Ensayos sobre Economía de Transporte

Pelayo González Arbués

Departamento de Economía

Universidad de Oviedo
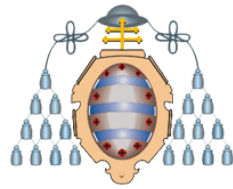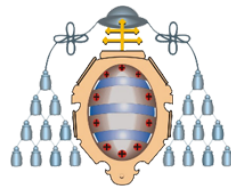
Essays on Transport Economics

Pelayo González Arbués

Departamento de Economía

Universidad de Oviedo

Programa de doctorado

Economía: Instrumentos del Análisis Económico

Oviedo, 2013

Acknowledgements

Contents

List of Figures

Chapter 1

Introduction

Transportation systems have a critical function in many different aspects of contemporary societies. It is difficult to imagine a developed economy with an inadequate transportation network. Public policies have traditionally aimed to improve the efficiency of such systems by pursuing a variety of goals. For instance, a more efficient transportation network lowers firms' transportation costs, which reduces the total cost of products and might lead to a reduction of the final prices paid by consumers. Although lowering logistical costs is an important consequence of improving the quality of a transportation system, the positive after effects to the economy extend further. Other major consequences of a high-quality transportation system that have a substantial impact on the development of societies include fostering specialized labor markets along with urban agglomerations and increased trade levels, among others. Land use shapes transportation planning, which simultaneously determines the spatial configuration of accessibility to public and private services that is, of course, important to economic development.

Nevertheless, providing these transportation services requires substantial re-

sources that can burden the economy with substantial opportunity costs. The construction of transportation infrastructure largely depends on the considerable use of a skilled labor force, heavy machinery and financial resources. Additionally, the negative externalities imposed on society by transportation networks are wide-ranging and may include not only environmental damage, such as air pollution and climate change, but also traffic noise and increased risk of accidents. Most citizens are familiar with some of the ineluctable consequences of traffic congestion, such as the considerable loss of time caused by slow-moving traffic that inevitably entails inconveniences to commuters and consequently reduces their welfare. Furthermore, one of the greatest challenges faced by developed economies, and particularly their transportation sectors, is their overwhelming dependence on fossil fuels.

As a result of all these factors, we believe that we should undertake an exhaustive analysis of transportation activities. Transportation consists of a complex system in which system-level characteristics arise from the behavior and interactions of a large number of decision-makers. In this context, the analysis is typically undertaken through a multidisciplinary approach in which economics is one of the core fields of research. Thus, the transportation sector presents a wide range of economic problems at both the micro and macro levels that must be studied through the lens of economics.

Economic analysis might be useful to examine the oligopolistic structure of the aviation market, the mechanisms that attempt to mitigate the environmental costs associated with freight transportation and/or the welfare effects of public infrastructure provision. All these problems are similar to those that are typically addressed with tools that are commonly used in economic analyses. Nevertheless,

the existence of a specific branch of economics related to the study of transportation is justified by two fundamental features of this sector: space and time.

Transportation is necessary because activities are located at different points in space. The role of space affects economic analysis because it influences market accessibility, land price and, therefore, land usage. Moreover, land usage and transportation networks influence one another simultaneously and define the spatial structure of activities. In that sense, spatial information is an essential input in the economics of transport.

The second specific feature is time. Spatial separation of activities and agents makes it necessary to spend time moving from one place to another. The evaluation of the different characteristics of transportation systems, including reliability, speed and comfort, requires that time is considered as an important variable. A transportation service (like any service) is non-storable, which means that production and consumption must be simultaneous; this makes short-term decisions about when to travel or ship products key factors in the demand for transportation. Moreover, these short-term decisions are also affected by long-term decisions such as car ownership, residence location or the outsourcing of transportation services.

In this regard, I would like to make a short comment on my own evolution as a researcher and the impact of space and time in the process of writing this dissertation. In the first essay, a dynamic specification of the gravity model aims to explain certain interaction data, road freight flows, and explores the importance of the temporal dimension on road transportation systems. In that specification, the effect of space is simply considered by including fixed effects and road distance between different pairs of regions, which is a common approach in the

empirical literature that applies standard econometric techniques. However, the research on this topic, and particularly a seminar by Carlos Llano hosted jointly by the Department of Economics and the Department of Applied Economics of the University of Oviedo, impressed upon me the importance of spatial information in this field of economics. Since that time, I have updated and broadened my knowledge of spatial econometrics, which complemented my previous education in standard econometrics and opened my mind to consider space as a fundamental aspect of the analysis.

The second essay focuses on the choice of transport mode and exploits the spatial characteristics of the data through standard econometric models but with the particularity that enables the analysis to account for spatial heterogeneity. This feature of the model permits spatial variation with respect to certain estimated parameters, which allows for a more flexible specification. The influence of time is not explicitly considered in this essay, although it is approximated by the inclusion of distance among the variables shaping transport mode choice.

In the third essay, the importance of space is taken into account by considering spatial dependence among regions as measured by means of the Spatial Durbin Model, which enables us to study the spatial dependence of the explicative variables and the dependent variable. In our specification, the spillover effect of the latter indicates that the business cycles of various provinces affect how public and private capital is utilized in other regions.

Before presenting the outline of this dissertation, it should be noted that the essays collected in it are adaptations of articles that have been sent to specialized academic journals. Each paper is self-contained and may be read independently of the others. The first chapter is a version of an article co-authored by my co-

director, Dr. José F. Baños. A version of the second essay was prepared jointly with my two supervisors and with the fundamental contribution of our colleague Patricia Suárez. Finally, the third essay is an adaptation of an article that I prepared with my two supervisors, Dr. José F. Baños and Dr. Matías Mayor.

The overall goal of Chapter 2 is to find the determinants of freight transport flows by road. More specifically, the study focuses on the effect of lagged transport operations on current flows. The thesis to be tested posits that businesses open distribution channels and create networks in some regions and thereby incur sunk costs that generate inertia in freight flows. The methodology adopted is a gravity model, which is a tool that has been the workhorse of international trade in recent decades and that enables the study of the underlying transport flows among regions. In its basic specification, the model permits estimating the effects of scale variables, such as gross domestic product or population (as drivers of trade), and the effects of distance as a deterrent of commercial interchange. In this essay, the model assumes a dynamic functional form that includes the lag of the dependent variable, which permits us to test the dynamic characteristics of transportation flows. This dynamic approach, which has recently been implemented in international goods trade models in place of the traditional static specification, is applied to the case of Spain using panel data that consist of the 15 NUTS-3 regions of the Peninsula between 1999 and 2009. The estimation procedure is the System General Methods of Moments methodology and includes the quality of road transport infrastructure as an additional explicative variable. Finally, the estimated equation is used to predict traffic flows among regions and calculate total emissions associated with such predictions.

Chapter 3 presents another study related to the demand for transportation

services, although the focus is now on passengers (not freight carriers). The essay analyzes modal choice determinants in long distance travel by three modes of transport: private car, public bus and train. Information about the profile of potential users is a major concern for producers and public agencies in every market. With respect to transportation economics, modeling modal choice is a fundamental component for policy-makers trying to improve transportation system sustainability. However, the empirical literature typically focuses on short-distance trips within urban systems. This essay contributes to the limited number of studies on transport mode choice in medium- and long-distance travel. We employ data from the 2006 Spanish National Mobility Survey to illuminate how socio-economic factors and trip attributes affect the selection of the primary mode of transportation. In addition to the traditional explicative variables used in inter-city modal choice analysis, two additional variables have been included, travel distance and the out-of-home duration of the trip. Travelers can be nested in provinces depending on the province of origin of their trips. The hierarchical nature of the data allows a multilevel approach to be employed. In particular, a multilevel multinomial model with random intercepts is estimated. This methodology presents an important feature for this purpose because it permits accounting for the potential spatial heterogeneity of travelers. The spatial heterogeneity of customers can be related to non-homogeneous preferences, infrastructure characteristics of the location of the origin of the trip or to area characteristics of the modes of transport such as price or quality.

The last essay is found in Chapter 4 and addresses a traditional topic on the supply side of transportation services by offering new insights into the contribution of public infrastructure to regional economies using a production function

approach. Although there is a vast body of work analyzing the effects of public investment on the performance of the economy, new developments in the spatial econometrics literature allows this topic to be explored from a new perspective. The empirical literature that studies the effects of public capital on the performance of private enterprises remains inconclusive more than 20 years after the pioneering studies in this genre. Although aggregated models have generated results that show significant effects of public capital on productivity, disaggregated applications have produced results that have shown little or no significant effects. The traditional explanation for the differences in these results is the existence of regional spillovers and network effects that are caused by transportation and communication infrastructures.

According to this view, the aggregate effect seems to consist of the direct and indirect effects of public capital investment. The essay contributes to the literature in at least three ways. First, it explicitly tests the existence of spatial dependence among geographical units and models transportation infrastructure spillovers. The specification applied is known as the Spatial Durbin Model, which includes the spatial lags of both the independent and dependent variables. Second, a theoretical justification for using the Spatial Durbin Model in regional production function studies is provided. Finally, another contribution includes the proposal of an alternative variable to substitute for traditional stock measures and overcome the caveats of these forms of measurement. The proposed model is applied to the 47 mainland Spanish provinces during the 1986-2006 period that witnessed heavy investments in public capital.

Chapter 2

A Dynamic Approach to Road Freight Flows Modeling in Spain

2.1   Introduction

Freight transport plays a key role in economic activity and enables the exchange of goods between economic agents while boosting regional development. However, the social impact in terms of negative external effects on both security and the environment is not negligible. In recent decades, population growth and greater economic activity have caused a steady increase in goods traffic worldwide.

However, as statistics indicate, these new transport operations are not carried out uniformly by the different modes of transport. In the case of Europe, road transport is the most frequently used medium for the movement of goods. For example, in 2009, 79% of the volume of goods deployed within the borders of the EU-15 was transported by road compared to 14% by rail, with inland waterways accounting for the remaining 7%. In Spain, the pattern is very similar: according

to the National Statistics Institute, road transport accounts for 77 of every 100 tons transported. The prevalence of road transport was consolidated over the period 1999-2009, when its use increased by 107% with a growth rate of 11%. To put this into context it should be noted that 80% of the products manufactured in Spain are destined for domestic trade (Llano *et al.*, 2010).

This enormous volume of road traffic is distributed differently across the transport network of the origin and destination regions. This research aims to estimate elasticities associated with the variables that determine the different patterns of flows between regions using a gravity model. Additionally, once we have estimated the parameters of the model, we make predictions of national transport operations and compute the carbon emissions levels associated with these transport movements.

Despite the apparent simplicity of the gravity model from a theoretical point of view, in empirical applications the model is usually misspecified because unobserved heterogeneity is not controlled for and the dynamic nature of road transport flows are ignored (Egger & Pfaffermayr, 2003). In order to control for bias due to the unobserved heterogeneity of the regions, panel models are used since they are able to capture the individual effects of the territories under study. In recent years, panel models have replaced cross-section settings in the empirical trade and transport demand studies. However, the dynamic role of the internal transport of goods is usually completely ignored. Once established, trade relations between regions can last for long periods of time (Eichengreen & Irwin, 1998). Thus, current trade flows are likely to be dependent on past trade flows, with that the underlying freight movements also depend on variations in traffic flows in previous periods.

Different methodologies have been used to analyze the factors that govern the flow of goods among regions. An early classification of the models used to analyze the traffic of goods between cities (Harker, 1985) distinguished among econometric models, spatial equilibrium models, pricing models, and freight network equilibrium models. In this study, the analysis of freight flows is carried out by applying an econometric model. This methodology is based on using econometric techniques to obtain structural relations between system variables. These models usually ignore formal aspects of the transport network such as, for example, specific nodes, and the spatial dimension of the process is included only by the distance between the regions of origin and destination (Harker, 1987).

These models, also known as microeconomic models, can be classified into aggregated and disaggregated models according to the level of data aggregation (Winston, 1983). The basic unit of observation in the aggregated data model is the total share or the total volume corresponding to a mode of transport for a given geographic level. From a theoretical point of view they are often the solution to a cost-minimization problem, though in empirical applications this is not explicitly recognized. In disaggregated models, the basic unit of observation is the decision-maker's choice for a specific mode of freight. These models are more focused on behavioral aspects of decision-making and have a firmer theoretical grounding, but have the disadvantage that they require extensive data.

