ROAD FREIGHT TRANSPORT DEMAND IN SPAIN: A PANEL DATA MODEL

Rosa Aza, Department of Economics, University of Oviedo, riaza@uniovi.es
José Baños, Department of Economics, University of Oviedo, jbanos@uniovi.es
Pelayo González Arbués, University of Oviedo, rgonzalezpelayo@gmail.com
Manuel Llorca, University of Oviedo, llorcamanuel@gmail.com

ABSTRACT

International trade has been widely studied in the economic literature. However, in spite of being an important sector, domestic trade has not been analysed in a comprehensive way. This paper presents a model of the demand of road transport of goods in the domestic traffic of the Spanish Autonomous Communities between 1999 and 2008. Demand functions based on the gravity model are estimated. The panel model with fixed effects estimated in this paper is used to predict future trade flows as well as CO_2 emissions.

Keywords: Transport demand, gravity model, panel data, fixed effects, Spanish Autonomous Communities.

1. INTRODUCTION

In the last decades, international trade has been a fertile topic in the economic literature generating an important number of scientific papers. However, the flows of goods within a particular country have not been thoroughly analysed, even when in some cases (like Spain) they account for the transportation of 80% of the manufactured products (Llano *et al.*, 2008). Interregional traffic covers almost 60% of these national flows, whereas intraregional traffic reaches 40%.

In the case of Spain, the first scientific contributions to the field of domestic trade analysed the commercial exchange between a single region and the rest of the world (Castells and Parellada, 1983). The lack of information clearly underlined the need for a database that included all the statistics related to interregional flows (Llano, 2004). Since these studies used a descriptive approach, there was not a deep analysis of the variables that trigger the changes in the trade flows in Spain. An alternative source aims to verify the international trade models by using trade data among the Spanish regions (Artal *et al*., 2006). The relative importance of distance as a hurdle for domestic trade has been considered for the case of Spain by means of a gravity model similar to the one presented in this paper (Hernández, 1998). Nevertheless, there is a gap in the economic literature as regards the study of the determinants of regional trade flows in Spain.

This study reports an analysis of the demand of the flows of goods transported by road in the Spanish Autonomous Communities by means of the econometric estimation of a demand function based on the gravity model. This methodology -developed in the field of physics- was applied to economy for the first time in a study on international transport (Tinbergen, 1962). Subsequently, the theoretical model was optimized and become more complex (Anderson 1979, Rose and Wincoop 2001 or Anderson and van Wincoop 2003), and more recently it has been used in order to explain the existent flows among the different regions of the world (McCallum, 1995, Egger and Pfaffermayr, 2004 or Martínez-Zarzoso *et al.,* 2008). In the case of Spain, it is interesting to point out the research on tourism demand by means of a dynamic model (Garín, 2007), as well as a study that explains the originating factors of the flows between Andalusia and the rest of the regions (Borra, 2004).

Road transport of goods does play a key role in the Spanish economic activity by promoting the exchange of goods between the economic agents. According to the National Statistics Institute (*Instituto Nacional de Estadística*), road transport was the most used means of transportation in 2008, carrying 77 out of 100 transported tons; it was followed by sea transport, accounting for 22% of the total. In addition, these figures show the tendency to consolidate the relevance of road haulage between 1995 and 2008. In this period, the use of this mode of transport increased in 185%, growing at an average annual rate of 9%.

The correct evaluation of road transport of goods contributes to minimise certain problems -such as the congestion or the low use of infrastructures- arising when transport offer and demand are not balanced. Once the imbalance has been defined, several methods and tools can be applied according to the considered deadlines. In the short term, it is possible to use price discrimination, while in the long term investments will allow to match the capacity of infrastructures with their future use. The predicted future level of demand on infrastructures will determine the amount of this investment. This study reports the prediction of trade flows among Spanish regions for a period of ten years (2009 - 2018), taking into account the polluting emissions associated with road transport of goods.

The paper is organised as follows: section 2 describes the methodology of the gravity model of transport demand. Section 3 presents the data and the specification of the applied models. Section 4 expounds the results obtained for the model estimation. Finally, section 5 includes the conclusions of the research and its possible future lines.

2. GRAVITY MODEL-BASED TRANSPORT DEMAND

In this study, a gravity model is applied in order to explain the demand of road transport of goods among different Spanish regions. Therefore, this model will be used to analyse the first three steps of the classical model: trip generation and distribution by means of a mode of transport given in an exogenous way.

Similarly to Newton's law of universal gravitation, this type of models can be generally represented as:

$$T_{ij} = G \frac{Y_i^{\alpha_i} Y_j^{\alpha_j}}{D_{ij}^{\beta}} \qquad (1)$$

Where

T_{ij} is the trade flow with origin in i and destination in j.

G is the gravity or proportionality constant.

Y_i and Y_j stand for economic variables that will measure the scale of the origin and destination zone, respectively.

D_{ij} represents the inhibitory effect of the distance between i and j.

The gravity model shows the relationship between different places, such as homes and workplaces. In the related literature there is a common agreement stating that the interaction between two locations is reduced according to the increase in distance, time and cost among them. In addition, it is positively linked to the amount of activity in every location. In this sense, one of the most significant contributions by Tinbergen's (1962) pioneer specification of the gravity model was the definition of the variables that determined international trade flows. On the one hand, Tinbergen verified that the trade volume between two countries is proportional to the GDP of importer and exporter countries, whereas trade was reduced due to impediment factors such as distance. This approach has been subsequently supported by several scholars (Anderson 1979, Anderson and Wincoop 2003) and it has been widely used in most studies at both, regional and international levels (Helpman, 2007).

3. MODEL SPECIFICATION

Empirical research has confirmed that results obtained with gravity models are relatively linked to the adopted specification (Mátyás 1998, Egger and Pfaffermayr 2003 and Baldwin and Taglioni 2006). The gravity model was originally developed to be estimated by means of cross-sectional data analysis of different countries in a given moment. However, in order to avoid a defective model specification, the high heterogeneity of trade flow patterns in many countries has been considered in the specification (Westerlund and Wilhelmsson, 2009). A solution for this problem was implemented in the 80s, when the first applications of gravity equations using panel data structures appear. This section describes the different model specifications considered in this study: on the one hand, a cross-sectional model and on the other hand, five different models with fixed effect structure panel.

The panel data approach provides several advantages regarding the cross-sectional model: first, panel data improves the degrees of freedom and allows for capturing the relationship among variables for a long period of time. In addition, it is possible to spot the role the economic cycle has on these relationships. On the other hand, the use of panel data structures contributes to collect the specific time invariant effects of the different regions. The cross-sectional specification is likely to suffer from omitted variable bias since it does not include the unobserved effects of the regions and it ignores the time effects of trade traffic (Harris and Mátyás, 1998).

In spite of the theoretical superiority of the panel data model over the cross-sectional one, the latter has been widely used in the last decades in order to explain the existing trade flows among different regions. Hence, a cross-sectional model has been included in this study with the aim of contrasting results.

The following specification will include the demand of road transport of goods based on a gravity model:

$$X_{ijt} = \alpha_{ij} Pa_t^{\beta_1} Dist_{ijt}^{\beta_2} Border^{\beta_3} Intra^{\beta_4} GDP_{it}^{\beta_5} GDP_{jt}^{\beta_6} KHCNpc_{it}^{\beta_7} KHCNpc_{jt}^{\beta_8} e^{u_{ijt}} \quad (2)$$

Where:

X_{ijt} represents the flows of goods transported from the region of origin i to the region of destination j, measured by thousands of tons transported and transport operations. In the case of intraregional trade flows, i will be equal to j.

α_{ij} varies depending on the model specification. In the cross-sectional model, it represents the constant. In panel data models, it will report every fixed effect.

Pa_t is the average price of road transport in the year t. Calculated as an index based on the first quarter of 1999, it includes the annual price change.

$Dist_{ijt}$ is the distance between origin and destination. It is calculated by dividing the ton-kilometre variable by the dependent variable of transported tons in the case of every observation.

Border is a dummy variable with value 1 in the case of adjacent Autonomous Communities and 0 in the rest of the cases.

Intra is a dummy variable with value 1 when the Autonomous Communities of origin and destination are the same, and 0 in the rest of the cases. In the ultimate model this dummy variable is replaced with 15 dummy variables, one for each region.

GDP_{it} is the real GDP of Autonomous Communities of origin i in the year t.

GDP_{jt} is the real GDP of Autonomous Communities of destination j in the year t.

$KHCNpc_{it}$ are the kilometres of high capacity network *per capita* of the community of origin i in the year t.

$KHCNpc_{jt}$ are the high capacity network kilometres *per capita* of the community of destination j in the year t.

u_{ijt} is the usual term of the random perturbation.

The econometric representation of the gravity model, in linear logarithmic form, is determined by:

$$\begin{aligned} \ln X_{ijt} &= \alpha_{ij} + \beta_1 \ln Pa_t + \beta_2 \ln Dist_{ijt} + \beta_3 Border + \beta_4 Intra + \beta_5 \ln GDP_{it} + \\ &+ \beta_6 \ln GDP_{jt} + \beta_7 \ln KHCNpc_{it} + \beta_8 \ln KHCNpc_{jt} + u_{ijt} \end{aligned} \quad (3)$$

When taking logarithms, β parameters are interpreted as elasticities of the variables they are associated to. The expected sign for the coefficients β_1 and β_2, estimated in each model is negative due to the effect of the variables general price and distance, which stand for a hurdle for the flow of goods. It is expected that the rest of the estimated coefficients yield positive results: flows among adjacent communities (and within the same community) should to be larger, *ceteris paribus*, than among distant communities; the GDP variable includes the scale of communities in economic terms, so a larger flow is expected among Autonomous Communities with a more important economic activity; the ratio of high capacity network kilometres *per capita*, which stands for a measure of density of high capacity roads understood as a proxy of the road quality of the Autonomous Communities, is expected to have a positive influence on the flows among communities.

Within the framework of panel data methodology, the model can be estimated with fixed and random effects. These effects compile the unobservable heterogeneity of individuals, that is to say, issues such as geographic, historical or political characteristics that are not represented by the variables of the models. Since these effects are probably correlated to the independent socio-economic variables of the individuals, the fixed effects estimation is more suitable in these models (Egger, 2000). In order to support this idea, this study verifies that for the final model, the specification of the fixed effects model is more suitable than the one of random effects. This has been done by means of the Hausman Test, included in the Appendix A.I of this paper.

