PARAMETRIC VS. NON PARAMETRIC TRADE GRAVITY MODELS: AN APPLICATION TO THE FREIGHT CORRIDOR BETWEEN ITALY AND CHINA

Mariano Gallo, Dipartimento di Ingegneria, Università del Sannio (Italy)

Vittorio Marzano, Dipartimento di Ingegneria dei Trasporti, Università di Napoli (Italy)

Fulvio Simonelli, Dipartimento di Ingegneria dei Trasporti, Università di Napoli (Italy)

ABSTRACT

This paper proposes and compares several parametric and non-parametric trade gravity models for estimating freight transport volumes of international connections. In detail, the kernel of the research focuses specifically on different types of estimation gravity models, following both panel-data parametric and non parametric regression approaches: the outcomes of the two different approaches are presented and contrasted, providing for interesting results both from the theoretical and the practical standpoints. The models are calibrated and tested on the case of freight corridor between Italy and China and, in order to obtain a regional focus, also a port choice model is proposed and calibrated.

Gravity model, kernel regressions, regression trees, Italy - China freight flows

1. INTRODUCTION

This paper deals with the activities carried out within the context of an Italian research project aimed at analyzing trade exchanges in quantity between Italian regions and China, together with the corresponding impacts on the freight container flows between the two Countries, with specific reference to the entry/exit ports from the Italian side, and on the economy of each Italian region. For this aim, a system of models has been implemented with the general framework reported in Figure 1.
Inside that general framework, this paper focuses on models for forecasting freight volumes between Italy and China (i.e. the import/export flows model), exploring two different approaches: parametric and non-parametric models. Both kinds of models are applied to the specific problem and compared in terms of goodness of results and forecasting capacity.

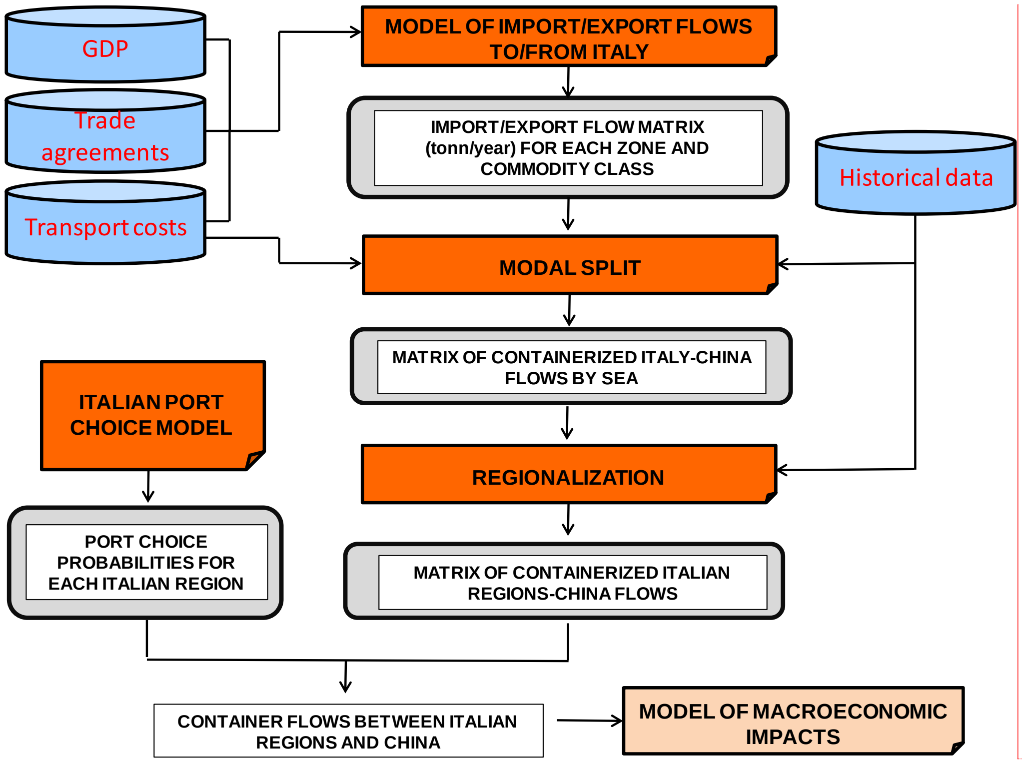

Figure 1 – Structure of the system of models for Italy-China trade flow analysis

Comment [MG1]: Correggere CHNIA in CHINA in Matrix of Containerized… Correggi: regionalization con regionalisation

As parametric models, gravity models are specified and calibrated; they are able to reproduce import/export trade flows, for 10 NST/R 1-digit commodity nomenclature classes, between Italy and 13 world economic zones (included China as a single zone), as a function of impedances (transport cost, tariffs), origin and/or destination masses (GDP, total trade) and dummy variables representing economical and other kind of agreements. Notably, this kind of model allows reproducing both the demand generated by a decrease of trade impedances and the effect of competition among economic zones. Even if the focus regards freight volumes between Italy and China, for applying the gravity models we need to consider all concurrent economic zones. Parametric models are widely adopted in the literature for estimating passenger and freight transport volumes.
As non-parametric models in this paper we have tested the Kernel regression and the regression trees. All models were calibrated using the same data.
Then, flows in tons/year between Italy and China for each commodity class, coming from the gravity model, are in turn disaggregated by mode, using temporal series of modal shares (taking into account that the sea mode is the prevailing mode) and then regionalized among the 20 Italian regions. This allows calculation of the freight flows by container from/to each Italian region to/from China.
Finally, for each Italian region, the probability of choosing one of the 21 Italian ports with direct or transshipment connections with China is calculated through a port choice model.

The functional form is a Multinomial Logit, and the systematic utility of each port is expressed as a function of the transport costs of passing through that port and of the capacity of the direct/transshipment services calling at it. Notably, the structure of the model is such that the impacts of future scenarios with different shares of transshipment services vs. direct services can be explicitly modeled.

Therefore, the proposed system of models can be applied to future scenarios involving changes both in the transport sector (e.g. variation of the supply of maritime services to/from China, change in oil price, trends in shipping fares) and in the economic sector (e.g. new economic agreements and/or GDP changes impacting on the competition among world economic zones), providing as a result the matrix of container flows to/from each Italian region and China, with explicit indication of the entry/exit port from the Italian side. This matrix can be in turn adopted as input for a MRIO model predicting macroeconomic impacts, i.e. regional GDP change.