Broadly speaking, the data used to calibrate a model will depend on the purpose of the study and the availability of the data, as recognized by Oum (1989). Thus, if the main purpose of the analysis is to predict aggregate traffic, the use of aggregate data would be preferable because otherwise an aggregation method would need to be applied. If the main goal of the paper is simulate how decision-

makers respond to changes in regulation, on the other hand, disaggregated data would be more appropriate. In terms of cost, disaggregated models require an extensive database and these are usually difficult to obtain because of the confidentiality of private information.

Thus, the empirical approach depends on the objectives of the study and the availability of suitable data. This paper uses an aggregate database composed of a balanced panel of the 15 Spanish peninsular regions for the period 1999-2009. Our study has several aims. First, we try to determine whether the correct specification of the gravity model applied to road transport should include a dynamic component. In addition, we attempt to measure the importance of infrastructure quality and also to compute the evolution and prediction of CO_2 emissions caused by road transportation. Moreover, thus is the first study, as far as we are aware, to use a time-varying variable for the distance between regions as a proxy for transportation costs.

2.2   Freight flow models based on gravity

Gravity models have been widely used in the literature of international trade[1].In line with Newtons law of universal gravitation, they are based on the idea that bilateral trade flows are directly related to the size of regions and inversely to the distance that separates them. In general, the variables included in the models to represent the size of the regions are the GDP or population, while among the friction factors, distance is often used as a proxy for transport costs. The augmented gravity model specification can include other variables to test whether they have

---

[1]See Anderson (1979), Helpman & Krugman (1985) and Anderson & van Wincoop (2003)

a significant effect on trade flows. In the international trade literature, dummy variables usually appear to capture the effect of shared language, adjacency or belonging to a monetary union. In our case, we include a proxy for the quality of road infrastructure. By including this variable we try to capture the effect of the heavy investments made by the European Union in Spain through the Structural Funds[2].

Although the theoretical support for the gravity model was originally very poor, since the second half of the 1970s, several theoretical developments have filled this gap. Anderson (1979) made the first formal attempt to derive the gravity equation from a model that assumed product differentiation. Bergstrand (1985, 1989) also explored the theoretical determination of bilateral trade in a series of papers, in which gravity equations were associated with simple monopolistic competition models. In a recent study Anderson & van Wincoop (2003) applied gravity model to bilateral trade including transport costs and other specific trade factors as explanatory variables. Fidrmuc (2009) appoints other popular applications such as the study of the effects of the currency unions on international trade (Rose, 2000), intra-firm trade (Egger & Pfaffermayr, 2005) and bilateral exports and Foreign Direct Investment relationships (Egger, 2001). Gravity models have also been widely applied in the trip distribution step of complex national and international freight transport models (Jong et al., 2012).

Initially, the gravity model was developed for cross-sectional analysis of countries at a particular time. However, to avoid a poor specification of the models, the relative heterogeneity of patterns of trade flows across countries has to be considered (Glick & Rose, 2002).The panel approach has several advantages over

---

[2]See, among others, Cantos et al. (2005)

cross-sectional analysis: apart from increasing the degrees of freedom, it also captures the relationships between variables over a long period of time and identifies the role of the business cycle on these relationships. On the other hand, the use of panel data allows taking into account the time-invariant specific effects of the different regions. The cross-section setting is likely to suffer from omitted variable bias as is does not include the unobserved effects of regions and ignores the temporal effects of trade (Harris & Mátyás, 1998). The comparisons of panel estimators (Egger & Pfaffermayr, 2003) show, however, that instead of using one dummy variable per country, individual country pair dummies (fixed effects) should be included to get efficient estimators. While these authors recommend panel data models, another caveat may be related to the possible non-stationarity of analyzed data (Fidrmuc, 2009).

Furthermore, studies using the gravity model with panel data typically ignore the possibility that trade has a dynamic dimension. The importance of past trade values has been stressed already by Eichengreen & Irwin (1998), who included lagged trade in repeated cross-sections for selected years. Lagged values of trade are usually a highly important determinant of current trade flows. De Benedictis & Vicarelli (2005) highlight that, despite the importance of taking into account unobserved heterogeneity and the probable persistence effect, few studies have used a dynamic specification of a panel model when estimating the gravity equation:Bun & Klaassen (2002) Nardis & Vicarelli (2003), Martínez-Zarzoso & Nowak-Lehmann D. (2003) and Martínez-Zarzoso et al. (2009a,b). To our knowledge, so far no studies in the field of freight flows have exploited this type of model. [3]

---

[3]See Chow et al. (2010) and Jong et al. (2012) for recent surveys of freight transport models.

According to these authors sunk costs incurred by the companies to open distribution channels and service networks in new markets can generate inertia in bilateral trade. Bun & Klaassen (2007), in a study that examines the impact of the European Monetary Union on trade, add that consumers in a country get used to products of another nation, creating a long-term relationship. On the other hand, De Benedictis & Vicarelli (2005) justify this idea by emphasizing the importance of the accumulation of invisible political, cultural and geographical assets that have some influence on the trade between regions. Martínez-Zarzoso et al. (2009b) evaluate the link between German development aid and trade with the recipient country.

The literature on road freight flows modeling makes no reference to dynamics. However, in this study we empirically test for the existence of such effects. These could take the form of a partial adjustment process in which the current level of (observed) transport flows would be adjusted as a proportion of the difference between the desired level in a period and the levels achieved in recent periods. Thus, a coefficient of persistence in the flows variable is considered and identified in the estimation by including the lagged dependent variable as an additional regressor. According to the results, it appears that a second-order setting for a dynamic model is appropriate in line with the work of Bun & Klaassen (2007) [4].

$$
\begin{aligned}
Y_{ijt} &= Y_{ijt-1}^{\beta_1} Y_{ijt-2}^{\beta_2} Dist_{ijt}^{\beta_3} GDP_{it}^{\beta_4} GDP_{jt}^{\beta_5} KHCNpc_{ijt}^{\beta_6} e^{\beta_7 border} e^{\beta_8 intra} e^{u_{ijt}} \\
u_{ijt} &= \mu_{ij} + \lambda_t + \nu_{it}
\end{aligned}
\tag{2.1}
$$

---

[4]Introducing more than two lags of the dependent variable produces non-significant estimates. A specification with one lag was also tested but an econometric test failed to accept it.

where

Y_{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 ($Flows$) and transport operations ($TransportOperations$). In the case of intraregional trade flows, i will be equal to j.

Y_{ijt-1} and Y_{ijt-2} represent the flow of goods transported from the region of origin i to the region of destination j lagged one and two time periods respectively. As with the dependent variable, these are measured alternatively by thousands of tons carried and transport operations.

$Dist_{ijt}$ is the distance between origin and destination. It is calculated by averaging the real distances traveled by all vehicles transporting goods from an origin i to a destination j.

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.

$KHCNpc_{ijt}$ are the kilometers per capita of high capacity network of the regions where the traffic of goods runs.

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

$Intra$ is a dummy variable with value 1 when the Autonomous Communities of origin and destination are the same, and 0 otherwise.

μ_{ij} it will pick up the individual effects defined as pairs of regions by origin and destination, λ_t stands for the unobservable time effect and ν_{it} is the random disturbance term.

The econometric representation of the gravity model is in Equation 2.1 in log-linear functional form and is used as highlighted in Oum (1989). This specification presents a number of advantages. The estimated coefficients are directly interpreted as elasticities. In addition, the log-linear function is able to model non-linear.

The β_1 coefficient indicates the degree of response of current flows to the flows of the previous year, whereas β_2 is associated with the flows of two years before. If the domestic trade of a country experiences inertia over time, the expected sign of these coefficients will be positive. On the other hand, the expected sign of the estimated β_3 coefficients in each model will be negative because, in accordance with theory, the distance variable acts as an impediment to the flow of goods. Finally, the remaining coefficients are expected to have positive signs: the amount of goods carried within the region itself should be higher, ceteris paribus, than flows to other regions; flows between adjacent regions are also expected to be greater than between distant communities; the GDP variable includes the size of communities in economic terms, so larger flows are expected among Autonomous Communities with greater economic activity; the number of kilometers of high capacity network per capita, which is a measure of density of high capacity roads, and can be understood in turn as a proxy of the road quality of the Autonomous Communities, is expected to have a positive influence on the flows among communities. This variable is constructed from the addition of the kilometers of high-capacity network corresponding to the region of origin, of destination and the regions with the shortest route between the capitals of the regions of origin and destination. This sum was then weighted by the total population of the regions involved in order to obtain a measure of density.

2.3   Database

The database used represents a balanced panel for the 15 peninsular Autonomous Communities[5] between 1999 and 2009. This study does not include the Balearic Islands, the Canary Islands and the Autonomous Cities of Ceuta and Melilla due to their geographical locations. Hence, we have a balanced panel with dimension N = 225 (all possible bilateral combinations of Autonomous Communities) and T= 11. Thus, the total number of observations is NT =2,475.

The flows of road transport of goods - the dependent variable - is measured by two alternative variables available in the Permanent Survey of Road Transport of Goods (PSRTG), issued by the Spanish Ministry of Development: these variables are flows and transport operations. Their main objective is to measure the level of activity in road freight transport. The PSRTG is aimed at Spanish transport companies and does not include international shipments entering Spain. Since 1999, this survey uses a homogeneous methodology in accordance to the European Union Regulation 1172/98. Since that date, the survey has been regularly updated on a yearly basis. These are the basis for the sample period of the current study. As may be observed in Table 2.1, the rest of the explanatory variables are provided by different statistical sources.

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

---

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

Table 2.1: Definition of variables

	Description	Source
Flows	Thousands tons	
Transport Operations	Number of trips	Ministry of Development
Dist	Average Km. traveled	
Population	Standing population	
KHCN	Km.	National Statistics Institute
GDP	Constant €	

Since the Permanent Survey of Road Transport of Goods also includes the ton-kilometer variable, it is possible to retrieve the mean travel distance of the goods by dividing the total ton-kilometers by the transported tons for each origin-destination pair. The distance variable measured in this way reports information on the routes chosen by the transport companies. The resulting variability, in contrast to the fixed character of the previously mentioned approach, will capture the cost-minimizing behavior of the companies. We firmly believe that measuring distance in this fashion is a better approximation to the cost of freight and provides a more accurate estimation of the intraregional flows included in the model.

Table 2.2 shows the descriptive statistics of the quantitative variables used in the estimated models. The dependent variables (transport flows and operations) contain few values equal to 0 throughout the whole sample (13 and 8 respectively). In order to avoid problems when taking logarithms in these variables, 0 values have been replaced by the value 1 since this is the minimum non-zero value among those taken by the independent variable. This is usually the most suitable statistical manner to deal with this sort of problems, aside from adding

the provincial data with the aim of getting positive values per Autonomous Community ((Burger *et al.*, 2009)[6]

Table 2.2: Descriptive statistics of the variables

	Mean	Std. Dev.	Min.	Max.
Flows	7,248.48	30,584.93	1	358.245
Transport Operations	993,657.90	4,498,450	1	50,973,200
Dist	481.93	293.91	11.47	4,834.07
Population	2,677,281	2,299.15	264,178	8,059,460
KHCN	815.55	594.36	130	2,631
GDP	43,229,025	40,864,631	3,949,932	151,444,000

2,475 Observations. Dummy variables Border and Intra not included

2.4 Estimation and results

The inclusion of the dynamic component of trade through the lagged dependent variable leads to endogeneity problems that complicate the estimation of panel models (Baltagi, 2008). As an example, we consider a basic dynamic panel model to illustrate these issues:

$$y_{it} \; = \; y_{it-1}^{\beta_0} x_{it}^{\beta} + (\alpha_i + \epsilon_{it}) \tag{2.2}$$

where

[6]If the proportion of observations with 0 values were significant, alternative estimation methods should be considered. The Poisson fixed effects estimator, proposed by (Westerlund & Wilhelmsson, 2011), could be applied here.

y_{it} is the dependent variable, x_{it} is a vector of current and past values of independent variables, α_i represents an unobservable time-invariant specific effect and $\epsilon_{it} \sim IID(0, \sigma^2)$ is a serially uncorrelated error term[7]. The econometric problem arises from the correlation between y_{it} and α_i. One possible solution is to apply a within transformation of the fixed effects estimator to eliminate the individual effect i. However, the correlation between the errors and the transformed regressors still exists, yielding an inconsistent estimator unless there is a long time period available (Nickell, 1981). This has been explored in a research paper on trade among European Monetary Union countries by Bun & Klaassen (2007). The database has a time horizon of 36 years, and the authors find that the Least Squares Dummy Variable (LSDV) estimator performs better than the GMM method in this case.