3.1. Database

The used database represents a balanced panel for the 15 peninsular Autonomous Communities[1] between 1999 and 2008. This study does not report the data of the Balearic Islands, the Canary Islands and the Autonomous Cities of Ceuta and Melilla due to their geographic locations.

The measure of the demand of road transport of goods -a dependent variable- is carried out by means of two alternative variables available in the Permanent Survey of Road Transport of Goods, issued by the Spanish Ministry of Development: these variables are flows and transport operations. Since 1999, this survey uses a homogeneous methodology according to the European Union Regulation 1172/98. These are the basis for the sample period of the current study. As it may be observed in Table 1, the rest of the explanatory variables are provided by different statistical sources.

Usually, this kind of model considers distance as an indicator of the kilometres between the capital cities of the regions or between their centroids. However, this has a clear disadvantage since it is necessary to select the points of origin and destination of the regions in a discretionary way (Hernández, 1998).

---

[1] Andalusia, Aragón, Asturias, Basque Country, Castile and León, Castile-La Mancha, Cantabria, Catalonia, Extremadura, Galicia, La Rioja, Madrid, Murcia, Navarre and Valencian Community

Since the Permanent Survey of Road Transport of Goods also includes the ton/kilometre variable, it is possible to retrieve the mean travel distance of the goods by dividing the total ton-kilometres by the transported tons in each origin-destination pair. The distance variable, measured in this way, reports information on the routes chosen by the transport companies. The resulting variability, as opposed to the fixed character of the previously mentioned approach, will report the cost minimizing behaviour of the companies. On the other hand, by measuring distance in this way we can get a more accurate estimation of the intraregional flows included in the model.

Besides, the Permanent Survey does also provide additional information about the price associated to public transport operations. Relying on these data, an index that includes the prices of transport has been developed since 2005 in order to monitor the tendency of the price per kilometre.

Previous estimations suggested that other variables, such as the monetary value of road stock, were not statistically significant in order to explain the flows among regions. Therefore, it was rejected as an explanatory variable. Similarly, it has been confirmed that other variables, such as the Gross Added Value and the road kilometres had a lower explanatory capacity than the GDP and the high capacity network kilometres, respectively. The data for this last variable have been collected until 2007; the geometric mean of the growth reported in the available years has been used in order to calculate the value in 2008.

On the other hand, the technical change was modelized by means of time effects in the form of dummies. These models were rejected due to the correlation of the effects with the General Price variable; since it reports the time tendency (as it is formed by indexes that vary every year for the whole group of Autonomous Communities), this was the selected variable. In addition, it is a determinant variable in a demand estimation study.

Table 1. Definition of variables

	Variable	**Description**	**Source**
Dependent Variables (Demand)	Flows	Transported tons (thousands)	Spanish Ministry of Development
	Transport operations	Number of trips	Spanish Ministry of Development
Explanatory variables	General Price	Average general price of transport of goods (1999 base index)	Spanish Ministry of Development
	Distance	Kilometres travelled by road between origin and destination (average)	Spanish Ministry of Development
	GDP	Gross Domestic Product (Thousands of constant Euros)	National Statistics Institute
	High Capacity Network	Km. High Capacity Roads	National Statistics Institute
	Population	Standing population in the Autonomous Communities	National Statistics Institute

Table 2 shows the descriptive statistics of the quantitative variables used in the estimated models. Dependent variables (transport flows and operations) present a few values equal to 0 throughout the whole sample (13 and 8 respectively). In order to avoid certain problems when taking logarithms in these variables, 0 values have been replaced by 1, since it is the minimum value larger than 0 found among the different values taken by the independent variable. This is usually the most suitable statistical way to deal with this sort of problems after having added the provincial data with the aim of getting positive values per Autonomous Community (Bergkvist and Westin, 1997)[2].

Table 2. Descriptive statistics of the data

Variable	Mean	Typical deviation	Minimum	Maximum
Flow	7,270.93	30,925.80	1	358,245
Transport operations	995,056	4,498,450	1	50,973,200
General Price	119.15	12.57	100.80	138.30
Distance	481.79	293.91	11.47	4,834.07
GDP	43,927,100	41,338,600	4,515,850	151,444,000
Km. High Capacity Network	806.43	594.36	130	2,631
Population	2,661,150	2,299,150	264,178	8,059,460

*Number of observations: 2,250

4. ESTIMATION AND RESULTS

The first model estimated in the study is a pooled panel data. This methodology estimates parameters by grouping all the individuals and time periods, so it is equivalent to a cross-sectional cut and it does not deal with the problem of unobservable heterogeneity among individuals. Table 3 shows that the explanatory variables included in the two models exhibit the expected signs and are statistically significant. The adjusted determination coefficients show that almost 90% of the flow variation and 88.2% of the variation in transport operations are explained by the variables of the model.

As it has been remarked in Section 3, the results of these models are likely to suffer from omitted variable bias (since variables that include the unobservable heterogeneity have been omitted). In order to solve this problem, panel data models are estimated in this study. In the Appendix A2 of this paper, we apply the homogeneity test aiming to confirm that the fixed

---

[2] If the rate of observations with 0 values was significant, alternative estimation methods should be considered. The Poisson fixed effects estimator, proposed by Westerlund and Wilhelmsson 2009, could be applied here.

effects estimated for the final model report important differences among the groups. Therefore, the use of the panel data model seems to be more suitable than the cross-sectional one for this particular case.

Table 3. Results in cross-sectional models

	Cross-section			
Variables	**Flows**		**Transport Operations**	
Constant	-8.991		-2.506	
Pa	-0.439		-0.561	
	(-3.117) ***	(-3.420) ***	(-3.309) ***	(-3.849) ***
Dist	-1.235		-1.430	
	(-36.803) ***	(-34.680) ***	(-35.395) ***	(-33.253) ***
Border	0.4751		0.594	
	(9.643) ***	(9.271) ***	(10.032) ***	(8.739) ***
Intra	1.457		1.529	
	(12.910) ***	(12.740) ***	(11.261) ***	(11.506) ***
GDP_O	0.869		0.894	
	(51.128) ***	(42.779) ***	(43.725) ***	(31.241) ***
GDP_D	0.896		0.884	
	(52.986) ***	(46.613) ***	(43.401) ***	(36.281) ***
KHCNpc_O	0.416		0.409	
	(10.841) ***	(8.995) ***	(8.862) ***	(6.213) ***
KHCNpc_D	0.273		0.348	
	(7.066) ***	(6.612) ***	(7.484) ***	(6.677) ***
Obs.	2,250		2,250	
R^2 adjust.	0.899		0.882	

* Significant at 10%. **Significant at 5%. ***Significant at 1%
The shadowed columns show the values of the Student's t after White's correction

Next, the different panel models estimated in this paper are shown. As it has been already mentioned, they have different specifications regarding their fixed effects. Table 4 presents the results of two models. The first one estimates 120 fixed effects that represent the intraregional trade as well as every pair of Autonomous Communities with a trade flow, independently of their position as the origin or destination of the transit. This is, when $\alpha_{ij} = \alpha_{ji}$.

The fixed effects used in the second model account for 225. They represent the trade flows of every Autonomous Community as well as the pairs of Autonomous Communities (in this case, according to their position as the origin or the destination of trade flows). To illustrate this with an example, the associated effect of the flows from Madrid to Catalonia will not be the same than the flows from Catalonia to Madrid. This method is intended to differentiate between several tendencies to import and export among specific communities.

In the case of the fixed effect panel for the origin-destination pairs, the *Border* variable is implicitly reported by these effects, so it is not included in the model specification. For the same reason, the dummy variable *Intra* is not included in the specification of the two models.

Table 4. Results in the panel models with origin-destination pair fixed effects

	Origin-destination fixed effects panel $\alpha_{ij} = \alpha_{ji}$				**Origin-destination fixed effects panel** $\alpha_{ij} \neq \alpha_{ji}$			
Variables	**Flows**		**Transport Operations**		**Flows**		**Transport Operations**	
Pa	-1.423		-2.167		-1.417		-2.163	
	(-2.787) ***	(-0.008)	(-3.107) ***	(-0.009)	(-2.962) ***	(-0.006)	(-3.133) ***	(-0.007)
Dist	-0.596		-0.407		-0.682		-0.477	
	(-7.042) ***	(-0.077)	(-3.518) ***	(-0.039)	(-8.084) ***	(-0.062)	(-3.919) ***	(-0.031)
Border	0.535		0.005		-		-	
	(2.645) ***	(0.041)	(0.017)	(0.000)	-	-	-	-
GDP_O	1.735		2.058		0.326		0.033	
	(6.208) ***	(0.017)	(5.390) ***	(0.015)	(0.415)	(0.000)	(0.029)	(0.000)
GDP_D	1.761		2.041		3.165		4.062	
	(6.303) ***	(0.017)	(5.345) ***	(0.015)	(4.031) ***	(0.003)	(3.586) ***	(0.002)
KHCNpc_O	0.173		0.217		0.084		0.131	
	(2.666) ***	(0.0439	(2.445) **	(0.040)	(1.012)	(0.008)	(1.086)	(0.009)
KHCNpc_D	0.014		0.140		0.099		0.224	
	(0.214)	(0.003)	(1.578)	(0.026)	(1.188)	(0.032)	(1.858) *	(0.052)
Obs.	2,250		2,250		2,250		2,250	
R^2 adjust.	0.943		0.913		0.950		0.915	

* Significant at 10%. **Significant at 5%. ***Significant at 1%
The shadowed columns show the values of the Student's t after White's correction

As it can be observed in Table 4, when applying White's correction, the values of the Student's t decrease, avoiding the rejection of the null hypothesis of lack of significance of all the variables. Therefore, due to the heteroskedasticity problem they present, these models are excluded. In order to tackle this issue, alternative methods have been considered aiming to achieve the stratification of the fixed effects.

Table 5 presents the results of two new models with 15 fixed effects estimated: the first case shows the effects associated to the Autonomous Communities of origin of the trade flows and the second case presents the effects associated to the regions of destination.

