravity relation is derived from the intra-temporal maximization of the following CES utility function:

$$U_i^k = \sum_{j=1}^{C} \beta_j^{1/\theta} \left(c_{ij}^k\right)^{(\theta-1)/\theta}, \qquad (5)$$

where country *i's* utility derives from imported consumption of the k^{th} variety from the rest of the world (C is the number of countries). Utility maximization is bounded to the following budget constraint:

$$Y_i^k = \sum_{j=1}^{C} p_{ij}^k c_{ij}^k \Leftrightarrow p_i^k y_i^k = \sum_{j=1}^{C} p_j^k T_{ij} c_{ij}^k, \text{ with trade costs.} \qquad (6)$$

Furthermore, a market clearing condition for any k^{th} variety is imposed:

$$Y_i^k = \sum_j X_{ij}^k . \qquad (7)$$

We continue our analysis only for a one sector economy. Consequently, we omit the superscripts *k* from now on. We wrote eq.(6) and the following ones with and without trade costs, the reason of this appears clearly afterwards.

In eq. (5) c_{ij} is country *i* 's imported consumption from *j*, β is a parameter different across countries (that we can set equal to country *j*'s relative size $[Y_j / Y_w]$), and θ is the intra-temporal elasticity of substitution between goods (which is constant given a CES utility function). If $\theta > 1$, preferences are biased in favour of home consumption. Prices are equal across varieties but not across exporters. This implies that every variety has got the same intrinsic utility. Then, consumers do not prefer a variety to another on the mere base of its specific characteristics, but only on the base of which country produces that variety. Indeed, the import price is source-dependent because

[17] An approach which takes to similar results is the *Ideal Variety Approach* by Lancaster (1980).

trade costs are different for different sellers. Since we sum across countries, the budget constrain (6) is a General Equilibrium condition which implies that every variety market is cleared.
By maximizing eq. (5) with respect to country *j,* we derive country *i* 's optimal amount of import from *j*:

$$M_{ij}=\left(\frac{p_{ij}}{P_i}\right)^{1-\theta}\frac{Y_iY_j}{Y_w}\Leftrightarrow M_{ij}=\left(\frac{T_{ij}p_j}{P_i}\right)^{1-\theta}\frac{Y_iY_j}{Y_w}\text{, with trade costs,}\qquad(8)$$

where:

$$P_i=\left(\sum_{j=1}^{N}\beta_j p_{ij}^{1-\theta}\right)^{\frac{1}{1-\theta}}\Leftrightarrow P_i=\left(\sum_{j=1}^{N}\beta_j (T_{ij}p_j)^{1-\theta}\right)^{\frac{1}{1-\theta}}\text{, with trade costs.}\qquad(9)$$

The variables p_{ij} and P_i play an important role in eq. (8) and make it different and richer than its simplest version (1). Indeed, we stress that eq. (8) belongs to a different kind of gravity equations which are derived from the solution of a theoretical model, and not just assumed as true macroeconomic relations.
p_{ij} is the exporter-dependent import price which relies on the presence of trade costs. Trade costs cause a gap between the import and the export price ($P_{imp}>P_{exp}$); if trade costs are ruled out from the model, utility maximization takes to a gravity equation similar to eq. (1). P_i is country *i* 's overall price index which has a substitution effect into the structure of the preferences. If trade costs faced by *i* are high on average (P_i is large), then the specific trade costs paid by *i* to import from *j* will be weighted less, and it will import more from *j* than from more highly weighted countries. This enforces the concept of gravity in this model: a country imports more from its neighbours than from farther countries because this reduces trade costs.
The import price is $p_{ij}=T_{ij}p_j$, where p_j is the price set by the exporter and $T_{ij}>1$. Trade costs (T_{ij}) are assumed to be *Iceberg kind of Costs*. Then to import a unit of good from *j* to *i*, *T* units of good must leave country *j* since *T-1* melts away in the shipping; *T-1* is the *ad valorem* tax equivalent of trade costs.[18] When $T_{ij}=1$, trade costs have not restrictive effects because trade is *domestic trade* (for instance, the cost of exchanging different currencies is zero between countries which belong to the same currency union).
The estimation of theory-founded gravity relations (such as eq. (8)) is not straightforward due to the presence of the price index and the price term. As regards eq. (8), Bergstrand (1985, 1989) estimates it directly by using statistical proxies (GDP deflator) for the variable P_j and p_i, and by

[18] If supply is monopolistic, the export price will be the sum of the marginal product cost and the mark-up. As long as the mark-up is invariant over destinations, T_{ij} contains only trade costs. Otherwise the tax equivalent must be interpreted to contain mark-ups.

breaking the trade costs term into tariffs and transport costs. Anderson and van Wincoop (2004) criticize Bergstrand's strategy. They gauge published data to be inaccurate because evidence from observed prices does not reflect distortions in the commodity markets.[19] To overcome this problem, Feenstra (2004) estimates these trade costs through a regression of the kind:

$$\log(T_{ij}) = \tau_{ij} + \rho \log D_{ij} + \varepsilon_{ij} \ , \qquad (10)$$

This is problematic because, by substituting eq.(10) into (8), the dependent variable in eq. (8) depends upon $(1-\theta)\times\tau_{ij}$ and $(1-\theta)\times\rho$. Consequently, a non linear estimator is required but it is not likely to outperform that one which uses published data in the manner of Bergstrand. Moreover, eq. (8) explains why countries close to each other trade more, but it does not explain why trade between equally distant country-pairs can be different. For example, the pair UK-Greece trades much less than the pair New Zealand-Australia despite their distance is the same. Then, the model needs to be improved. In the section dedicated to Anderson and van Wincoop's (2003) paper, we discuss both how the simple assumption of symmetric costs makes simpler the estimation of theory-founded gravity equations and how the model can be improved to differentiate across equally distant pairs.

III.B. Further Considerations about the Identification Issue.

Trade economists have demonstrated that gravity relations can be derived from many theoretical trade models, in which the causes of trade are not IRS and monopolistic competition. Indeed, recent studies by Deardorff (1998), and Feenstra, Markusen, and Rose (1998) show how to derive gravity relations from theoretical frameworks where the causes of trade are endowment/ technological differences or strategic market policies. This complicates the explanation of trade patterns successfully described by gravity relations. In fact, if they could be derived only from the IRS/ MC model, their good empirical performance would be a clear sign in support of that model. An alternative strategy consists in finding a correspondence between empirical results and theoretical underpinnings in order to find out what theory supports the success of the estimated gravity relation in a specific sample of data. Here we discuss the paper by Evenett and Keller (2002) which uses this approach.[20]

---

[19] For example, distortions due to non-tradables, or local taxes and subsides, affect the computation of the theoretical price indexes.

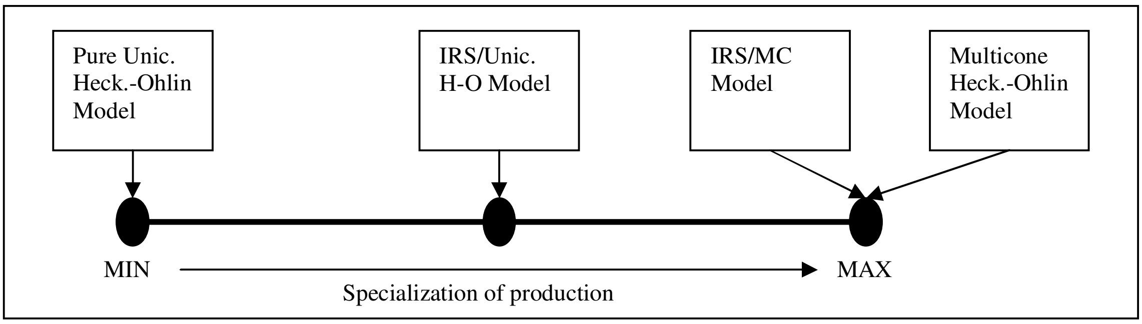
[20] The sample approach dates far back in the past. An example is Helpman (1987) who finds that the theoretical implications of the IRS model are consistent with data of the OECD countries, for which trade is mainly of the intra-industry kind. However, later on Hummel and Levinsohn (1995) repeated Helpman's analysis with a set of non OECD countries whose bilateral trade was not expected to be IRS-base trade, but the correlations found by Helpman held for the group of non OECD countries as well.

Then empirical results seemed to be misleading, and even the idea of a sample to sample reconciliation among the theoretical models, from which gravity equations may be derived, appeared not conclusive. Although it is worth to point out that this kind of papers rely deeply on the data used, then results may be biased due to a misunderstanding of the

Evenett and Keller designate four theoretical models which can explain trade flows well-described through gravity equations: two models of *perfect specialization* and, innovatively, two of *imperfect specialization*. The two models of perfect specialization are the *Multicone Heckscher-Ohlin model* and the *IRS/MC model*. In the former, factor endowment differences are supposed to be large and the assumption of constant return of scale (CRS) holds. This model explains a kind of trade where products traded differ in their factor requirements, therefore inter-industry trade is expected. In the latter, there is no account for factor endowment differences and the assumption of Increasing Returns of Scale (IRS) holds. This model explains a kind of trade where countries specialize in the production of varieties, therefore intra-industry trade is expected.[21]

The two models of imperfect specialization are the *IRS/Unicone Heckscher-Ohlin model* and the *Pure Unicone Heckscher-Ohlin model*. In both models the degree of specialization is a function of relative factor abundance, a key exogenous variable. Imagining these four models as points on a line, the more a model predicts specialization of production, the more it is located rightwards. Then, the models which predicts no specialization at all lie on the left hand-side extreme, while the models which predict complete specialization lie on the right hand side; see Figure 1. We highlight that specialization of production determines trade to be either of the Intra-Industry or of the Inter-Industry kind.