In order to overcome these difficulties, Anderson and Hsiao (1981) propose the two-stage least squares estimator (2SLS). In the first stage, differences are taken to eliminate the origin-destination fixed effect, α_i. However, by taking differences, Δy_{it} and $\Delta \epsilon_{it-1}$ are now correlated. This problem can be overcome by using as an instrument the second lag of the dependent variable, y_{it-2} given the assumption that there is no serial autocorrelation. This instrument fulfills the requirements that it be correlated with Δy_{it-1} and not correlated $\Delta \epsilon_{it-1}$ Nevertheless, this estimator is inconsistent when the number of periods T is finite and the number of cross-section observations, N, is large (Baltagi, 2008). This is usually a common feature in gravity models estimated with panel data, so it tends to be a common problem.

---

[7]We do not consider spatial dependency in transport between regions, so the spatial correlation is ignored if it exists.

As an alternative approach, GMM estimators can deal with the endogeneity of the explanatory variables. Arellano (1989) states that efficiency can be increased by GMM difference estimation using all available instruments, Δy_{it-k} with $k > 1$. The GMM method is based on the idea that given a set of instrumental variables correlated with the regressors but orthogonal to the errors, some moment conditions can be defined and solved using the true value of the estimated parameters. Fixed effects are eliminated by taking first differences and endogenous variables in levels lagged two or more periods are used as instruments to solve simultaneity issues (Arellano & Bond, 1991). However, the use of data in differences also involves removing all time-invariant variables which may be of interest.

The GMM model is one of the methodologies most used to estimate dynamic gravity models (Jung, 2009). However, as argued in Blundell & Bond (1998), when the data are highly persistent - as in the case of bilateral trade flows - this procedure can be improved through the estimation by System GMM (De Benedictis & Vicarelli, 2005). It has been shown that when the lagged values of the series approach a unit root, these instruments contain little information about the endogenous variables in differences. When applying the system-GMM equations in first levels, instead of transforming the regressors to eliminate the fixed effects, differences are taken of the instruments to make them orthogonal to the individual effects. Lagged first differences are used for equations in levels while lagged instrumental variables are used for equations in first differences (Roodman, 2009). The consistency of this estimator depends on no first-order autocorrelation in the errors and an array of truly exogenous instruments. System GMM is the chosen methodology. In Table 2.3, results and estimation tests are displayed.

Table 2.3: System GMM estimations

	Flows		Transport Operations	
	Coef.	t-stat	Coef.	t-stat
y_{it-1}	0.296***	5.29	0.179**	2.01
y_{it-2}	0.292***	6.41	0.135***	3.23
$Dist_{ijt}$	-0.741***	-4.53	-1.069***	-5.46
GDP_{it}	0.337***	4.25	0.539***	4.83
GDP_{jt}	0.392***	4.06	0.519***	4.65
$KHCNpc_{ijt}$	0.660***	3.19	0.542***	2.64
Intra	-0.046	-0.20	0.856***	2.67
Border	-0.656	-0.77	0.320**	2.60
Yr2002	-0.093	-1.41	-0.208**	-2.21
Yr2003	0.011	0.32	0.022	0.47
Yr2004	0.009	0.31	0.110***	2.75
Yr2005	-0.06	-1.56	0.289	0.61
Yr2006	-0.120**	-2.27	-0.029	-0.60
Yr2007	-0.082	-1.35	0.005	0.08
Yr2008	-0.273***	-4.23	-0.233***	-3.47
Yr2009	-0.335***	-7.30	-0.187***	-3.95
Number observations	2,475		2,475	
Number of instruments	178		178	
AR(1) in first diff. (p value)	-3.31 (0.001)		-2.53 (0.012)	
AR(2) in first diff. (p value)	-2.37 (0.018)		-1.08 (0.282)	
Hansen (p value)	184.66 (0.107)		178.17 (0.247)	
F statistic (16, 225)	8,448.88		18,576.95	

** Significant at 5%. *** Significant at 1%.

The estimated coefficients have the expected signs according to the theory of the gravity model. In addition, the coefficients of the principal variables of the model (GDP and distance) are significant at the 1% level. The positive and significant coefficients associated with the proxy of the quality of infrastructure, $KHCNpc_{ijt}$, indicates that the availability of high capacity roads directly affects

the transport of goods. Of particular note is that the lagged transport variables positively affect the present flows and transport operations, as indicated by the estimated coefficients of y_{it-1} and y_{it-2}. The size of these coefficients - around 0.30 in the case of the flows and 0.15 for transport operations - indicate that inertia is not very important in domestic road transport in Spain. However, the statistical significance of the estimated coefficients indicates that not taking the dynamic component into account would cause a model misspecification. In this case, the Hansen test for detecting over-identification does not reject the null hypothesis of the validity of the instruments. Autoregressive tests for AR (1) and AR (2) show the consistency of the GMM estimator and the inconsistency of estimators based on OLS[8].

Including lagged variables in the equation yields short term estimated elasticities. To recover the long-run elasticities it is necessary to divide each of the estimated values by $(1-\beta_0-\beta_1)$. The result of this transformation is shown in Table 2.4.

Table 2.4: Short term and Long run elasticities

	Short Term		Long Term	
	Flows	Transport Operations	Flows	Transport Operations
GDP_{it}	0.337	0.539	0.818	0.786
GDP_{jt}	0.392	0.519	0.951	0.757
$Dist_{ijt}$	-0.741	-1.069	-1.799	-1.558
$KHCNpc_{ijt}$	0.660	0.542	1.602	0.790

---

[8]Arellano & Bond (1991) propose the hypothesis test of no second-order serial correlation in perturbations of the equation in first differences: $E[\Delta\epsilon_{it}, \Delta\epsilon_{it-2}]$. This is a necessary condition for the validity of the instruments. The rejection of the null hypothesis in the AR (1) test indicates the inconsistency of the OLS estimator.

Among the results, we highlight the magnitude of the long-run elasticities associated with the distance variable. In accordance with the model - where it acts as a proxy for transport costs - and noting also the elasticities associated with high-capacity network, we emphasize the importance of quality infrastructure in order to assist the road transport of goods. The coefficients that accompany the GDP variables are inelastic but close to unity.

2.5 Application of the Model: Forecast and Estimation of CO_2 Emissions

The calculation of greenhouse gas emissions is one of the most interesting applications of the estimated models. In this section, we present calculations and predictions of CO_2 emissions resulting from domestic road transport of goods in Spain for the period 2001-2009. The calculation procedure relies on the exploitation of the data provided by the Permanent Survey on Road Transport of Goods and the high explanatory capacity of the estimated models. The analysis does not include other greenhouse gases because CO_2 emissions are proportional to fuel consumption whereas the contamination by other gases such as CH_4 and N_2O depends largely on the emissions control systems (of Transportation & Quality, 2006) of vehicles and such information is not available.

The different modes of transport in Spain dedicated to both passenger and goods transport consume 32.5% of total energy demand. Road transport accounts for just over 70% of the total gases emitted to the atmosphere. To our

knowledge, no studies as yet have computed disaggregated data for internal traffic of goods and we strongly believe it is important to know this information for policy decision-making. The following equation has been used to obtain the total amount of CO_2 emissions in the year t:

$$E_{CO_2}^t = \sum_{i=1}^{n} \sum_{j=1}^{n} (D_t \cdot Q \cdot Dist_{ijt} \cdot O_{ijt}) \qquad (2.3)$$

where

D_t are the liters of diesel per kilometer.

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

$Dist_{ijt}$ are the kilometers 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 liters), 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 amounted to 0.385 liters per kilometer. 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, 2010). 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.

Figure 2.1: Average age of road transport vehicles
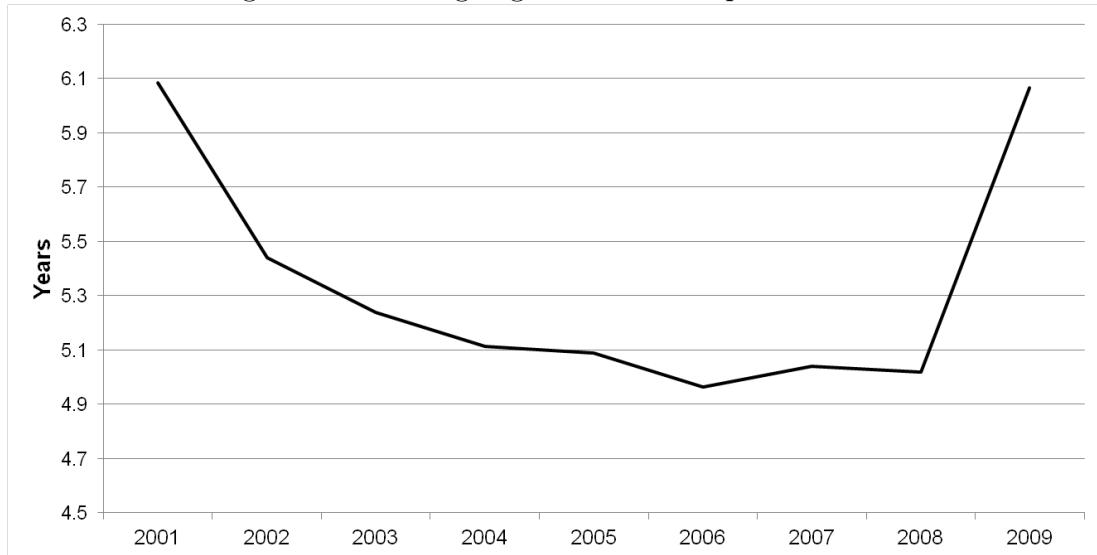
Figure 2.1 shows the evolution of the average age of heavy goods vehicles in Spain. The average age of vehicles declined to 5 years in 2006. After that, due to a slower pace in the renovation of older vehicles the average age began to rise back to 2001 levels.

Regarding the amount of CO_2 kilograms emitted to the atmosphere per liter of diesel, it accounted for 2.71 $kg./liter$; this figure is included in the inventory of Greenhouse gas emissions in Spain (Spanish Ministry of Environment & Areas, 2009).

The total amount of CO_2 gas emissions obtained using this methodology is illustrated in Figure 2.2. Calculations are performed ex-post from 2001 to 2009 using the observed values of the transport operations coming from the survey and, alternatively, predictions are computed using the forecasted values of transport operations from the model estimated in Section 4.

The increasing trend in total CO_2 emissions from domestic freight traffic by

26

Figure 2.2: Estimation of CO_2 emissions of road transport of goods
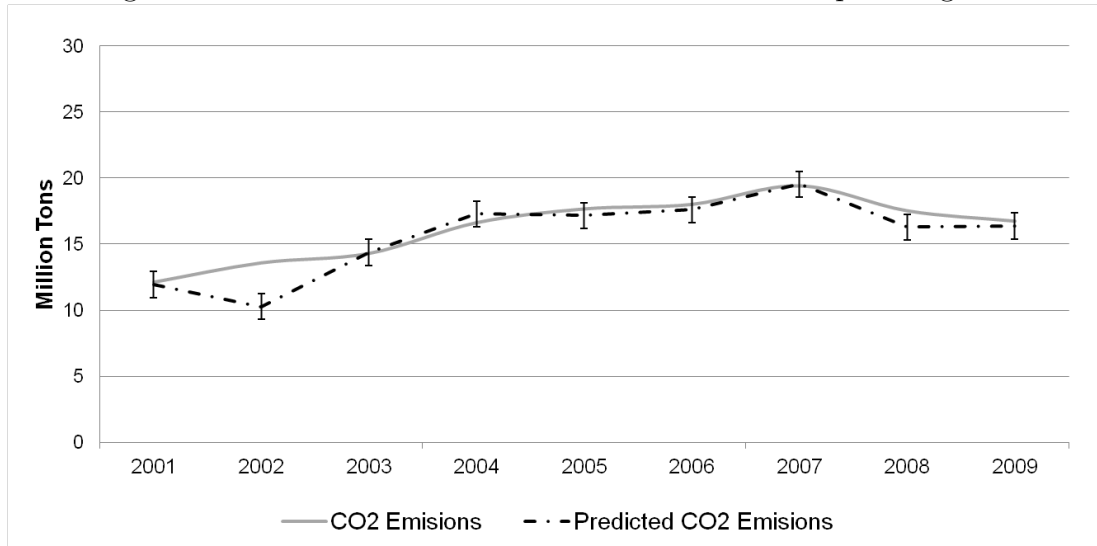
road reaches its peak in 2007 with nearly 20 million tons emitted into the atmosphere. From that moment there was a reduction in the amount of CO_2 due to the fall in the number of transport operations in recent years because of the economic slowdown. However, in 2009 and 2010 there seems to be a slight upward trend in emissions, partially caused by the lower rate of renewal of the fleet.