Within this modelling framework, the most significant theoretical contribution has been achieved with reference to the gravity model, whilst the remaining modelling steps have been faced through models already available in the literature. Notably, the performances of panel-data parametric and non parametric regression approaches for the implementation of gravity models have been extensively explored, leading to interesting outcomes from both a theoretical and a practical perspective: indeed, the motivation for the analysis of such different approaches comes mainly from the need of exploring their elasticities and the perspective different forecasts they may provide. Consistently, the paper is organised as follows: section 2 provides for a brief literature review of parametric and non-parametric gravity models; parametric and non-parametric gravity models are examined in section 3 and are specified, calibrated and compared in section 4; section 5 summarises discussion, conclusions and further research.

Comment [MG2]: La parte evidenziata credo debba essere spostata altrove (ad esempio dove si parla dei modelli parametrici), mettendo nell'introduzione solo una sintesi delle altre parti del modello generale.

2. LITERATURE REVIEW

Modelling international trade flows is one of the most consolidated and significant research topics in transport and geography. In that respect, gravity models are normally regarded as the most efficient and effective modelling tool for reproducing trade exchanges between countries (e.g. Porojan, 2001). Mimicking Newtonian physics, they express trade flow between two zones in a study area as a direct function of masses of origin and destination zones (e.g. GDP, total trade, population) and as an inverse function of impedances between origin and destination (e.g. transport costs, custom duties). Such regression is normally defined in log-linear form, for the sake of simpler analytical tractability, leading to a constant elasticity model.

Several gravity models have been proposed to date, with remarkable variety in the reference context (e.g. geographical and commodity coverage), specification (e.g. choice of explanatory variables) and estimation. Normally, most of the models and studies proposed to date adopted a parametric approach. A detailed analysis of the topic goes beyond the aim of this paper, however the reader may refer to the thorough review recently reported in Kepaptsoglou et al. (2009) with the related bibliography. In detail, starting from the naive assumption of uncorrelated disturbances across countries and years, typical of the seminal

specifications, nowadays panel data estimation approaches are widely explored in the theory and applied in the practice. That is, explicit correlation is assumed across O-D pairs in the study area (cross-sectional dimension) and across years within the time horizon (temporal dimension), leading to substantial advantages in modelling the phenomenon (Washington et al., 2003). Furthermore, recent contributions show how it is important to account also for correlation across commodity classes, leading to a Seemingly Unrelated Regression Estimation (SURE) problem, initially proposal by Zellner (1962) and subsequently adapted to the panel-data context by Wan et al. (1992). A recent specification of a panel-data SURE gravity model in the Euro-Mediterranean context has been proposed in the already mentioned work by Kepaptsoglou et al. (2009).
Notably, very few researches have proposed to date non-parametric regression approaches within the context of gravity models. The most recent contribution in that respect has been proposed by Coulibaly (2007), who proposed a semi-parametric approach for estimating a gravity model explicitly taking into account international trade agreements.

3. PARAMETRIC AND NON-PARAMETRIC GRAVITY MODELS

The present section, representing the kernel of the paper, deals with the specification and the analysis of the performances of parametric and non-parametric gravity models, applied to the context of the international imports and exports of Italy. In more detail, the target of the gravity model to be specified is to provide for reliable forecasts of the Italian import/export flows in quantities from/to 13 macroeconomic zones in the World, with the China treated as a single zone (Table 1). This choice allows for building a detailed and wide estimation dataset, potentially leading to correct elasticity estimates, and also allows for carrying out all requested scenario simulations. In more detail, policy requirements lead to the explicit inclusion, among the explanatory variables, of transport costs, customs tariffs and duties, trade agreements.

Comment [MG3]: Riportare in questo paragrafo solo la descrizione dei vari modelli, con le caratteristiche in termini di vantaggi e svantaggi. Nel paragrafo 4 riportare i dati utilizzati, la calibrazione ed i confronti

Table 1 – World zonization

Zone no.	Zone name
1	EU15
2	EU members since 2004
3	Other EU countries
4	North Africa
5	Other African countries
6	North America
7	Central/South America
8	Middle East
9	Central Asia
10	Far East
11	Australia
12	China
13	EU members since 2007

In more detail, Section 3.1 provides for a brief description of the estimation database, Section 3.2 deals with the estimation of parametric models, Section 3.3 is focused on non-parametric models, and finally Section 3.4 provides for a results assessment.

3.1 Estimation database

Accordingly with the model to be specified, the estimation database should contain the following set of variables:

1. dependent variable, that is the output variable reproduced by the gravity model;

2. mass variables, representative of the generation and attraction capabilities of each zone;

3. impedance variables, representative of physical (e.g. transport costs, distances) and immaterial (e.g. custom duties, tariffs) limitations to trade between zones;

4. dummies expressing the incidence of further factors, related to either a single zone (origin or destination) or to both zones (i.e. o-d specific dummies).

Taking into account the evolution of trade agreement and freight transport between Italy and China, the time horizon 1996-2006 has been chosen as reference. Furthermore, the NST/R 1-digit commodity nomenclature has been adopted as maximal disaggregation for goods clusterization (Table 2).

Table 2 – NST/R 1-digit commodity nomenclature

NST/R 1-digit commodity classes
0 - Food and live animals
1 - Other food products
2 - Solid mineral fuel
3 - Oil products
4 - Minerals and raw iron materials
5 - Other iron products
6 - Other minerals/materials and construction products
7 - Chemical products for agriculture
8 - Other chemical products
9 - Manufactured goods, vehicles and machines

The dependent variable is made up by trade flows in quantity (tonnes/year for each year in the reference time horizon) between Italy and each of the zones reported in Table 1, disaggregated by direction (import/export), collected from the main national data source of ISTAT COEWEB. A further disaggregation by prevailing transport mode is also available from the same source: it has been adopted in the context of the implementation of the modal split model (Figure 1).
With reference to mass variables, in accordance with suggestions in the literature, GDP expressed in current billion US$ has been firstly collected, for each year of the period 1996-2006, from EUROSTAT source for European countries, from Arab Monetary Fund for most of the African countries and from World Bank data for the remaining countries. Notably, a disaggregation of the GDP by commodity group (Table 2) was not possible for all zones, therefore in the application of the gravity model the overall GDP has been used as mass