Figure 1



Evenett and Keller's procedure consists in writing a gravity equation conform to each of the four theoretical models. After that, they estimate all the different equations on the same data sample and assess which one fits better the data. A description of the models tested by Evenett and Keller follows:

---

data or their bad quality. This is why the contradiction emerging from Helpman (1987), and Hummel and Levinsohn (1995) has not discredited this approach.

[21] We point out that the gravity equations used by Evenett and Keller are similar to gravity equation (1), and not to equation (7). However, we decided to insert this topic in this section because we treat the micro-foundations on the basis of the IRS/ MC model here and not elsewhere. In fact, we discuss Evenett and Keller's paper for their contribution to a better understanding of how gravity models can be used to find the cause of observed trade patterns.

A) Perfect Specialization of Production: IRS/MC model (Helpman and Krugman 1985) - M_S.

The model is a two countries (*i* and *j*), two products (X and Z produced in differentiated varieties) model where perfect specialization of production and IRS hold. The gravity equation for this model is the simplest one:

$$M_{ij} = \frac{Y_i Y_j}{Y_w}. \quad (11)$$

B) Perfect Specialization of Production: Multicone Heckscher-Ohlin model - M_S.

Evenett and Keller affirm that eq. (11) can be derived even from a Heckscher-Ohlin framework when factor endowment differences are so large that countries' relative endowments lie outside the cone of diversification.[22] Then, when they estimate eq. (11) over a data sample of countries which have large endowment differences, the success of the gravity equation accounts for the Heckscher-Ohlin model. But, if they estimate eq. (11) over a data sample of countries which have not relevant endowment differences, the success of the gravity equation accounts for the IRS/ MC model. In this way eq. (11) accounts either for the IRS/ MC or the Multicone H-O model.

C) Imperfect Specialization of Production: IRS/Unicone Heckscher-Ohlin model - M_{IH}.

In this theoretical framework, they suppose one sector Z producing a homogenous good under CRS, and a second sector X producing a differentiated good under IRS. The model is a two countries (*i* and *j*), two factors (K and L) model. The homogenous good Z is more labour-intensive, country *i* is capital abundant, and $\gamma_c = Z_c / (p_X X_c + Z_c)$ is the share of good Z in country *c*'s GDP. Given the assumptions, the Heckscher-Ohlin model predicts that country *i* exports only the capital intensive X varieties. Country *i* 's share of X varieties in GDP is equal to $(1-\gamma_i)$, so the amount of its production available for imports by country *j* is given by $(1-\gamma_i)Y_i$. Country *j* purchases the X varieties abroad according to its share in world GDP (Y_j / Y_w). Assuming balanced trade, this means that country *i*'s import from *j* are:

$$M_{ij} = (1-\gamma_i)\frac{Y_i Y_j}{Y_w}. \quad (12)$$

For any $\gamma_i > 0$, the amount of imports is smaller than when both goods are differentiated. As the share of the homogenous good on GDP declines, the predicted level of imports rises.

D) Imperfect Specialization of Production: Pure Unicone Heckscher-Ohlin model - M_H.

This is the most classical 2 x 2 x 2 Heckscher-Ohlin model (two homogenous goods -X and Z- are produced in both countries -*i* and *j*- under CRS) in which factor endowment differences are the

[22]The cone of diversification is a fundamental notion of the Heckscher-Ohlin model. In an Edgeworth's Box, it is the area representing all the possible combinations of country A and B's relative factor endowments which allow both countries not to specialize in the production of only one of the two products in the economy (more specifically, in the production of the good in which its relative abundant factor is used more intensively).

cause of trade. Country i is relatively capital abundant and good Z is relatively labour intensive. Country i's import is equal to $p_X[X_i - (Y_i / Y_w)X_w]$, where X_w is good X's world production. Then the gravity equation is:

$$M_{ij} = [(1-\gamma_i) - (1-\gamma_j)]\frac{Y_i Y_j}{Y_w} = (\gamma_i - \gamma_j)\frac{Y_i Y_j}{Y_w}, \tag{13}$$

when the capital-labour ratios of the two countries converge, so do γ_i and γ_j. When $\gamma_i = \gamma_j$ there is no trade as the Heckscher-Ohlin model predicts.

Ceteris paribus, the following inequalities hold $M_S > M_{IH} > M_H$ (11>12>13). This means that the volume of bilateral trade is higher when more specialization occurs. Evenett and Keller affirm that observed trade flows are unlikely determined uniquely by any of the four archetypal models discussed so far. However they expect that, under different circumstances (such as different degrees of product differentiation), observed trade is better described by one particular model than by the others.[23]

They use the Grubel and Lloyd (1975) index to indicate the extent of IRS-based trade (hereinafter, GL index), the GL index is equal to 1 when all trade is of the intra-industry kind.[24] If they consider intra-industry trade, the candidate models are the IRS/ MC and the Multicone Heckscher-Ohlin model. They expect that the IRS/ MC model accounts for the performance of the gravity equation in data samples with a high GL index (so indicating that a relatively large portion of bilateral trade is two-way trade in differentiated products). While the Multicone Heckscher-Ohlin accounts for in those data samples with a relatively low GL index.

The organization of the data sample is fundamental in Evenett and Keller (2002). An arbitrarily chosen critical value ($\overline{GL}$) splits the data in two sub-samples which are sorted in a bi-dimensional way: for different values of the GL index and for different levels of FEDs. The pairs for which $GL_{ij} \leq \overline{GL}$ belong to what is referred to as the *Low-GL Sample*, while the remaining observations are part of the *High-GL Sample* (first dimension). Within the high-GL sample they expect a substantial amount of trade based on product differentiation and IRS. Furthermore, they sort the observations within the high-GL sample in 5 classes according to different level of FEDs (second dimension). The higher FEDs are, the lower Intra-Industry trade is. Denoting by V the number of

---

[23] Consider cross-sections of country pairs with little specialization due to IRS, but where the degree of factor endowment differences (FEDs) increases across countries. If Heckscher-Ohlin forces cause the pattern of trade, they will expect more specialization in country-pairs where FEDs are larger than in country pairs where they are smaller. This allows them to identify the Heckscher-Ohlin motivation for specialization and the gravity prediction.

[24] Grubel-Lloyd index measures the extent of IRS-based trade over total trade between country i and j. Evenett and Keller compute the GL index for every country which experiences a positive amount of intra-industry trade. This is the case for 2,870 observations; Bolivia has the lowest average GL index (value of 0.0006), while the United Kingdom has got the highest (value of 0.1495).

classes in which the sorted observations are located, the sample is organized given $\overline{GL}$ =0.05 (2240 observations in the low-GL sample and 630 in the high) and V =5 ($1 \le v \le V = 5$). Differences in factor proportions are lowest for v =*1* (more credit for the IRS/ MC model) and highest for v =5 (more credit for the Multicone Heckscher-Ohlin model). The results of the estimation of the four archetypal models are:

a) Perfect Specialization Model: *IRS/MC (High-GL sample)*. Evenett and Keller's estimation shows that the IRS model substantially overpredicts the level of bilateral trade. This conclusion is drawn from the estimation of the respective gravity equation for each class v :

$$M_{ij}^{v} = \alpha^{v} \frac{Y_i^{v} Y_j^{v}}{Y_w} + \varepsilon_{ij}^{v} .$$

Indeed, the theory-consistent value of alpha is 1, while the estimated value ranges from 0.0116 to 0.139 (highest value obtained for v = 3). Estimating over the 630 observations all together (all high-GL sample and not just for one class v), they obtain an estimated alpha of 0.087. Since the estimated values of alpha are much lower than their theoretical value, they conclude that the theoretical model overpredicts the level of bilateral trade and that there is no evidence for the IRS/ MC model.

b) Perfect Specialization Model: *Multicone Heckscher-Ohlin (Low-GL sample)*. In this model all trade is in homogenous perfectly specialized products. The multicone Heckscher-Ohlin model grossly overpredicts the volume of bilateral trade by the same token discussed above. Evenett and Keller assert that "in summary the predictions of the Heckscher-Ohlin model are rejected by the data…and that models of perfect specialization do not appear to be important in explaining the success of the gravity equation", page 297.

c) Imperfect Specialization Model: *IRS/Unicone Heckscher-Ohlin (High-GL sample)*. When country *i* is capital abundant relative to *j,* the gravity equation is:

$M_{ij}^{v} = (1-\gamma_i^{v}) \frac{Y_i^{v} Y_j^{v}}{Y_w} + \varepsilon_{ij}^{v}$ and $M_{ij}^{v} = (1-\gamma_j^{v}) \frac{Y_i^{v} Y_j^{v}}{Y_w} + \varepsilon_{ij}^{v}$ when country *j* is capital abundant.

The estimate of $(1-\gamma_i^{v})$ quantifies the average size of the differentiated good sector. This value varies by class v from 0.053 to 0.128. Some support for this model emerges since the average value of the coefficient is not negative, and higher values of the GL index (and hence class v) are associated with higher estimates of the coefficient as expected. Indeed, the model predicts that a higher share of the differentiated good is associated with higher estimates of alpha.

d) Imperfect Specialization Model: *Unicone Heckscher-Ohlin (Low-GL sample).*

The model estimated is the Unicone Heckscher-Ohlin model with two homogenous goods and no specialization at all. When country *i* is capital abundant, the relevant equation for this model is:

$M_{ij}^{v} = (\gamma_j^v - \gamma_i^v)\frac{Y_i^v Y_j^v}{Y_w} + \varepsilon_{ij}^v$ and $M_{ij}^{v} = (\gamma_i^v - \gamma_j^v)\frac{Y_i^v Y_j^v}{Y_w} + \varepsilon_{ij}^v$ when country j is capital abundant.

