International Conference on Regional and Urban Modeling

Brussels, June 1-3, 2006

EMPIRICAL EXAMINATION OF THE GRAVITY MODEL IN TWO DIFFERENT CONTEXTS: ESTIMATION AND EXPLANATION

Ana Sargento

School of Technology and Management
Polytechnic Institute of Leiria (Portugal)

sargento@estg.ipleiria.pt

Empirical examination of the gravity model in two different contexts: estimation and explanation.

1. Introduction

Gravity model belongs to the family of spatial interaction models, which form a substantial part of regional science and geography research. The goal of these models is to explain and/or estimate spatial interaction flows, broadly defined as the movement or communication between different spaces. This interaction implies a decision which is taken after a cost-benefit analysis, in which the individual evaluates the trade-off between the benefit from the movement (related to the motivation that causes it) and the cost of that same movement (which corresponds to the traveling across the spatial separation between his / her origin and the several destinations) (Fotherigham and O'Kelly, 1989). Spatial interaction models deal with a diverse collection of flows, such as: international and interregional trade, migration, information flows, traffic flows and commuting movements, among others. In presence of such a variety of applications, there is no specific type of model which is superior to all the remaining, whatever the topic that it is applied to. Being so, in each particular circumstance, the researcher must decide which is the most adequate, among all the proposed models (Isard, 1998).

Yet, due to its simplicity and capacity to produce reasonable results, the gravity model continues to be the most attractive among spatial interaction models, especially in trade empirical applications. By analogy to Newton's gravity law, the application of gravity model to trade flows states that trade increases with the dimension and proximity between trade partners.

The consolidation and enlargement of economic integration blocs in the end of 20^{th} century led to a renovated interest in trade flow gravity model applications, to study specific issues, such as: searching for non-institutional regional trading blocs, computing of trade creation and trade diversion effects and estimation of trade potential between former and new EU members (Porojan, 2001).

Gravity models, as well as other spatial interaction models, can be used, in two different information contexts:

(a) Spatial interaction flows are **known** *a priori*; in this case, the model is used to explain trade flows' behaviour, through econometric modelling;

(b) Spatial interaction flows are **unknown** *a priori*; here, the model is applied in order to assess the unknown flows.

Very little attention has been paid to the second context of application. In fact, most of the literature on gravity models is dedicated to its theoretical underpinnings and to empirical applications with explanatory purposes (information context (a)). Being so, the main objective of the present work is to discuss and test the practical applicability of gravity model in studying trade flows, in both above referred information contexts. Additionally, this work also intends to fill another gap in trade flow gravity model uses: the fact that the majority of studies consider trade in an aggregate manner. Realizing the specificity of each product, this study is applied separately to different trading products.

This paper is organized as follows: section 2 reviews the gravity model basic specification and its extensions in recent trade flows applications. Section 3 describes the empirical use of gravity model in this work, in the two distinct information contexts. Alternative methodologies are presented and results are discussed. Section 4 summarizes the main conclusions.

2. The gravity model.

1.1. Basic equation.

Analytically, the basic equation that is used to express the gravity hypothesis on trade flows between origin *i* and destination *j* is:

$$X_{ij} = G \frac{P_i^{\alpha_1} P_j^{\alpha_2}}{d_{ij}^{\alpha_3}} \tag{1}$$

in which: X_{ij} represents exports from origin *i* to destination *j*, G is a constant of proportionality, P_i and P_j express the sizes of origin *i* and destination *j*, with weights α_1 and α_2, respectively, d_{ij} represents spatial separation between each origin *i* and each destination *j* and α_3 is the so-called distance decay parameter, measuring the flow sensibility to spatial separation.

This equation comprises some quite vague concepts, such as size and spatial separation (Sen and Smith, 1995). These concepts allow for different interpretations. Spatial separation, for instance, can be expressed by physical distance or other concepts of separation, like political or cultural distance. Also, the whole of specific formulations that are consistent with the gravity hypothesis is very vast, being equation (1) a particular case[1]. The debate over the different variables to express each of the above referred concepts and the different formulations to gravity model is beyond the scope of the

[1] For example, some researchers consider the exponential functional form to represent spatial separation, instead of the power functional form. If that was the case, equation (1) would be: $X_{ij} = G \frac{P_i^{\alpha_1} P_j^{\alpha_2}}{\exp(\alpha_3 d_{ij})}$.

present paper, being minutely discussed in Isard (1998). For this work purposes, equation (1) is taken as the basic gravity equation and the correspondent basic variables are: GDP, for masses and physical distance, for spatial separation.

In practical applications, equation (1) is usually taken in logarithmic form, as:

$$\ln X_{ij} = \ln G + \alpha_1 \ln P_i + \alpha_2 \ln P_j - \alpha_3 \ln d_{ij} \tag{2}$$

From equation (2) it is clear that each parameter α can be seen as elasticity of exports with respect to: the exporting country's GDP, the importing country's GDP and the distance between *i* and *j*.

In type a) information contexts, equation (2) can be the starting point to a regression equation like:

$$\ln X_{ij} = \alpha_0 + \alpha_1 \ln P_i + \alpha_2 \ln P_j + \alpha_3 \ln d_{ij} + \varepsilon_{ij} \tag{3}$$

in which $\ln X_{ij}$ is the endogenous variable, α_0, α_1, α_2, and α_3 are the coefficients to be estimated[2], P_i, P_j and d_{ij} are the explanatory variables and ε_{ij} is the error term.

[2] In this equation, the expected α_3 sign is negative.

1.2. Gravity model extensions to trade applications.

1.2.1. Augmenting the gravity equation with additional explanatory variables.

Numerous studies have used an augmented version of gravity model basic equation, in order to address some specific issues. One of the most important motivations to gravity model extensions is the study of preferential trade agreements effects. Some examples of this kind of exercises can be found in Martinez-Zarzoso (2003), Soloaga and Winters (1999) and Piani and Kume, (2000). The common feature of these three works is the addition of specific bloc-related dummy variables to equation (3), in order to capture the effects of preferential trade agreements, especially those concerning trade creation and trade diversion. Further dummy variables are also included to isolate the effects of other determinants of trade, such as: sharing the same language or sharing a common border. The work of Blavy (2001) is another example of extending the gravity model to answer some specific trade issues. In this case, the author starts by applying the basic gravity model to trade patterns in specific region composed by six Middle East countries, reaching to the conclusion that it overestimates intranational and international trade in that region. To overcome such problem, the model is extended with specific explanatory variables, to assess the effects of: over-appreciation of exchange rates, trade barriers and political uncertainty. It is shown that the augmented model has a better performance in estimating trade flows.

1.2.2. Formal specification of spatial dependence.

One of the frequent criticisms pointed out to gravity model is the fact that it generally assumes that observations collected at different points of space are completely independent, which is not true. There are well known diffusion processes among different locations that must be taken into account, through a specific modeling of space.