variable for each commodity class. Furthermore, other mass variables have been also added to the estimation database for potential inclusion in model specification. In more detail, the total trade volume in import/export for each zone, expressed in current thousand US$, has been determined from UNCTAD COMTRADE source for the whole 1996-2006 period and for each commodity class.
With reference to impedance variables, most of the specifications available in the literature include the straight distance between zone centroids as a proxy of the overall impedance in trade. More properly, transport costs and custom duties/tariffs should be explicitly and separately taken into account in the specification. For this aim, in this study transport costs have been calculated on the basis of previous studies carried out by the research group (e.g. Marzano et al. 2008), with substantial integration from studies available from the Central Bank of Italy (2009), providing detailed information often disaggregated by commodity group and by transport mode. However, a significant armonization effort has been spent in transforming raw data accordingly with the gravity model estimation requirements. As a result, transport costs expressed in US$/tonn per year and per commodity class have been calculated. With reference to custom duties and tariffs, the UNCTAD TRAINS dataset has been used as reference. It provides tariffs data between countries in three different ways: MFN (most favoured nations), i.e. nominal tariffs applied by WTO members; PRF (preferential rates), normally lower than the corresponding MFN tariffs, they account for the presence of formal preferential agreements; AHS (effectively applied tariffs), i.e. those actually applied in trade. Therefore, for the purposes of the study, the AHS tariffs have been adopted. Furthermore, since tariffs are partly expressed as percentage of the value of the traded goods, partly as fixed amounts over a certain trade threshold, an equivalent ad-valorem rate (AVE) has been applied. That is, tariffs are expressed always as percentage of the traded value. Notably, since tariffs are remarkably different among commodity classes with disaggregation much higher than the NST/R 1-digit, an aggregation has been performed through average of tariffs for commodities within each NST/R 1-digit class weighted with the corresponding trade value.
Finally, some dummy variables have been inserted into the estimation database for possible inclusion in model specification. In more detail, three different groups of dummy variables have been taken into account: cultural, historical, political linkage dummies; presence of trade agreements and other kind of preferential trade relationships; relevant geographical characteristics (e.g. common border, island, landlocked and so on).
In conclusion, the implemented database is made up by 2860 records, resulting from the combination of 13 zones, 11 years, 10 commodity classes and 2 flow directions.

3.2 Parametric models

The first step of the model implementation was the estimation of parametric log-linear gravity models on the basis of the estimation database described in Section 3.1. In more detail, a different gravity model has been specified for each commodity class and for trade direction, leading to 20 different models, each one estimated on 143 records (combination of 11 years and 13 zones). A first OLS estimate has been performed in order to define a base reference, then both panel data fixed effects and random effects estimations have been performed as well, not improving the goodness of fit of the regressions with respect to the OLS estimate. A Durbin-Watson test also provided for negative results in data autocorrelation. Finally, a SURE estimation across commodity classes has been carried out, leading to the best estimates reported in the following Tables 3 and 4 respectively for import and export flows, together with some goodness of fit measures usually adopted in the literature (R^2, R^2_{adj} and RMSE).

Table 3 – Parametric SURE across NST/R 1-digit commodity classes: import flows

IMPORT	Explanatory variable											Statistics		
NST/R class	constant	transport cost	dummy Africa	dummy America	dummyChina	dummy EU15	dummyEurope	custom tariffs	GDP world zone	GDP Italy	total trade	R2	R2 adj	RMSE
0 - Food and live animals	6.172	-2.922	-	0.683	-	-	-	-0.194	-	1.360	0.714	0.6595	0.6471	0.8874
	2.321	*-8.272*		*2.139*				*-1.598*		*3.791*	*10.42*			
1 - Other food products	11.949	-0.764	-	-	-	-	-	-0.358	0.332	-0.347	0.389	0.5893	0.5744	0.9224
	4.178	*-2.179*						*-3.005*	*4.486*	*-0.8993*	*5.135*			
2 - Solid mineral fuel	-18.308	-3.289	6.199	-1.313	-	-7.888	-	-	-	-	2.887	0.4529	0.4329	3.2254
	-2.748	*-2.145*	*4.214*	*-1.038*		*-3.664*					*9.148*			
3 - Oil products	22.076	-5.808	3.150	1.150	-	-4.084	-	-2.806	-	-	1.215	0.3921	0.3653	3.1155
	3.442	*-3.9*	*2.678*	*1.236*		*-2.173*		*-2.296*			*7.279*			
4 - Minerals and raw iron materials	21.478	-1.446	1.750	4.836	-	-	-	-1.280	-	-	-	0.3774	0.3594	2.0071
	5.896	*-1.83*	*2.148*	*7.597*				*-5.703*						
5 - Other iron products	11.428	-0.777	0.778	-	-	-	2.084	-0.360	-	-	0.333	0.5409	0.5242	1.0078
	5.022	*-1.631*	*2.105*				*7.316*	*-1.78*			*6.329*			
6 - Other minerals/materials and construction products	12.155	-2.150	-	-	-	-	-	-0.341	-	0.485	0.464	0.1545	0.1300	1.5182
	3.151	*-3.427*						*-1.155*		*0.9456*	*5.937*			
7 - Chemical products for agriculture	33.862	-7.162	3.141	3.937	-	-2.468	-	-2.195	-	1.736	-	0.4333	0.4083	2.8246
	4.652	*-7.67*	*3.426*	*5.13*		*-1.68*		*-2.716*		*1.905*				
8 - Other chemical products	15.761	-1.311	-	1.496	-	-	0.970	-0.767	0.279	-	0.156	0.6686	0.6540	0.9082
	6.394	*-2.856*		*4.67*			*2.73*	*-4.469*	*2.664*		*1.765*			
9 - Manufactured goods, vehicles and machines	16.129	-1.919	-	-	2.777	0.564	-	-0.515	0.478	0.449	0.036	0.7979	0.7874	0.5901
	7.722	*-13.75*			10.88	*1.306*		*-2.002*	*8.771*	*1.649*	*0.9764*			