The median value of the estimated import parameter ($\alpha_v = [\gamma_i^v - \gamma_j^v]$) can be interpreted as the difference between the size of the labour-intensive sector respectively in the labour and in the capital abundant country. From the theory we expect that α_v is positive and that it tends to rise when bilateral differences in factor proportions increase. The estimated values support this prediction.

Evenett and Keller conclude in favour of both models of incomplete specialization. Moreover, they demonstrate that the Unicone H-O model outperforms the IRS/ Unicone H-O by means of country-specific estimations of the imperfect specialization models.[25, 26] It is to notice that the two relations found (positive between increasing trade and increasing FEDs, inverse between relative size of the labour-intensive sector and relative capital endowment across countries) fit the North-South pattern of trade particularly well.

IV. The Border Puzzle.

In this chapter we discuss the Border Puzzle (or Home Bias Puzzle) in International Trade which originally emerged from the work by McCallum (1995). This will take us to consider the recent and relevant work by Anderson and van Wincoop (2003) who solve the puzzle through the estimation of a specific gravity equation. Their gravity equation is a simplification of eq. (7) derived under the assumption of symmetric trade costs, it belongs to the group of micro-founded gravity equations. In the last section of this chapter, we will discuss how Obstfeld and Rogoff explain the home bias in international trade by using the interaction between trade costs and the elasticity of substitution.

---

[25] We assumed that the differentiated good is relatively capital intensive, then the model implies that the share of the differentiated good in GDP increases with the relative abundance of capital to labour. This is tested checking if the correlation between $(1-\gamma_i)$ and (K_i / L_i) is positive. 32 out of 35 estimates satisfy the restriction of being between zero and one, but Evenett and Keller do not find the positive correlation between $(1-\gamma_i)$ and (K_i / L_i); this is a negative result against a complete support of the Heckscher-Ohlin model.

[26] Evenett and Keller point out that their results and conclusions rely deeply on the quality of the data used . They gauge this as a weakness of their work which is more likely to affect the results for the low-GL sample. Indeed, the low-GL sample includes all the under-developed or developing countries (countries with a low level of intra-industry trade) whose statistics are not as trustworthy as those of the OECD countries (which are all in the high-GL sample).
Moreover, Evenett and Keller cast some doubts about the failure of the IRS/ Unicone Heckscher-Ohlin model. They affirm that a world where all countries occupy the same cone of diversification is unlikely, and conclude that the rejection of a model in which each country occupies a distinct cone of specification does not mean that perfect specialization due to FEDs is irrelevant . In fact, "it leaves open the possibility of a relatively small number of, say, three of four diversification cones", Evenett and Keller (2002) page 311. In a comment to Evenett and Keller, Feenstra (2004) affirms that their results depend on having just two countries, otherwise the Heckscher-Ohlin model makes no prediction at all about bilateral trade flows.

IV.A. The Border Puzzle in International Trade.

McCallum (1995) studies domestic trade among the Canadian provinces compared to international trade between the Canadian provinces and the U.S. states. He wants to assess the effect of the border on a region (North America) which is supposed to be very integrated. He uses an extremely simple gravity equation -eq. (1) plus a trade term- with intra-provinces trade data and international trade data between provinces and states dated 1988.[27] McCallum's equation is:

$$x_{ij} = a + b * y_i + c * y_j + d * dist_{ij} + e * DUMMY_{ij} + u_{ij} . \quad (14)$$

The dependent variable is the logarithm of exports from region *i* to region *j*, the regressors are respectively: region *i* and *j's* logarithm of GDP, the distance between province/ state and province capital cities, the National Trade dummy variable (equal to 1 for inter-provinces trade and 0 for province to state trade) and an error term. The estimated parameters for the 1988 and for other years are reported in Table 1. The most striking value is the high coefficient of the *Border Effect* for Canada (column 1), which predicts that trade flows among the Canadian provinces are 22 times higher than those between the U.S. states and the Canadian provinces. This unexpected high value (which captures all the factors restricting trade between the US and Canada, what we call border effect*)* gave birth to the issue of the *Border Puzzle* in International Trade. Indeed, the magnitude of this effect seems to be excessive and not to reflect the reality of the US-Canadian exchanges.[28]

Their main criticisms are: first, the estimation from the Canadian perspective tends to overestimate the effect of the border because borders have an asymmetric effect on countries of different size (particularly large on the side of small countries); second, the use of a theoretically unfounded equation (such as McCallum's) produces biased estimations due to omitted variables. In the next section we discuss the gravity equation used by Anderson and van Wincoop (2003) to solve the Border Puzzle.

Table 1 - Comparison of gravity equations.

Source: Feenstra (2004), Anderson and van Wincoop (2003).	McCallum's equation for the original and other samples.			Anderson van-Wincoop
	1	**2**	**3**	**4**
Year of data Sample	1988 original	1993	1993	1993
Regions included	CA-CA CA-US	CA-CA CA-US	CA-CA CA-US US-US	CA-CA CA-US US-US
ln Y_i	1.21	1.22	1.13	1
ln Y_j	1.06	0.98	0.97	1
ln d_{ij}	-1.42	-1.35	-1.11	-0.79
Indicator Canada	3.09	2.80	2.75	
Indicator US			0.4	

[27] This is the last year before the enforcement of the Free Trade Agreement between Canada and the U.S.A. which was effective from the 1st of January 1989.

[28] The term puzzle is used when the prescriptions of a certain theory are not supported by empirical evidence, but McCallum's paper is merely descriptive and not devoted to assess the validity of any theory. Indeed, McCallum simply concludes affirming that national borders still matter a lot even in highly integrated regions such as North America.

Indicator border				-1.65
Border Effect Canada	22.0 (=exp3.09)	16.4 (=exp2.8)	15.7 (=exp2.75)	10.5 (see eq. 17)
Border Effect US			1.5 (=exp0.4)	2.6 (see eq. 17)
Border Effect Average			4.8	5.2
R^2	0.81	0.76	0.85	
Observations	683	679	1511	1511

Column 1: McCallum's equation for the original sample.
Column 2: McCallum's equation for the 1993.
Column 3: McCallum's equation for the 1993 including data of the U.S. states.
Column 4: Anderson and van Wincoop's regression for the 1993 including data of the U.S. States.

IV.B. Anderson and van Wincoop's Gravity Equation.

Anderson and van Wincoop solve the puzzle by using a different gravity equation in which *Terms of Multilateral Resistance* are introduced as explanatory variables. The theoretical framework of their model is not different from what we discussed earlier in section III.B., but they use a slightly different utility function for consumers in region j who import consumption from i : [29]

$$U_j = \left(\sum_i \beta_i^{(1-\theta)/\theta} c_{ij}^{(\theta-1)/\theta}\right)^{\theta/(\theta-1)}, \tag{15}$$

where c_{ij} is region j's consumption from i , β_i is an undefined distribution parameter and θ is the elasticity of substitution. Utility maximization is bounded to the budget constraint:

$$Y_j = \sum_i p_{ij} c_{ij} , \tag{16}$$

where the import price is $p_{ij} = p_i T_{ij}$ (T_{ij} >1 is the trade costs term, p_i is the export price). Eq. (16) implies that the nominal value of exports from i to j (j's payments to i) can be written as $X_{ij} = p_{ij} c_{ij}$. The nominal demand of region i goods by region j 's consumers, which satisfies maximization of (15) subject to (16) , is:

$$X_{ij} = Y_j \left(\frac{\beta_i p_i T_{ij}}{P_j}\right)^{1-\theta}, \tag{17}$$

P_j is the consumer price index, it is:

$$(P_j)^{1-\theta} = \sum_i \left(\beta_i p_i T_{ij}\right)^{1-\theta}, \tag{18}$$

The general equilibrium structure of the model imposes market clearance:

$$Y_i = \sum_j X_{ij} = \sum_j \left(\beta_i p_i T_{ij} / P_j\right)^{1-\theta} Y_j = \left(\beta_i p_i\right)^{1-\theta} \sum_j \left(T_{ij} / P_j\right)^{1-\theta} Y_j, \qquad \forall i \tag{19}$$

To derive their gravity equation, Anderson and van Wincoop use the market clearance condition (19) to solve for the scaled prices $(\beta_i p_i)$ and substitute them in the demand equation (17). After having defined world nominal income as $Y_w \equiv \sum_j Y_j$ and national income share as $\varphi_j \equiv Y_j / Y_w$, they achieve the gravity equation:

$$X_{ij} = \left(\frac{T_{ij}}{G_i P_j}\right)^{1-\theta} \frac{Y_i Y_j}{Y_w}, \tag{20}$$

where $(G_i)^{1-\theta} = \sum_j \left(T_{ij} / P_j\right)^{1-\theta} \varphi_j$. (21)

Eq. (18) can be rewritten as:

[29] To derive eq. (5), we used a utility function for consumers in region i who import consumption from j. We change to be as close as possible to Anderson and van Wincoop (2003).

$$(P_j)^{1-\theta} = \sum_i \left(T_{ij} / G_i\right)^{1-\theta} \varphi_i \ . \tag{22}$$

The problem of eq. (20) is that its estimation requires G_i and P_j which are functions of the export prices. Instead of using an explicit solution for G_i and P_j (quite complicated to achieve), Anderson and van Wincoop employ an implicit solution achieved under the assumption of symmetric trade barriers ($T_{ij} = T_{ji}$). Then they demonstrate that a solution to (21)-(22) is $G_i = P_i$, with:

$$\left(P_j\right)^{1-\theta} = \sum_i \varphi_i P_i^{\theta-1} T_{ij}^{1-\theta} \qquad \forall j. \tag{23}$$

Eq. (23) is an implicit solution of the price indexes as functions of bilateral trade barriers and income shares. Then the gravity equation becomes:

$$X_{ij} = \left(\frac{T_{ij}}{P_i P_j}\right)^{1-\theta} \frac{Y_i Y_j}{Y_w}, \tag{24}$$

which is Anderson-vanWincoop Gravity Equation for balanced bilateral trade. It is a simplification of the gravity equation (8) and (20) because it can be directly estimated, given the implicit solutions for the indexes P_i and P_j. In section IV.C. we will explain the estimation of eq. (24) in more details.