Due to the stratification of the fixed effects, the first model presents a correlation between the explanatory variables of origin and the fixed effects. Similarly, in the second model, there is a correlation between the explanatory variables of destination and the fixed effects. This problem causes that the elasticities obtained for the origin and destination GDP are too high from the economic point of view. In addition, it leads to the non significance of the variable kilometres of high capacity network *per capita* in origin and destination in the models where the flows are the dependent variable. On the basis of these facts, these models have also been rejected.

Table 5. Results in the panel models with fixed effects by origin and destination

	Panel with fixed effects by origin				Panel with fixed effects by destination			
Variables	**Flows**		**Transport Operation**		**Flows**		**Transport Operation**	
Pa	-1.037		-1.548		-1.694		-2.461	
	(-1.734) *	(-2.108) ***	(-2.055) **	(-3.149) ***	(-2.696) ***	(-0.006)	(-3.248) ***	(-0.006)
Dist	-1.281		-1.449		-1.354		-1.560	
	(-36.694) ***	(-33.170) ***	(-32.948) ***	(-29.561) ***	(-35.247) ***	(-4.147) ***	(-33.686) ***	(-3.423) ***
Border	0.429		0.589		0.355		0.470	
	(8.765) ***	(8.297) ***	(9.559) ***	(8.561) ***	(6.646) ***	(0.727)	(7.303) ***	(0.688)
Intra	1.324		1.479		1.107		1.146	
	(11.531) ***	(10.368) ***	(10.222) ***	(9.522) ***	(8.819) ***	(1.096)	(7.570) ***	(0.813)
GDP_O	1.926		2.340		0.882		0.907	
	(2.958) ***	(3.630) ***	(2.853) ***	(4.502) ***	(52.226) ***	(2.430) **	(44.564) ***	(1.789) *
GDP_D	0.901		0.884		2.510		3.170	
	(56.589) ***	(49.528) ***	(44.080) ***	(36.423) ***	(3.670) ***	(0.008)	(3.844) ***	(0.007)
KHCNpc_O	0.081		0.126		0.426		0.416	
	(0.737)	(0.587)	(0.910)	(0.704)	(11.238) ***	(0.5689	(9.112) ***	(0.397)
KHCNpc_D	0.279		0.346		0.073		0.191	
	(7.682) ***	(7.230) ***	(7.545) ***	(6.629) ***	(0.635)	(0.013)	(1.369)	(0.024)
Obs.	2,250		2,250		2,250		2,250	
R^2 adjust.	0.912		0.887		0.903		0.885	

* Significant at 10%. **Significant at 5%. ***Significant at 1%
The shadowed columns show the values of the Student's t after White's correction

In order to overcome the correlation problems between fixed effects and variables, the former were grouped into sets of Autonomous Communities, as it is explained in Table 6.

Table 6. Groups of Autonomous Communities that form the fixed effects.

Set	Autonomous Community
1	Asturias, Cantabria, Galicia
2	Navarre, Basque Country, La Rioja
3	Castile and León, Castile-La Mancha, Madrid
4	Andalusia, Extremadura, Murcia
5	Aragon, Catalonia, Valencian Community
6	Set 1 - Set 2
7	Set 1 - Set 3
8	Set 1 - Set 4
9	Set 1 - Set 5
10	Set 2 - Set 3
11	Set 2 - Set 4
12	Set 2 - Set 5
13	Set 3 - Set 4
14	Set 3 - Set 5
15	Set 4 - Set 5

This grouping is intended to include the unobservable, shared and specific characteristics of each set: geographic, historical and political issues, similar productive structure or orographic issues. These characteristics are not reported by other variables and are expected to have an influence on the transport flows among the Autonomous Communities forming each set (and the ones belonging to different groups). The aim is to capture part of the modelling carried out in the spatial econometric models for interregional flows introduced by Lesage and Pace (2005), which highlight the importance of the areas formed by the adjacent regions close to the points of origin and destination of the flows.

These groups are mutually exclusive. The trade among the communities that integrate the first five groups is associated with the first five fixed effects; on the other hand, trade among communities included in different groups will be related to fixed effects 6 to 15. For instance, the trade between Asturias and Cantabria or the interregional trade in Galicia will be related to the fixed effect 1, whereas the commerce between these regions and Aragón will be included in the fixed effect 9. Summarizing, there is not an overlapping of fixed effects.

The estimated coefficients in this model -for flows and transport operations- present the expected signs and are significant at 1%, as it is shown in Table 7. Both models have an explanatory capacity of the dependent variable close to the 90%, although as it can be observed in the table, the adjusting is slightly higher in the case of the flows. The coefficient obtained for the dummy variable border, shows that the expected flow for adjacent Autonomous Communities is 1.71 times higher than for distant regions. Regarding transport operations, this value increases until 1.95. As for the coefficients estimated for the dummy variables *Intra*, they indicate that intraregional trade flows are higher than interregional ones. This coefficient is non-significant only for the case of Madrid (Intra11). It suggests that in this particular community the relative weight of intraregional commercial transport (compared to interregional transport) is lower than in other regions.

Table 7. Results in the panel models with fixed effects by groups

	Panel with fixed effects by groups			
Variables	**Flows**		**Transport Operations**	
Pa	-0.878		-0.769	
	(-6.170) ***	(-7.224) ***	(-4.209) ***	(-5.374) ***
Dist	-1.198		-1.366	
	(-27.425) ***	(-23.610) ***	(-24.330) ***	(-21.993) ***
Border	0.535		0.666	
	(9.966) ***	(8.611) ***	(9.667) ***	(8.123) ***
GDP_O	0.972		0.954	
	(48.436) ***	(41.285) ***	(37.004) ***	(28.921) ***
GDP_D	1.000		0.944	
	(50.027) ***	(46.054) ***	(36.755) ***	(35.519) ***
KHCNpc_O	0.536		0.459	
	(13.330) ***	(11.280) ***	(8.885) ***	(6.720) ***

KHCNpc_D	0.392		0.395	
	(9.594) ***	(9.750) ***	(7.543) ***	(7.856) ***
Intra1	1.211		1.358	
	(5.438) ***	(7.209) ***	(4.748) ***	(6.593) ***
Intra2	1.708		1.872	
	(7.577) ***	(10.031) ***	(6.465) ***	(8.835) ***
Intra3	2.191		2.295	
	(10.103) ***	(12.399) ***	(8.238) ***	(9.739) ***
Intra4	2.210		2.377	
	(10.008) ***	(12.088) ***	(8.383) ***	(9.824) ***
Intra5	2.439		2.442	
	(11.535) ***	(16.825) ***	(8.991) ***	(13.523) ***

Table 7. Results in the panel models with fixed effects by groups (cont.)

Variables	Flows		Transport Operations	
Intra6	1.925		1.996	
	(8.962) ***	(12.659) ***	(7.233) ***	(10.372) ***
Intra7	0.666		1.020	
	(3.007) ***	(3.883) ***	(3.582) ***	(4.965) ***
Intra8	1.326		1.583	
	(5.947) ***	(7.939) ***	(5.526) ***	(7.735) ***
Intra9	2.409		2.511	
	(11.073) ***	(14.039) ***	(8.988) ***	(12.182) ***
Intra10	1.674		2.045	
	(7.736) ***	(10.133) ***	(7.358) ***	(9.324) ***
Intra11	0.136		0.251	
	(0.629)	(0.772)	(0.904)	(1.272)
Intra12	2.065		2.067	
	(9.264) ***	(11.425) ***	(7.221) ***	(9.991) ***
Intra13	1.432		1.399	
	(6.896) ***	(10.473) ***	(5.245) ***	(8.208) ***
Intra14	0.881		1.015	
	(4.356) ***	(7.186) ***	(3.910) ***	(6.702) ***
Intra15	1.696		1.739	
	(8.056) ***	(10.633) ***	(6.434) ***	(9.467) ***
Obs.	2,250		2,250	
R^2 adjust.	0.923		0.897	

* Significant at 10%. **Significant at 5%. ***Significant at 1%
The shadowed columns show the values of the Student's t after White's correction

Table 8 presents the fixed effects calculated for this model.

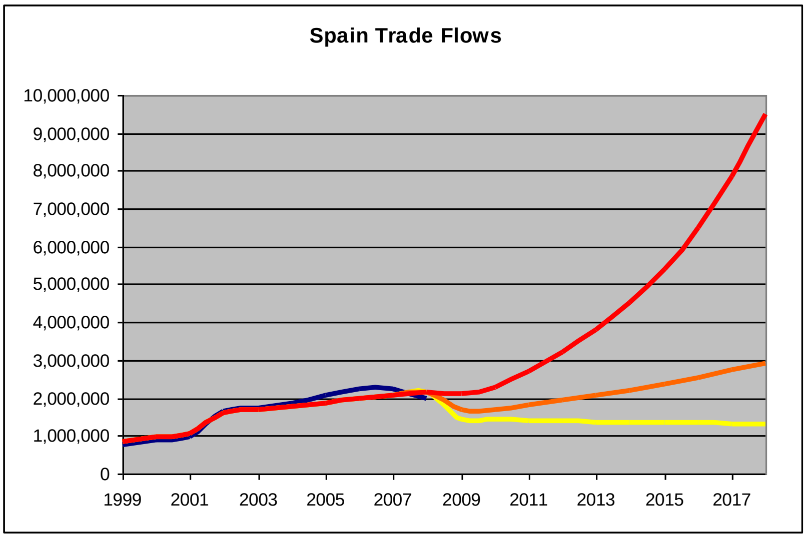
Table 8. Fixed effects

Set	Autonomous Community	Fixed effects by groups	
		Flows	Transport
1	**Asturias, Cantabria, Galicia**	-8.475	-3.026
2	**Navarre, Basque Country, La Rioja**	-8.593	-3.013
3	**Castile and León, Castile-La Mancha,**	-9.438	-3.705
4	**Andalusia, Extremadura, Murcia**	-8.770	-3.105
5	**Aragon, Catalonia, Valencian**	-8.783	-3.258
6	**Set 1 - Set 2**	-8.370	-2.896
7	**Set 1 - Set 3**	-8.727	-3.282
8	**Set 1 - Set 4**	-9.221	-3.658
9	**Set 1 - Set 5**	-8.501	-3.045
10	**Set 2 - Set 3**	-9.014	-3.386
11	**Set 2 - Set 4**	-8.829	-3.276
12	**Set 2 - Set 5**	-8.623	-3.040
13	**Set 3 - Set 4**	-8.921	-3.185
14	**Set 3 - Set 5**	-8.881	-3.261
15	**Set 4 - Set 5**	-8.754	-3.088

After considering the models previously estimated, this one has been chosen in order to explain the current flows of transport and predict the future ones. The selection obeys to its statistical goodness and its higher explanatory capacity. In addition, from an economic point of view, this model presents the most suitable results according to the gravity model theory.