Table 4 – Parametric SURE across NST/R 1-digit commodity classes: export flows

EXPORT	Explanatory variable									Statistics		
NST/R class	constant	transport cost	dummy America	dummyChina	dummy EU15	GDP world zone	GDP Italy	custom tariffs	total trade	R2	R2 adj	RMSE
0 - Food and live animals	-2.514	-2.299	-1.679	-	0.119	-	1.989	-	0.682	0.6597	0.6473	0.9438
	-0.891	*-7.56*	*-8.118*		0.2518		*4.305*		*11.87*			
1 - Other food products	2.576	-1.007	-	-2.596	-	-	1.023	-0.171	0.575	0.7235	0.7135	0.9627
	1.017	*-3.956*		*-7.987*			*2.605*	*-3.718*	*12.01*			
2 - Solid mineral fuel	-30.437	-6.924	-	-4.434	0.261	-	9.761	-	-	0.3153	0.2954	3.2483
	-3.274	*-6.146*		*-4.409*	0.1837		*6.112*					
3 - Oil products	-3.169	-4.405	-	-3.751	-3.040	-	4.178	-	0.478	0.2632	0.2363	2.1030
	-0.5477	*-6.239*		*-5.885*	-3.536		*4.328*		*4.555*			
4 - Minerals and raw iron materials	-14.797	-3.134	-1.986	1.559	-0.927	0.735	4.764	-	-	0.2999	0.2691	1.9550
	-2.39	*-3.916*	*-3.396*	*2.11*	-0.7924	*5.575*	*4.359*					
5 - Other iron products	-4.230	-1.934	-	-0.765	-	-	2.634	-0.235	0.464	0.5574	0.5413	1.0092
	-1.45	*-6.569*		*-2.499*			*5.76*	*-3.367*	*8.595*			
6 - Other minerals/materials and construction products	8.814	-0.258	-	-	-	-	-	-0.600	0.408	0.6069	0.5984	0.8668
	6.217	*-1.078*						*-7.011*	*6.173*			
7 - Chemical products for agriculture	-14.330	-0.722	-	-7.506	-	-	2.112	-	0.636	0.4556	0.4399	2.3110
	-2.214	*-1.339*		*-9.061*			*2.189*		*5.778*			
8 - Other chemical products	3.928	-1.204	-	-0.482	-	-	1.481	-0.264	0.295	0.5723	0.5567	0.8502
	1.627	*-5.708*		*-1.982*			*4.14*	*-4.457*	*8.148*			
9 - Manufactured goods, vehicles and machines	8.131	-1.387	-	-0.890	-	0.305	1.452	-0.195	0.038	0.7423	0.7310	0.5859
	5.235	*-13.47*		*-4.671*		*8.429*	*6.701*	*-6.641*	*1.266*			

Notably, not all variables were significant for each commodity, and for a given commodity different sets of explanatory variables have been introduced for import and export respectively. There is a remarkable heterogeneity in R^2 values among commodities, with satisfactory values (compared to the average model performances in the literature) for the most significant commodities (e.g. classes 0, 1 and 9). On the contrary, the very poor values of some commodities (e.g. classes 4 and 6) can be explained by considering the specificity of such traded goods. In general, all variables have the expected sign. With reference to model elasticity, considering the commodity class 9, elasticity of trade flows to GDP is approximately 1.42 for export and 0.86 for import flows, to tariffs about 0.30 and 0.20 for import and export respectively, while there is a remarkable difference in the elasticity to transport costs for import (about 2.00) with respect to export (about 0.27). This may be explained by the inherent structural characteristics of the Italian economy, characterized by high production costs.
Finally, in spite of the satisfactory results underlined above, it should be noted that there is an unsatisfactory model performance in terms of MAPE (Mean Absolute Percentage Error) on the absolute values of flows (i.e. not their logarithms). For instance, for the commodity class 9 and for import direction, the MAPE calculated on the first one hundred o-d pairs in decreasing order by trade is close to 50%, meaning that the model is able to capture only the magnitude order of the traded flows. This should be normally accommodated in the practical use of the model in prediction by means of pivot applications.

3.3 Non-parametric models

The present section proposes the estimation of non-parametric gravity models, on the basis of the estimation database described in Section 3.1, respectively belonging to the family of kernel regressions (Section 3.3.1) and regression trees (Section 3.3.2). Estimation results are then compared with the outcomes of the parametric estimation results (Section 3.2) in Section 3.4.

3.3.1 Kernel regression

Comment [MG4]: Indicherei il software utilizzato

Non-parametric or smoothing techniques, such as Kernel regression, are based on estimating the dependent variable as a weighted average of the observed realization of that variable within an appropriate neighbourhood of the independent variables; different calculation methods of this weighted average normally lead to different types of models. That is, the estimate is carried out through the relationship:

$$\hat{y}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} W_{ni}(\mathbf{x})\, Y_i$$

where y is the dependent variable to be estimated as a function of $\mathbf{x}$ and W_{ni} represents the vector of weights to be assigned to the n realizations Y_i in correspondence of the point $\mathbf{x}$ wherein the estimation is required. In scalar Kernel regressions, the Nadaraya-Watson weight definition can be adopted as reference:

$$W_{ni}(x) = \frac{n\, K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)}$$

wherein the kernel function K_h can be specified in different ways, and depends on a depending on a bandwidth factor *h*, influencing the width of the neighbourhood and therefore on the weight values. Notably, the Nadaraya-Watson estimator is an order-0 estimator; in substitution, order-1 and order-2 estimators can be applied as well, corresponding to a polynomial regression within the estimation neighbourhood. In this case, the polynomial regression within each neighbourhood is obtained through a WLS estimation using the kernel as weight, and different neighbourhoods correspond to different polynomial regressions. In the following a comparison of the performances of estimators of various order will be proposed.

However, as reported in the literature, there is substantial robustness across choice of the kernel function, therefore in the following reference will be always made to the Epanechnikov kernel function:

$$K(u) = 0.75\,(1 - u^2) I_{|u| \le 1}$$

On the contrary, specific attention should be paid on the choice of the bandwidth factor, since for larger values there is very likely oversmoothing, while for shorter bandwidths there is the risk of not considering any experimental point or of encountering overfitting issues. Therefore, in the applications of the paper the bandwidth has been linked to the standard deviation of the training set, and fixing an upper bound of *p* points to be taken into account, i.e considering a so-called *p*-nearest neighbour estimate.

Due to the inherent nature and characteristics of the kernel regressions, a sample holding approach has been followed for the analysis of estimated model performances. That is, the whole estimation database has been split in three subsets: the learning set, i.e. the set used for model estimation, the test set, i.e. the set used for the calculation of model performance indicators, and the evaluation set, i.e. a set which the model will be applied to in prediction in order to explore further its forecasting capabilities. In more detail, with reference to the database described in Section 3.1, the 2006 year data have been chosen as evaluation set (about 9% of the total number of rows), and the remaining data have been randomly assigned to the test and learning sets respectively in a ratio 30/70. Finally, since formal tests on estimated model parameters are not implementable for kernel regressions, a stepwise methodology has been followed, that is adding at the generic step the explanatory variable leading to the highest increase of model performances.