One possible way to acknowledge spatial structure effects is through the inclusion in the gravity model of additional variables that, in some way, illustrate the map pattern of the observations under study. For example, Hu and Pooler (2002) use an augmented gravity model (applied to explain international trade flows between *i* and *j*), in which spatial structure effect is captured through the inclusion of an additional variable designed to measure accessibility of destination *j* (given by the weighed sum of the distances between all the origins and *j,* in which each origin's mass is the relevant ponderer). They compare the performance of this model with the traditional gravity model, showing that the addition of the accessibility variable contributes to a better predictive capacity of the model. Some previously referred exercises also attempt to include in their gravity models a variable that expresses the relative locations of the different observations. That is the case, for example, in Piani and Kume, (2000) and in Soloaga and Winters (1999), which consider a relative distance (or remoteness) indicator, to control for the stronger trade intensity that usually exists between remote pairs of countries, when compared with trade between neighbors that have many other close trading partners[3]. A simpler and more common way used to illustrate map pattern of the observations is the inclusion of a dummy variable that indicates the presence (or not) of a common border between the trading partners.

However, spatial effects are often more comprehensive, making unavoidable the use of more sophisticated modeling techniques, that fall in the spatial econometrics field. If this is the case, it is very important, not only to find the proper way to formally express the spatial effects, but also to use the adequate techniques to estimate the model. Standard regression methods (as Ordinary Least Squares) are no longer acceptable when spatial effects are definitely present[4] (Anselin and Griffith (1988)). The paper of Anselin and Griffith (1988) is crucial to systematize the nature of spatial effects. These are associated

[3] Soloaga and Winters (1999) illustrate this with the examples of Australia and New Zealand that tend to trade more with each other than Portugal and Spain.
[4] Yet, most of the practical applications completely ignore the possibility of spatial effects, using OLS as the single estimating method.

to **spatial dependence**, on the one hand, and to **spatial heterogeneity**, on the other hand. Spatial dependence may exist due to spillover effects across space and occurs whenever the dependent variable is "affected by the values of the dependent variable in nearby units, with nearby suitable defined" (Beck, Gleditsch and Beardsley 2005, p.9). It can be discovered by the presence of autocorrelated error terms (originating the spatial error models) and/or autocorrelation in the dependent variable (resulting in the spatial lag model). Spatial heterogeneity may be due to structural instability, meaning that functional forms and/or parameters differ from one observation to another[5], or to model misspecification that leads to non-constant error term variances (heteroskedasticity).

Spatial heterogeneity can be undertaken by means of the typical solutions in traditional econometrics. To formally account for spatial dependence, however, it is necessary to introduce the concept of spatial lag. Let *W* be the spatial lag operator, also called connectivity matrix (Beck, Gleditsch and Beardsley 2005). This matrix represents spatial morphology and is composed by non-stochastic w_{ij} elements, based on the geographic arrangement of observations. One of the most popular criteria to express geographic arrangement is contiguity; following this criterion, w_{ij} assumes the value 1 if *i* and *j* are contiguous locations and the value 0, otherwise[6]. In short, the spatial lag operator can be seen as a "weighted average (with the w_{ij} being the weights) of the neighbors, or as a spatial smoother" (Anselin, 1999, p.6).

Analytically, a spatial error model is expressed by an equation like:

$$Y = X\beta + u$$
$$u = \lambda W u + \varepsilon$$

[5] Formally, this would mean that, for each observation *i*, there would be a function $y_i = f_i(x_i, \beta_i) + \varepsilon_i$, in which x_i is a 1**m* row of *m* explanatory variables and β_i stands for the correspondent coefficients.

[6] The matrix W is row-standardized, as usual in this type of models; each row sums 1, so that there is no need to worry about the units used to measure connectivity.

(4)

in which Y represents the vector of dependent variables, X is the matrix of explanatory variables, W is the lag operator, β is the vector of parameters that reflect the influence of explanatory variables on Y and λ expresses the degree of spatial correlation among the model disturbances. In this model it is assumed that the only source of interdependence among observations is in the error formation process, more precisely, the fact that some omitted variables are spatially correlated (Beck, Gleditsch and Beardsley 2005).

The spatial lag model can be formally expressed by:

$$Y = \rho WY + X\beta + \varepsilon \tag{5}$$

in which ρ illustrates the degree of the dependent variable spatial autocorrelation and the remaining variables have the above referred meaning. This model implies the assumption of feedback effects among observations / locations: variations in the explanatory variables of location i affect the dependent variable of that location and of neighboring locations (because of the lag operator). Consequently, location i will be affected for a second time (again, because of the spatial link with its neighbors) and this process will be successively repeated as in a multiplier effect.

The emergence of new software tools and theoretical contributions to deal with spatial dependence has facilitated the empirical application of these models. Some examples can be found in Beck, Gleditsch and Beardsley (2005) and Porojan (2001). Beck, Gleditsch and Beardsley (2005) propose an alternative connectivity measure to include in the spatial lag model. Their objective is to explain democracy level. They argue that instead of geographical notion of proximity other measures can be used. So, they propose a W matrix with elements given by the "volume of the dyadic trade flow between i and j as a proportion of country *i*'s total trade" (p.13); the empirical exercise proves that the spatial autocorrelation coefficient associated to this connectivity matrix is statistically significant, suggesting that "countries that trade more with democracies are more likely to be democratic (…)" (p. 17). The empirical application carried out in Porojan (2001) aims to find the most proper version of gravity model to explain international trade. Several alternative equations are tested, including the gravity traditional specification, leading the author to the conclusion that the most adequate equation is the one which explicitly considers the existence of two spatial effects: spatial heterogeneity (adapting the model to account for heteroskedastic error) and spatial autocorrelation of the dependent variable.

3. Empirical application.

1.3. Type (a) information context.

When the researcher has previous access to a known trade matrix, the objective is to calibrate the model, i.e., to estimate the model parameters. An immediate question emerges: "if we already have the interaction matrix, why do we need to calibrate an interaction model?" (Fotheringham and O'Kelly (1988), p.43). In fact, the calibration process in useful to forecasting purposes (admitting that the parameters remain the same in different points of time and/or space) and to draw conclusions on the behavior patterns of the subject in study (for example, to assess the degree of elasticity of exports with respect to the distance between the trading partners and to evaluate how this varies from one product to another).

In this section, attention will be given to the econometric application of gravity model to explain bilateral trade flows among the 15 EU countries[7] (before enlargement). This

[7] In fact, the number of origins (equal to the number of destinations) is only 14, since Belgium and Luxembourg are considered jointly, as one country.

application was carried out in a stepwise fashion, testing several alternative equations and analyzing the results of each one.

All the equations were calibrated using LeSage's *Econometrics MATLAB toolbox*, of which functions are available at http://www.spatial-econometrics.com/.