4.1. Predictions of flows and transport operations

Relying on the last model of fixed effects by sets of Autonomous Communities, predictions for the period 1999-2018 have been calculated (as it is shown in Figure 1). These predictions represent the added values of the flows of goods for the whole of the regions, as well as the added values of the transport operations of the total transit within the country. They include the commercial activity in every region as well as the interregional flows.



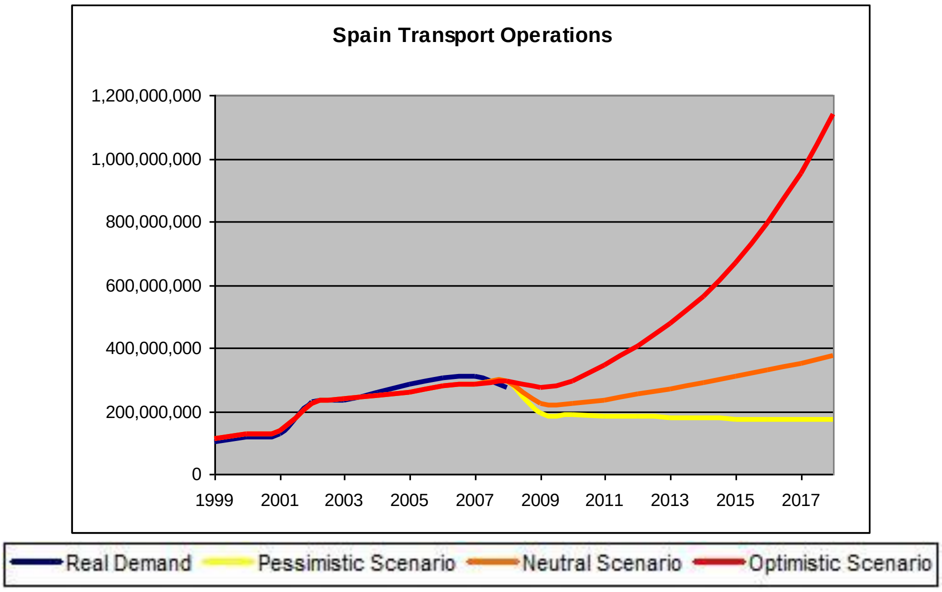


Figure 1. Predictions of flows and transport operations

Additionaly, the appendix includes some graphs with the estimations for every Autonomous Community. These graphs show the evolution of the transport flows and operations originated in every community for the period of study.

Ex ante predictions are calculated from 2009 to 2018. In order to obtain the values of trade flows in these years, it is necessary to introduce in the equation of the model the predicted values for each of the explanatory variables. Dummy variables Border and Intra are time-variant so they do not present any problem.

However, due to the high volatility of the economic environment, three possible scenarios for the evolution of the GDP have been considered in order to carry out the *ex ante* prediction. These scenarios have been designed by taking geometric measures of the annual growth rates from different periods.

Therefore, the pessimistic scenario takes into account the geometric mean of the annual growth rate of 2008 and the annual growth rates predicted by the European Union for Spain. In the case of the neutral scenario, the growth rate results from the application of the geometric mean to the annual growth rates in the available data period (from 1996 to 2010), including the predictions for 2009 and 2010. Regarding the optimistic scenario, we apply the rate resulting form the geometric mean of the rates of the previous period to the decrease in the growth rate of the GDP started in 2008 (that is, from 1996 to 2007). Since there are not regional data, the growth rates shown in Table 9 are calculated for the whole country, so the same rates have been applied to all the Spanish Autonomous Communities.

Table 9. GDP average growth predictions calculated for Spain (2011-2018)

Scenario	Rates	Geometric mean of the period
Pessimistic	0,99%	2008-2010
Neutral	1,026%	1996-2010
Optimistic	1,035%	1996-2007

In the case of the prediction of the general prices of transport, three scenarios have also been considered taking into account that official forecasts have not been released: a deflationary scenario, where the minimum annual rate of the whole available series was used as the rate of price change; a neutral scenario, where the rate was calculated as the geometric mean of the available rates of the series, and finally, a inflationary scenario with the maximum rate of the series as the change rate.

The predicted values of the variable kilometres of high capacity network *per capita* have been calculated in the following way: in the numerator, for the change in kilometres in the 3 years of the prediction, three scenarios have been calculated applying the same methodology used for the general price variable. In the denominator, we used the predicted value in the short-term projections of the population of the National Statistics Institute.

As it has been mentioned in Section 3.1, the distance variable is obtained by dividing the ton-kilometre variable by the tons. The timeline of both variables reaches 2008. From 2009, the source of the values was the calculation of the geometric mean of the distance variable for each origin-destination pair in the period 1999-2008.

The following equation has been applied to deduce the predicted values of the trade flows:

$$\begin{aligned}\ln X_{ijt} = {} & \alpha_{ij} - 0.878 \ln Pa_t - 1.198 \ln Dist_{ijt} + 0.535 Border + 1.211 Intra1 + \\ & + 1.708 Intra2 + 2.191 Intra3 + 2.210 Intra4 + 2.439 Intra5 + 1.925 Intra6 + \\ & + 0.666 Intra7 + 1.326 Intra8 + 2.409 Intra9 + 1.674 Intra10 + 0.136 Intra11 + \\ & + 2.065 Intra12 + 1.432 Intra13 + 0.881 Intra14 + 1.696 Intra15 + 0.972 \ln GDP_{it} + \\ & + 1.000 \ln GDP_{jt} + 0.536 \ln KHCNpc_{it} + 0.392 \ln KHCNpc_{it} + e_{ijt}\end{aligned} \tag{4}$$

Regarding the prediction calculation for transport operations, the following equation has been used:

$$\begin{aligned}\ln X_{ijt} = {} & \alpha_{ij} - 0.769 \ln Pa_t - 1.366 \ln Dist_{ijt} + 0.666 Border + 1.358 Intra1 + \\ & + 1.872 Intra2 + 2.295 Intra3 + 2.377 Intra4 + 2.442 Intra5 + 1.996 Intra6 + \\ & + 1.020 Intra7 + 1.583 Intra8 + 2.511 Intra9 + 2.045 Intra10 + 0.251 Intra11 + \\ & + 2.067 Intra12 + 1.399 Intra13 + 1.015 Intra14 + 1.739 Intra15 + 0.954 \ln GDP_{it} + \\ & + 0.944 \ln GDP_{jt} + 0.459 \ln KHCNpc_{it} + 0.395 \ln KHCNpc_{it} + e_{ijt}\end{aligned} \tag{5}$$

Flow and transport operation predictions for each Spanish Autonomous Community are shown in Appendix A3.

4.2. Application of the model to the calculation of CO_2 emissions

The calculation of Greenhouse gas emissions is one of the most relevant applications of the estimated models. In this section, we present a preliminary approach to the calculation and prediction of CO_2 emissions resulting from domestic road transport of goods in Spain for the period between 1999 and 2008.

The calculation procedure relies on the exploitation of the potential of the data provided by the Permanent Survey on Road Transport of Goods and the high explanatory capacity of the estimated models.

The following equation has been used to obtain the total amount of CO_2 emissions in the year t:

$$E_{CO_2}^{t} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(D_t \cdot Q \cdot Dist_{ijt} \cdot O_{ijt} \right) \qquad (6)$$

Where:

D_t are the litres of diesel per kilometre.

Q are the CO_2 kilograms emitted to the atmosphere per litre of diesel.

$Dist_{ijt}$ are the kilometres separating the region of origin i from the region of destination j for each year t.

O_{ijt} are the transport operations between the region of origin i and the region of destination j for each year t.

In order to calculate the fuel consumption (in litres), we relied on the information provided by the Transport Costs Observatory, derived from the Permanent Survey of Road Transport of Goods. According to this report, the average consumption of an articulated freight vehicle in 2001 accounted for 0.385 litres per kilometre. Due to the long timeline of this study, the technological improvements applied to the vehicles have to be taken into account. The mean gain of energy efficiency of new vehicles in the last 40 years accounts for 0.8 - 1% per year (McKinnon, 2008). In addition, the Permanent Survey provides the average age of the fleet of heavy vehicles for each year. On the basis of these data, a fuel consumption efficiency index was generated for heavy vehicles in Spain.

Regarding the amount of CO_2 kilograms emitted to the atmosphere per litre of diesel, it accounted for 2.71 CO_2 Kg. / litre; this figure is included in the inventory of Greenhouse gas emissions in Spain (Ministry of Environment, Rural and Seaside Areas, 2009).

The representation of the total amount of CO_2 gas emissions obtained by means of this methodology is shown in Figure 2, considering the same scenarios included in the prediction of transport operations.