Given these premises, in order to test the effect of database and variables aggregation on the performances of the kernel regression, different types models have been specified and estimated, as reported as follows.

Firstly, a unique model for import and export reproducing total trade has been estimated, i.e. summing up trade flows and averaging explanatory variables over the commodities. The overall database is therefore made up by 286 observations, split as follows: 182 learning set, 78 test set, 26 evaluation set.

In more detail, a simple specification has been firstly adopted, by considering as predictors the GDP of origin and destination as proxies of zone importance and the straight distance as proxy of the impedance. Consistently with the specification of the parametric gravity models (Section 3.2), a log-linear transformation of the explanatory variables and of the dependent variable has been performed. As mentioned above, the Epanechnikov kernel function has been adopted, and order-0, order-1 and order-2 estimators have been tested in order to compare their performances. Estimation and validation results are reported in the following Table 5. The table reports the MAPE (Mean Absolute Percentage Error), as further aggregated validation indicator; notably, since it suffers from the presence of outliers which may bias significantly its interpretation, the MAPE distribution is also presented, i.e. MAPE x% means the percentage of dataset rows ordered by increasing percentage error to be included for obtaining a x% MAPE value.

Table 5 – Unique model for import and export reproducing total trade: estimation results for the base specification.

Statistics	Degree of the estimator					
	order-0		order-1		order-2	
	Learning Set	Test Set	Learning Set	Test Set	Learning Set	Test Set
R^2	0.810	0.840	0.689	0.665	0.708	0.645
SSE	68.371	22.346	116.266	42.599	107.556	46.837
MSE	0.376	0.286	0.639	0.546	0.591	0.600
RMEQ	0.613	0.535	0.799	0.739	0.769	0.775
MAPE	56.40%	54.80%	75.20%	78.40%	69.80%	63.60%
MAPE 20%	42.31%	47.44%	30.22%	28.21%	26.37%	34.62%
MAPE 30%	50.55%	57.69%	42.31%	42.31%	43.41%	46.15%
MAPE 40%	60.99%	67.95%	50.00%	48.72%	53.30%	53.85%
MAPE 50%	65.93%	75.64%	55.49%	60.26%	58.79%	60.26%
MAPE 60%	77.47%	82.05%	62.09%	61.54%	64.84%	62.82%
MAPE 70%	81.32%	84.62%	67.58%	65.38%	69.78%	69.23%
MAPE 80%	83.52%	84.62%	75.82%	70.51%	76.37%	73.08%
MAPE 90%	84.62%	85.90%	79.67%	74.36%	80.77%	80.77%

The main result is that the order of the estimator does not affect remarkably model results, and that the simplest Nadaraya-Watson estimator leads to very satisfactory results in terms of MAPE (30% for the 60% of testing database rows and 50% for more than the 75% of testing database rows). Moreover, it is worth mentioning that, in spite of the good R^2 values, the MAPE errors are very high in overall terms, leading to the conclusion that accounting for R^2 and MAPE values contemporarily in model analysis is a crucial point.
Starting from the base specification reported in Table 5 (e.g. only GDP and distance as predictors), enhanced specifications have been estimated by introducing other predictors with the stepwise method previously mentioned. The adopted specifications and the corresponding estimation and validation results are reported in the following Tables 6.
Interestingly, the specification offering the most effective results takes into account GDP, straight distance, transport costs and custom duties as predictors. Furthermore, the sensitivity of the choice of the *k* points for neighbourhood approximation towards estimation results has been checked, leading to the definition of the optimal value of five points.

On the basis of estimation results reported in Tables 5 and 6, the three specifications with best results - i.e. the no. 6, 9 and 10 respectively in Table 6 - have been also used as basis for estimation of a model on a database encompassing all commodities. Firstly, a commodity-specific dummy has been added to each specification, leading to the estimation results reported in the following Table 7. That is, the whole dataset has been used for estimation, trying to capture differences among commodities only by means of such specific dummies.

Table 6 – Unique model for import and export reproducing total trade: estimation results.

Model no.	k-points	Specification							Estimation statistics									
									Learning					Testing				
		custom tariffs	total trade	GDP destination	GDP origin	straight distance	transport cost	dummyEU15	R²	MSE	RMEQ	MAPE	MAPE 50%	R²	MSE	RMEQ	MAPE	MAPE 50%
1	5			x	x	x			0.878	0.243	0.493	41.20%	80.77%	0.887	0.199	0.446	38.30%	85.90%
	10								0.810	0.376	0.613	56.40%	65.93%	0.840	0.286	0.535	54.80%	75.64%
	20								0.711	0.566	0.752	75.40%	56.04%	0.779	0.407	0.638	72.80%	61.54%
2	5			x	x	x	x		0.888	0.203	0.203	45.10%	84.07%	0.826	0.382	0.618	37.70%	70.51%
3	5	x		x	x	x			0.924	0.137	0.370	31.00%	86.81%	0.955	0.101	0.318	27.90%	89.74%
4	5	x		x	x	x			0.863	0.257	0.507	40.00%	78.57%	0.916	0.173	0.415	39.00%	76.92%
5	5			x	x	x		x	0.908	0.175	0.418	27.10%	81.87%	0.859	0.278	0.527	35.30%	82.05%
6	5	x		x	x	x	x		0.938	0.117	0.342	28.10%	86.26%	0.940	0.123	0.351	24.20%	92.31%
7	5	x	x	x	x	x			0.919	0.158	0.398	35.40%	85.71%	0.921	0.145	0.381	35.70%	88.46%
8	5	x		x	x	x		x	0.944	0.101	0.318	24.60%	90.11%	0.962	0.083	0.289	21.10%	89.74%
9	5	x		x	x	x	x	x	0.936	0.132	0.363	25.90%	85.17%	0.933	0.109	0.331	26.30%	83.33%
10	5	x	x	x	x	x	x	x	0.917	0.145	0.381	30.80%	85.71%	0.938	0.144	0.379	28.40%	80.77%

Table 7 – Unique model for import and export reproducing total trade for each commodity: estimation results (aggregated)

Model no.	Estimation statistics									
	Learning					Testing				
	R²	MSE	RMEQ	MAPE	MAPE 50%	R²	MSE	RMEQ	MAPE	MAPE 50%
6 + commodity specific dummy	0.860	1.481	1.217	1417%	62.58%	0.898	9.985	3.160	263%	68.46%
9 + commodity specific dummy	0.861	1.357	1.217	1719%	65.93%	0.913	10.206	3.195	193%	68.72%
10 + commodity specific dummy	0.878	1.235	1.111	2012%	68.41%	0.906	10.221	3.197	179%	69.10%

In aggregated terms, estimation results seem to be satisfactory for all specifications in terms of R^2, while MAPE values are totally unsatisfactory and unfeasible. In order to understand their blowing up and count for the possible presence of outliers, it is worth analyzing the performances of the estimated models with respect to each commodity class: the corresponding MAPE values are reported in the following Table 8.