1.3.1. The data.

The set of data used in this work is composed by:

- Export data from each of the 14 countries to each of the others, for year 2001, in USD and current prices; source: OECD Bilateral Trade Database 2002;
- Population, year 2001, in thousands; source: OECD member countries' population 1981-2004 (thousands and indices: 2000=100). Labour Force Statistics, 2005 Edition;
- Gross Domestic Product, year 2001, in USD and current prices; source: OECD Annual National Accounts database.
- Great circle distances between capital cities; source: http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/Data/Gravity/dist.txt.

Origin-destination flow data have specific characteristics, which must be emphasized before explaining the practical application that was carried out. First, the number of observations, N, is equal to 196 (14 origins multiplied by 14 destinations), in spite of being only 14 countries in study[8]. Second, the vectors of explanatory variables have a particular feature: in the origin related variables (as, for example, GDP of origin i), the same value is repeated n times: once to each destination country; in the destination

[8] Let n be the number of origins, which in this case is equal to the number of destinations.

related variables (for example, GDP of destination *j*) the same sequence of values is repeated *n* times: once to each origin country. Finally, distance and contiguity matrices are symmetric (ex.: if Germany is contiguous to France, the opposite is also true; the same reasoning applies to the distance between these two countries). The particular features of origin-destination flow data, and its implications, are the main subject of LeSage and Pace (2005).

As previously referred, all the alternative equations were estimated ten times: once to each of the ten next manufactured products (followed by the correspondent abbreviate designation):

1	FOOD PRODUCTS, BEVERAGES AND TOBACCO	FBT
2	TEXTILES, TEXTILE PRODUCTS, LEATHER AND FOOTWEAR	TEX
3	WOOD AND PRODUCTS OF WOOD AND CORK	WOO
4	PULP, PAPER, PAPER PRODUCTS, PRINTING AND PUBLISHING	PPP
5	CHEMICAL, RUBBER, PLASTICS AND FUEL PRODUCTS	CHE
6	OTHER NON-METALLIC MINERAL PRODUCTS	OTH
7	BASIC METALS AND FABRICATED METAL PRODUCTS	MET
8	MACHINERY AND EQUIPMENT	MAQ
9	TRANSPORT EQUIPMENT	EQT
10	MANUFACTURING NEC; RECYCLING	NEC

The use of gravity model with individual products is less common than aggregate trade applications. However, some exceptions exist. For example, Feenstra, Markusen and Rose (1998) distinguish two groups of products: differentiated and homogeneous, expecting to find a higher value of domestic income exports elasticity in manufactured / differentiated products, when compared to the correspondent value in primary, homogeneous, resource based goods. Their results confirm the initial expectative. In the present work, the objective is also to ascertain the variability of the several estimated coefficients in different products, devoting special attention to the distance parameter. However, it should be noted that there is an a priori limitation that must be taken into account when inferring the results: the level of aggregation involved in the above list of ten products is still very high.

1.3.2. Model calibration through alternative gravity equations.

Model calibration was done in a stepwise fashion, testing succeeding formulations for gravity model. The first regression, named *Model 1*, was based on the traditional specification of gravity model, expressed before in equation (3). Yet, the estimated equation was a bit different from that equation, since GDP was decomposed in two separate factors, in order to capture two distinct effects on trade: population, as a size explanatory variable, and per capita income, as an indicator of development. Being so, *Model 1* is expressed by:

$$\ln X_{ij} = \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \varepsilon_{ij} \qquad \textbf{(6)}$$

in which N stands for per capita GDP and *POP* is population. The remaining variables and parameters have the meaning referred before. It is expected that the coefficients associated to *N* and *POP*, have positive signs, because these are the traditional propulsion (for origins) and attraction (for destinations) variables in the gravity model. On the contrary, it is expected that the distance parameter, β_5, has a negative sign.

Equation (6) was applied recursively to the ten manufactured products. The dependent variable vector was different to each product, but the explanatory variables remained the same, since these variables are not related to any particular product. Ordinary Least Squares was the methodology used to calibrate *Model 1*. After running the first regression to this model, the White test was applied, in order to investigate the possibility of heteroskedasticity. The results of this test proved that heteroskedastic errors existed in *Model 1*, when applied to four of the ten products: FBT, WOO, PPP and EQT. Regression results were then adjusted through the White procedure (using the correspondent function in *Econometrics MATLAB toolbox*). The results remained the same, except for one parameter (β_3 in PPP) that became insignificant in consequence of this correction.

The display and analysis of the results will be targeted to some specific issues, since the complete list of results for the ten products would be too extensive[9]. The following Table 1 sums up the more relevant results of *Model 1*:

Table 1 – *Model 1* principal results.

	R-bar squared	Statistically insignificant coefficients (5%)	Coeff. signs equal to expected?	β_5 (distance coeff.)
FBT	69%	β_0 and β_3	yes	-1,1
TEX	58%	β_0 and β_3	no: negative β_1	-1,21
WOO	42%	β_3	yes	-1,17
PPP	64%	β_3	yes	-0,83
CHE	82%	β_3	yes	-1,22
OTH	77%	β_0 and β_1	no: negative β_1	-1,26
MET	79%	none	yes	-1,13
MAQ	84%	none	yes	-0,75
EQT	69%	β_0 and β_3	yes	-1,09
NEC	76%	β_3	yes	-1,2

[9] If the reader is interested in getting some specific result that is not published in this paper, please contact the author.