2.6   Conclusions

In this paper a dynamic panel model has been estimated to explain the flow of goods by road among the different Spanish autonomous regions and within each of them. Transport flows are measured by both the tons carried and the number of transport operations. The study period chosen is from 1999 to 2009 due to the availability of different explanatory variables.

The results obtained in the model estimated by system-GMM, show that the

theory of the gravity model is verified and that there is inertia in the internal traffic of goods in Spain. While it is true that the values that accompany the lagged variables are not large in magnitude, it seems that this is the most appropriate specification for this model. The persistence of freight can be justified by the sunk costs incurred by the companies to settle in different regions.

Other empirical results to note are the statistical significance of two variables which are measured in a way that have not been used so far. On the one hand, the distance variable is not a fixed measure between two points in the regions and is instead constructed by dividing the variable ton-kilometers by the total tons transported. Thus, the distance between two regions varies from year to year and the variable captures the route choice decisions made by transport operators. On the other hand, as this study attempts to estimate structural relationships that direct the flow of goods transport, the model includes a measure of the quality of infrastructure. Therefore, a variable that reflects the density of high-capacity road network in the route of the goods is introduced. According to our specification, the quality of infrastructure measured in this fashion has a positive and significant effect on the flows of goods by road transport.

The appropriateness of the model specification and goodness of fit to the data analyzed allows worthwhile predictions to be made. Specifically, we calculate and predict the tons of CO_2 emitted in that period by domestic transport of goods by road in Spain. In the case of polluting emissions, there is a reduction of the gases after 2007 provoked by the economic slowdown. However, predictions of CO_2 emitted into the atmosphere show a slight upward trend in 2009 and 2010, partly explained by a lower rate of renewal of older vehicles (which are the most polluting ones). Among the policy actions that might be taken we suggest

an intensification of policy incentives to replace older vehicles by new ones that are more energy-efficient and allow for a stronger control of greenhouse gases emissions.

Some of the main future research lines that could be followed are the use of the estimated demand for other modes of transport and the construction of a system of equations to study the interaction between different modes and estimate cross-elasticities between them. We also expect to be able to obtain a measure of the price of transport for each route, in which case we would be able to run simulations with different fiscal policy actions, such as taxes that internalize the environmental cost of transport, to measure the impact on road freight transport.

Chapter 3

Determinants of ground transport modal choice in long-distance trips in Spain

3.1 Introduction

Long-distance travel has rapidly increased in recent decades. Technological innovation, car ownership and economic growth are the major factors behind the rising demand for inter-city mobility. Improvements in highway, train and airport infrastructure have reduced travel costs and times while increasing safety. Schafer (1998) noted that travel time budgets have remained relatively constant, thus allowing people to travel further. Economic growth has also led to higher average disposable income; this favorable economic environment, along with less expensive car ownership, has contributed to the increase in long-distance business and personal trips.

Increased mobility implies economic, social and environmental consequences. According to Limtanakool *et al.* (2006), it permits a higher integration among regions, provides better accessibility to public services and social networks and extends the potential market for tourism activities. However, it can also have a negative environment effect, as longer trips involve increased energy consumption and greater emissions of pollutants. Furthermore, the investments required for enhancing long distance travel are substantial, implying high opportunity costs to the economy. In this sense, a deeper understanding of travel behavior in long-distance travel may ease transport policy challenges such as making mobility more sustainable and reducing negative externalities (Bhat, 1998). The demand for transportation services is the result of interactions between short- and long-term individual decisions. *Long term* travel planning typically involves making decisions concerning car ownership and residential or work locations. The final outcome of these decisions impact trip behavior, affecting short-term aspects such as modal choice, departure time and choice of route (Ben-Akiva & Bierlaire, 1999). To date, transportation and geographic researchers have typically studied the impact of these factors on daily and short-distance trips. Nonetheless, factors determining daily travel decisions and their interdependencies may have different impacts on less frequent events such as medium- and long-distance trips (Limtanakool *et al.*, 2006).

This paper contributes to the overall discourse by increasing the knowledge in the determinants of modal choice in medium- and long-distance trips in Spain. More specifically, our primary goal is determining the influence of socioeconomic, land use and trip attributes on the selection of principal modes of transport among three possible choices: private car, bus and train. For this purpose, we use a

database obtained from a survey carried out by the Statistical Office of Ministry of Public Works on the mobility patterns of Spanish residents. The data has a hierarchical structure, with travelers nested in their provinces of origin. This feature of the data permits us to incorporate unobservable variables through the application of multilevel analysis, which also enriches the existence knowledge in long-distance trips. In particular, we estimate a multilevel multinomial logit model with random intercepts to study the determinants of mode choice among the three alternatives.

In line with previous research, our findings indicate that long distance modal choices are influenced by a combination of traveler socioeconomic characteristics, trip attributes and geographical factors. A key result is that the geographical context where the traveler begins the trip affects the alternative mode of transportation utilities in different ways. In addition, our analysis provides evidence of the positive effect of trip duration, measured as overnight stays, on railway demand; this has not been previously documented.

We initially review the existing literature on long-distance travel behavior, with an emphasis on modal choice studies. In Section 3, we explain the database and present some descriptive results of the included variables. In Section 4, we explain the foundations of rational choice relying on random utility theory in a multilevel framework. Section 5 displays the results of the estimated microeconometric models. Finally, Section 6 contains a summary of the major results and conclusions.

3.2 Previous research

Modal choice is the result of a complex process that includes objective and subjective determinants stemming from different disciplines and interrelated to a greater or lesser extent. In a recent survey, modal choice was defined as "the decision process to choose between different transport alternatives, which is determined by a combination of individual socio-demographic factor and spatial characteristics, and influenced by socio-psychological factors" (De Witte *et al.*, 2013). These authors also provided a useful list of 26 determinants typically found in an empirical analysis of modal choice. In another survey, Buehler (2011) compared transport mode choice determinants in the case of Germany and the USA. Although both surveys were comprehensive examinations of the literature, they did not clearly distinguish between short- and long-distance studies.

Long-distance or inter-city trips are usually differentiated from short trips through the use of a distance threshold. Although there is no standard definition, trips are usually defined as long distance if they are longer than a threshold between 50 to 100 Km. (Axhausen *et al.*, 2003). Long distance trips involve more time and out-of-pocket cost, so the traveler facing the modal choice decision is in a different situation than an individual making a short-distance trip. In addition, modal availability and travel purposes are also different. While the principal motive of short-distance travel is commuting, long distance transit is dominated by pleasure and business pursuits. Therefore, long-distance trips are less frequent, making travelers less familiar with available transportation alternatives. Decisions on mode choice can be affected by similar variables in short and long-distance trips, but the impact of the same variables can be conditional

to the distance travelled. The purpose of this literature review is not to find these differences but to focus our attention on the determinants typically found in long-distance travel studies.

Sociodemographic factors play a significant role in transport mode decisions. Bhat (1997) applied an endogenous segmentation model to the estimation of inter-city travel mode choices in the Toronto-Montreal Corridor. The author found that women were more responsive than men to rail frequency improvements in Canadian inter-city travel. Limtanakool *et al.* (2006) estimated binary logit models that distinguished between private cars and trains, finding that women were more likely than men to use trains. Georggi & Pendyala (2001) and Mallett (1999) also found that women were slightly less car dependent in long-distance trips. The effect of the travelers age on modal choice was not as clear. Limtanakool *et al.* (2006) indicated that senior commuters were more likely to use private cars than middle aged and young travelers. In an analysis of elderly and low-income mobility, Georggi & Pendyala (2001) found that the elderly were more bus dependent. As for the impact of education on modal choice, only Limtanakool *et al.* (2006) included it as an explanatory variable for long-distance models. They found that highly educated commuters tended to use public transportation more often. It is important to note that age, education and occupation are related to income and car ownership. High income travelers have higher opportunity costs and values of time, which implies the selection of faster transportation modes such as airplanes because they are more sensitive to travel time improvements (Bhat, 1997). In another study by Mallett (2001), lower-income individuals were slightly more dependent on buses and other public transport modes.

Spatial configuration indicators such as population density, diversity of land

34

use and accessibility to transportation infrastructure are typically included in short-distance mobility studies (De Witte *et al.*, 2013). Few papers studying inter-city modal choice include land use factors. An exception is Limtanakool *et al.* (2006), which found that higher population densities and higher degrees of mixed land use around public transport stations make these modes more attractive in long-distance travel. In a recent descriptive study, Garmendia *et al.* (2011) found that travelers from cities less than 10 Km. away from a high-speed rail station more frequently chose trains than cars for their trips to metropolitan areas. However, it is possible that there may be a self-selection effect; individuals with a preference for public transport may move to areas with an abundance of these services (Buehler, 2011).

Journey characteristic indicators such as purpose, distance, frequency and travel time also impact the mode selection in medium- and long-distance travel. Moeckel *et al.* (2013) crafted an exhaustive description of long-distance mode choice studies focusing on trip attributes. Among the principal travel motives, existing literature usually distinguishes between commuting, business and leisure. The results in Limtanakool *et al.* (2006) indicated that private car use was very prominent for business trips, while commuters relied to a greater extent on trains. Georggi & Pendyala (2001) compiled a descriptive cross tabulation analysis of the 1995 American Travel Survey, showing that modal distribution changed across trip purposes. In a paper that attempted to assess the impact of high-speed train investments on the mobility of Spanish residents, Martín & Nombela (2007) found that public transport modes were more attractive in commuting to work. A drawback of this study was that they did not support their conclusion through a multivariate analysis. Faster travel modes are usually preferred in the case of

longer distances. Martín & Nombela (2007) found distance to be a positive effect in selecting trains and a negative for buses. Koppelman & Sethi (2005) estimated models with different methodologies and concluded that distance discouraged travel by automobile. The empirical analysis carried out by Bel (1997) indicated the importance of travel times in modeling modal choice. Longer railway travel times negatively affect rail demand, while travel times by road have a positive relationship on rail demand.

Overall, the existing empirical evidence confirms that socioeconomic indicators, trip characteristics and land use factors affect mode choice decisions. Another source of variation in long-distance travel behavior arises from traveler location. Bricka (2001) showed that traveler compositions are different among states in terms of household income and race. While trip purposes seem to be quite stable for the analyzed states, dissimilar mode choices are explained by availability of modes and the urban form in the location of origin. The results from Lapparent *et al.* (2013) suggested disparities between European countries when heterogeneous preferences were taken into account. Spatial features seem to play an important role in modal decisions, not only through the impact of land use configurations but also through the spatial heterogeneity in traveler composition and preferences.

Many statistical methods assume that relationships are constant over the space of the sample, i.e., all coefficients are forced to be identical (or stationary) for all individuals, locations or zones. This hypothesis is likely to be violated in the case of mode choice due to the influence of the geographical and socioeconomic context, as mentioned above. Páez (2006) offered a review of the different alternatives proposed in existing literature to overcome this problem: market

segmentation, the introduction of dummy variables in the model, the Casetti's expansion method and multilevel models. After evaluating the database information in Section 4.4 and the theoretical implications discussed in Section 4.2, we selected a multilevel multinomial model.