Results in Table 8 show substantially heterogeneous model performances across commodity classes, with unsatisfactory results for commodities no. 2, 3, 4 and 7 (see Table 2 for a description). However, it should be noted that the mentioned commodities are usually traded with means other than the traditional transport systems (e.g. pipelines) and that their trade is normally determined by drivers not encompassing, or taking into account only marginally, transport costs and other explanatory variables explicitly introduced into the model. Notably, the high MAPE values cannot be explained only by the presence of outliers, i.e. the specification with only a commodity specific dummy is too naive and unable to provide for a reliable representation of the phenomenon, and in addition the inherent nature of the aforementioned commodities leads to a weak explanation capability of transport costs. For this reason, disaggregated estimations will be also performed in the following of the section, in order to enhance model performances.

Table 8 – Unique model for import and export reproducing total trade for each commodity: estimation results (disaggregated by commodity class)

Model no.	Commodity class	MAPE Learning	MAPE Testing
6 + commodity specific dummy	0	43.00%	46.50%
	1	64.00%	35.40%
	2	-	-
	3	-	-
	4	-	-
	5	73.00%	63.90%
	6	47.00%	60.30%
	7	-	-
	8	25.00%	25.50%
	9	32.00%	27.00%
9 + commodity specific dummy	0	47.80%	30.60%
	1	63.70%	46.90%
	2	-	-
	3	-	-
	4	-	-
	5	67.90%	73.00%
	6	43.20%	61.70%
	7	-	-
	8	21.40%	25.00%
	9	27.10%	26.70%
10 + commodity specific dummy	0	32.20%	30.60%
	1	42.40%	51.00%
	2	-	-
	3	-	-
	4	-	-
	5	90.30%	41.50%
	6	47.80%	31.00%
	7	-	-
	8	18.80%	19.60%
	9	26.70%	26.90%

Note: missing values mean MAPE>100%.

In more detail, accordingly with the outcomes of estimations in Table 8, a separated model has been estimated for each commodity. That is, a unique model for import and export, but separated for each commodity class, has been taken into account, leading i.e. specification of 10 kernel regressions, one per commodity. The database for each kernel regression is made up by 286 observations, split as follows: 182 learning set, 78 test set, 26 evaluation set. That is, while in the preceding estimation trial (Tables 7 and 8) the whole estimation

dataset has been taken into account, with only a commodity specific dummy in order to capture differences among commodities, in this estimation step a separate dataset, made up by a single specific commodity, has been applied for the estimation of a specific model for that commodity: results are reported in Table 9.
Notably, the adoption of a disaggregated model for each commodity does not help in increasing the overall goodness of fit for the commodities providing for unsatisfactory results.

Table 9 – Unique model for import and export, but separated for each commodity class: estimation results (disaggregated by commodity class)

Model no.	Commodity class	R^2 Learning	MAPE Learning	R^2 Testing	MAPE Testing
6	0	0.896	46.80%	0.933	44.10%
	1	0.802	70.30%	0.888	86.80%
	2	0.838	-	0.851	-
	3	0.738	-	0.782	-
	4	0.802	-	0.791	-
	5	0.793	86.70%	0.88	36.80%
	6	0.866	48.00%	0.898	45.80%
	7	0.832	-	0.779	-
	8	0.961	23.70%	0.947	26.10%
	9	0.891	28.30%	0.958	22.60%
9	0	0.908	42.60%	0.902	50.70%
	1	0.812	74.50%	0.918	75.20%
	2	0.845	-	0.818	-
	3	0.762	-	0.828	-
	4	0.774	-	0.814	-
	5	0.811	79.00%	0.826	42.00%
	6	0.871	47.40%	0.854	48.90%
	7	0.821	-	0.798	-
	8	0.957	23.30%	0.96	21.20%
	9	0.924	24.90%	0.881	28.40%
10	0	0.938	33.30%	0.901	34.90%
	1	0.948	29.20%	0.847	46.00%
	2	0.873	-	0.858	-
	3	0.79	-	0.854	-
	4	0.849	-	0.875	-
	5	0.871	50.60%	0.725	101.00%
	6	0.885	42.40%	0.794	44.30%
	7	0.826	-	0.692	-
	8	0.968	19.90%	0.968	19.50%
	9	0.911	28.70%	0.939	26.00%

Note: missing values mean MAPE>100%.

Finally, a separated model for import and export and for each commodity class has been taken into account, i.e. specifying 20 kernel regressions, one per commodity and flow

direction. The database for each kernel regression is made up by 143 observations, split as follows: 91 learning set, 39 test set, 13 evaluation set. Notably, no further enhancement of the goodness of fit of the model has been observed for the remaining commodities, and the corresponding results have been therefore not reported.
Finally, a validation of the best estimated models has been performed through the already mentioned evaluation set, that is checking their capability of reproducing trade data related to each zone in Table 1 for the year 2006, which have not been used either in the learning or in the testing datasets. In more detail, model specification no. 9 in Table 6 has been chosen as reference for kernel regressions, because it has found to be the more robust in both learning and in testing. Notably, both the aggregated (i.e. the same for all commodities, estimation in Table 6) and the disaggregated (i.e. one for each commodity, estimation in Table 9) versions of the model no. 9 have been checked. Results are respectively proposed in Table 10 and Table 11 respectively.
It should be noted that kernel regressions are not always able to provide for a forecast: notably, this happens when the dependent variable should be estimated on the basis of values of explanatory variables falling outside the boundaries of the learning set. Importantly, this circumstance occurs very often in the practice, and represents an inherent limit of the non parametric methods.
The same results occur when dealing with values of the explanatory variables with no dataset values falling into their neighborhood. With specific reference to the models under analysis, this happens for some applications of the disaggregated model, whose database encompasses only 143 records per commodity and per flow direction.

Table 10 – Percentage error on the evaluation dataset of the aggregated estimation of the model specification no. 9 in Table 5.