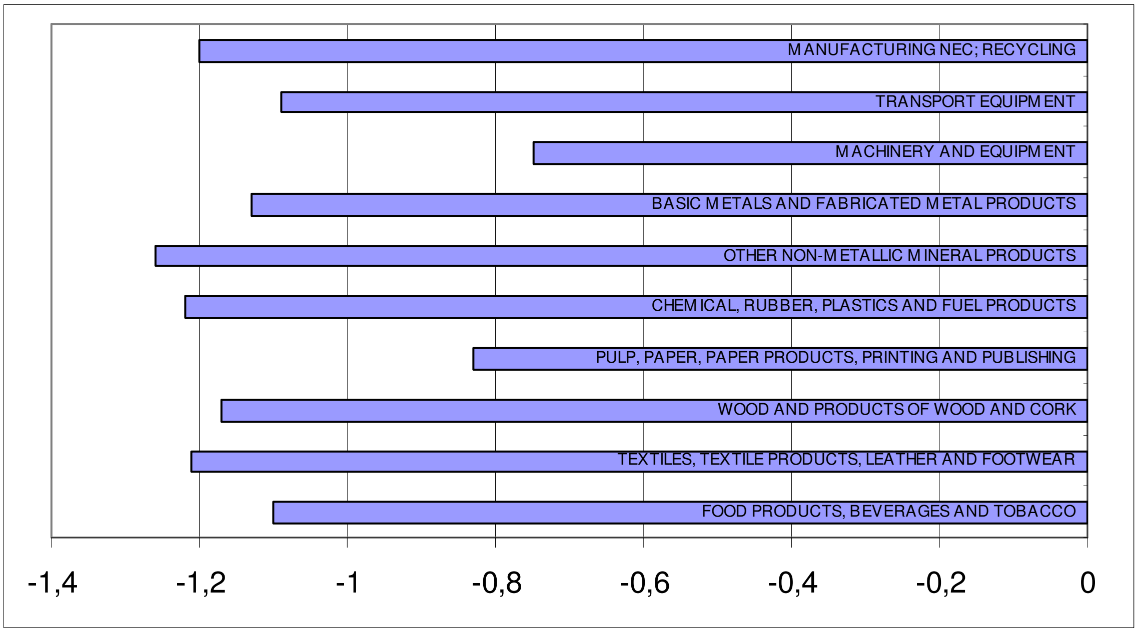
The first column of Table 1 expresses the explicative power of the model, by means of its R-bar Squared. In spite of being extremely variable between the different products, it should be emphasized that this indicator assumes relatively low values for some of them, like WOO and TEX. The second column refers to the statistical relevance of the variables included in this model, through the t-statistic value (at 5% significance level). The most evident observation is that the estimated parameter associated to per capita GDP of the importing country is not significant in six of the ten cases. This may be a sign that size matters more than development level as an attraction measure to international trade (since the origin's population parameter is always statistically significant). Column 3 indicates the coincidence (or not) between the estimated parameters' signs and the expected ones. That coincidence is not verified in the cases of TEX and OTH, in what respects to β_1. In fact, the traditional interpretation of gravity model in international trade applications is that trade tends to be greater between larger countries. However, since GDP effect was decomposed in two indicators (size and development), it could be argued that the sign of per capita GDP is more an empirical issue, i.e., it may be positive or negative, according to the specific case. One plausible explanation to the negative sign found in those two products is that they belong to a class of low-technology industries, in which less developed countries are more specialized[10]. Being so, countries with a smaller per capita GDP would be expected to export more of these products and *vice-versa*. Finally, the last column presents the distance parameter estimated value, to each of the products. It is clear that, as expected, distance produces a negative effect on international trade flows. However, the estimated elasticity is extremely variable among the different products. Even though the level of product aggregation avoids more accurate conjectures, it seems that there is a direct relationship between the export's elasticity with respect to distance and the typical product weights for monetary unit. For example, WOO, which is typically a heavy and low value product, reveals more resistance to distance than the average.

Figure 1 illustrates distance parameter variability:

---

[10] See, for example, the case of textiles and related products, in which Portugal, with a per capita GDP below the average, shows a great specialization.

Figure 1 – Estimated distance parameter in *Model 1*, for the ten products.



The awareness of potential spatial dependence effects motivated a spatial autoregressive model application, named *Model 2*, which may be generally expressed by:

$$\ln X_{ij} = \rho W \ln X_{ij} + \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \varepsilon_{ij} \quad (7)$$

According to the previous classification of spatial dependence models, this is a spatial lag model. The statistical significance and value of ρ will allow inferring about the presence and degree of spatial dependence in the dependent variable.

As referred before, origin-destination data have particular features. This has implications on how to construct an adequate weight matrix *W*. It should be noted that, when dealing with spatially collected data, usually each observation corresponds to one region. This is nor the present case: in origin-destination data, the observations vector is composed by the flows generated by every possible combination of origin and destination, in both directions (LeSage, 2005). In this particular case, we have 196 rows in the observations vector. In this context, the *W* operator must have a compatible dimension. As it will be referred, matrix *W* can be constructed in three different ways, resulting of distinct types of spatial dependence under consideration. One possible way of assembling matrix *W* is by repeating the typical contiguity matrix *n* times (14 times, in this case) in the diagonal of an *N*N* block matrix, with blocks of zeros in all the off-diagonal matrices (LeSage, 2005). In the present work, the most common concept of contiguity was used: sharing of a common border. Thus, an initial 14*14 *w* matrix was produced, with elements w_{ij} that were given the value 1 if *i* and *j* shared a common border and the value 0, otherwise. Matrix *w* was then row-standardized, becoming matrix *C*. Finally, *W* was formed spreading out the 14*14 matrix *C* on a 196*196 matrix *W*. The following diagram illustrates the process of matrix *W* construction:

Figure 2 – Construction of matrix W

(I)

$$w = \left[w_{ij}\right]_{14*14}; \quad w_{ij} = \begin{cases} 1, \text{ if } i \text{ and } j \text{ contiguous} \\ 0, \text{ if } i \text{ and } j \text{ not contiguous} \end{cases}$$

(II)

$$C = \left[c_{ij}\right]_{14*14}, \text{ such as } \sum_j c_{ij} = 1$$

(III)

$$W = \begin{bmatrix} [C] & 0 & \cdots & 0 \\ 0 & [C] & 0 & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & [C] \end{bmatrix}_{196*196}$$

The spatial lag $W \ln X_{ij}$ included in equation (7) captures a "destination-based" spatial dependence (LeSage, 2005). Let us consider a specific element X_{ij} in the observations vector. The inclusion of the spatial lag in the regression equations means that flows from *i* to *j* are influenced, among other factors, by the average of flows from *i* to all the neighbors of *j*. Using a concrete example, this is to say that the France to Germany value of exports is influenced by the average of exports coming from France to all the neighbors of Germany. Hence, matrix *W* will be named W_d, in order to specify a destination-based dependence. Note that the matrix W_d can also be obtained by simply applying the *Kronecker* product between a *n*n* identity matrix and matrix *C* ($W_d = I_{14} \otimes C$)[11]. In this case, the general SAR model expressed in equation (7) can be written as a *destination-based spatial autoregressive* (DSAR) model:

$$\ln X_{ij} = \rho_d W_d \ln X_{ij} + \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \varepsilon_{ij}$$

(8)

It is expected that, if significant, the autoregressive coefficient ρ_d as a positive sign. As to the remaining variables, the expected signs are the same as previously mentioned.

After creating the adequate W_d matrix, the calibration of *DSAR* was done using the SAR (Spatial Autoregressive) function in *Econometrics MATLAB toolbox* (LeSage, 1998), which comprises maximum likelihood estimation method. Table 2 presents the principal results of this application.

[11] It should be noted that the way of constructing the *W* matrix depends on the way in which the Origin-destination (O-D) flow matrix is organized. In LeSage (2005), for example, this matrix is organized putting origins as columns and destinations as rows. In the present work's case, the O-D matrix is assembled in a reversed way (origins in rows and destinations in columns). That is the reason why matrix W_d is constructed in a different way here.

Table 2 – *DSAR Model* principal results.