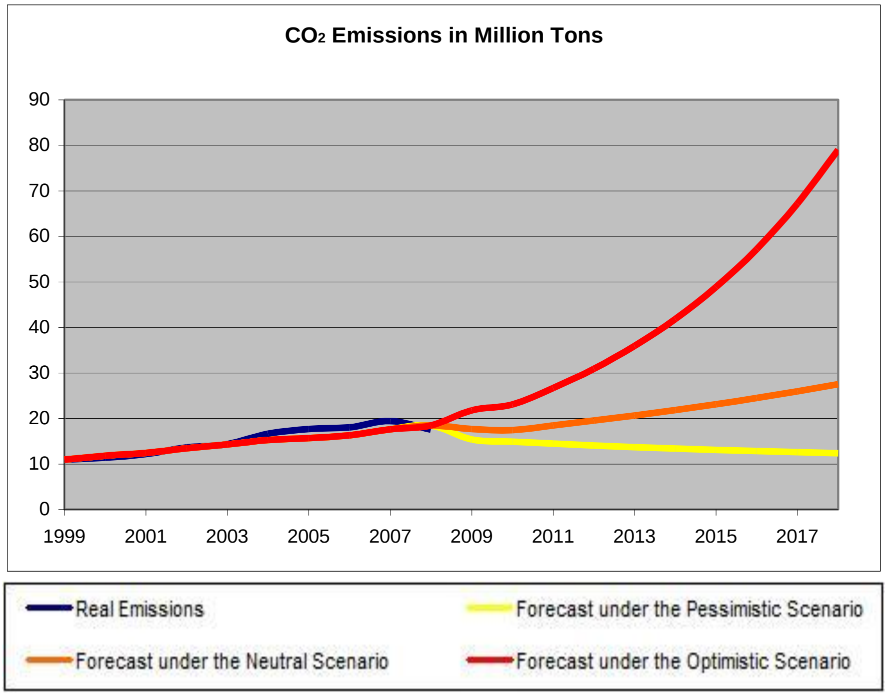


Figure 2. Estimation of polluting emissions of road transport of goods

4. CONCLUSIONS

This study reports the estimation of six gravity models intended to explain the flows of road transport of goods among the different Autonomous Communities in Spain (and within each region). The flows of transport are measured according to the transported tons and to the number of transport operations. The timeline of the research goes from 1999 to 2008 due to the availability of the different explanatory variables. The first estimated model is a cross-sectional method that presents coherent results with the economic theory. In addition, other four fixed effect panel models have been estimated in order to avoid any possible bias caused by the heterogeneity of the trade patterns among the different areas. The fixed effects of the first two models were the pairs of communities, whereas in the other 2 cases the fixed effects were the communities of origin and destination of the flows. The first two models have been rejected due to heteroskedasticity problems, meanwhile the other two were ignored on the basis of a correlation of variables and fixed effects. Finally, we present a model where fixed effects are determined by groups of communities that report the common, specific and unobservable characteristics of the regions.

From the results obtained in the final model, it is interesting to underline that all the coefficients are larger, in absolute value, than the ones presented in the cross-sectional model. In addition, these results are reasonable from an economic point of view and similar to the ones obtained in related studies. The theory of the gravity model is verified by means of the elasticities obtained for the transported tons and the transport operations in the variables distance (-1.198 and -1.366) GDP in origin (0.972 and 0.954) and GDP in destination (1.000 and 0.944). These results are to be included within the highest and lowest elasticity interval considered in other studies. The estimated coefficients for the variable “distance” in the works of Gil et al. (2010) and Llano et al (2009) vary between -0,63 and -1,12 for the first one and between 0,792 and -1,456 for the second. Martin and Pham (2008) obtain a value of 0,711 for Exporter GDP; this value goes from 0,69 to 0,76 in the estimations of Gil-Pareja et al. (2010). For the GDP variable in destination, the results obtained by Martin and Pham (2008) vary between 0,8 and 0,9 being lower than the unitary coefficient obtained by Anderson and van Wincoop (2003) as it happens in other studies.

Since we are dealing with demand models, the general price variable yields negative coefficients with close values to the unit in absolute value (-0.878 and -0.769). In spite of being a price inelastic demand, economic policy decisions cannot be deduced since this variable represents the sum of costs per kilometre instead of a unitary price that can be fixed by a particular customer. Besides, the results obtained for the flows of transport allow for the estimation of the CO_2 emissions associated to the road transport of goods.

The future research lines derived from this study rely on the use of gravity models in the estimation of demands for other types of transport. In addition, further research could include the construction of a system of equations that allows for the analysis of the different modes and the estimation of cross elasticity among them, applying these models to the analysis of environmental impact.

REFERENCES

Anderson, J.E. (1979). A theoretical foundation for the gravity model, American Economic Review, 69, 106-116.

Anderson, J.E, and van Wincoop, E. (2001). Borders, trade and welfare, NBER Working paper No. 8515.

Anderson, J.E. and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle, American Economic Review, 93, 170-192.

Artal, A., Castillo, J. and Requena, F. (2006). Contrastación empírica del modelo de dotaciones factoriales para el comercio interregional en España, Investigaciones económicas. vol. XXX, 539-576.

Bergkvist, E. and Westin, L. (1997). Estimation of gravity models by OLS estimation, NLS estimation, Poisson and Neural Network specifications, CERUM, Working Paper No.6

Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations. Working Paper Nº 12516, NBER, Cambridge.

Borra, C. (2004). La estimación de la demanda de transportes de mercancías. Una aplicación para Andalucía, Ed. University of Seville.

Castells, A. and Parellada, M. (1979). Els fluxos econòmics de Catalunya amb la resta d'Espanya i la resta del món, Borsa d'Estudi Jaume Carner i Romeu, Barcelona.

De Rus, G., Campos, J. and Nombela, G. (2003). Economía del transporte, Antoni Bosch, Barcelona.

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

Egger, P. and Pfaffermayr, M. (2004). The panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects, Empirical Economics, 28, 571-580.

Egger, P. and Pfaffermayr, M. (2004). Distance, trade and FDI: a Hausman-Taylor SUR approach, Journal of Applied Econometrics, 19(2), 227-246.

Eichengreen, B., Irwin, D. (1996). The role of history in bilateral trade flows, NBER Working paper 5565.

Garín, T. (2007). German demand for tourism in Spain, Tourism Management, 28, 12–22.

Gil-Pareja, S., Llorca-Vivero, R. and Martínez-Serrano, J.(2006) The border effect in Spain: The Basque Country case, Regional Studies, 40: 4, 335 — 345

Greene, W. (1999). Econometric Analysis, 3rd edition, Prentice Hall.

Harris, M. and Mátyás, L. (1998). The Econometrics of Gravity Models, Melbourne Institute Working Paper No. 5/98

Haynes, K. E. and Fotheringham, A. S. (1984). Gravity and Spatial Interaction Models, SAGE, 9-13.

Helpman, E., Melitz, M. and Rubinstein, Y. (2007). Estimating trade flows: Trading partners and trading volumes, NBER Working Paper No. 12927.

Hernández Muñiz, M. (1999). Dinámica espacial de la economía española e infraestructuras de transporte. Ph. Thesis.

Isard, W. (1956). Location and Space-economy; a General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure, Cambridge, Cap.9.

Jorge-Calderón, J.D. (1997). A demand model for scheduled airline services on International European routes, Journal of Air Transport Management, Vol. 3, 23-35.

Lesage, J.P. and Pace, R.K. (2008). Spatial econometric modelling of origin-destination flows, Journal of Regional Science, 48, 941-967.

Levinson, D. and Kumar, A. (1995). Activity, Travel, and the Allocation of Time, Journal of the American Planning Association, Vol. 61, 548–470.

Llano, C. (2004). Economía espacial y sectorial: el comercio interregional en el marco Input-Output, Instituto de Estudios Fiscales, Ministerio de Economía y Hacienda, No.1.

Llano, C., Esteban, A., Pérez, J. and Pulido, A. (2008). La base de datos C-intereg sobre el comercio interregional de bienes (1995-06): metodología, Documento de trabajo del Instituto Klein.

Llano C., Minondo, A. Requena, F. (2009) "Is the border effect an artefact of geographic aggregation?". Forthcoming

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

McKinnon, A. (2008). The potential of economic incentives to reduce CO_2 emissions from goods transport, Paper prepared for the 1st International Forum on “Transport and energy: The challenge of climate change”, Leipzig.

Martín, J. C. and Nombela, G. (2008). Impacto de los nuevos trenes AVE sobre la movilidad en España, Revista de Economía Aplicada, Vol. 16, 5-23.

Martin, W. and Pham, C. (2008) Estimating the Gravity Model When Zero Trade Flows Are Frequent, Economics Series 2008_03, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Martínez-Zarzoso, I., Felicitas, N and Horsewood, N. (2008). Are regional trading agreements beneficial?, Static and dynamic panel gravity models", North American Journal of Economics and Finance, Vol. 20, 1, 46–65.

Mátyás, L. (1998). The gravity model: som econometric considerations. The World Economy, 21, 397-401.

Ministerio de Medio Ambiente, Medio Rural y Marino (2009). Inventario de emisiones de gases de efecto invernadero de España Años 1990-2007.

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

Tinbergen, J. (1962). Shaping the world economy. Suggestions for an International Economic Policy, New York: The Twentieth Century Fund.

Westerlund, J. and Wilhelmsson, F. (2009). Estimating the gravity model without gravity using panel data, Applied Economics, Forthcoming

APPENDIX

A.1 Hausman Test

The application of this test is intended to verify if there are significant and systematic differences between the two estimators. This methodology has been chosen in order to test if it is preferable to use a fixed effect panel model or a random effect panel. Therefore, the contrast of the following null hypothesis is suggested:

$$H_0 : Cov(u_i, X_{it}) = 0 \qquad (7)$$

The test is intended to observe if there is a correlation between the variables included in the model and its random perturbations. If this is not the case, the random effect estimator will be more efficient and therefore it will be chosen for the definitive model.

This contrasting study is carried out by means of the comparison of coefficient estimations and their variations with the distribution of the chi-squared:

$$H = \left(\hat{\beta}_{EF} - \hat{\beta}_{EA}\right)' \left[V\left(\hat{\beta}_{EF}\right) - V\left(\hat{\beta}_{EA}\right)\right]^{-1} \left(\hat{\beta}_{EF} - \hat{\beta}_{EA}\right) \square\ \chi_k^2 \qquad (8)$$

Where $\hat{\beta}_{EF}$ represents the estimation vector of the fixed effect model, $\hat{\beta}_{EA}$ the estimation vector of the random effect model, $V\left(\hat{\beta}_{EF}\right)$ the variation and covariation matrix of the fixed effect estimator, and $V\left(\hat{\beta}_{EA}\right)$ the variation and covariation matrix of the random effect estimator. χ_k^2 is a squared-chi with k degrees of freedom that represents the number of variables introduced in the models including the constant.

The critical level value obtained in this study is 14.04, with a p-value lower than 0.1. With this result, and for a level of confidence of 90%, we conclude that the perturbations and the variables included in the models are correlated. This leads us to consider the use of the fixed effect estimator as the best solution (as theory suggests).

A.2. Fixed effects test of homogeneity

This section is intended to verify the hypothesis that fixed effects are equal among them, that is to say, there are not relevant differences among unobservable heterogeneities of the different groups considered in the definitive model. In that case, it would not be suitable to apply a panel data model and therefore the use of a cross-sectional model would be the right choice (since a single constant is estimated in this model).