P_i and P_j are named *Indexes of Multilateral Resistance* because they measure the overall restrictions to trade depending on trade costs; a general raise of trade barriers will increase the index for every country.[30] P_j has the same effect as P_i on imports from i to j. If j is on average faraway from its trading partners (high P_j) as well as i (high P_i) but i and j are close to each other (low T_{ij}), bilateral trade will be relatively high.[31] On the contrary, if bilateral trade barriers are high (high T_{ij}), ceteris paribus, the amount of bilateral trade will be low. This suggests that countries which are faraway from all their potential partners tend to be autarkic.

Indeed, trade between two countries depends upon relative trade costs (defined as the ratio of bilateral trade costs to multilateral average trade costs). If two countries are relatively close with respect to any other country, their nearness is an incentive (relative lower costs) to trade with each other. This explains why trade between Australia and New Zealand is higher than trade between UK and Greece.

From eq. (20)-(24), Anderson and van Wincoop (2003) draw some fundamental conclusions about bilateral trade which provide an indirect explanation of the border puzzle. Those conclusions are:

1) Trade barriers reduce size-adjusted trade more between two big countries than between two small countries. This happens because big countries are large markets themselves, so they are less

[30] Leamer and Stern (1970) note that bilateral trade is indeterminate in the absence of trade costs. In that case, they suggest assuming that countries essentially draw their trading partners out of a hat, according to various probabilities.

[31] Note that the bracketed term is small. But to a negative power (θ >1) the lower it is, the higher bilateral trade between i and j gets.

dependent on foreign demand. When trade barriers increase, big countries succeed to divert their production towards their domestic market more easily than small countries do.

2) Trade barriers raise size-adjusted trade more within two small countries than within two big countries. A rise in multilateral resistance causes a drop in relative resistance for international trade. The drop is larger for small countries because they bear a bigger increase in multilateral resistance.

3) Trade barriers raise the ratio of size-adjusted trade within country 1 relative to size-adjusted trade between country 1 and 2 by more the smaller is country 1 and the larger is country 2.[32, 33]

IV.C. Solution of the Border Puzzle.

As we said in section IV.A., Anderson and van Wincoop deem McCallum's estimation of the border effect to be too high for two reasons:

1) McCallum's estimation is based on a regression with omitted variables: the multilateral resistance terms. By estimating McCallum's equation for the 1993 (McCallum estimates his equation for the 1988), Anderson and van Wincoop find a value of the border effect of 16.4 (column 2, Table 1). While using their asymptotically unbiased eq. (22) for the 1993, they find a value of the border effect of 10.5 instead of McCallum's 22 (column 4, Table 1).

2) McCallum's border effect is made greater by the small size of the Canadian economy. Indeed, McCallum estimates the border effect from the Canadian perspective (using data of the Canadian inter-provinces trade). Consequently the magnitude of the border effect appears excessively high when he computes the ratio of domestic to foreign trade. On the contrary, by estimating McCallum's regression with U.S. data (from the U.S. perspective), they find that trade among the

---

32 These conclusions are proved by means of a sensitive-analysis exercise. Given a uniform marginal increase in trade barriers across all countries ($dT_{ij} = dT$ for $i \neq j$), it is demonstrated that multilateral resistance rises more for a small country than for a large one. In fact, the multilateral resistance of the large country does not change a lot because an increase in trade barriers does not affect trade within its borders.

33 The following example makes this statement clearer:

1st case - Identical Countries: Consider a two-county world economy, where country A and B are identically constituted by 51 regions each. Any region is of the same size in terms of GDP in both countries. Under borderless trade, every region sells one unit of product to all the other 102 regions (including itself).

Now suppose that a border is erected between A and B, and that it reduces original international trade of the 20% (because every region in each country cuts its exports and sells more goods domestically). Assuming that markets are cleared in both countries, domestic trade rises by the same amount in both countries as well as the ratio of domestic over international trade. In the next case, we show that this increase is higher for the relative smaller country of the pair if country A and B are not identical.

2nd case - Asymmetric Countries: Consider a two-county world economy, where country A is small and constituted by 2 regions, while country B is big and constituted by 100 regions (this recalls the U.S.- Canada case). Under the same assumptions, if a border which reduces original international trade of the 20% is erected between A and B, region 1 in the small country will reduce its exports to the big country by 20 (it sells 10 more goods to itself and 10 more goods to region 2 in the same country).

Trade between the two regions in country A rises by a factor of 11, and it is now 13.75 times higher than international trade between country A and B. While trade among the 100 regions in country B is only 1.255 times higher than international trade (conclusion 3).

U.S. states is only 1.5 times larger than trade between the U.S. states and the Canadian provinces (column 3, Table 1). The values in column 3 show that the border effect for the U.S. states is 1.5, while that one for the Canadian provinces is 15.7.

In point 1 we mentioned the results from the estimation of eq. (24), now we discuss how Anderson and van Wincoop estimate that equation. First of all, Anderson and van Wincoop suppose that trade costs are a linear function of observables:

$T_{ij} = b_{ij} d_{ij}^{\rho}$,

where d_{ij} is bilateral distance and b_{ij} is equal to one if region i and j are located in the same country. Then the log-linear form of eq. (24) is:

$$\ln X_{ij} = k + \ln Y_i + \ln Y_j + (1-\theta)\rho \ln d_{ij} + (1-\theta)\ln b_{ij} - (1-\theta)\ln P_i - (1-\theta)\ln P_j . \qquad (25)$$

In this form, it is very close to McCallum's gravity equation except for the presence of the multilateral resistance terms (whose absence biases McCallum's estimation). Eq. (25) is rearranged for estimation:

$$\ln Z_{ij} = \ln(X_{ij} / Y_i Y_j) = k + \alpha \ln d_{ij} + \gamma(1-\delta_{ij}) + \ln(P_i)^{\theta-1} + \ln(P_j)^{\theta-1} + \varepsilon_{ij} , \qquad (26)$$

where $\alpha = (1-\theta)\rho$, $\gamma = (1-\theta)\ln b$, $b_{ij} = b^{1-\delta_{ij}}$ (so that *b-1* is the equivalent-tax of the U.S.-Canada border) and δ_{ij} is the dummy variable equal to 1 if i and j are in the same country, zero otherwise.
To estimate eq. (26) , we need to compute the multilateral resistance terms. As we said in section III.A., the estimation of a gravity equation which includes price indexes is problematic because an explicit solution for them is difficult to find. As abovementioned, an alternative technique is the use of statistical proxies (such as the GDP deflator or consumer price indexes), but Anderson and van Wincoop deem this practice incorrect. Thus, under the assumption of symmetric trade costs, they demonstrate that the multilateral resistance terms can be solved as an implicit function -eq. (23)- of observables (trade costs and income shares) and of the parameter α and γ . After having substituted such implicit solutions in eq. (26), their gravity equation is written as a function of observables and it can be estimated more easily. Nevertheless, they need to estimate a system of equations (eq. (26) and the implicit solutions for the price indexes) in which eq. (26) is a non log-linear function of the parameters.[34]

[34] Another way to estimate eq.(24) (which is the easiest to implement) consists in using *Fixed Effects*; Rose and van Wincoop (2001), Redding and Venables (2004), and Hummels (2001) use this technique. This is implemented computing the variation of trade when country i trades with j (which is our pair of interest) relative to when it trades with any other country. In the regression this is done through the use of two dummies: δ_1^i which denotes any indicator variable which is unity if country i is the exporter, and zero otherwise; δ_2^i which is unity if country j is the importer, and zero otherwise.

The border effect estimate is $\gamma = (1-\theta)\ln b =$ -1.65. To obtain an indicator of the border effect *in terms of price,* it is necessary to assign a value to the elasticity of substitution (θ). Estimates for different values of the elasticity of substitution are shown in Table 2.[35]

Table 2 – BE for different values of θ.

Indicator Border Effect = exp [$\hat{\gamma}$ /(1-θ)]			
If θ =	5	10	20
Border Effect	1.5	1.2	1.09

To turn the coefficient $\hat{\gamma} = 1.65$ into an estimate of the border effect *in terms of volumes of trade*, Anderson and van Winccop need a more complicated procedure than taking the exponent of $\hat{\gamma}$ (what we did with McCallum's estimation) because the multilateral resistance terms are influenced by the presence of the border through the term T_{ij}. Assuming $(\overline{P}_i)^{\theta-1}$ as the multilateral resistance term without the border effect (the denominator of eq. (27) where T_{ij} is only distance since we rule out the border effect), Anderson and van Wincoop demonstrate that the ratio of the exponent of eq. (26) with the border effect over the exponent of eq. (26) without the border effect is:

$$\frac{X_{ij}}{\overline{X}_{ij}} = e^{\hat{\gamma}(1-\delta_{ij})} \left[\frac{(P_i)^{\theta-1}(P_j)^{\theta-1}}{(\overline{P}_i)^{\theta-1}(\overline{P}_j)^{\theta-1}} \right]. \tag{27}$$

Eq. (27) provides us with an estimate of the border effect in terms of volumes of trade (ratio of trade within Canada over trade between Canada and U.S. with and without the border). For intra-Canada trade ($\delta^{ij} = 1$) the ratio is equal to 4.31 when θ = 5, then intra-Canada trade is 4.3 times larger with the border than without. For intra-U.S.A. trade, trade is 1.05 times larger with the border than without. This means that cross-border trade is 0.41 times smaller with the border effect than without (value in terms of absolute volumes, not from a national perspective). Furthermore, intra-Canada trade is 10.5 (= 4.3/ 0.41) times higher than cross-border trade. While intra-U.S.A. trade is 2.6 (=1.05/ 0.41) times higher than cross-border trade. The geometric mean of the border effect is 5.2 (=$[10.5*2.6]^{1/2}$); all these values are in Table 1 and taken by Anderson and vanWincoop (2003).