	R-bar squared	Statistically insignificant coefficients (5%)	Coeff. signs equal to expected?	β_5 (distance coeff.)
FBT	80%	β_0, β_3 and ρ_d	yes	-0,13
TEX	85%	β_0, β_1 and ρ_d	no: negative β_1 and ρ_d	-0,13
WOO	61%	β_3	yes	-0,11
PPP	79%	ρ_d	no: negative ρ_d	-0,11
CHE	92%	ρ_d	no: negative ρ_d	-0,14
OTH	86%	β_0 and ρ_d	no: negative ρ_d	-0,13
MET	89%	ρ_d	no: negative ρ_d	-0,14
MAQ	94%	ρ_d	no: negative ρ_d	-0,11
EQT	80%	β_0, β_3 and ρ_d	yes	-0,13
NEC	86%	ρ_d	no: negative ρ_d	-0,11

There are three main observations to make on these results. Firstly, it seems that spatial dependence is negligible, since autoregressive coefficient ρ_d is statistically insignificant in all but one product[12]. Secondly, further than being insignificant, ρ_d also presents a negative sign in seven of ten cases, which is contrary to what would be expected. Finally, the distance resistance coefficients have now much smaller values than in *Model 1*, being far from the commonly obtained values in similar gravity trade studies[13]. The presence of a new explanatory variable (the spatial lag) capturing some spatial effects previously

[12] Yet, it assumes a low value. In WOO case, ρ_d =0,13.

[13] Usually the distance coefficient is around unity.

captured by β_5 is not an acceptable reason to explain this difference, because this variable does not appear to be statistically significant.

One possible reason for the non-significance of the auto-regressive coefficient may rely on the type of dependence considered in the *W* matrix. Thus, another hypothesis was taken into account: the possibility of "Origin-based" spatial dependence between the observations. In fact, "it seems plausible that forces leading to flows from any origin to a particular destination may create similar flows from neighbors to this origin to the same destination." (LeSage, 2005, p. 7). In order to capture this potential effect, an origin-based weight matrix can be computed by: $W_o = C \otimes I_{14}$. The resulting *origin-based spatial autoregressive* (OSAR) model is:

$$\ln X_{ij} = \rho_o W_o \ln X_{ij} + \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \varepsilon_{ij} \qquad \textbf{(9)}$$

The results of this model's calibration are summarized in the following Table 3. Two main features distinguish this table from Table 2. First, the origin-based auto-regressive coefficient is now statistically significant in half of the cases under consideration; this implies that the origin spatial dependence hypothesis is more robust than the destination dependence one. Yet, even in those cases, the ρ_o assumes very low, negligible values, except for the WOO case, illustrating a rather weak spatial dependence. Second, there is now better agreement between the excepted signs to the parameters and the obtained ones. In short, these two features support the choice of this *OSAR Model*, in detriment of the *DSAR Model*.

Table 3 - O*SAR Model* principal results

	R-bar squared	Statistically insignificant coefficients (5%)	Value of ρ_o, if significant	Coeff. signs equal to expected?	β_5 (distance coeff.)
FBT	82%	β_0	0,09	yes	-0,13
TEX	84%	β_0 and β_1	0,05	no: negative β_1	-0,13
WOO	66%	β_3	0,22	no: negative β_3	-0,11
PPP	80%	none	0,08	yes	-0,12
CHE	92%	ρ_o	---	no: negative ρ_o	-0,13
OTH	90%	ρ_o, β_0 and β_1	---	yes	-0,11
MET	94%	ρ_o	---	yes	-0,13
MAQ	85%	ρ_o	---	yes	-0,14
EQT	82%	β_0	0,09	yes	-0,13
NEC	86%	ρ_o	---	yes	-0,12

To complete this SAR analysis, a third type of W matrix was considered: $W_{od} = W_o \cdot W_d$. This matrix aims to capture an "origin-destination" mixed effect of spatial dependence. The inclusion of spatial lag $W_{od} \ln X_{ij}$ in the regression equations means that: flows from *i* to *j* are influenced, among other factors, by the average of flows from all the neighbors of *i* to all the neighbors of *j*. This weight matrix can be computed as the *Kronecker* product: $W_{od} = C \otimes C$ (LeSage, 2005). The regression equation, corresponding to *ODSAR Model*, becomes:

$$\ln X_{ij} = \rho_{od} W_{od} \ln X_{ij} + \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \varepsilon_{ij} \tag{10}$$

The principal results of this model are exhibited in Table 4.

Table 4 - *ODSAR Model* principal results

	R-bar squared	Statistically insignificant coefficients (5%)	Value of ρ_{od}, if significant	Coeff. signs equal to expected?	β_5 (distance coeff.)
FBT	83%	β_0 and β_3	0,13	yes	-0,12
TEX	85%	β_0, β_1 and ρ_{od}	---	no: negative β_1	-0,13
WOO	70%	none	0,30	no: negative β_3	-0,08
PPP	82%	β_3	0,15	yes	-0,10
CHE	92%	ρ_{od}	---	yes	-0,13
OTH	86%	β_0 and β_1	0,08	yes	-0,10
MET	87%	β_0	0,04	yes	-0,12
MAQ	95%	β_3	0,08	yes	-0,13
EQT	83%	β_0 and β_1	0,13	yes	-0,11
NEC	87%	β_0	0,09	yes	-0,11

The comparison between these results and the previous *SAR* results allow us to conclude that, when a mixed effect is considered, the spatial dependence hypothesis obtains a superior support, since the autoregressive coefficient is now statistically significant in eight of ten cases. However, once again, its absolute value is very low in all but one product (the same as before: WOO)[14].

Given that the spatial autoregressive model didn't seem a suitable model to explain international trade flows in all the products being considered, a third model was tested. *Model 3* uses the same explanatory variables as in *Model 1*, added with two new ones.

The first added variable is product specific and reflects the effect of each country's specialization on the volume of exports. Note that in *Model 1*, the vector of explanatory variables was the same, independently of the specific product in study. However, in some cases, de degree of specialization of some country in exporting a specific product *k* has an influence that may even prevail over the distance effect. At the limit, if one country had the monopoly in the international market supply of some specific product *k*, distance would not matter at all; all the product *k* demand would be satisfied by exports from that country. Consider, for example, the product "WOOD AND PRODUCTS OF WOOD AND CORK". The weight of this product exports on total exports of Finland is pretty above the average. More precisely, is almost 7 times the correspondent weight in the whole of countries being considered. Formally, this can be expressed by a Degree of Specialization (DS) indicator, given by:

$$DS_i^k = \frac{\dfrac{X_i^k}{\sum_{k=1}^{10} X_i^k}}{\dfrac{\sum_{i=1}^{14} X_i^k}{\sum_{i=1}^{14}\sum_{k=1}^{10} X_i^k}} \qquad (11)$$

[14] Besides, the asymptotic t-statistic is relatively low in all cases, being close to the limit of statistical significance.

The numerator of this index represents the weight of product *k* on origin *i*'s total exports[15]; the denominator indicates the weight of product *k* on all origins' exports. Values above (below) 1 indicate a higher (lower) than average specialization of country *i* in exporting product *k*. Table 5 shows the maximum values of this index obtained for each product and the correspondent highly specialized country.