	Percentage Error	
Zone	**Import**	**Export**
North Africa	26.41%	9.18%
Other African countries	34.31%	62.99%
Central/South America	27.82%	44.91%
North America	10.11%	19.82%
Central Asia	3.61%	56.10%
Far East	34.97%	45.77%
China	15.33%	7.90%
EU15	2.49%	5.21%
EU members since 2004	5.40%	23.70%
EU members since 2007	37.59%	23.87%
Middle East	17.83%	0.92%
Australia	15.60%	10.87%
Other EU countries	13.35%	15.54%

Table 11 - Percentage error on the evaluation dataset of the disaggregated estimation of the model specification no. 9 in Table 5 for export (top) and import (bottom) flows.

	Percentage Error Export									
Zone	**Commodity 0**	**Commodity 1**	**Commodity 2**	**Commodity 3**	**Commodity 4**	**Commodity 5**	**Commodity 6**	**Commodity 7**	**Commodity 8**	**Commodity 9**
North Africa	60.85%	28.30%	49.56%	1.31%	-	47.39%	18.37%	15.77%	20.94%	15.61%
Other African countries	4.11%	63.33%	-	23.72%	80.08%	11.49%	32.40%	24.42%	5.08%	44.68%
Central/South America	86.85%	81.37%	-	90.62%	35.04%	24.38%	9.12%	93.82%	20.53%	42.57%
North America	5.15%	45.12%	48.16%	50.11%	99.44%	53.46%	34.59%	81.36%	12.63%	4.21%
Central Asia	14.91%	47.66%	70.35%	-	22.34%	45.33%	24.95%	-	36.89%	28.40%
Far East	12.89%	15.20%	-	-	40.67%	78.28%	-	72.57%	10.55%	1.88%
China	34.43%	82.37%	68.21%	58.79%	73.78%	61.13%	25.96%	52.10%	23.31%	11.62%
EU15	1.27%	49.74%	0.16%	9.61%	70.87%	17.45%	-	29.91%	8.59%	5.66%
EU members since 2004	25.49%	25.45%	7.76%	-	72.02%	30.46%	35.58%	30.98%	15.04%	20.12%
EU members since 2007	52.37%	28.57%	96.49%	57.94%	86.11%	49.17%	19.66%	55.48%	19.56%	-
Middle East	42.02%	16.45%	8.69%	19.16%	96.98%	11.22%	4.24%	41.50%	8.40%	8.47%
Australia	23.51%	31.17%	13.53%	-	55.49%	89.05%	55.38%	44.58%	33.27%	46.25%
Other EU countries	39.48%	53.37%	89.96%	16.03%	93.32%	40.13%	49.38%	-	17.57%	65.83%

	Percentage Error Import									
Zone	**Commodity 0**	**Commodity 1**	**Commodity 2**	**Commodity 3**	**Commodity 4**	**Commodity 5**	**Commodity 6**	**Commodity 7**	**Commodity 8**	**Commodity 9**
North Africa	38.11%	53.33%	93.39%	9.34%	10.96%	33.82%	38.86%	4.91%	0.14%	25.79%
Other African countries	35.34%	3.52%	99.46%	44.13%	80.99%	52.90%	6.91%	83.15%	18.60%	12.35%
Central/South America	54.45%	50.56%	28.67%	92.85%	64.39%	37.73%	51.07%	54.95%	41.36%	49.79%
North America	12.41%	179.80%	7.11%	17.48%	-	93.26%	5.63%	-	8.44%	15.06%
Central Asia	61.84%	-	-	94.20%	-	65.97%	23.51%	87.64%	26.22%	4.18%
Far East	18.33%	10.79%	93.97%	-	-	81.20%	5.06%	8.21%	38.16%	8.67%
China	-	-	-	-	-	76.32%	50.62%	36.09%	59.04%	53.57%
EU15	5.92%	0.45%	28.46%	0.53%	11.07%	6.27%	1.69%	1.04%	3.10%	6.43%
EU members since 2004	43.86%	33.63%	13.17%	-	28.29%	-	4.11%	44.11%	11.05%	28.49%
EU members since 2007	5.78%	-	-	19.83%	28.66%	38.49%	60.90%	70.78%	6.84%	9.88%
Middle East	62.97%	-	-	16.67%	23.21%	-	-	31.32%	0.11%	-
Australia	68.89%	29.07%	11.78%	99.72%	42.10%	85.12%	6.71%	12.03%	62.86%	53.94%
Other EU countries	22.12%	32.91%	34.55%	23.90%	60.86%	42.03%	35.62%	15.20%	2.86%	15.34%

Note: missing values mean MAPE>100% or model inapplicability due to lack of experimental points.

3.3.2 Regression trees

In order to provide for a further insight on the performances of non parametric regression models for gravity trade, the category of regression tree models has been taken into account for possible estimation.
Regression trees (e.g. Loh, 2008) aim at classifying data in order to build homogeneous groups with reference to the response (i.e. dependent) variable. The more widespread is the set of explanatory variables within the estimation dataset, the more reliable is the result of the regression. A number of software packages (XLSTAT has been used for this study) allows for easy application of regression trees. Similarly with the kernel regression, aggregated and disaggregated models with respect to commodities to be reproduced have been estimated, leading respectively to the results reported in Table 12 and Table 13, wherein all explanatory variables have been used for carrying out the regression tree estimation.

Table 12 – Regression tree estimation of the aggregated model (compare with Table 5)

	R^2	MAPE	MAPE 50%
Learning	0.960	18.60%	93.75%
Validation	0.204	69.01%	12.82%

Table 13 – Estimation results of regression tree models for the disaggregated model (compare with Table 9) for the export (top) and import (bottom) flows respectively

Commodity class	R^2 Learning	MAPE Learning	R^2 Validation	MAPE Validation
0	0.793	-	0.102	-
1	0.977	28.39%	0.839	-
2	0.945	-	0.897	-
3	0.933	-	0.704	-
4	0.881	-	0.041	-
5	0.988	52.03%	0.752	-
6	0.987	20.42%	0.844	-
7	0.857	-	0.448	-
8	0.990	27.41%	0.774	-
9	0.994	23.44%	0.951	-

Commodity class	R^2 Learning	MAPE Learning	R^2 Validation	MAPE Validation
0	0.966	-	0.980	-
1	0.967	59.99%	0.956	-
2	0.858	-	0.512	-
3	0.662	-	0.458	-
4	0.966	-	0.886	-
5	0.968	79.14%	0.641	-
6	0.829	60.80%	0.677	-
7	0.780	-	0.272	-
8	0.994	-	0.877	-
9	0.990	25.86%	0.897	-

Note: missing values mean MAPE>100%

The main result is that, with reference to the aggregated model, there is a substantially similarity with the results of the kernel regression in terms of learning, but a significant failure in validation, since R^2 falls to some 0.20 and MAPE increases from 18% to 69% about. That is, regression trees are effective classification methods but not useful for prediction. This result is dramatically amplified by the estimation of the disaggregated models, which does not reproduce correctly any commodity.