Feenstra (2004) compares the value of the coefficient of the border effect obtained with Anderson-van Wincoop's approach (implicit price indexes) and using fixed effects. The estimates are nearly the same even in terms of efficiency, but the use of fixed effects is much easier to implement and it produces consistent estimates of the average border effect. Feenstra concludes that, given its simplicity ,"it might be considered the preferred empirical method", page 162. A detailed discussion of this technique is in *the second paper* of this thesis.

[35] Anderson and van Wincoop estimate their model both in a two-country framework and in a multi-country one (namely, considering cross-sections data for other country-pairs and not just U.S.-Canada). We review only their results of the two-country estimations which allow a closer comparison to McCallum (1995).

Anderson and van Wincoop's results support the theoretical implications of their model (see conclusions 1, 2 and 3 in section IV.B.). On the whole, their paper seems to explain accurately and to solve the Border Puzzle since they obtain a much smaller (and conform to expectations) estimate of the border effect than McCallum's.

IV.D. Obstfeld and Rogoff 's Explanation of the Home Bias in International Trade.

Even though Anderson and van Wincoop solve the puzzle, a home bias for domestic consumption still emerges. Obstfeld and Rogoff (2000) provide a demand-side explanation of *The Home Bias Puzzle in Trade*. They focus on trade costs and their interaction with the elasticity of substitution (between foreign and domestic products) in order to show how trade costs skew consumption in favour of home produced goods. Obstfeld and Rogoff use a typical two-country endowment model in which the representative agent has to maximize his CES utility function:

$$C = \left(C_H^{\frac{\theta-1}{\theta}} + C_F^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}} . \qquad (28)$$

C_H is home consumption of the home produced good, C_F is home consumption of the foreign good, and θ is the elasticity of substitution. Preferences are homothetic internationally, then foreigners have got identical utility functions in C^*_H and C^*_F. They suppose iceberg kind of costs T_{HF} ($T_{HF} > 1$, $1 - T_{HF}$ melts away) and define P_H (P_F) as the home price of the home (foreign) good and P^*_H (P^*_F) as the corresponding foreign price. Obstfeld and Rogoff derive the ratio of home expenditure for home goods over home expenditure for imports (foreign expenditure for foreign goods over foreign expenditure for home goods):

$$\frac{C_H}{pC_F} = \frac{p^* C_F^*}{C_H^*} = (1 - T_{HF})^{1-\theta} , \qquad (29)$$

where $p = \frac{P_F}{P_H}$ and $p^* = \frac{P_F^*}{P_H^*}$.

If there were no trade costs ($T_{HF} = 0$), eq.(29) is equal to 1 and this is the case of no home bias in international trade. While if $T_{HF} = 0.25$ and $\theta = 6$, it is equal to 4.21, this means that the expenditure for domestic goods is much higher than for foreign goods in presence of trade costs (namely, there is home bias).[36] Furthermore, they show that the higher trade costs are, the greater the impact of a 1 percent reduction in their value on the home bias. This is shown through the elasticity of the home bias with respect to trade costs:

[36] Obstfeld and Rogoff argue that an elasticity of substitution equal to 6 is a reasonable value since it is found in many empirical studies.

$$\frac{d\ln(C_H / pC_F)}{d\ln T_{HF}} = \frac{T_{HF}}{1-T_{HF}}(\theta - 1)\,. \qquad (30)$$

For T_{HF} = 0.35 and θ = 6, it is 2.7; this means that when T_{HF} = 0.35 a 1 percent reduction of trade costs reduces the home bias by 2.7 percent. For T_{HF} = 0.25 and θ =6, it is equal to 1.67; then a 1 percent reduction of trade costs reduces the home bias by 1.67 percent. The elasticity (30) is directly proportional to θ, then a high constant elasticity of substitution can explain a high home bias even with relatively low trade costs. Trade flows highly sensitive to trade costs is the empirical result which we observe both in McCallum (1995) and in Anderson and van Wincoop (2003). In this way Obstfeld and Rogoff explain the home bias which emerges from gravity estimations of international trade flows.

From this insight into the nature of the preferences, we can say that there are two dimensions of home bias in the gravity model (see eq. (24)). The first operates through the constant elasticity of substitution term (θ>1), since the higher θ *is,* the higher the home bias gets as shown by Obstfeld and Rogoff. The second dimension operates through the Multilateral Resistance Terms because higher trade costs imply lower imports from every source. This implicitly makes consumption biased for home goods because countries tend to become autarkic in order to save on trade costs (we made this point clear at page 20). How these two dimensions interact and in what measure each of them contributes to determine the amount of foreign consumption is expressed in eq. (24).

V. Criticisms of the Gravity Approach.

A common assumption in the majority of macroeconomic models has always been zero trade costs, and the idea of a more and more integrated world charmed many economists and policy makers in the last decade. But the results coming from gravity estimations, that show simply and clearly the important role played by trade costs in shaping the modern economy, have moved many of them to reconsider the traditional assumptions of International Economics.[37] Nevertheless, the results of gravity studies have not always been accepted without resistance, the gravity approach has often received strong and theoretically founded criticisms. In the rest of this chapter we discuss some of the most important ones and their possible solutions.

Anderson and van Wincoop (2004) collect and summarize the main criticisms of the gravity approach. These are: i) estimates of the distance elasticity are unrealistically high and have not dropped over time (Grossman 1998, Coe et al. 2002) ; ii) there are non import competing sectors or non-tradable sectors that only supply to the domestic market (Engel 2002); iii) elasticity of

[37] Discussing the missing contribution of Gravity in rethinking International Economics, Leamer and Levinsohn (1995) affirm that one possible explanation is the lack of a relation between theory and empirical work. Indeed, gravity models are strictly descriptive and the lack of a theoretical underpinning, or the complexity to catch that one which suits best, does not allow turning the obtained description of trade flows into an explanation of why such patterns exist.

substitution between domestic and foreign goods should be different from the elasticity of substitution between domestic goods (Engel 2002); iv) in contrast to the predictions of the model, the substantial increase in U.S.-Canada trade during the 1990s was not accompanied by a big drop in intra-provincial trade (Helliwell 2003); v) the model implies trade among all countries for each sector, while the reality is dominated by zeros (Haveman and Hummels 2001); vi) estimated trade barriers are unrealistically high (Balistreri and Hillberry 2002); vii) estimates of the gravity equation have the unrealistic implication that consumer prices are much higher in Canada than in the U.S. (Balistreri and Hillberry 2002).

For a detailed explanation of the above mentioned criticisms which concern estimation techniques or empirical issues, we refer to Anderson and van Wincoop (2004). Here we discuss some possible modifications to the analytical framework of the model in order to tackle only some of the previous criticisms. Such possible modifications are:

1) Possible Improvements to the CES structure. Tchamourliysky (2002) blames homothetic preferences to make consumption too sensitive to trade barriers. He affirms that a non homothetic structure of the preferences would soften the effect of distance on bilateral trade, so providing a lower value of the estimated elasticity. He considers non homothetic preferences by modifying a standard CES consumption index, where he adds a constant δ^i to consumption of goods from country *i*. Thus he derives a different gravity equation, where the estimated value of δ^i is negative and it is interpreted as substance requirements which make trade flows less sensitive to trade barriers.

2) Different Preferences and Technology. In the gravity equations derived so far, preferences and technology are the same for all the agents. It is possible to relax this assumption in different ways without compromising the simplicity of the gravity approach. Indeed, it is possible to define specific preferences for different countries to allow a bias for home consumption or for consumption from selected sources. Nevertheless, the effect of a different structure of the preferences is not distinguishable from the effect of trade costs in terms of estimation output. An example of an augmented structure of the preferences is the following utility function:

$$U_j = \left(\sum_i \beta_i^{(1-\theta)/\theta} (c_{ij} / \gamma_{ij})^{(\theta-1)/\theta}\right)^{\theta/(1-\theta)},$$

where c_{ij} is consumption of goods produced in country *i* by country *j 's* consumers. Then the utility of the representative agent in country *j* differs across countries for different values of the γ_{ij} term.[38]

3) *Fixed Costs of Trade*. Some researchers believe that selling into a foreign market involves fixed costs. Consequently, firms exporting in a foreign market in the present are likely to continue

[38] According to Evan (2003), what really matters is the location of the firm producing the good and not its nationality.

exporting in that market in the future (Bernard and Jensen 1997, Bernard and Wagner 2001). Then, fixed costs (usually ignored in traditional gravity analysis) might explain the *missing trade* puzzle (critic v), because firms could limit themselves to not penetrate a foreign market that is not profitable given certain fixed costs (to wit, menu costs). By the same token, a reduction in fixed costs could explain trade growth.[39]

These are examples of how the traditional gravity model can be modified to deal with some criticisms. Some changes complicate considerably the gravity approach, but trade economists have demonstrated that the model remains manageable. A more complicated supply side of the model allows considering firms' operational decisions (as in the case of fixed costs which limit firms' profitability). The study of firms' operational decisions from the perspective of New Economic Geography (NEG) is the topic of the next chapter. NEG will facilitate the understanding of medium-term trade flows, which we believe to be influenced by the dynamics of the agglomeration of economic activity. Moreover, we will discuss the foundations of NEG in order to show how much Gravity and NEG are linked, and to support our proposal to enrich gravity analysis with some fundamentals of NEG.