The second added variable seeks to reflect the map pattern of the observations in a simpler way than that used in *Model 2*: through the inclusion of a dummy variable that indicates the presence (or not) of a common border between the trading partners (*DBor)*. *Model 3* may, thus, be expressed by:

$$\ln X_{ij} = \beta_0 + \beta_1 \ln N_i + \beta_2 \ln POP_i + \beta_3 \ln N_j + \beta_4 \ln POP_j + \beta_5 \ln d_{ij} + \beta_6 \ln DS_i + \beta_7 DBor + \varepsilon_{ij}$$

(12)

Table 5 – Maximum value of DS, for each product, and correspondent specialized country.

	Max SD	Country
FBT	1,746	Spain
TEX	4,741	Portugal
WOO	6,757	Finland
PPP	6,599	Finland
CHE	1,489	Ireland
OTH	2,078	Portugal
MET	1,772	Greece
MAQ	1,730	Ireland
EQT	1,746	Spain
NEC	2,233	Italy

[15] $X_i = \sum_j X_{ij}$

From the previous explanation, it is clear that the expected signs for the new variables estimated coefficients are both positive.

Table 6 - *Model 3* principal results.

	R-bar squared	Statistically insignificant coefficients (5%)	Coeff. signs equal to expected?	β_5 (distance coeff.)	β_6 (DS coeff.)	β_7 (dummy coeff.)
FBT	90%	β_7	yes	-0,89	1,43	0,31
TEX	67%	none	yes	-0,83	0,92	0,60
WOO	86%	β_3	yes	-1,24	1,32	0,57
PPP	89%	none	yes	-0,97	1,32	0,40
CHE	87%	none	yes	-0,92	1,29	0,41
OTH	81%	none	yes	-0,97	0,97	0,75
MET	86%	none	yes	-0,86	0,85	0,75
MAQ	89%	none	yes	-0,67	1,34	0,43
EQT	90%	β_7	yes	-0,89	1,42	0,32
NEC	85%	none	yes	-0,95	1,00	0,59

The main estimation results, obtained by means of OLS, are shown in Table 6[16]. This model exhibits a better performance than *Model 1* (to which it can be compared). Several reasons support this statement:

[16] The White test application to *Model 3* results shows that, for all products, the error terms are homoskedastic.

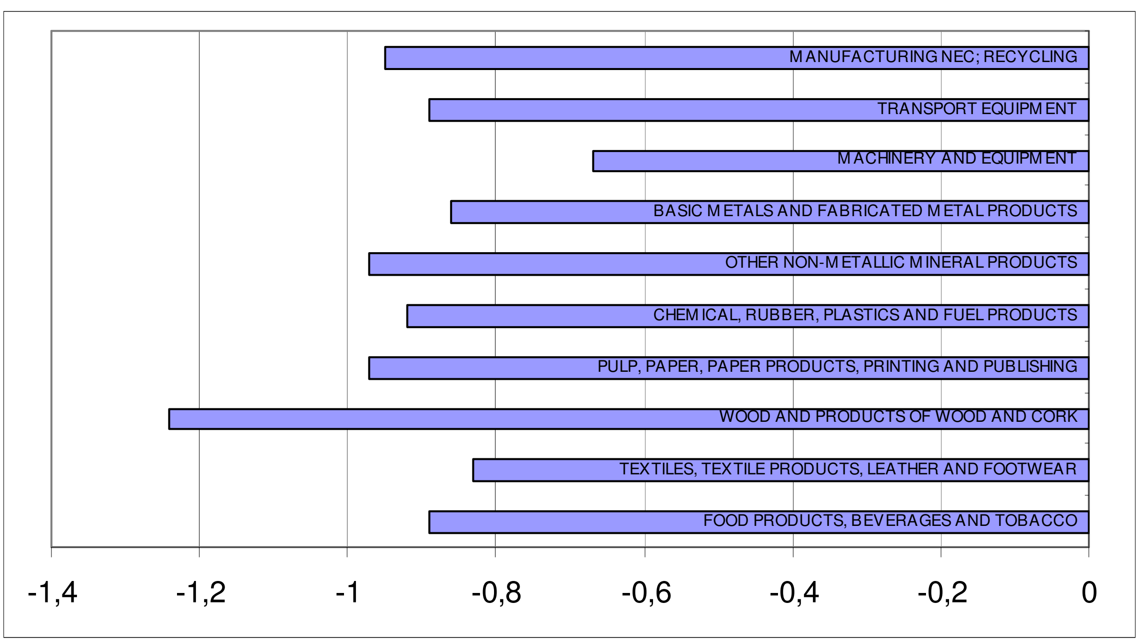
- Predictive capacity indicated by R-bar squared is rather superior in *Model 3* than in *Model 1*; the improvement is remarkable, for example, in WOO, in which this indicator increased from 42% to 86%.

- Statistically insignificant coefficients are rare, unlike in *Model 1*. Yet, it should be noted that, in two cases (FBT and EQT) the dummy variable coefficient is not significant, indicating that map pattern is not a determinant factor to export flows of these products.

- All the estimated coefficient signs correspond to what was expected *a priori*.

Additional comments on these results are pertinent. First, in most of the cases, the distance resistance coefficient assumes smaller values than in *Model 1*. This may be due to the fact that, in *Model 3*, there are additional variables that control for factors previously embodied in the distance coefficient. Once again, this coefficient exhibits great variability, assuming superior values in products that typically have a high weight for monetary unit. This is clearly the case in Wood products, as can be seen in Figure 3:

Figure 3 - Estimated distance parameter in *Model 3*, for the ten products.



Second, estimated *DS* coefficients show a high sensibility of exports with respect to origin's specialization on the specific product under study. All coefficients are close to or greater than unity. Finally, except for the cases in which β_7 is statistically insignificant, the presence of a common border between trading partners has a considerable effect on the correspondent level of exports. Since the dependent variable is in logarithmic form (as well as the remaining explanatory variables), a coefficient β_7 represents a common border additional impact of $\left[\exp(\beta_7)-1\right]$[17]. This means that in OTH, for example, the presence of a common border increases the level of exports in 112%.