3.3   Database description

In this section, we discuss the data employed in the estimation of the model explained in section 3.4. The primary source of information is a mobility survey (Movilia 2007) carried out by the Statistical Office of the Ministry of Public Works [1]. The objective of the survey was to study the basic characteristics of Spanish resident travel to better understand Spanish population mobility habits. The sample unit only considered the movements of Spanish residents; it did not include trips by tourists and non-resident immigrants that might have affected certain locations. In this survey, long-distance trips were defined as those longer than 50 Km.. They collected the microdata information on long-distance trips longer than 50 Km. through quarterly telephone surveys between February 2007 and January 2008. The dataset revealed preferences about the individual characteristics of the trip maker, land use factors and attributes of the trip. For each trip, we know the gender of the interviewed person, their age group, employment status and educational level. The trip characteristics included the province of origin, the trip purpose, the selected mode of transportation, the distance and its duration. Despite the robust information provided by the survey, we did not have information on individual income, household characteristics such as num-

---

[1]Methodology and definitions applied in this survey comply with the requirements set by the European Commission for long-distance travel surveys.

ber of children, car ownership, disposable income and time or cost of the trip. Moreover, the geographical information was insufficient for computing accessibility to public transportation infrastructure. While there was information on the province where the trip began, there was no detailed information on the specific town of origin. As for trip destinations, information was available on whether the destination province was adjacent or of the same region as the province of origin.

We overcame these caveats by using the available information and including unmeasured and unobservable characteristics. In our analysis of the variability of ground mode choice, we not only controlled for the effect of the individual and trip characteristics but also for the spatial context where they belong. This feature of the multilevel model (that will be explained in Section 3.4) allowed us to include random intercepts containing specific information on the trips province of origin. This information could be related to specific area characteristics that are not specifically included as explanatory variables but may be relevant for modal choice. The individuals in the same geographical unit are likely to be similar in some ways due to these unobserved characteristics (Hong *et al.*, 2013). For instance, the spatial environment where the traveler makes the choice is bounded by the borders of the trip province of origin, which is linked to the access to transportation infrastructure and the public services provides within this area.

The empirical model to be estimated requires some database modifications [2]. We constructed a cost variable by multiplying the distance reported in the survey by the average price per kilometer for each mean of transportation [3]. The

---

[2] We have disregarded observations on plane trips and other means of transport because our objective is the study of ground transportation modes and their interactions.

[3] The average prices used in these computations are 0.19 /Km., 0.0877 /Km. and 0.09202 /Km. for car, bus and train, respectively. These prices were obtained from different sources. In the case of cars, the source is the law that sets the trip compensation for public workers in case

motivation to use a cost (rather than distance) variable can be found in Whalen *et al.* (2013), where the authors criticized the use of a distance variable because it was an individual-specific attribute and equivalent all modes. In contrast, the cost variable is treated as a mode-specific variable and (according to these authors) superior to distance because it is more easily interpreted and more accurately represents the characteristics of the trip.

After the data were prepared, the analyzed database includes 19,514 observations as displayed in Table 3.1. Each observation represents a trip and collected information on the individual characteristics of the traveler, variables describing the trip origin and specific attributes such as purpose, duration and distance to build the cost variable. Personal information showed a similar proportion of men and women and that the age of the respondents in almost 69% of the sample observations was below 50 years. Two thirds of the reported cases had completed secondary education. As previously explained, there was no information on personal income. We constructed a proxy for a personal income variable relying on educational level and labor information, the latter being primarily composed of employed workers (66%). The low-income group was composed of the unemployed, housewives, retirees, students and unschooled children and employed people with pre-primary education. The medium income group was comprised of employed people with primary and secondary educations. The high-income group consisted of workers holding a university degree or vocational training.

As explained in Section 3.2, among the common explaining variables of mode choice found in existing literature were geographical variables with information

---

of travel. The average bus price comes from a report of the National Competence Commission in 2006, while the train price per kilometer was obtained from the Railway Yearly Report by the Spanish Railway Foundation

about land use, population density and accessibility to transportation infrastructure. Available information in this survey included the province of origin, a categorical variable for the city size and a variable to determine if the origin was located in a metropolitan area. More than two thirds of the trips began in a non-metropolitan area. The size of the city of origin, measured in population, was less than 50,000 habitants in 54% of the observations, while trips beginning in a city larger than 500,000 residents accounted for 7.1%. We expected that larger cities located in metropolitan areas would have a higher population density. Higher densities are related to improved public transportation in the sense of higher frequencies of public transport and better connections.

Finally, two variables containing information about attributes of the travel were available. The distance travelled during the trip indicated that almost half of the observations were trips shorter than 100 kilometers. The duration of the trip was measured as the number of overnight stays. In half of the sample observations, the traveler did not stay overnight; in 40% of the cases, overnight stays ranged from 1 to 6 nights.

Table 3.1: Descriptive Statistics

	Frequency	Percentage	Cum. Perc.
Gender			
Female	8,858	45.39	45.39
Male	10,656	54.61	100.00
Age group			
15 to 29	4,960	25.42	25.42

Continued on next page

Table 3.1 – *Continued from previous page*

	Frequency	*Percentage*	*Cum. Perc.*
30 to 39	4,042	20.71	46.13
40 to 49	4,428	22.69	68.82
50 to 64	4,437	22.74	91.56
65 or more	1,647	8.44	100.00
Labour situation groups			
Employed	12,997	66.60	66.60
Housewife	1,977	10.13	76.73
Unemployed	635	3.25	79.99
Retired	2,116	10.84	90.83
Students	1,789	9.17	100.00
Educational level			
Pre-primary education	870	4.46	4.46
Primary education	5,223	26.77	31.22
Secondary education	6,910	35.41	66.63
Vocational training FP	1,655	8.48	75.12
University degree	4,856	24.88	100.00
Income level			
Low Income	6,764	34.66	34.66
Medium Income	7,410	37.97	72.64
High Income	5,340	27.36	100.00
Municipality size			
Less than 10000	6,081	31.16	31.16

Table 3.1 – *Continued from previous page*

	Frequency	Percentage	Cum. Perc.
10000 to 50000	4,590	23.52	54.68
50000 to 500000	7,445	38.15	92.84
500000 and more	1,398	7.16	100.00
Type of area			
Not metropolitan	13,470	69.03	69.03
Metropolitan area	6,044	30.97	100.00
Purpose			
Pleasure	14,594	74.79	74.79
Business	2,706	13.87	88.65
Second residency	2,214	11.35	100.00
Distance			
50-99 Km.	9,244	47.37	47.37
100-249 Km.	6,887	35.29	82.66
250-499 Km.	2,650	13.58	96.24
500 Km. and more	733	3.76	100.00
Nights			
0	9,771	50.07	50.07
1 to 6	8,190	41.97	92.04
7 to 14	1,086	5.57	97.61
15 and more	467	2.39	100.00

In table 3.2, we conduct a descriptive cross tabulation analysis of mode choice by trip purpose with some of the variables included in the survey. Trip purposes are composed of pleasure, business and second residence. Pleasure includes holiday trips, visits to relatives and leisure trips. Business comprises those trips made for professional reasons. Trips to a second residence are in a different category. Although different motives such as holidays or leisure might be motivating travel to a second residence, these are folded into this category. Table 3.2 shows the predominant role of private car usage, with a minimum modal share of almost 84% in the case of pleasure trips, 89% of business trips and a similar share in those heading to a second residence. The remaining two modes of transport were chosen differently, depending on the trip motivation. While trains were hardly used for second residence trips, they were used slightly more than buses for business trips. Bus shares for pleasure purposes were almost twice the share for business trips and slightly larger than the shares in case of visiting a second residence.

A descriptive analysis of sociodemographic variables provides a useful insight in mode choice. Male travelers appear to be more dependent than females on private car usage, as noted in previous studies. The largest difference in gender mode choice appears in the case of pleasure trips, where 87% of male travelers select cars while only 79% of females do. The effect of age on mode choice is a little more difficult to distinguish through descriptive analysis. In general, car usage increases until middle age and then decreases in the senior years. Buses and trains reach their maximum market shares in people aged over 65 for pleasure and business purposes, respectively, although the modal split is totally different. In trips motivated by pleasure, senior respondents rely heavily on buses, while in business trips the final choice is the train. As for income, car usage is the

highest in medium income individuals, closely followed by high-income travelers. Descriptive data on bus usage conditional to income appear to support the idea of a bus as an inferior good; for all trip purposes, the market share decreases when income rises.

Another expected fact found in the data is that public transport was more intensively used in larger cities. Particularly in the case of rail, there was a strong positive correlation between city size and rail demand. This was likely caused by improved access to principal railway stations with connections to a wider range of destinations and more frequent services. The bus mode shares were not as closely related to city size as train demand. Travelers with origins in a metropolitan area also depended more on railway transport than those living in non-metropolitan zones. Private car usage did not differ much between metropolitan and non-metropolitan cities. Public buses were more important for non-metropolitan travelers. Trip characteristics such as distance and overnight stays also appear to have had an impact on primary transport mode choice. While car usage decreased with distance travelled, bus and train importance grew in longer distance trips. Additional nights spent travelling hindered car usage and favored the use of trains, particularly for pleasure purposes, with no clear pattern emerging for the other purposes.

Exploratory analysis using descriptive statistics is a helpful tool for identifying the stylized facts and characteristics of the data. This information provides important insights in the description of mode choice, but the estimation of a multivariate model allows us to assess the impacts of the different variables on probabilities of choosing a particular mode of transport.

Table 3.2: Selected mode of transport by purpose of the trip

| | Pleasure | | | Business | | | Second residency | | |
	Car	Bus	Train	Car	Bus	Train	Car	Bus	Train
Gender									
Female	79.22	14.12	6.66	83.09	9.43	7.48	85.60	9.75	4.65
Male	87.36	8.48	4.17	91.01	3.97	5.02	91.72	5.67	2.60
Age group									
15 to 29	78.63	14.07	7.29	85.82	8.08	6.10	77.53	14.90	7.58
30 to 39	92.90	3.88	3.22	90.03	3.79	6.18	94.93	2.90	2.17
40 to 49	90.75	5.15	4.09	90.71	4.04	5.25	93.60	3.93	2.48
50 to 64	82.66	12.03	5.31	91.67	4.96	3.37	90.94	5.94	3.12
65 or more	61.29	31.04	7.67	61.29	9.68	29.03	79.35	17.00	3.64
Income									
Low Income	71.35	21.02	7.63	82.00	9.00	9.00	81.76	13.13	5.12
Medium Income	91.36	5.10	3.54	91.19	5.29	3.52	93.00	5.18	1.83
High Income	90.56	4.86	4.58	88.11	4.46	7.43	93.47	3.04	3.50

Continued on next page

Table 3.2 – *Continued from previous page*

Municipality size									
Less than 10000	84.10	12.28	3.62	92.12	4.82	3.06	88.02	9.58	2.40
10000 to 50000	84.08	11.01	4.92	91.59	4.59	3.82	87.96	8.10	3.94
50000 to 500000	82.84	10.56	6.60	86.34	6.11	7.55	90.48	6.35	3.17
More than 500000	80.58	10.52	8.91	82.58	3.87	13.55	84.57	8.68	6.75
Metropolitan area or not									
Not metropolitan	83.21	12.11	4.68	90.11	5.64	4.25	88.52	8.44	3.05
Metropolitan area	83.84	9.14	7.02	86.87	4.11	9.02	88.65	6.85	4.50
Distance									
50-99 Km.	87.39	9.03	3.58	91.01	5.18	3.81	89.73	7.60	2.66
100-249 Km.	83.53	11.51	4.96	91.35	3.73	4.93	89.09	7.43	3.48
250-499 Km.	74.75	14.92	10.33	76.35	8.45	15.20	85.55	8.37	6.08
500 Km. and more	68.40	20.03	11.56	75.93	11.11	12.96	75.93	11.11	12.96
Nights									
0	86.48	10.18	3.34	91.15	4.72	4.13	89.74	7.69	2.56

Table 3.2 – *Continued from previous page*

1 to 6	81.44	11.49	7.07	79.96	7.59	12.45	89.52	6.96	3.51
7 to 14	73.57	17.78	8.65	95.45	4.55	0.00	81.82	11.48	6.70
15 and more	76.73	12.36	10.91	85.71	0.00	14.29	86.49	10.27	3.24
No. of Observations	12,171	1,638	785	2,414	141	151	1,961	172	81
Total Percentage by purpose	83.40	11.22	5.38	89.21	5.21	5.58	88.57	7.77	3.66

3.4 Utility framework and multilevel analysis

The mainstream approach to study modal choice assumes that travelers make rational decisions by selecting the alternative that maximizes their utility. From the optics of a rational choice perspective, subject n would choose the alternative with the highest utility. In our exercise, travel mode m is chosen by individual n if the utility of this alternative is greater than the utility of any other transportation mode t:

$$U_n^{(m)} > U_n^{(t)} \text{ for all } m \neq t \tag{3.1}$$

The workhorse tool for travel behavior analysis relies on the random utility framework. In these models, the decision rule is deterministic but the utility function includes a random component. The deterministic component includes information on the transport mode attributes, socioeconomic characteristics of the traveler, land use factors and other variables as in Section 3.2. The error term is included because the analyst assumes that there is incomplete information on the selection process faced by the individual (Manski, 1977).