VI. Economic Geography Issues.

The increasing interest about the effect of distance and of other variables on trade, which has been mainly studied through the gravity approach, fostered a stream of research on how geography affects the distribution of economic activity across countries. It is called New Economic Geography. According to us, the link between NEG and Gravity is substantial but not straightforward.

We showed that distance affects trade by means of trade costs borne to move goods across borders. Gravity studies have quantified how much such costs affect bilateral trade, but trade costs affect even firms' location as well as countries' welfare. Indeed, why should country A firms (which sell to country B) pay high trade costs instead of moving into country B or establishing a production branch in that country? And if they would do so, what happens to bilateral trade flows between country A and B?

[39] Fixed costs are paid by either exporters or importers. Evans (2003) supposes that exporters pay a sunk cost which allows them to export to all foreign markets (internationalization cost). She finds that in the 1992 only 25% of all the U.S. firms exported abroad. Klenow and Rodriguez (1997) assume that fixed costs are paid by importers, so firms export only to those markets where they find a buyer who accepts to pay the fixed cost.

All these questions are linked and it is unequivocal that to understand how medium-term trade patterns evolve, we need to analyse location decisions conditional on trade costs.[40] NEG models explain why some regions attract a disproportionate share of economic activity by using the interaction between trade costs and firm-level scale economies as source of agglomeration. NEG starts with the observed pattern of agglomeration and postulates a process through which it might have emerged, which typically is: producers and consumers co-locate to exploit plant-level scale economies and to minimize trade costs. NEG models are basically those which follow the approach by Krugman (1991a, 1991b); it is not a case that he is the same author of the IRS/ MC model. Head and Mayer (2003) affirm that five essential ingredients distinguish NEG models from other approaches to the geography of economic activity, these are:

i) *Increasing Returns of Scale (IRS) internal to the firm.* NEG models assume a fixed, indivisible amount of operational costs which are source of IRS, while they *do not* assume any pure technological externality that would lead directly to external scale economies.

ii) *Imperfect competition.* With internal IRS, marginal costs are lower than average costs. Hence, perfect competition does not hold, otherwise firms would be unable to cover their costs. In the majority of NEG models, the market structure is monopolistic competition in the manner of Dixit and Stiglitz (1977).

iii) *Trade costs.* Outputs and inputs used by firms are tradable over distances, but they incur in trade costs. Such costs are often assumed to be proportional to the value of the goods traded.

iv) *Endogenous location of firms.* Firms enter and exit in response to profitability at each possible location, but they incur in sunk costs which limit their mobility.[41]

v) *Endogenous location of demand.* Expenditure in each region depends upon the location of firms since workers employed in a country demand consumption goods in that country.[42]

Assumptions i-iii) are present in the New Trade literature and they are foundations of the IRS/ MC model which we used to micro-found gravity equations. Indeed, it is important to stress that Gravity (at least in its most used version) and NEG are built on the same foundations of monopolistic competition.

The key innovation of NEG is assumption iv and v, Head and Mayer (2003) state that without assumption v) "symmetric initial conditions can be expected to lead to symmetric outcomes, while with all five assumptions, initial symmetry can be broken and agglomerations can form through a

[40] This argument is strongly linked to another branch of trade theory, which is the study of Multinational Enterprises.

[41] The assumption of IRS implies that firms have an incentive to select a single production site and to serve the majority of their customers at distance. If plant-level fixed costs were negligible, the firm would replicate itself everywhere.

[42] Two mechanisms for the mobility of demand have been proposed: 1) Mobile workers who consume where they work (Krugman 1991a); 2) Firms that require the outputs of their sector as intermediate inputs (Krugman and Venables 1995).

process of circular causation" (page 11). Then, by considering all the five assumptions contemporaneously, dynamic analysis can show how factors' movement and profitability cause reallocation of production over time (reallocation is the asymmetric outcome). The dynamics of allocation is characterized by some fundamental theoretical prescriptions, these are:

a) *Market potential raises local factor prices.* A location whose access to major markets and suppliers is not impeded by large trade costs will tend to reward its factors with higher wages and land rentals.

b) *Market potential induces factor inflows.* Capital is drawn to areas with good access to major markets for final goods and major suppliers of intermediate inputs (backward linkages). Workers favour locations with good access to suppliers of final goods (forward linkages).

c) *Home market/magnification effect* (HME). Regions with large demand attract IRS industries, so increasing their production over time. This means that the larger of two regions will be a net exporter to the smaller region in industries characterized by plant-level increasing returns.

d) *Trade induces agglomeration*. In an industry featuring IRS and partially mobile demand, a reduction in trade costs facilitates spatial concentration of producers and consumers.

e) *Shock sensitivity*. A temporary shock to economic activity in a location can permanently alter the pattern of agglomeration.

The key mechanism at work in NEG's models is the HME which is triggered by the size of the market; this mechanism was formalised by Krugman (1980). It entails that the country with the larger market is appealing because it allows the producer to economize on trade costs. If wages do not rise to eliminate this advantage, then a disproportionate share of producers will locate in the large market. This mechanism is usually referred to as *Home Market Effect*.

In the following section we discuss the HME in details. In the last section we will discuss two fundamentals of NEG-based models, which are the Trade-Freeness parameter (ϕ_{ij}) and the Real Market Potential. In both sections we will discuss papers which use both Gravity and NEG to test the significance of some NEG's prescriptions.

VI.A. The Home Market Effect.

As we explained, the HME basically predicts that large countries tend to attract firms which want to exploit plant-level IRS, with the consequence that they eventually become exporters. Models entailing the HME overcome Armington's assumption supposed so far.[43] Indeed, in these models the number of varieties produced in each country is not fixed because firms relocate according to the HME. Krugman (1980) formalizes the HME in a model of monopolistic competition with two

[43] We have always assumed that the number of varieties produced in each country is fixed with the consequence that domestic demand of more varieties causes an import increase.

goods: a homogenous good, and a differentiated good produced in different varieties.[44] The homogenous good is a numeraire and its circulation does not involve transportation costs, while the differentiated good incurs in positive transportation costs. There is only labour as factor of production, and workers-consumers spend a fraction λ of their total wage L in consumption of the differentiated good. Their demand of the differentiated good is:

$$c_{ij} = \left(\frac{p_{ij}}{P_j} \right)^{-\theta} \left(\frac{\lambda L_j}{P_j} \right) . \tag{31}$$

Given trade costs, the price index makes consumption of the foreign varieties relatively more expensive, therefore it diminishes the volume of trade. Once consumption and prices are assumed fixed, and wages are equalized across countries (Factor Price Equalization holds due to no trade costs in the numeraire sector), the model is solved for the *change* in the number of products produced in each country as country size L_j rises.[45]

Considering two countries (1 and 2) which start as identical, trade leads them to export an equal number of varieties to each other. But if we suppose that the labour endowment of country 1 grows with no change in country 2 $(\Delta L_1 > 0 \text{ and } \Delta L_2 = 0)$, the number of varieties in the larger country will increase by more than the increase in its size (the new work-force is employed with IRS in the production of new varieties). While the number in the smaller will shrink. Since exports from a country to another $(X_{ij} = N_i \overline{p}_{ij} \overline{c}_{ij})$ change in proportion to the number of varieties (N_i), with N_1 growing and N_2 falling, country 1 becomes a net exporter of the differentiated good. On the contrary, if we assume fixed and freely-mobile world labour force ($\Delta L_1 > 0 = \Delta L_2 < 0$), reallocation of labour between the two countries will be a likely outcome because, given trade costs, real wages in country 1 are higher and workers wish to move in.

Feenstra, Markusen, and Rose (1999) test if the HME depends on the type of demanded good by estimating different gravity equations. They demonstrate that the HME emerges when the elasticity of domestic income exceeds the elasticity of partner income, so it can be tested and measured through a gravity equation. The HME is expected to be higher when only export of differentiated goods (and no export of homogenous goods) is taken into account. Indeed, manufactured goods (differentiated goods) are likely to have low barriers to entry, while resource-based products (homogenous goods) have usually high entry barriers.

[44] Davis (1998) and Feenstra, Markusen, and Rose (1999) show that the HME does not arise exclusively in the IRS/ MC model, but it arises in a model of reciprocal dumping with a homogenous good as well. Davis (1998) affirms that the differentiated good has higher costs than the numeraire one, otherwise the HME does not arise.

[45] Wages are normalized to one, so $w^*L=L$ is a proxy of the country size.

Using a classification scheme for goods, they sort the exports of each country into categories of homogenous, intermediate and differentiated goods. Subsequently, they estimate the following gravity equation for each category to check the magnitude of the HME:

$\ln X_{ij} = \beta_0 + \beta_1 \ln Y_i + \beta_2 \ln Y_j - \beta_3 \ln D_{ij} + \beta_4 Cont_{ij} + \beta_5 Lang_{ij} + \beta_6 FTA_{ij} + \beta_7 Rm_{ij} + \varepsilon_{ij}$. [46]

The estimated values of the elasticity confirm their expectations. The highest HME is for differentiated goods (domestic GDP elasticity>1, partner GDP elasticity= 0.65), it has a medium value for intermediate goods (domestic GDP elasticity= 0.9, partner GDP elasticity= 0.65), and it reverses for homogenous goods (domestic GDP elasticity= 0.5, partner GDP elasticity= 0.8). A reversed HME means that exports, as percentage of a country output, are higher when a country is smaller than its trading partner.