The hypothesis to be tested is the following:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_{15} \qquad (9)$$

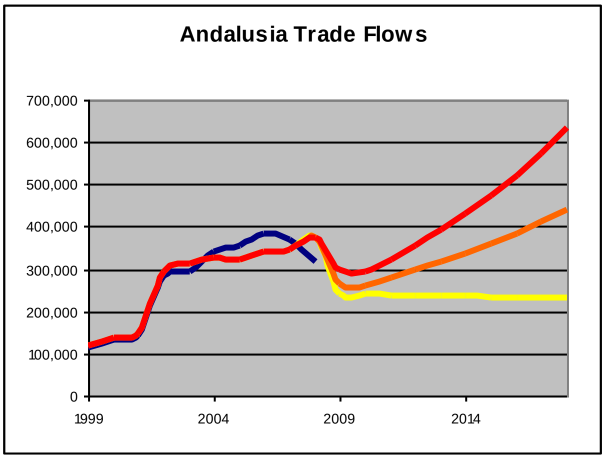
reflecting the imposition of 14 linear restrictions. In order to verify that these conditions are met, both models are estimated, computing the existent difference between the sum of the squares of the errors of the two models. By means of the comparison with a statistic distribution -Snedecor’s F distribution-, it can be observed whether the difference between the models is large enough as to support the use the panel data model:

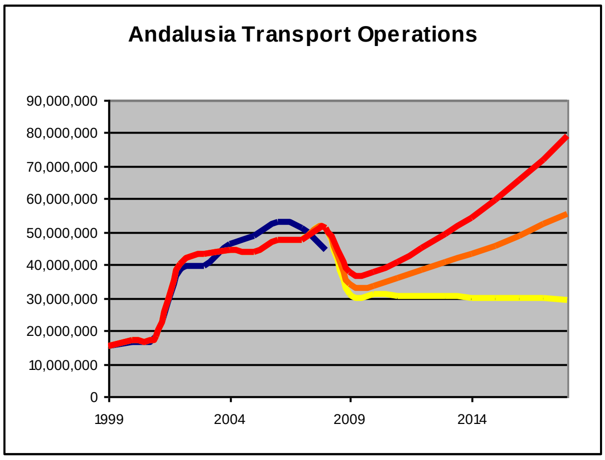
$$F = \frac{SCE_R - SCE_U \big/ r}{SCE_U \big/ n-k} \sim F_{n-k}^{r} \qquad (10)$$

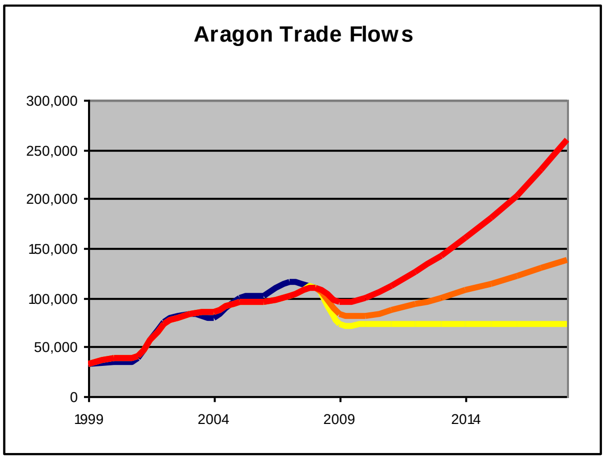
Where SCE_R represents the sum of the squares of the errors of the restricted model (cross-section), SCE_U is the sum of the squares of the non-restricted model (panel), r is the number of imposed restrictions and n-k is the number of observations minus the number of estimated parameters. The value resulting from the previous quotient is 31.199, and using the statistical tables we know that $P\left(F_{2213}^{14} \leq 2.090\right) = 0.99$.

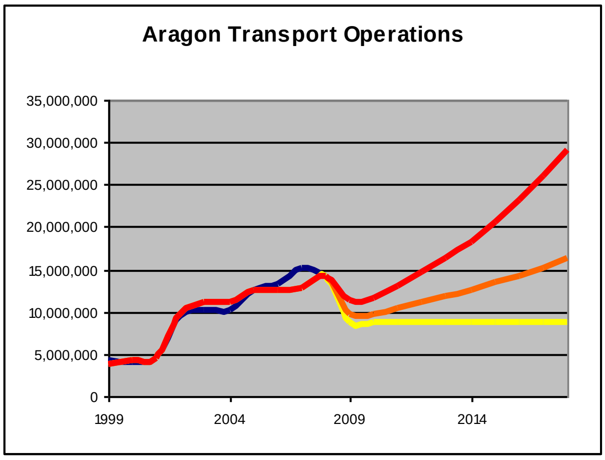
Therefore, the null hypothesis formulated in (9) is rejected, concluding that the proposal of a fixed effects panel is meaningful for the groups considered in the final model.

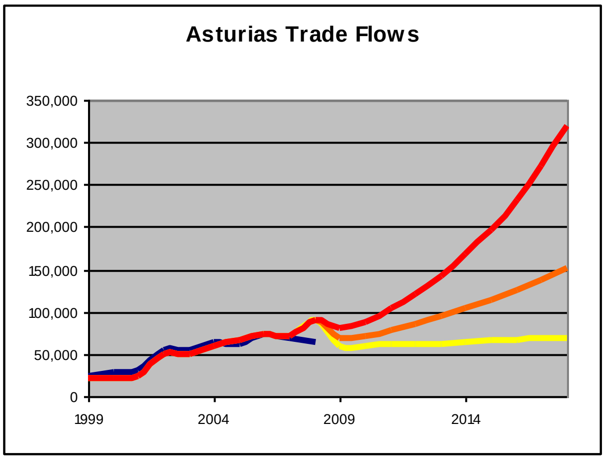
A.3 Flow and transport operation predictions for each Autonomous Community

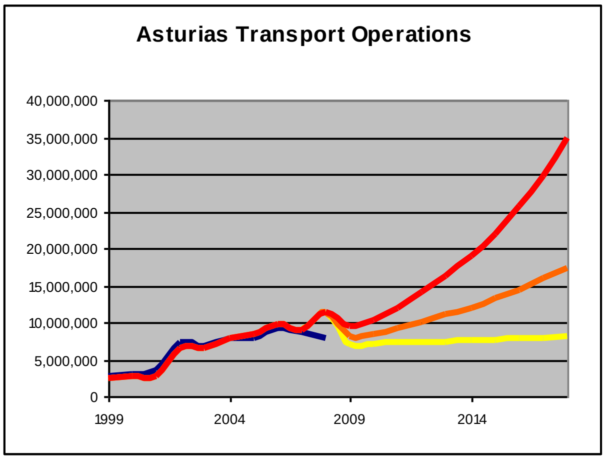

Andalusia Trade Flows


Andalusia Transport Operations

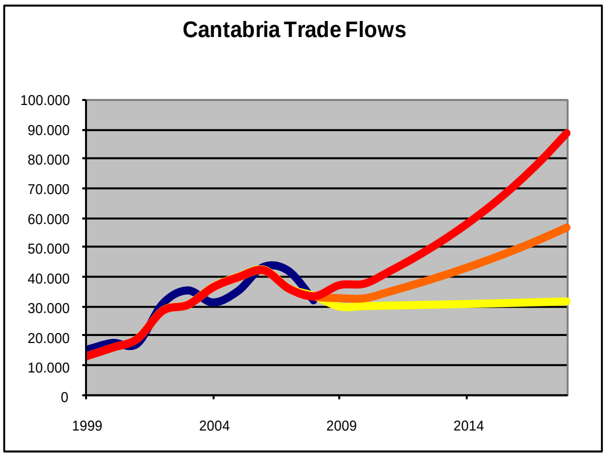

Aragon Trade Flows

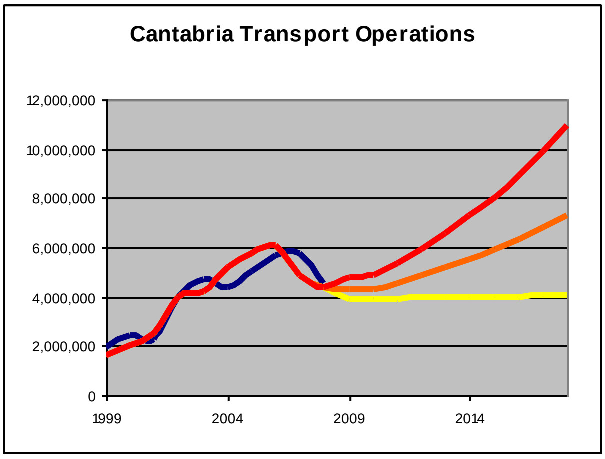

Aragon Transport Operations

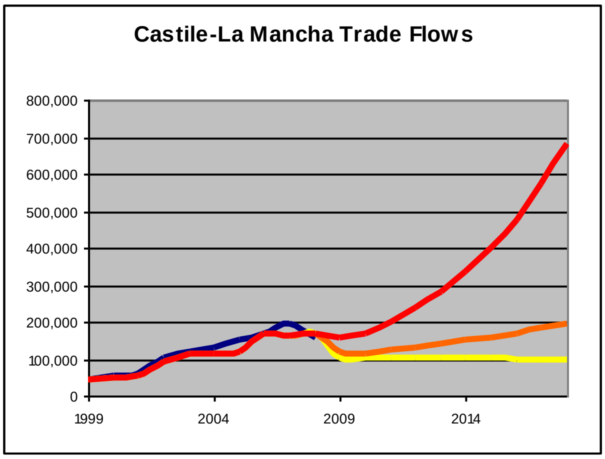

Asturias Trade Flows

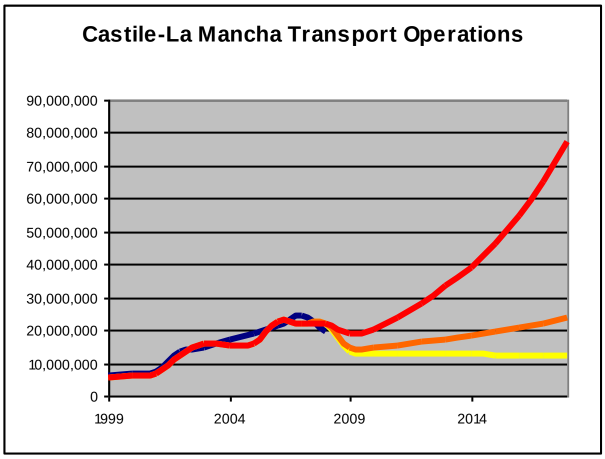

Asturias Transport Operations

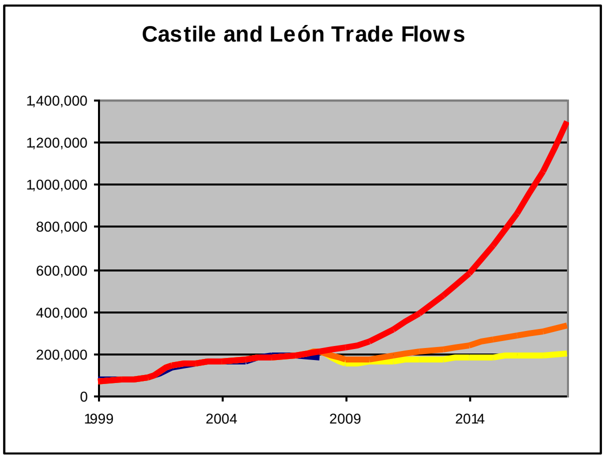
A.3 Flow and transport operation predictions for each Autonomous Community (cont.)