In a multilevel framework, the utility $U_{nj}^{(m)}$ of an alternative m for individual n nested in cluster j is assumed to consist of a deterministic part $V_{nj}^{(m)}$, and a random component called the error term $\epsilon_{nj}^{(m)}$ (Grilli & Rampichini, 2007; Skrondal & Rabe-Hesketh, 2003) as follows:

$$U_{nj}^{(m)} = V_{nj}^{(m)} + \epsilon_{nj}^{(m)} \tag{3.2}$$

where $m = 1, 2, ..., M$ denotes the response category (mode of transport), $j = 1, 2, ..., J$ denotes the cluster (province of origin) and $n = 1, 2, ..., n_j$ denotes the traveler of the j-th province of origin. The deterministic part $V_{nj}^{(m)}$ represents the fixed part of the utility and is linearly related to the linear predictors of the model.

$$V_{nj}^{(m)} = \alpha^{(m)} + \beta^{(m)} X_{nj} + \beta X_{nj}^{(m)} + \zeta_j^{(m)} \tag{3.3}$$

where $\alpha^{(m)}$ is a fixed alternative-specific intercept, X_{nj} is a set of explanatory covariates that vary over travelers and $\beta^{(m)}$ is the set of associated coefficients to be estimated. Alternative specific covariate $X_{nj}^{(m)}$ is the attribute that varies between response categories m and travelers n nested in j and has a coefficient β that does not vary over alternatives m. The single level multinomial logit model would be solely composed of these components. The multilevel version of the model, where travelers are nested in the province of origin, also includes random alternative-specific intercepts $\zeta_j^{(m)}$ to account for unobserved heterogeneity at the province level. This setting allows for relaxing the multinomial logit assumption of independence of the irrelevant alternatives (IIA), which might be inappropriate

in some choice situations, as discussed by Hausman & McFadden (1984)[4]. In the case of the multilevel multinomial logit model, the error terms $\epsilon_{nj}^{(m)}$ of the utility functions have Gumbel distributions and are independent over transportations modes, travellers and provinces (Rabe-Hesketh & Skrondal, 2012).

Discrete choice models estimate the probability that an individual will select a mode of transport from a given set of alternatives, based on the attributes of the alternatives and on his preferences (Ben-Akiva & Bierlaire, 1999). The number of model equations equals the number of mode choice alternatives (three in our analysis), while the utility maximization rule specified in Equation 3.1 helps in computing the probability of choosing an alternative, e.g., Alternative 2:

$$
\begin{aligned}
Pr(Y_{nj} = 2 \mid X_{nj}, X_{nj}^{(m)}, \zeta_j^{(m)}) &= Pr(U_{nj}^{(2)} > U_{nj}^{(1)}, U_{nj}^{(2)} > U_{nj}^{(3)}) \\
&= Pr(U_{nj}^{(2)} - U_{nj}^{(1)} > 0, U_{nj}^{(2)} - U_{nj}^{(3)} > 0) \qquad (3.5) \\
&= Pr(\epsilon_{nj}^{(2)} - \epsilon_{nj}^{(1)} > V_{nj}^{(2)} - V_{nj}^{(1)}, \epsilon_{nj}^{(2)} - \epsilon_{nj}^{(3)} > V_{nj}^{(2)} - V_{nj}^{(3)})
\end{aligned}
$$

where

$$
V_{nj}^{(2)} - V_{nj}^{(1)} = \alpha^{(2)} - \alpha^{(1)} + \zeta_j^{(2)} - \zeta_j^{(1)} + (\beta^{(2)} - \beta^{(1)})X_{nj} + \beta(X_{nj}^{(2)} - X_{nj}^{(1)}) \qquad (3.6)
$$

---

[4]Rabe-Hesketh & Skrondal (2012) showed the distribution of the vector of random intercepts and derived the correlations between utility differences where ψ denoted the covariances between them

$$
\begin{bmatrix} \zeta_j^{(2)} \\ \zeta_j^{(3)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_2 & \\ \psi_{32} & \psi_3 \end{bmatrix} \right) \qquad (3.4)
$$

and

$$V_{nj}^{(2)} - V_{nj}^{(3)} \quad = \quad \alpha^{(2)} - \alpha^{(3)} + \zeta_j^{(2)} - \zeta_j^{(3)} + (\beta^{(2)} - \beta^{(3)})X_{nj} + \beta(X_{nj}^{(2)} - X_{nj}^{(3)}) \quad (3.7)$$

Analogously similar expressions can be obtained for alternative *1* and *3*. To obtain an identifiable model, we must select an alternative as a base category whose fixed and random parameters are set to 0. Utilities and their differences are unobservable, but we do have information on the mode choice Y_{nj}^*. It can be shown that the resulting choice probability, when the base category is alternative *1*, is given by the multinomial logit model in a Generalized Linear Model formulation McFadden (1974).

$$Pr(Y_{nj} = m \mid X_{nj}, X_{nj}^{(m)}, \zeta_j^{(m)}) \quad = \quad \frac{exp[V_{nj}^{(m)}]}{1 + \sum_{l=2}^M exp[V_{nj}^{(l)}]} \qquad (3.8)$$

where the denominator is the sum of the numerators of the probabilities of the three alternatives, guaranteeing that the probabilities sum to one. The model estimates presented in Section 3.5 were obtained through the Stata program gllamm [5].

In this context, a multilevel model has a number of advantages over a traditional single-level multinomial model. Bhat (2000) proposed a multilevel analysis as a useful tool for incorporating a spatial context where individuals must make decisions while allowing simultaneous consideration of the spatial heterogeneity between higher level units (provinces). This spatial heterogeneity can be ex-

---

[5]See Rabe-Hesketh *et al.* (2005) and Skrondal & Rabe-Hesketh (2003)

plained through different reasons. First, the existence of spatial autocorrelation is typical in any regional analysis Tobler (1970). In this case, individuals belonging to the same higher level units exhibited similar behavior (modal choice), but there is no complete individual information to explain this pattern. Second, it is possible to denote important differences in terms of structural behavior and relationships between higher level units. Both phenomena must be included in the modeling issue to avoid estimation and testing errors and parameter instability. Bhat (2000) and Jones & Duncan (1996) claimed to have adequately differentiated the heterogeneity among higher level units and individual heterogeneity, which may be related to socio-psychological factors. All of these issues were satisfactorily addressed through the multilevel framework.

3.5   Model Results

In this section, we present the empirical results obtained from applying the multilevel multinomial model to the mode choice sample [6]. We also estimate a standard multinomial model as a benchmark. Table 3.3 presents the results of these two models. The baseline mode of transport is the private car, and the results show odds ratios. Odds ratios, also referred to as relative risk ratios, indicate the ratio of the probability of choosing one outcome category (train or bus) over the probability of choosing the baseline category (private car) [7]. The results were derived by Maximum Likelihood Estimation using gllamm program in Stata.

---

[6]Similar results were obtained using a constrained sample were trips shorter than 100 Km. were disregarded

[7]Odds ratios are the ratio of probabilities of events and take the form: $\frac{Pr(Y_{nj}=m|X_{nj},X_{nj}^{(m)},\zeta_j^{(m)})}{Pr(Y_{nj}=1|X_{nj},X_{nj}^{(m)},\zeta_j^{(m)})}$.

The effects of all explanatory variables are very similar in the standard multi-nomial model and in its multilevel counterpart, although significant testing of the multilevel model based on the likelihood ratio test, Bayesian information criterion (BIC) and Akaike information criterion (AIC) shows that the model has significant spatial heterogeneity; therefore, this model is preferred to the conventional regression model. As expected, socio-demographic variables are important explaining the mode choice outcomes. Being male, *ceteris paribus*, decreases the odds of selecting a bus over a car by 40%, and the odds of taking the train compared with driving are reduced by an estimated 36%. According to this result, men are more car dependent than women in long-distance trips. Limtanakool *et al.* (2006) appointed different factors in explaining these gender differences found in the literature such as inequality in monetary rewards and different household task allocations. Categorical variables referred to age indicated that elderly trip makers were more likely to choose a bus over a car. The relative risk ratio associated with the age category 30-39 years implies that their odds of choosing a bus over a car relative to the base age of 15-29 years is multiplied by a factor of 0.332 (considerably less than 1), which sharply reduces their chances of selecting a bus over a car. This reduction in the probability lessens with older travelers, as shown by the estimated odds ratios for the middle aged, which are still below 1 but larger than 0.332. In the case of travelers between 40-49 years, the reduction would be 60%, while travelers aged between 50 to 64 years would reduce the odds to 30%. When switching from the base age category to the one including 65 years or older, travelers would increase their chances of selecting a bus over a car by 64%. This result concurs with the findings in Georggi & Pendyala (2001), as shown in Section 3.2.

A similar pattern can be observed in the case of studying train and car mode choice relationships where the factor that multiplies the odds of choosing a train over a car is below 1 but increases in the subsequent categories of older people. It is important to note that the odds ratio associated with the category of 65 years or more reaches a value that is not significantly different from 1, meaning that the probability of train selection is similar in the youngest and oldest participants. A likely explanation of these results relates to the impact of age on public transportation demand; students and the elderly have reduced access to car ownership and enjoy similar discounts in travel fares in both train and bus services, thus favoring the use of these modes.

In addition, income variables also show expected effects. Switching from the low-income group to the medium- or high-income groups decreases the odds of the traveler choosing a bus over a car. A higher disposable income also discourages the use of trains, though the decrease of the odds when switching from low to high income is smaller than in the case of buses. The train mode of transport also includes high-speed trains that can reduce travel times for some destinations and become more attractive for higher income travelers.

Land use variables included in the models consist of the size of the city of origin and an indicator of whether this location is a metropolitan area. Travelers beginning their inter-city trips in non-metropolitan areas are more likely to select a bus over a car than those users in metropolitan locations. The model results also show that, when all else is equal, the impact of the size of the city is insignificant in determining the odds of bus vs. car usage. The influence of land use indicators is different when confronting train and private car usage. In the multilevel version of the model, the metropolitan characteristic of a city of origin has no effect on

the mode choice decision. However, city size has a significant impact. The odds of choosing a train over a car when departing from a large city are almost twice the odds of a city of origin of less than 10,000 habitants. Larger cities are expected to have greater accessibility to high-speed rail and other important rail stations. Travelers from these cities enjoy a better supply of public transport services, including more frequent departures and a larger variety of destinations.