VI.B. Measuring Access to Markets and Prospective Profitability.

Prospective Profitability measures firms' incentive to enter a specific market, Krugman (1991a, 1991b). In the original model consumers' utility is a CES function of differentiated varieties in which the elasticity of substitution θ represents an inverse index of product differentiation. To purchase a variety produced in country *j*, country *i* 's consumers spend the amount:

$$p_{ij} q_{ij} = \frac{p_{ij}^{1-\theta}}{\sum_{k=1}^{N} n_k p_{ik}^{1-\theta}} \mu_i Y_i , \qquad (32)$$

where $\mu_i Y_i$ is the exogenous expenditure by region *i* on the representative industry as a fraction of its income. p_{ij} is the import price paid by consumers in *i* for the quantity q_{ij} of product imported from *j* (N is the number of countries which supply the i^{th} market, note that both *j* and *k* belong to the set of N countries exporting to *i*), p_{ij} is the product of the export price (p_j) times the *ad valorem* trade cost ($T_{ij} > 1$) paid by consumers. The value of country *i's* imports from all the n_j firms settled in country j is:

$$m_{ij} = n_j p_{ij} q_{ij} = n_j p_j^{1-\theta} \phi_{ij} \mu_i Y_i P_i^{\theta-1} , \qquad (33)$$

where $P_i = \left(\sum_{k=1}^{N} n_k p_k^{1-\theta} \phi_{ik} \right)^{1/(1-\theta)}$ is the price index in each country.[47] It is an average of suppliers' delivered costs to country *i* which assigns increasing weights to sources that have either a large

[46] Where X_{ij} denotes the value of exports from country *i* to country *j*, Y_i is country *i* 's real GDP, Y_j is country *j* 's real GDP, D_{ij} is the distance between *i* and *j*, $Cont_{ij}$ is a control for geographic contiguity between *i* and *j*, $Lang_{ij}$ is a control for language common to *i* and *j*, FTA_{ij} is a control for Free Trade Agreements common to *i* and *j*, Rm_{ij} denotes the remoteness of *j* with respect to *i*, and ε_{ij} is the disturb term.

number of suppliers n_k or good access to market *i* (measured by ϕ_{ik}). Thus a country served by a large number of nearby and low-price sources will have a low P_i, consequently it will be a market where it is difficult to obtain a high market share.

As abovementioned, ϕ_{ij} is a proxy for market *i 's* accessibility to firms from the *j*th country-source that we call Trade-Freeness parameter. Assuming free trade within countries ($\phi_{ii} = \phi_{jj} = 1$) and symmetric bilateral barriers ($\phi_{ij} = \phi_{ji}$), it is possible to obtain an estimate of ϕ_{ij} from eq. (33):

$$\hat{\phi}_{ij} = \sqrt{\frac{m_{ij} m_{ji}}{m_{ii} m_{jj}}}, \qquad (34)$$

where the denominator is each region's imports from itself. $\hat{\phi}_{ij}$ ranges from zero to one, with zero denoting prohibitive trade costs.[48] The trade-freeness parameter ϕ_{ij} is fundamental in NEG models because it is part of firms' profit equation, whose maximization boosts firms to relocate in the most profitable regions.

To derive the prospective profitability of location *j*, we have to consider the following fundamental relations of the model:

A) $tc_j = c_j q_j + F_j$

B) $\pi_{ij} = (p_j - c_j) T_{ij} q_{ij}$

C) $p_j = \dfrac{c_j \theta}{\theta - 1}$

Eq. (A) is the total cost of production for a variety produced in region *j,* where *F* is the fixed cost; eq. (C) is the export price of the variety, it is higher than the marginal cost of production; eq. (B) is the gross profit earned in each market *i* through the sell of a variety produced in region *j.* Substituting eq.(B) in eq. (33) and summing for all the *M* markets where it is possible to sell the variety produced in *j*, we achieve the equation for the net profit to be earned producing in each potential location *j* (by selling in the domestic and in all the other *M* markets):

$$\prod_j = \frac{1}{\theta} c_j^{1-\theta} RMP_j - F_j . \qquad (35)$$

Where the Market Potential in real terms is equal to:

---

[47] Note that the only difference between this price index and Anderson-van Wincoop's multilateral resistance term is ϕ_{ik}.

[48] Estimates of the Trade-Freeness parameter for the pair U.S.-Canada and France-Germany are available in Head and Mayer (2003). For selected industries, the estimated value is low so meaning that trade cots are high within regions (North America and Europe) which are supposed to be very integrated. An over-time analysis shows an upwards trend of the estimated parameter for both pairs, this is likely to reflect decreasing trade costs due to the enforcement of regional trade agreements (NAFTA and European Single Market). It is interesting that the estimated parameter is higher for the U.S.-Canada pair than for France-Germany where we reckon that a more balanced relation between France and Germany has an important role.

$$RMP_J = \sum_{1=i}^{M} \phi_{ij} \mu_i Y_i P_i^{\theta-1} \ , \tag{36}$$

or in nominal terms:

$$NMP_J = \sum_{1=i}^{M} \phi_{ij} \mu_i Y_i \ .$$ [49]

From eq. (35) we understand on which base firms take their location decisions. The more competitors are in a given market i , and the lower the access cost of that market is, the smaller any firm's share of the market i will be.[50] Then the profitability of penetrating a certain market j is low since the small magnitude of the RMP_j term. In other words, a large market that is extremely well-served by existing firms might offer considerably less potential for profits than a smaller market with fewer neighbouring competitors.

How does this affect country j 's trade in this multi-country framework? If the RMP_j is low, we may reasonably expect that country j 's trade with the rest of the world does not change because the intensity of economic activity within its borders remains constant. On the contrary, if market j 's potential profitability is high, firms could decide to move into the country. Then, country j is likely to reduce its imports from, and to increase its exports to, any other country because more varieties are now produced within its borders.[51]

Redding and Venables (2004) use the gravity approach and some of the NEG relations just derived to test one prediction of NEG, they study how geography influences per capita income by applying a two stage procedure. Firstly, they estimate a trade equation from which they construct a *Market Access* term (our Real Market Potential term) and a *Suppliers Access* term, secondly, they estimate a wage equation in which the regressors are these Market and Suppliers Access terms, and test the relation between geography and per capita income.

The *Market Access* term of each exporting country i (MA_i) is the distance-weighted sum of each country j 's market capacity for all its trade partners:

$$MA_i = \sum_{j=1} E_j G_j^{\theta-1} T_{ij}^{1-\theta} = \sum_{j=1} T_{ij}^{1-\theta} m_j \ , \tag{37}$$

---

[49] The concept of NMP is intuitively appealing and not difficult to implement in empirical works. But the omission of the price index adjustment ($P^{1-\theta}$) in the NMP severs the link with the underlying profit maximization problem. Davis and Weinstein (2003a) use a variant of the NMP in which they set $\phi_{ij} = d_{ij}^{\delta}$. d_{ij} is the distance and δ is its coefficient obtained by estimating a gravity equation. Since the estimates of δ do not differ greatly from minus one, $1/\ d_{ij}$ is a reasonable approximation for ϕ_{ij} .

[50] Remember that the product realized in j is sold in i, and that we are evaluating the profitability of market j.

[51] Through the relations just derived, it is possible to explore how firms and workers' location decisions depend upon market and supply potential. Such analysis can be interpreted as an empirical test of the existence of *Backward linkages* (are firms attracted to locations with large demand of their products?) and *Forward linkages* (are consumer-workers attracted to locations with high industrial production?).

where the *Market Capacity* of the importing country j (m_j) is the product of the total expenditure in j (E_j) and the price index in j (G_j). While the *Suppliers Access* term of each importing country j (SA_j) is the distance-weighted sum of each country i 's supply capacity for all its trade partners:

$$SA_j = \sum_i n_i (p_i T_{ij})^{1-\theta} = \sum_i s_i (T_{ij})^{1-\theta} , \tag{38}$$

where the *Supply Capacity* of the exporting country i (s_i) is the product of the number of firms in i and the price which they set . Both in the Market Access and in the Supply Access term, T_{ij} measures bilateral transport costs between two countries.

To construct empirical values of the Market and Supply Access, Redding and Venables need estimates of the market and supply capacity. They can be estimated in two ways: either by using dummy variables or by using economic measures of supply and market capacity. With the dummies, the estimated gravity equation is:

$$\ln X_{ij} = \alpha + \beta * cty_i + \gamma * ptn_j + \delta_1 * \ln(dist_{ij}) + \delta_2 * bord_{ij} + u_{ij} , \tag{39}$$

where the distance ($dist_{ij}$) and the border coefficient ($bord_{ij}$) measure bilateral trade costs, cty_i and ptn_j are the country and partner dummy which capture respectively the market and supply capacity. The Market and Suppliers Access term are eventually computed through the estimated market and supply capacity. The wage equation shows that they explain up to 70% of the cross-country variation in per capita income.

VI.C. A Proposal for Further Research.

Papers such as Redding and Venable (2004), and Feenstra, Markusen, and Rose (1999) (which we discussed respectively in the previous section and in section VI.A.) are examples of how Gravity and NEG can be used together to check the validity of theoretical prescriptions. But, they do not provide us with an example of a unified model in which elements of both theories are included. Nevertheless, we believe that this kind of models would be valuable for a better understanding of trade patterns. For instance, trade flows could be studied through the intertemporal maximization of a country's utility function (and not through its intratemporal maximization, as seen so far) in order to study the dynamic property of the optimum path. Then, sensitivity analysis could serve to investigate how the variation of some NEG's variables (included in the model) affect the optimal level of bilateral trade. More specifically, we could study how bilateral exchanges between two regions change over time when the Real Market Potential of those regions change.