1.4. Type (b) information context.

It is very common to find situations in which the required data are not directly available and the survey collection of them is beyond the researcher's means (in terms of money, time, human resources, etc). In this context, non-survey methods are used to estimate the missing data. The objective of this section is to test a gravity trade model application when the purpose is to generate nondisclosed data[18]. The concern in the context of missing trade data is especially pertinent when the researcher is dealing with regions, instead of countries. In fact, one of the main barriers to regional economic analysis is the

[17] This can be easily demonstrated, as follows.

$$DBor = 1 \Rightarrow \ln \hat{X}_{ij} = \hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i + \hat{\beta}_7 DBor \Leftrightarrow \hat{X}_{ij} = \exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i + \hat{\beta}_7\right)$$

; $DBor = 0 \Rightarrow \ln \hat{X}_{ij} = \hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i \Leftrightarrow \hat{X}_{ij} = \exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i\right)$ Being so,

$$\Delta\hat{X}_{ij} = \exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i + \hat{\beta}_7\right) - \exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i\right) =$$

$$\exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i\right) \cdot \exp(\hat{\beta}_7) - \exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i\right) =$$

$$\exp\left(\hat{\beta}_0 + \cdots + \hat{\beta}_6 \ln DS_i\right) \cdot \left[\exp(\hat{\beta}_7) - 1\right]$$

[18] In this paper, attention was exclusively given to the gravity model. However, it must be noted that there are alternatives that should be accounted for in further studies. Yet, even without dedicating special attention to these models, it may be argued that there are serious problems when it comes to empirical applications. The empirical use of the entropy maximizing model, for instance, requires the introduction of a cost constraint, for which data are seldom available. Another optimizing type spatial interaction model is the Minimum Discrimination of Information approach (minutely presented in Snickars and Weibull, 1977). But this model requires the access to previous information on the spatial interaction flows which is not the case considered in type (b) information context. Further details on these alternative spatial interaction models can be found, for example, in Roy and Thill (2004), Fotheringham and O'Kelly (1989) and Batten and Boyce (1986).

fact that, in most countries, there is no survey gathering of interregional trade flows data (Canning and Wang, 2003). The knowledge of this data is of fundamental importance, for example, in regional input-output models that include more than one region, since each region's demand is, in part, supplied from other national regions and each region's supply is, in part, directed to supply other national regions' demand. Being so, when an exogenous change occurs in one region's final demand, there are interregional feedback effects that can only be accounted for when interregional trade flows are known.

Regional input-output models are the main theme of investigation of this paper's author. This motivates a special concern in type (b) information contexts. Nevertheless, the following empirical application uses the same international trade data that was used in the previous section (also in a product disaggregated manner). This is because it would not be possible to evaluate gravity model performance in estimating interregional trade data, without having the real value of trade flows, to serve as benchmark. Being so, the European set of open economies is used as if it was a large country with several regions. Additionally, it will be assumed that, for every product, the only known information is the sum of all inflows into each region and of all outflows from each region.

Figure 4 – International trade matrix with known margins.

Destination / Origin	**Country 1**	**Country 2**	**…**	**Country 14**	**Sum**
Country 1	0	X_{12}	…	X_{114}	X_1
Country 2	X_{21}	0	…	X_{214}	X_2
…	…	…	0	…	
Country 14	X_{141}	X_{142}	…	0	X_{14}
Sum	M_1	M_2	…	M_{14}	$X = M$

Considering an n*n (14*14, in this case) trade matrix, as the one in Figure 4, this can be formally expressed by: $\sum_{i=1}^{14} X_{ij} = M_j$ and $\sum_{j=1}^{14} X_{ij} = X_i$ are previously known values. The total volume of flows is indicated below by $X = M = \sum_{i=1}^{14}\sum_{j=1}^{14} X_{ij}$.

Let $\tilde{X}_{ij}$ be the international trade flow from origin *i* to destination *j*, obtained by the gravity model application. Recalling the gravity equation (1), trade flow estimative could be obtained by:

$$\tilde{X}_{ij} = G\frac{P_i^{\alpha_1} P_j^{\alpha_2}}{d_{ij}^{\alpha_3}}$$

(13)

However, the problem here is that parameters G, α_1, α_2 and α_3 are unknown, which makes it impossible to directly apply the previous formula. On the other hand, in the information context described before, the model is doubly-constrained (Isard, 1998). That means that, independently of the particular version of the model that is used, the estimated values must verify the following additivity constraints:

$$\sum_j \tilde{X}_{ij} = X_i$$
$$\sum_i \tilde{X}_{ij} = M_j$$

(14)

Whenever this occurs, it is appropriate to use a gravity based model to generate a first proxy to international trade matrix and make use of biproportional techniques, like the popular RAS scaling algorithm[19], for example, to make that initial matrix as close as possible to the real matrix (this is done, for example, in Isard, 1998). One major question arise: how close are the obtained $\tilde{X}_{ij}$ flows from the real ones? The following experience aims to give an answer to this issue.

Let $\tilde{X}_{ij}^{0}$ designate the elements of the initial matrix. In this experience, $\tilde{X}_{ij}^{0}$ is obtained applying a particular version of equation (13), in which almost all the unknown parameters are arbitrarily set equal to one. Thus, we have:

$$\tilde{X}_{ij}^{0} = G_i \frac{P_i P_j}{d_{ij}};$$
$$G_i = X_i \left(\sum_j \frac{P_i P_j}{d_{ij}} \right)^{-1}$$

(15)

In this experience, only the basic gravity variables were used to determine the initial values of trade flows. The constant of proportionality G_i is a scalar that guarantees the exact observance of the *i*th row summing up constraint: $\sum_j \tilde{X}_{ij}^{0} = X_i$[20]; it is introduced in order to make the initial matrix comparable to the real one (if no scalar was introduced, the values of both matrices would have considerably different values). α_1, α_2 and α_3 are assumed to be unitary (similarly to *Model 1*, in information context (a)). The initial matrix was iteratively adjusted, making use of the known margins. The algorithm converged after six iterations.

[19] The application of RAS, as well as alternative biproportional techniques, is well described in Lahr and de Mesnard (2004), to which the interested reader is referred.
[20] The introduction of this scalar is equivalent to a first iteration of the RAS procedure, in which the row sums are the first to be adjusted.

To evaluate the performance of this model in assessing the real international trade values, the following measure of distance between matrices was used:

$$MAPE = 100 \cdot \frac{\sum_i \sum_j \left| X_{ij} - \tilde{X}_{ij} \right|}{\sum_i \sum_j X_{ij}} \tag{16}$$

in which *MAPE* stands for: Mean Absolute Percentage Error.

This measure was computed in two different stages of the process: before applying RAS (indicating the distance between the initial matrix and the real one) and after applying RAS (indicating the distance between the estimated final matrix and the real one). Table 7 presents the obtained results for all the ten products. These results show that the initial matrix is quite distant from the real one (with a mean error around or above 50%). The iterative adjustment allows for some improvements in the matrix, making it closer to the real one. In some products, like CHE, for example, the resulting error is rather low. However, generally speaking, the final matrix is still very distant from the real one[21]. The most obvious sources of error rely on the following assumptions of the model:

- all the unknown parameters are unitary;
- only the basic gravitational variables are included in the model.

[21] In previous works, namely in Ramos and Sargento (2003) and in Sargento and Ramos (2003), a different version of the gravity model was used to generate the initial matrix. In that case, the distance parameter was determined by way of minimizing a certain indicator of error of the initial matrix. Yet, the performance of this model remained unknown, because it was applied to interregional trade flows, to which there was no benchmark. This model was not sufficiently tested in the present international trade flow application.

Table 7 – Results from *Type (b) Gravity Experience*.