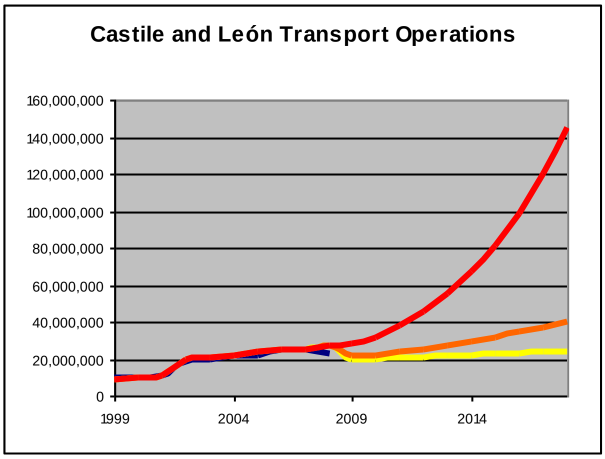






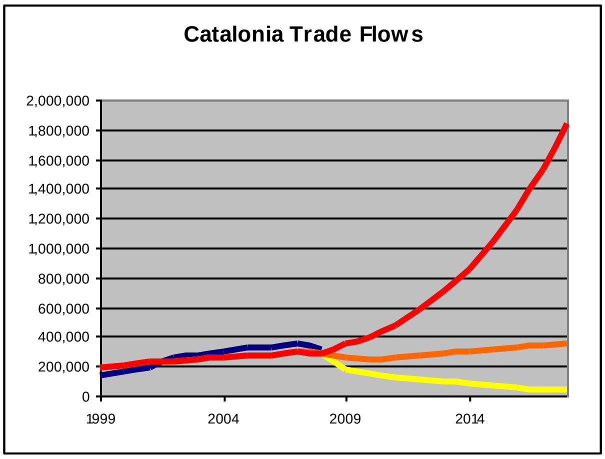


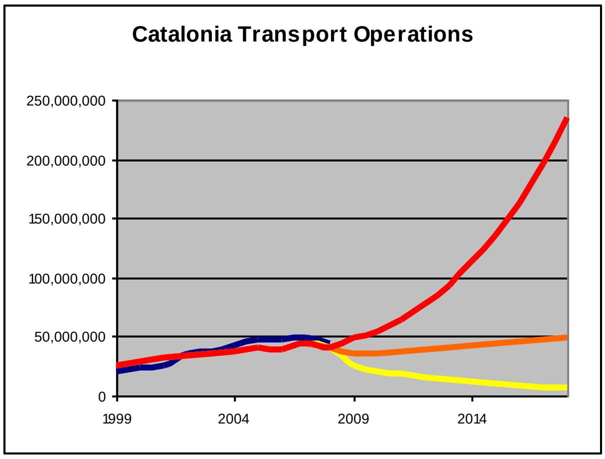


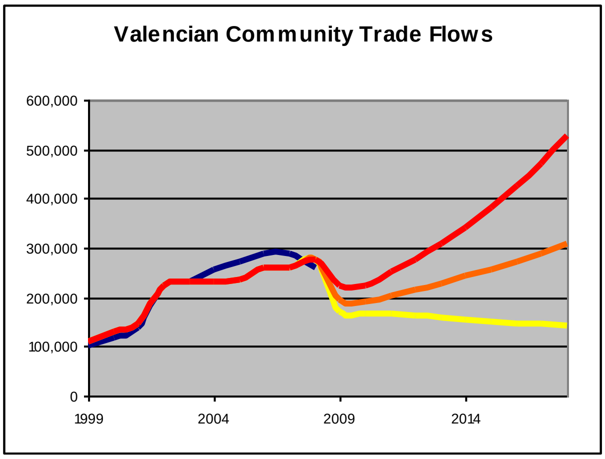


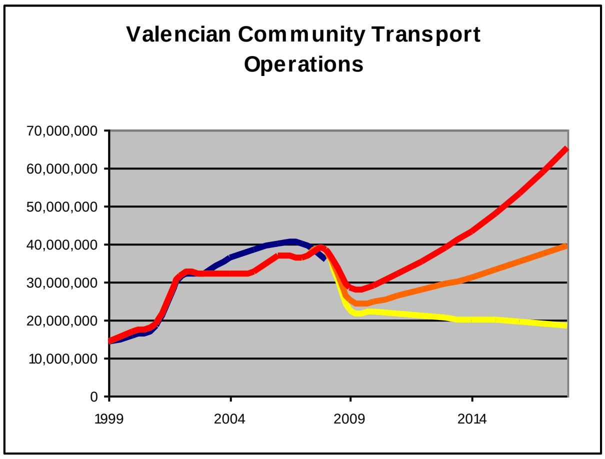
— Real Demand — Pessimistic Scenario — Neutral Scenario — Optimistic Scenario

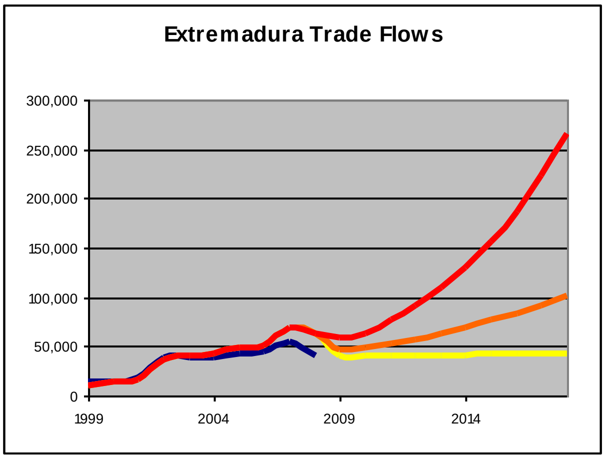
A.3 Flow and transport operation predictions for each Autonomous Community (cont.)