We could also check if trade costs lose significance in explaining trade, when we move from less to more accessible countries (on the basis of the Market and Suppliers Access parameter by Redding and Venables 2004). We expect that the more a country is accessible to any other, the higher its trade is. But, are trade costs more or less important in explaining trade between two countries? On the basis of gravity prescriptions, they should be very important because imports of very accessible countries (which have a low price index) are highly sensitive to trade costs. Indeed, when formal restrictions to trade fall (tariffs and NTBs), it is reasonable to suppose that the remaining restrictions (trade costs) are highly significant in explaining trade flows. To check this hypothesis, a sample of countries could be ordered in different groups for their degree of Market and Suppliers Access. Then we could estimate a gravity equation for each group in order to test this prediction.

Another possible object of further investigation is the explanation of estimated Border Effects through NEG and in particular through the Home Market Effect. Chen (2004) made a first move in this direction by explaining industrial Border Effect through Industrial Concentration Indices. We deem this strand of research very interesting and much more informative than proved yet.[52]

VII. Conclusions.

The aim of this paper was to show how space-related effects affect trade relations. For this reason we discussed the effect of trade costs on bilateral exchanges, how to quantify such effect through the estimation of gravity equations, and the micro-foundations of a typical gravity model. We argued about the descriptive nature of gravity models, and we showed that it does not restrict them to just quantify effects, but even to understand the causes of trade. We explained the Border Puzzle which emerges from gravity estimations, and how Anderson and van Wincoop solve it by improving both the micro-foundations of the model and its estimation. On the whole, we explained how space-related effects are embedded in gravity models and in the models of New Economic Geography, and we pointed out the role they play (or should play) in Modern International Economics.

In this paper we linked Gravity to New Economic Geography even though this link does not emerge vigorously in the present literature. We sought to make this link clear by discussing the assumptions of both Gravity and NEG. In the last chapter, we suggested ways of using such link to enhance trade analysis through a unified model which include both NEG and Gravity.

[52] This issue is investigated in *the third paper* of this thesis.

Appendix

Table 3 - The effect of Currency Unions on Bilateral Trade in Gravity Model (Frankel and Rose 2002).

Currency Union	1.53 (.18)	1.22 (.19)	1.59 (.18)	1.72 (.18)
Log Distance	-1.09 (.03)	-1.09 (.03)	-1.07 (.03)	-1.11 (.03)
Log Product Real GDP	.79 (.01)	.80 (.01)	.95 (.01)	.80 (.01)
Log Product Real GDP/capita	.64 (.01)	.66 (.01)	.46 (.02)	.63 (.02)
Common Language	.73 (.06)	.44 (.06)	.76 (.06)	.69 (.06)
Common Land Border	.37 (.12)	.43 (.12)	.62 (.13)	.39 (.12)
Common FTA	1.31 (.11)	1.25 (.10)	1.16 (.11)	1.28 (.11)
Common Colonizer		.65 (.08)		
Political Union		1.08 (.35)		
Ex-Colony/Colonizer		2.19 (.12)		
Number of landlocked in pair (0, 1 or 2)			-.36 (.04)	
Number of islands in pair (0, 1 or 2)			.04 (.04)	
Log of Product of Land Area			-.16 (.01)	
Currency Union/Non Union				.37 (.04)
R^2	.62	.63	.64	.63
RMSE	2.03	2.00	1.99	2.02

Regressand is log of bilateral trade in real American dollars.
Number of Observations = 31,101.
Year-specific fixed effects not reported.
Robust standard errors recorded in parentheses.

Table 4 – The Effect of Openness on GDP per capita (Frankel and Rose 2002).

	OLS	IV	IV	IV	OLS	IV	IV	IV
Openness	.79 (.18)	1.22 (.56)	1.27 (.69)	1.09 (.52)	.33 (.07)	.33 (.08)	.17 (.08)	.32 (.09)
Log Population	.14 (.06)	.19 (.09)	.18 (.10)	.12 (.09)	.07 (.02)	.07 (.03)	.10 (.03)	.06 (.09)
Log Area			.02 (.10)				-.06 (.03)	
Currency Union				-.86 (.25)				-.07 (.09)
Log '70 Real GDP/capita					.71 (.05)	.72 (.06)	.75 (.05)	.72 (.06)
Investment Ratio					.016 (.006)	.015 (.006)	.018 (.006)	.015 (.006)
Population Growth Rate					-.06 (.05)	-.05 (.06)	-.04 (.06)	-.05 (.06)
Primary Schooling					.002 (.002)	.002 (.002)	.002 (.002)	.002 (.002)
Secondary Schooling					.007 (.002)	.007 (.003)	.006 (.002)	.007 (.003)
Number of Observations	115	108	107	107	106	101	101	101
R^2	.11	.05	.04	.13	.94	.94	.94	.94
RMSE	1.02	1.05	1.06	1.01	.28	.28	.28	.28

Regressand is log of Real GDP/capita in 1990, PWT.
Intercepts not reported. Robust standard errors recorded in parentheses.
Instrumental variable from aggregated fit of gravity equation.

Bibliographical References.

Anderson J. E. (1979), "A Theoretical Foundation for the Gravity Equation", American Economic Review 69(1): 106-116.

Anderson J. E. and E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", American Economic Review s 93: 170-92.

Anderson J. E. and E. van Wincoop (2004), "Trade Costs", NBER Working Paper no.10480.

Armington P. S. (1969), "A Theory of Demand for Products Distinguished by Place of Production", IMF Staff Papers, v. 16: 159-176.

Baier S. L. and J. H. Bergstrad (2001), "The Growth of World Trade: Tariffs, Transport Costs and Income Similarity", Journal of International Economics 53, 2001, 1-27.

Bernard A. B. and J. B. Jensen (1997), "Why some firms export", Review of Economics and Statistics 2003.

Bernard A. B. and J. Wagner (2001), "Export Entry and Exit by German Firms", Weltwirschaftliches Archiv 2001, 137(2), pp.105-123.

Bergstrand J. H. (1985), "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence", Review of Economics and Statistics, 1985, 67(3), pp. 474-81.

Bergstrand J. H. (1989), "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade", Review of Economics and Statistics, 1989, 71(1), pp. 143-53.

Bergstrand J. H. (1990), "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade," The Economic Journal, 1990, 100, pp. 1216-29.

Chen N. (2004), "Intra-national versus international trade in the European Union: why do national borders matter?, Journal of International Economics, 63: 93-118.

Davis D. R. (1998), "The Home Market Effect, Trade and Industrial Structure", American Economic Review, 88:1264-76.

Deardorff A. V. (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?", in Jeffrey A. Frankel, Ed., The Regionalization of the World Economy, Chicago: University of Chicago Press for the NBER, 7-22.

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

Evans C. (2003), "Border effects and the availability of domestic products abroad", IMF Working Paper, Board of Governors 2003.

Evenett S. J. and W. Keller (2002), "On Theories Explaining the Success of the Gravity Equation", Journal of Political Economy, vol. 110, no.2.

Feenstra R. C. (2004), "Advanced International Trade", Princeton University Press.

Feenstra R. C., J. R Markusen, and A. K. Rose (1998), "Understanding the Home Market Effect and the Gravity Equation: the Role of Differentiating Goods", NBER Working Paper no. 6804.

Frankel J. A. (1998), "Regional Trading Blocs in the World Trading System", Washington DC: Institute for International Economics.

Frankel J. A. and A. K. Rose (2002), "An Estimate of the Effect of Common Currencies on Trade and Income", Quarterly Journal of Economics, 117: 437-466.

Harrigan J. (2001), "Specialization and the volume of trade: do the data obey the laws?", NBER Working Paper no. 8675.

Head K. and T. Mayer (2003), "The Empirics of Agglomeration and Trade", CEPII Working Paper no. 2003–15

Helpman E. and P. Krugman (1985), "Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy" Cambridge, MA: MIT Press, 1985.

Helpman E. (1987), "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries", Journal of the Japanese and International Economies, I(1987) 62-81.

Hummels D. (2001), "Toward a Geography of Trade Costs", working paper, Purdue University.

Klenow P. and A. Rodriguez (1997), "Quantifying variety gains from globalization", Working Paper.

Krugman P. (1991a), "Increasing Returns and Economic Geography", Journal of Political Economy 99:483–499.

Krugman P. (1991b), "Geography and Trade", MIT Press, Cambridge.

Lancaster K. (1980), "Intra-Industry trade under Perfect Monopolistic Competition" Journal of International Economics 10:151-75.

Leamer E.E. and J. Levinsohn (1995), "International Trade Theory: the evidence", Handbook of International Economics chap. 26.

Leamer E.E. and R.M. Stern, (1970), "Constant-market-share analysis of export growth", Chapter 7 in Quantitative International Economics, E.E. Leamer and E.M. Stern, Allyn and Bacon Inc.: Boston, pp. 171-183.

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

Melitz J. (2002), "Language and Foreign Trade", University of Strathclyde Working Paper.

Obstfeld M. and K. Rogoff (2000), " The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?", NBER Working Paper no. 7777.

Redding S. and A. J. Venables (2004), "Economic Geography and International Inequality", Journal of International Economics, Elsevier, vol. 62(1).

Rose A. K. and E. van Wincoop (2001), "National Money as a Barrier to Trade: The Real Case for Currency Union", American Economic Review Papers and Proceedings 91(2), pp. 386-390.

Silva S. and S. Tenreyro (2006), "The Log of Gravity", The Review of Economics and Statistics, 88(4): 641–658.

Tchamourliysky Y. (2002), "Distance and Bilateral Trade: the role of non-homothetic preferences", Boston College Working Paper.

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