Product	*MAPE* (before RAS)	*MAPE* (after RAS)
FBT	42%	30%
TEX	88%	48%
WOO	62%	56%
PPP	41%	39%
CHE	47%	23%
OTH	47%	37%
MET	47%	32%
MAQ	46%	29%
EQT	42%	30%
NEC	48%	42%

As to the first source of error, the alternative would be to consider the values estimated in the first part of this paper. Although, one must not forget that in the information context under consideration, no *a priori* data exists to calibrate the model. It could be argued that parameters estimated making use of international trade data could be used in interregional trade estimation. However, since regions and countries are quite different geographical units, this assumption needs to be empirically tested, before being adopted. One useful experience, to develop in further studies, would be to calibrate an interregional gravity trade model in a country with available interregional trade data (United States or Canada, for instance) and then, evaluate the proximity between its parameters and the ones obtained in this paper's model. The second source of error can be eliminated by considering additional variables in the model (similarly to *Model 3*). Nevertheless, in this case, the problem of the unknown parameters becomes more serious, since it would be necessary to assume more parameters as being unitary. Further experiences have to be done to test the improvement that may be generated by the inclusion of additional variables.

4. Conclusions.

In the present paper, gravity model performance was empirically tested, assuming two different contexts of information availability (previously referred as contexts a) and b)). The several experiences undertaken in this work confirm that gravity model generates quite good results when trade flows are known *a priori*, i.e., when the model is used with an explanatory purpose.

Contrarily to what has been done in most of the gravity trade model econometric applications, a spatial autoregressive model was tested, considering three different types of spatial dependence. The obtained results show that spatial dependence is statistically relevant in the majority of the products when a mixed origin-destination spatial effect is accounted for; but, even in this case, the autoregressive coefficient assumes very low values, indicating a very weak degree of spatial dependence. This led the author to search for a non-spatial model that performs better in this particular trade flow application.

The results obtained from two alternative non-spatial econometric equations allow for the conclusion that the augmented version of the gravity model is the best suited to explain trade flow behavior. This version comprises, besides the traditional variables of mass and spatial separation, two additional variables: one dummy to capture spatial contiguity effect and a product specific variable that represents the origin's relative specialization on the exportation of product *k*.

The fact that these econometric exercises were separately applied to distinct products also made clear that each different traded product has its own specificity, originating quite variable estimated coefficients. Special attention was given to the distance coefficient, showing that, as expected, this tends to be higher in products with high weight for monetary unit. Furthermore, the product disaggregated application is very important because, in most of the times, the gravity model is used in the construction of larger models, like input-output models, which require product specific estimated values.

As to the second information context, the results are not so satisfactory. In spite of being an incipient experience, which may be subject to some previously referred improvements, this empirical test suggests that the gravity model is not the most adequate to generate undisclosed trade data.

5. References.

- Anselin, L. 1999. "Spatial Econometrics". Department of Agricultural and Consumer Economics. University of Illinois at Urbana-Champaign
- Anselin, L. and Griffith, D. 1988. "Do spatial effects really matter in regression analysis?" Papers of the Regional Science Association, 65: 11-34.
- Batten, D. and Boyce, D. 1986. "Spatial interaction, transportation and interregional commodity flow models". In: Peter Nijkamp (ed.), Handbook of Regional and Urban Economics, Vol. 1, pp. 357-406. Free University, Amsterdam.
- Beck, N., Gleditsch, K. and Beardsley, K. 2005. "Space is more than Geography: using Spatial Econometrics in the study of political economy". Working Paper, www.nyu.edu/gsas/dept/ politics/faculty/beck/becketal.pdf, (2005, November, 15th).
- Blavy, R. 2001. "Trade in the Mashreq: an empirical examination", International Monetary Fund, IMF Working paper 01/163.

- Canning, P., Wang, Z. 2003. "A Flexible Modeling Framework to estimate interregional trade patterns and input-output accounts". Paper prepared for presentation at the Sixth Annual Conference on Global Economic Analysis, June 12-14, Scheveningen, The Netherlands.
- Feenstra, R., Markusen, J. and Rose, A. 1998. "Understanding the home market effect and the gravity equation: the role of differentiating goods". CEPR Discussion Paper no. 2035. London, Centre for Economic Policy Research. http://www.cepr.org/pubs/dps/DP2035.asp.
- Fotheringham, A. S. and M. E. O'Kelly. 1989. "Spatial interaction models: formulations and applications". Dordrecht, Netherlands: Kluwer Academic Publishers.
- Hu, P. and Pooler, P. 2002. "An empirical test of the competing destinations model", Journal of Geographical Systems, 4: 301-323.
- Isard, W. 1998. “Gravity and spatial interaction models”. In: Walter Isard et al., Methods of interregional and regional analysis, pp. 243-280. Ashgate Publishing Limited.
- Lahr, M. and de Mesnard, L. 2004. “Biproportional techniques in input-output analysis: table updating and structural analysis”. Economic Systems Research, 16(2): 115-134.
- LeSage, J.P., 1998. “Spatial Econometrics”. Department of Economics, University of Toledo. www.spatial-econometrics.com.
- LeSage, J.P. and Pace, R.K. 2005. "Spatial econometric modeling of origin-destination flows". Paper presented at the North American Meetings of the Regional Science Association International, Las Vegas, Nevada, November, 2005.
- Martinez-Zarzoso, I. 2003. "Gravity model: an application to trade between regional blocks", Atlantic Economic Journal, 31: 174-187
- OECD, Bilateral Trade Database, edition 2002.

- OECD, OECD member countries' population 1981-2004 (thousands and indices: 2000=100). Labour Force Statistics, edition 2005.
- OECD, Gross Domestic Product for OECD Countries, National Accounts of OECD Countries, Vol. 1, edition 2005.
- Piani, G. and Kume, H. 2000. "Fluxos bilaterais de comércio e blocos regionais: uma aplicação do modelo gravitacional", Instituto de Pesquisa Economia Aplicada. Texto para discussão nº749.
- Porojan, A. 2001. "Trade flows and spatial effects: the gravity model revisited", Open Economies Review, 12(3): 265-280.
- Ramos, P. and Sargento, A. (2003). “Estimating Trade Flows between Portuguese Regions using an Input-Output approach”. Paper presented at the 43rd Congress of the European Regional Science Association, Jyvaskyla (Finlândia), Agosto de 2003.
- Roy, J. R., Thill, J. 2004. "Spatial interaction modelling", Papers in Regional Science, 83: 339-362.
- Sargento, A. and Ramos, P. (2003). “Input-output based methodology for the estimation of interregional trade matrices in Portugal”. Paper presented at the 50th North American Meetings of the Regional Science Association International, Philadelphia, November, 2003.
- Sen, A. and Smith, T. 1995. "Gravity models of spatial interaction behavior", Springer-Verlag, Berlin.
- Snickars, F., Weibull, J. 1977. "A minimum information principle", Regional Science and Urban Economics, 7:137-168
- Soloaga, I. and Winters, A. 1999. "Regionalism in the nineties: what effects on trade?", Development Economic Group of the World Bank.