Table 3.3: Multinomial logit model estimates

	One level		Two levels	
	Odds Ratio	z- Statistic	Odds Ratio	z-Statistic
Fixed parameters				
Cost	0.978***	(-14.66)	0.978***	(-14.19)
Bus				
Gender: Male	0.596***	(-9.72)	0.594***	(-9.70)
Age: 30-39 years	0.333***	(-11.40)	0.332***	(-11.37)
Age: 40-49 years	0.405***	(-10.58)	0.406***	(-10.48)
Age: 50-64 years	0.700***	(-5.26)	0.705***	(-5.10)
Age: 65 years or more	1.644***	(6.53)	1.642***	(6.41)
Income: Medium	0.358***	(-14.67)	0.363***	(-14.36)
Income: High	0.316***	(-14.26)	0.316***	(-14.13)
Metropolitan Area: Yes	0.711***	(-4.94)	0.750***	(-3.56)
City size: 10001-50000	0.930	(-1.06)	0.963	(-0.53)
City size: 50001-500000	0.957	(-0.68)	0.940	(-0.91)
City size: 500000 or more	1.262	(1.86)	1.253	(1.63)
Purpose: Business	0.994	(-0.06)	1.010	(0.10)
Purpose: Secondary residence	0.674***	(-4.36)	0.666***	(-4.44)

Table 3.3 – *Continued from previous page*

Overnight stays: 1 to 6	0.987	(-0.23)	0.982	(-0.31)
Overnight stays: 7 to 14	0.948	(-0.51)	0.940	(-0.58)
Overnight stays: 15 and more	0.557***	(-3.51)	0.554***	(-3.51)
Intercept	0.296***	(-16.95)	0.287***	(-15.12)

Train

Gender: Male	0.643***	(-6.36)	0.640***	(-6.28)
Age: 30-39 years	0.530***	(-6.04)	0.546***	(-5.64)
Age: 40-49 years	0.608***	(-5.08)	0.658***	(-4.17)
Age: 50-64 years	0.648***	(-4.71)	0.663***	(-4.33)
Age: 65 years or more	1.053	(0.44)	1.060	(0.48)
Income: Medium	0.465***	(-8.35)	0.441***	(-8.74)
Income: High	0.577***	(-6.03)	0.557***	(-6.23)
Metropolitan Area: Yes	1.155*	(1.79)	1.040	(0.38)
City size: 10001-50000	1.251*	(2.20)	1.193	(1.63)
City size: 50001-500000	1.620***	(5.22)	1.683***	(5.38)
City size: 500000 or more	2.216***	(5.60)	2.059***	(4.39)
Purpose: Business	1.897***	(6.22)	1.979***	(6.48)
Purpose: Secondary residence	0.518***	(-5.33)	0.491***	(-5.66)
Overnight stays: 1 to 6	1.717***	(6.93)	1.820***	(7.51)
Overnight stays: 7 to 14	1.571***	(3.31)	1.641***	(3.54)
Overnight stays: 15 and more	1.394*	(1.71)	1.459*	(1.91)
Intercept	0.052***	(-27.35)	0.040***	(-23.17)

Random intercepts

		Est.	S.E.
ψ_{Bus}		0.091	(0.026)

Table 3.3 – *Continued from previous page*

ψ_{Train}		0.477	(0.082)
$\psi_{BusTrain}$		-0.126	(0.035)
No. Of observations	19,514	19,514	
Loglikelihood at convergence	-9,127.70	-8,933.33	
LR test			
Chi^2		388.7	
p value		0.000	
$ICC_{bus}(\%)$		2.755	
$ICC_{train}(\%)$		14.482	
BIC	18,639.6	18,283.8	
AIC	18,325.4	17,942.7	

* Significant at 10%. ** Significant at 5%. *** Significant at 1%

As for trip attributes, trip purpose and duration have a strong relationship with mode choice. If the purpose of the trip is visiting a second residence, a private car is preferred to both bus and train. For business travel, the odds of choosing a train over a car are almost twice that of leisure purposes. The effect of trip duration is quite different in the case of trains and buses. Overnight stays have no effect when the duration of the trip is below 14 nights, but in trips longer than 15 nights, the probability of choosing a bus over a car is considerably reduced. In the case of trains, increased nights spent away from home increase the odds of choosing a train over a car, although the effect seems to vanish in longer duration stays.

The only transport mode specific covariate is the cost variable; it has a unique estimated coefficient that does not vary over the alternatives. The odds ratio value for this variable is 0.978, which is below one and corresponds with a -0.02 estimated coefficient. A negative coefficient means that if the cost increases for one category, then the demand for that category decreases and increases for the other categories. The impact of variation in costs can be studied in more detail by computing the impacts on mode market shares caused by potential scenarios.

In this sense, we examine the results of four policy measures using the estimated model. The first policy is a 50 % increase in car usage cost that might be related to the evolution of gas prices or caused by the introduction of a congestion pricing measure. The second and the third are, respectively, a 25% decrease in public bus fares and railways costs. Finally, a fourth scenario includes a combination of the other three scenarios: a 50% increase in private car use costs jointly with 25% decreases in public transport fares. The effect of each policy measure is assessed by modifying the costs magnitudes to reflect a change, computing predicted outcome probabilities of mode choice using the estimated model, calculating predicted aggregate market shares of each mode and finally obtaining a percentage change from the baseline estimates.

Table 3.4 displays the results of the different scenarios. As expected, the model shows a decrease in private car market share in Scenario 1 when the cost associated with using the car rises. The bus market shares are more sensitive than railway shares, indicating a higher degree of substitutability between car and public bus. Scenario 2 and 3 show the effect of lower fares in public transport modes. In both situations, the market share of private cars decreases in a percentage close to 0.30%. The impact of reducing travel prices is larger in the

case of railway than in the case of public buses, which seem to be have a very low sensitivity to changes in their own fares compared to changes in car usage costs. Finally, the last scenario shows the effect of a simultaneous change in the costs of the three modes, which would induce a larger change in the modal split favouring the bus demand in a larger extent.

Table 3.4: Results of policy evaluation

| Scenario | Policy | Change in market shares | | |
		Car	Bus	Train
1	50% Increase car usage cost	-2.39%	+14.12%	+9.79%
2	25% decrease bus fare	-0.35%	+2.92%	-0.18%
3	25% decrease railway fare	-0.29%	-0.08%	+4.61%
4	Scenarios 1-3 simultaneously	-2.88%	+17.88%	+10.22%

Along with studying the effects of different independent variables on mode choice, multilevel analysis also allows researchers to analyze the effect of spatial heterogeneity. The estimated varying intercept can consider the correlation among people living in the same province. The correlation between individuals in the same area can be computed in terms of an intraclass correlation coefficient (ICC). This coefficient determines the proportion of variability that is accounted for by differences among areas [8]. The calculated ICC for buses (2.7%) indicates that there is a small correlation of travelers that choose a bus over a car. The associated ICC for trains, however, is significantly larger (14.5%). Relatively small ICCs indicate that individual factors included in the model can explain most

---

[8]See Grilli & Rampichini (2007) Rabe-Hesketh & Skrondal (2012) and Mercado & Páez (2009)

variance within a group. In the case of trains, differences between provinces seem to be more important in explaining the variation in train choice.

Spatial heterogeneity can be caused by very different factors such as non-homogeneous preferences in modes of transportation. Differences in preferences can be determined through historical reasons or regional policies that are not easily observable. Other unmeasured characteristics such as the supply level of public services or the accessibility to public transport stations can also play an important role in these differences. The variance of estimated random intercepts collects the odds variations caused by departing from different provinces. We obtained 94 realizations of these random intercepts, 47 for each mode of public transport (bus and train). The plotted maps differentiate between provinces with odds ratios below 1 and provinces with odd ratios above 1, indicating the areas that decreased and increased the probability of choosing a public mode over a car, respectively.

Figure 3.1 and Figure 3.2 display the exponential transformation of random intercepts associated with bus and train modes for each province, respectively. In the multinomial logit model, this transformation is interpreted as the odd ratio of choosing a public mode over a private car relative to the departure from a particular province. The plotted maps differentiate between provinces with odds ratios below 1 and provinces with odd ratios above 1, indicating the areas that decreased and increased the probability of choosing a public mode over a car.

The map represented in Figure 3.1 suggests a dominance of buses over cars in provinces in the north and in the west of the country, a route traditionally dominated by the largest bus companies in Spain. Figure 3.2 shows that the odds of choosing a train over a car are increased in trips departing from provinces

60

near the east coast of Spain. There also seems to be a strong effect in Madrid, Toledo, Ciudad Real, Córdoba and Sevilla; these five provinces enjoy a high-speed rail connection. However, provinces located along the southeast coast reduce the relative risk of choosing a train over a car. The radial structure of the railway network is easily recognizable, increasing the probability of choosing a train over a car in provinces covered by primary railway lines. The same map shows that certain areas in the north and south of the country not been favored by good access to train services reduce the use of trains.

Figure 3.1: Bus mode provincial random intercepts
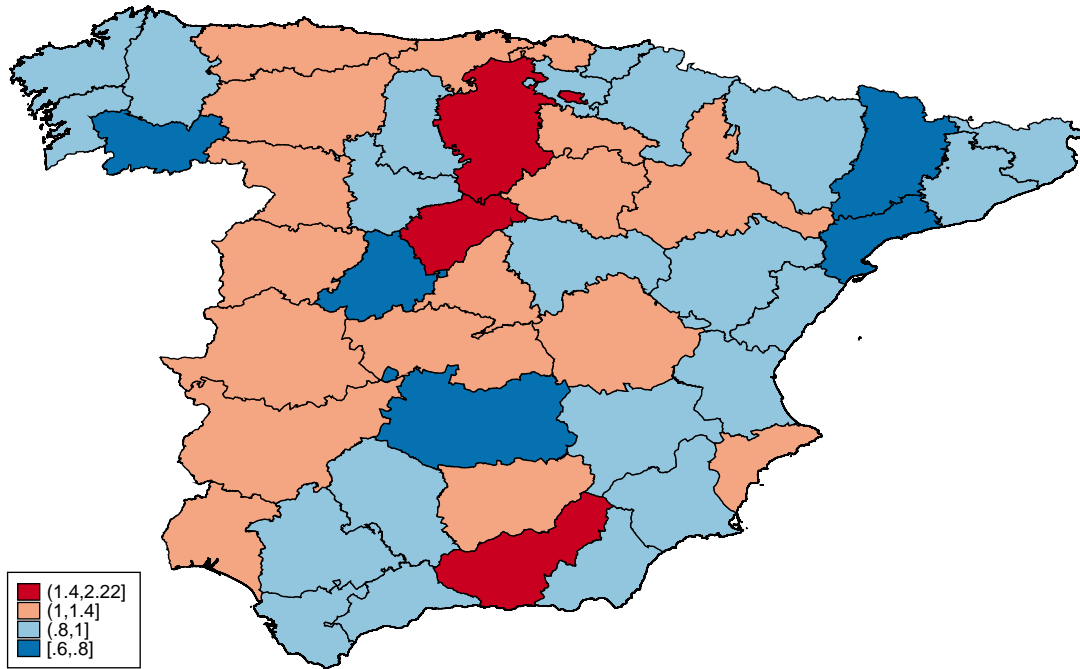
In comparing both figures, we can see that odds ratios associated with buses on the northeast coast and some landlocked provinces in the south are below 1. The same provinces appear to increase their probabilities of choosing a train over a car. This might indicate that in certain areas, travelers tend to choose

either buses or trains instead of cars. An insufficient provision of railway service might be appealing to bus companies, attracting them to operate in those markets where railway is less competitive. However, regions in the northwest (Galicia) of the country and in some landlocked provinces in the north middle east (Navarra, Alava, Guipuzcoa, Huesca, Teruel) and on the south coast (Mlaga, Almera, Murcia) show decreasing probabilities in both public transportation modes in relationship to cars. The dominance of cars in these areas might be related to different factors. The geographical remoteness of the provinces in the northwest and south of the country, along with a low supply of public transport in these provinces, might increase the need of private vehicle usage. The lack of a car ownership control variable might be a significant aspect affecting the probabilities of travelers from provinces in the Basque Country, one of the richest regions in the country, who are more likely to own a private vehicle than those in other areas.