4 ASSESSMENT OF RESULTS AND CONCLUSIONS

Comment [MG5]: Discutere i risultati in maggiore dettaglio; in particolare vantaggi e svantaggi dei metodi e quando si possono applicare i non parametrici.

A first general comment about the comparison of the performances of the parametric and non parametric models estimated in the previous Section 3 deals with the substantial heterogeneity in models performances across commodity classes. That is, all commodities whose trade is inherently driven by factors other than simple transport costs, i.e. by more complex and very often unquantifiable external factors, provide for bad results whatever adopted approach: this is particularly the case of commodities 2,3 4 and 7 of the NST/R 1-digit commodity class reported in Table 2. This should be taken into account also when estimating aggregated (i.e. summed up across commodities) models, whose performances may be conditioned by such outliers. However, for transport oriented applications, the preceding commodities often represent a limited share of the total trade. It is therefore worth summarizing the outcomes related to the commodity class 9 (manufactured goods, vehicles and other traded goods) which represents normally one of the main contributions to total trade.
In that respect, parametric models exhibited a goodness of fit in line with the state of the art, with a good significance for the explanatory variables, acceptable R^2 values and a not entirely satisfactory MAPE value, even if with more sophisticated regression techniques. With reference to non parametric regressions, kernel regression substantially outperform regression trees, which provide for a potentially effective way for data classification but are not effective in application and always dominated, in terms of goodness of fit, by kernel models. From the other side, kernel regressions do actually overcome standard parametric regressions in terms of R^2 and often also in terms of MAPE, i.e. they provide for a better representation of the current situation as expressed by the estimation dataset. However, some analyses carried out on the evaluation dataset (i.e. through an hold out sample approach) lead to some doubts about the efficacy in prediction of the kernel regressions.
For this aim, five different hypothetical scenarios have been considered for elasticity analysis:

1. 2020 GDP increase of both importing and exporting countries according to 2006 (i.e. pre-crisis) estimates;

2. 20% reduction of transport costs;

3. 20% reduction of custom tariffs and duties;

4. 5% GDP reduction of importing and exporting countries;

5. 5% GDP reduction of importing and exporting countries plus 30% custom tariffs increase.

That is, scenario 1 is a tendency scenario, scenarios 2 and 3 mimic interventions on transport supply, scenario 4 mimics a situation of economic recession and scenario 5 provides in addiction for protectionist policies.
For the sake of simplicity, results have been reported only for the Italy-China relationship, which was at the basis of the current study, and for commodity class 9 (Table 2). Models compared are the parametric SUR estimate (Tables 3 and 4) and the disaggregated kernel model estimates (model no. 9 in Table 9). Estimation results are reported in the following Table 14.

Table 14 – Elasticity analysis of parametric vs. non parametric regression models in forecasts for the corridor Italy-China (exports and imports are intended respectively from/to Italy).

	Parametric		**Non-Parametric**	
	Export	Import	Export	Import
Scenario 1	10.71%	9.26%	14.04%	4.01%
Scenario 2	5.39%	5.14%	6.22%	4.71%
Scenario 3	4.27%	8.08%	8.24%	7.28%
Scenario 4	-7.10%	-4.31%	-10.46%	7.40%
Scenario 5	-11.69%	-13.50%	5.70%	9.54%

Results show that kernel regressions sometimes provide for unexpected results and wrong signs, e.g. a GDP decrease (scenario 4) leads to an increase of imports while the corresponding parametric models predicts a decrease, as it can be normally expected. This result is further evidenced by the wrong forecast provided within the combined scenario 5 of GDP reduction and custom tariffs increase. However, when predictions are correct (e.g. in scenarios 1 and 2), the elasticities provided by the kernel regressions are very similar with those of the traditional regression techniques. This can be likely explained by the coverage characteristics of the database, which probably lacks of data coverage related to regression periods, and therefore providing for few support points for kernel regressions. Therefore, a main conclusion seems to be drawn, that is kernel models can normally outperform the regression techniques in the simulation of the current scenario (i.e. in the estimation dataset), but can be effectively applied in prediction only when the range of variation of the explanatory variables is sufficiently limited and however comprised within the estimation dataset boundaries, and when there is a very large amount of data covering sufficiently all the estimation dataset.

AKNOWLEDGEMENTS

This paper was developed under the FIRB Project "Intellilogic – Intelligence for Logistics" supported by Italian Ministry of University and Research (Project Number: RBIP06MMBW).

REFERENCES

Central Bank of Italy (2009). Indagine sul costo di trasporto internazionale di merci e passeggeri. Available at http://www.bancaditalia.it/.

Coulibaly S. (2007). Evaluating the trade effect of developing regional trade agreements: a semi-parametric approach. Policy Research Working Paper Series 4220, The World Bank.

Kepaptsoglou K., Tsamboulas D., Karlaftis M. G., Marzano V. (2009). Free trade agreement effects in the Mediterranean Region: an analytic approach based on SURE gravity model. Transportation Research Record – Journal of the transportation research board, 2097, pp. 88-96

Loh, W.-Y. (2008). Classification and Regression Tree Methods. In Encyclopedia of Statistics in Quality and Reliability, Ruggeri, Kenett and Faltin (eds.), 315–323, Wiley.

Marzano V., Papola A., Simonelli F. (2008). A large scale analysis of the competitiveness of new Short-Sea Shipping services in the Mediterranean. Atti della 36th European Transport Conference, Noordwijk, The Netherlands.

Porojan A. (2001). "Trade Flows and Spatial Effects: The Gravity Model Revisited". Open economies review, 12: 265-280.

Wan, G.H., Griffiths, W.E. and Anderson, J.R. (1992). Using Panel Data to Estimate Risk Effects in Seemingly Unrelated Production Functions, Empirical Economics, 17, 35-49.

Washington, S., Karlaftis, M.G. and Mannering, F.L. (2003). Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall / CRC Press, Boca Raton, FL.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias, Journal of the American Statistical Association, 57, 348-368.