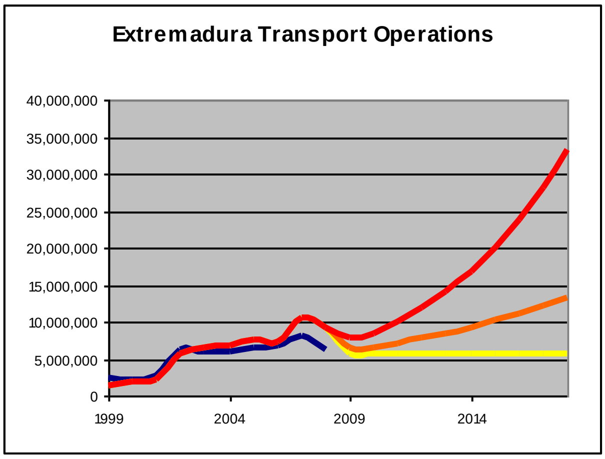

Catalonia Trade Flows


Catalonia Transport Operations


Valencian Community Trade Flows

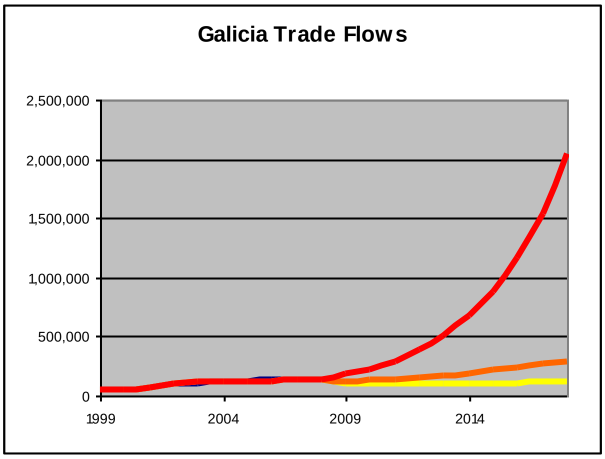

Valencian Community Transport Operations

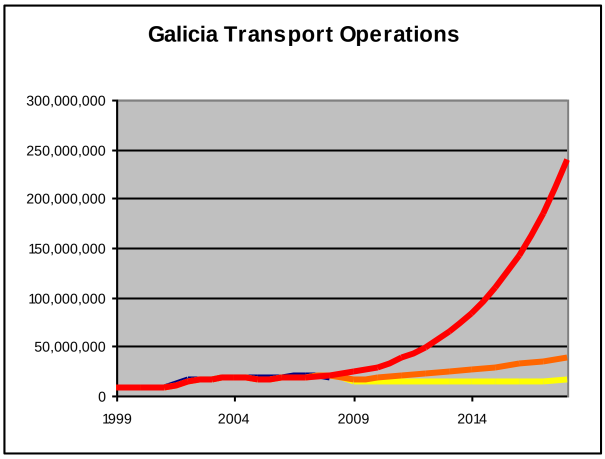

Extremadura Trade Flows

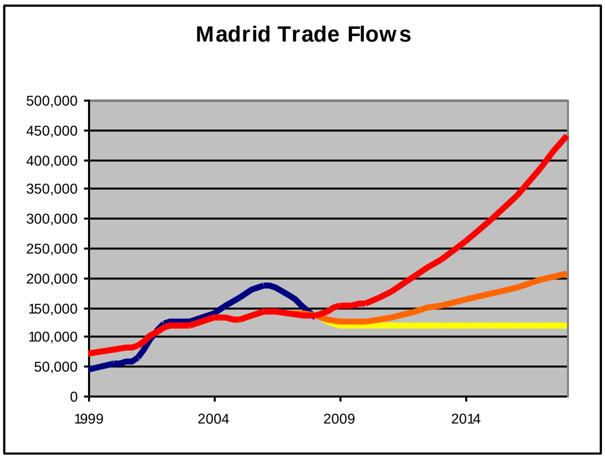

Extremadura Transport Operations

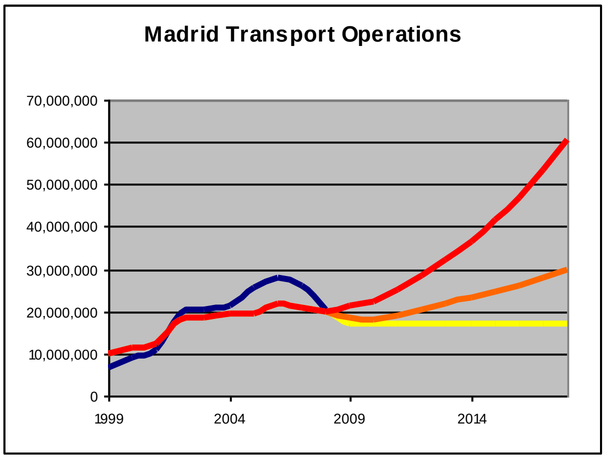
— Real Demand — Pessimistic Scenario — Neutral Scenario — Optimistic Scenario

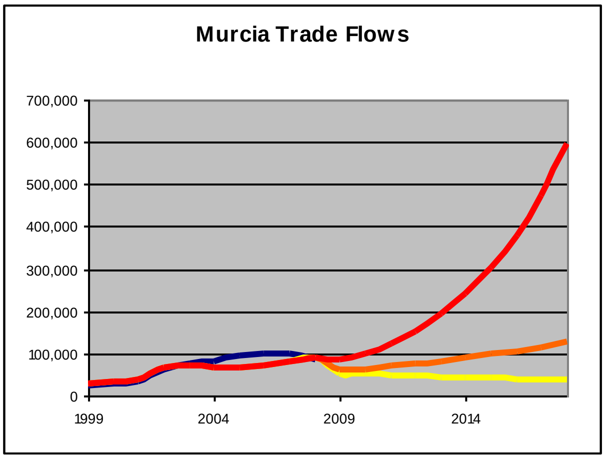
A.3 Flow and transport operation predictions for each Autonomous Community (cont.)

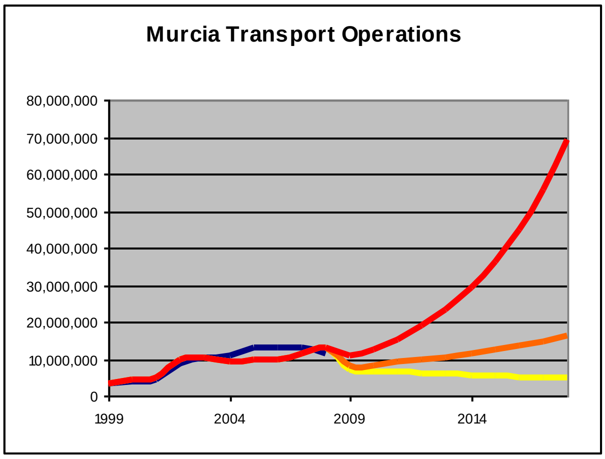

Galicia Trade Flows


Galicia Transport Operations


Madrid Trade Flows

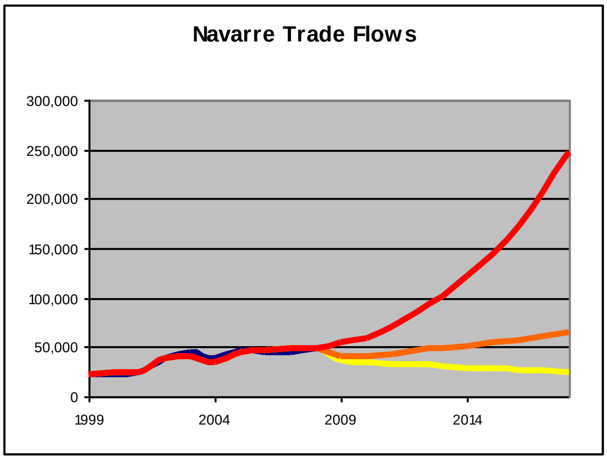

Madrid Transport Operations

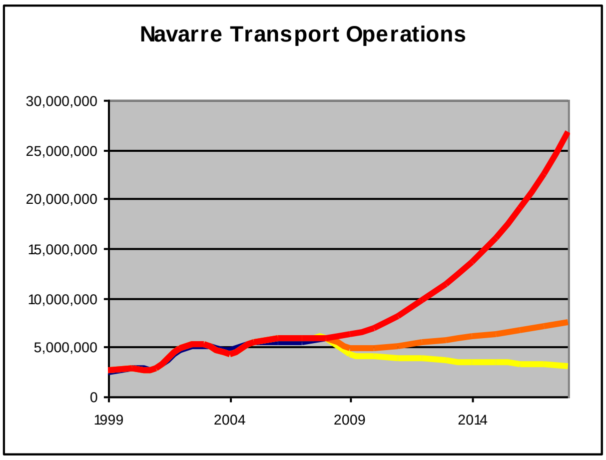

Murcia Trade Flows

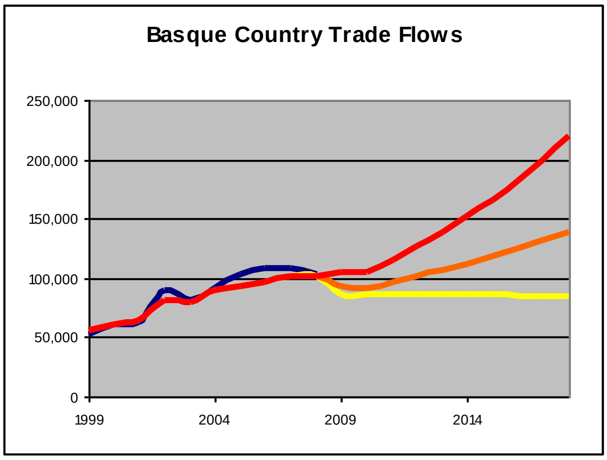

Murcia Transport Operations

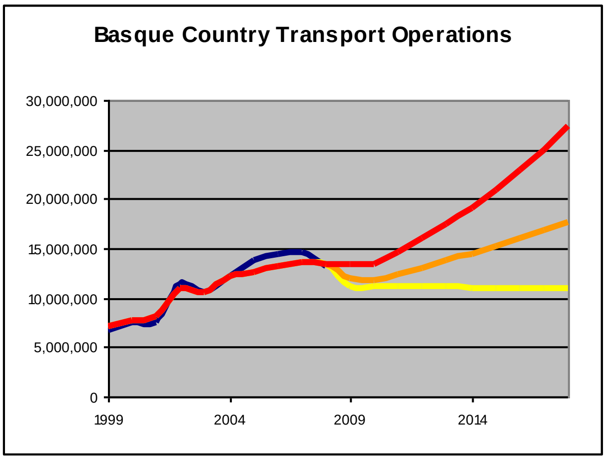
▬ Real Demand ▬ Pessimistic Scenario ▬ Neutral Scenario ▬ Optimistic Scenario

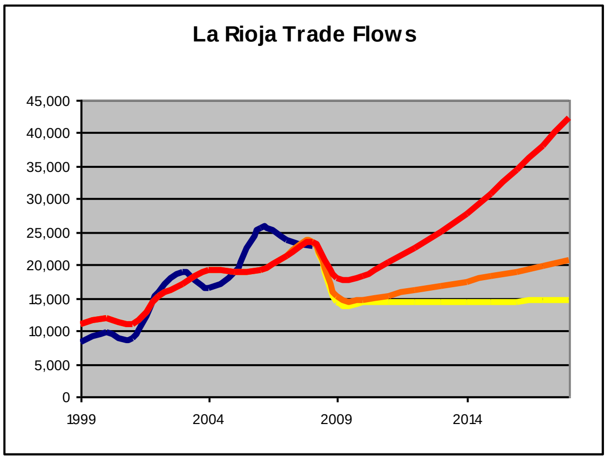
A.3 Flow and transport operation predictions for each Autonomous Community (cont.)

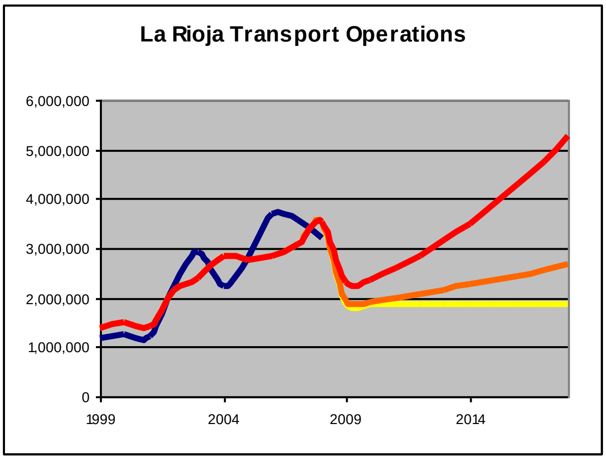












— Real Demand — Pessimistic Scenario — Neutral Scenario — Optimistic Scenario

ACKNOWLEDGEMENTS:

Previous versions of this study have been presented in 2009 in the 35th Meeting of Regional Studies and in a seminar at the University of Oviedo. The authors of the paper are grateful for the remarks and suggestions of the attendants to the congress and the seminar, especially to Juan Prieto and Manuel Hernández. We would like to express our gratitude to the Spanish Ministry of Science and Innovation for the financial support in the development of this research (included in the PSE-GLOBALOG Programme).