Figure 3.2: Train mode provincial random intercepts
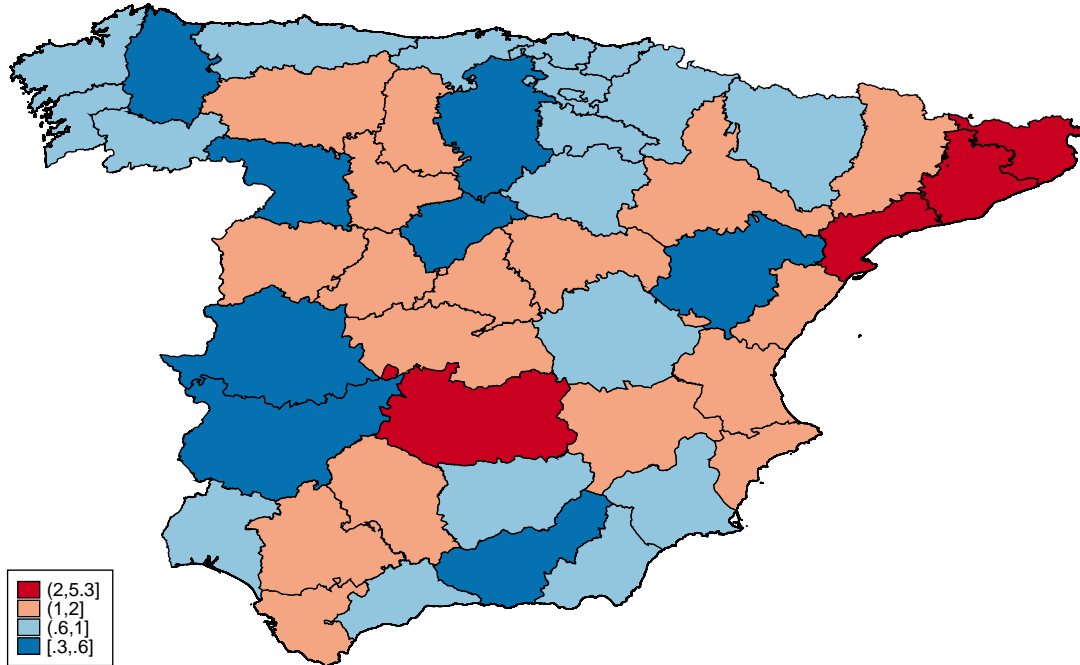
3.6   Conclusions

A better understanding of modal choice is important to take adequate policy measures to guide mobility behavior towards more sustainable modes of transport. The aim of this paper has been to study the determinants of mode choice in long-distance trips in Spain. In particular, we focused our attention on three ground modes of transport: private car, bus and train. For this purpose, we applied discrete model choice techniques and tested the impact of several sociodemographic, land use and trip purpose independent variables.

To conduct the analysis, we employed data from a 2006 Spanish mobility survey that attempted to evaluate the travel behavior of Spanish residents. It is worth emphasizing two characteristics of the database that might be common to

other long-distance databases in different parts of the world: missing variables and hierarchical structure. A literature review of the few papers dealing with inter-city travel helped in relating the most common explanatory variables in these previous studies. Comparing this information with the available indicators in the survey, we found some missing variables, including household income, car ownership and spatial indicators. Although all these variables may be important, our greatest concern was knowing the exact locations of origin and destination. Ignoring these geographical points prevented any attempt to include accessibility measures to transport infrastructure. As a solution, we proposed exploiting the hierarchical structure of the data. In this survey, trip observations might be nested in the provinces of origin for the trip. This characteristic feature of the data allowed us to estimate a multinomial multilevel model with random intercepts. Multilevel analysis permitted reducing the omitted variable bias and improving the estimation of standard errors through clustering of the observations.

As expected, the analysis confirmed some empirical evidence found in previous papers and added new insights in the determinants of long-distance travel mode choice. Socio-demographic variables such as gender, age and income play a significant role on mode choice. While women are less car dependent that men, young and elderly travelers rely more on buses and trains. Higher income levels also reduce the odds of using public transport over private cars. Characteristics of the origin of trip locations such as city size and belonging to a metropolitan area were also found to be significant. Differences on mode choice also arose depending on the motive of the trip. Trips for business purposes are more likely to be made by train instead of car, while travel for leisure preferred car usage. In the case of comparing trips to a second residence with leisure trips, the former are more car

dependent. Moreover, we tested the inclusion of a variable capturing the duration of the trip, overnight stays, that has not been used widely in previous literature on inter-city travel. A longer duration of inter-city trips appears to favor the use of railways, reducing the demand for buses. We also discussed some interesting significant relationships with the cost variable.

Random intercepts included in the model captured spatial differences on the probabilities of choosing transport modes, resulting in a more flexible specification. Individuals beginning their trip in the same province are affected by certain factors that might not be properly accounted for through explicative variables and may have important policy implications. Our results showed that, when controlled for individual characteristics (level 1), there is evidence of spatial differences (level 2). These differences can arise from different factors. The spatial distribution of preferences is not uniformly distributed, thus differently affecting the final decision depending on the departure province. The information plotted on the maps shows certain patterns that were explained in terms of the impact of important factors related to geography, institutions and transport policy.

This paper has clearly defined the dominance of cars over public transport modes in Spain. Regional heterogeneity in the design of transportation networks has yielded different spatial access to public transport modes. This factor, combined with inexpensive access to road networks by private vehicles, has promoted the predominant role of cars in mode choice decisions, especially in certain areas. While departing from certain provinces seem to increase the chances of using public transport modes (either bus or train), there are other areas where, according to our results, buses and trains are less likely to be used than private cars. Sustainability is an ineludible challenge of any transportation system, and transport

policy makers should consider if the current picture of the passenger transportation market in Spain, particularly in these areas, is on the path of reaching such a goal.

Our analysis provides interesting results and new insights that add to the existing knowledge on inter-city mode choice, and open new avenues for research on factors related to transport policy. Future research should further enrich this type of analysis by using more refined geolocated data on public transportation accessibility.

Chapter 4

Spatial Productivity of Road Transportation Infrastructure

4.1 Introduction

Measuring the economic effects of public infrastructure improvements on the productivity of private capital has been the center of academic debate for the last two decades [1]. The concept underlying these papers is that public capital plays a significant role as an input factor in the production process. The first empirical works addressing this issue appeared in the 1970s (Mera, 1973); however, it was only with the extraordinary results obtained by Aschauer (1989) that the research community showed a revived interest in the effects of public infrastructure improvements. In these early works, the authors found that public capital exerted a

---

[1] Different surveys on this topic compare different studies and focus on the methodologies adopted and the data used. For a detailed discussion, see: Gramlich (1994) and Pereira & Andraz (2010). Because of the high quality of the available data from Spain, many empirical works have recently been undertaken in Spain. **?** and Álvarez *et al.* (2003) provide an exhaustive review of international research, including studies in Spain .

large and significant effect on output. Aschauer estimated an output elasticity of approximately 0.4, and the results in the study by Munnell (1990) ranged from 0.31 to 0.39. In an era when the productivity growth of most OECD-countries experienced a significant slowdown, policy administrators and scholars wondered whether this might be caused, at least in part, by insufficient public capital. In this context, Aschauer's main findings were appealing because an increase in public investment in infrastructure seemed a straightforward solution to an alarming slowdown in productivity.

While there is little doubt that enterprises need a minimum level of public infrastructure to generate output to sell in markets, it should not be expected that the marginal output effect of extra public infrastructure will remain constant at every level. In the case of road transportation infrastructure, building one interstate network might cause a significant increase in productivity but building a second might not (Hulten, 2004). Following early articles, several methodological and conceptual objections on public capital productivity appeared in response. These critics pointed to problems involving spurious relationships (Garcia-Mila *et al.*, 1996), reverse causation bias (Chandra & Thompson, 2000; Fernald, 1999), the use of aggregate data at the country and industry levels (Gramlich, 1994) and problems with taking into account dynamic feedbacks in the relationship between public capital and private-sector performance (Pereira & Andraz, 2010). Subsequent research failed to find the significant positive effects of public capital on private output (Holtz-Eakin & Schwartz, 1995). Notably, regional data were used in the articles with models that obtained low estimates of marginal productivity, while the early studies obtained large effects with national data.

Certain researchers suspected the existence of spillover effects (Cohen & Mor-

rison Paul, 2004), which are also known as leakages; spillover effects indicate that the effects generated from public capital investment would not be confined to the region in which the infrastructure is located. If spillovers were present, part of the effect of public capital would be underestimated by using regional data. It should be noted that different categories of public capital may not have the identical spatial effects on private output, e.g., urban and water facilities projects may enhance economic activities that are confined to the local area, whereas communication and transportation infrastructure projects may cause important network effects.

Transportation infrastructure projects are the public capital projects that generate the greatest interest. The spillover effects seem particularly relevant to these types of projects because public investments in a region may affect other geographical units connected by a transportation network in addition to affecting that particular region (Boarnet, 1998). In fact, state highway projects are natural laboratories to test these effects because the interstate highway system is designed with interstate linkages in mind (Holtz-Eakin & Schwartz, 1995). Despite recent developments, the nature of infrastructure spillovers also remains inconclusive; positive and negative spillovers have been found. Positive spillovers are explained by the connectivity that is characteristic of transportation infrastructure; any piece of a network is related and subordinated to the entire system, which increases the interrelationships between areas (Moreno & Lopez-Bazo, 2007). If there were congestion, additional infrastructure development would enhance general economic activity (Cohen & Morrison Paul, 2003). Conversely, negative spillover might occur if migration processes arise that present evidence of leeching behavior; infrastructure improvements in neighboring areas enhance

that location and enable the region to attract productive resources, assuming that such resources are mobile (Boarnet, 1998).

We attempt to overcome another caveat in the empirical literature that involves measuring both private and public capital with stock indicators instead of with flow variables. Measuring capital as a stock fails to take into account utilization of the installed capacity. In the case of road transportation infrastructure, stock capital indicators are only a satisfactory measure of the quantitative properties of the infrastructure and not of the connective properties of the network. Using a Spatial Durbin Model, we attempt to correlate the use of private capital with business cycles studied in geographical units; and stock measures of road infrastructure investment are replaced by the interaction of vehicles in a region with available road infrastructure, as proposed by Fernald (1999). By applying this framework, we aim to solve previous problems in the literature that might have caused the current ambiguity in the empirical results, as noted by Mikelbank & Jackson (2000). These authors argue that any tool that does not consider an adequate geographic scale, the correct measures and the interactions between them will not capture the true relationship between spatial economy and public capital.

The resulting model is applied to Spain, where road transportation infrastructure projects have been promoted through the implementation of the Infrastructure and Transport Strategic Plan that raised the quality of Spain's road transportation network to European standards in a short period of time. Delgado & Álvarez (2007) studied Spanish highways and underlined the effect of European Union funds assigned to finance infrastructure projects in less-developed areas to promote growth and cohesion within the entire European Union. Most studies

conducted in Spain to date have focused on the Nomenclature of Territorial Units for Statistics 2 (NUTS-2) regional level [2]; however, following the recommendations of Rephann & Isserman (1994), we build a more disaggregated database using NUTS-3 level provinces in the Spanish territorial unit classification. In this context, the objective of this study is to measure the output effects of road transportation infrastructure projects in Spanish provinces in the period between 1986 and 2006. In particular, we aim to account for the marginal productivity effects of road transportation infrastructure services within a province and to document the existence of spillover effects outside the provincial boundaries through the use of spatial econometrics methodologies.

The structure of this paper is as follows. In the next section, we review the methodological issues of production function approach, the building of capital services variables and the treatment of the spillovers. In section 3, the empirical models are discussed along with the econometric issues. In section 4, we describe the data used and the source of the variables. In section 5, we present the estimation results. Finally, section 6 contains some conclusions and policy recommendations.

4.2   Theoretical background

In this paper, we focus on the changes in productivity that result from increased infrastructural investments by using a primal approach [3]. The main aim of this

---

[2]Several studies have utilized regional data to study this issue in Spain (for exhaustive information, see the recent survey by ?), whereas available papers using data on provinces are scarce, Álvarez *et al.* (2003), Delgado & Álvarez (2007) and Moreno & Lopez-Bazo (2007)

[3]Other papers have previously addressed this issue using a dual approach. Cost function models rely on duality theory and allow for a richer analysis through the estimation of the optimal input demand equation. However, one of the shortcomings of this type of model is

article consists of the estimation of the output elasticity of road infrastructures; to reach this objective, production function methodology is more useful than cost and profit function methodologies (Pfahler *et al.*, 1996). We suppose that there is a conventional output production function that relates real physical output, Y, to the quantity of variable inputs, X, quasi-fixed private capital input, K, and external factors as different types of public transportation infrastructure projects, G.

$$Y = f(X, K, G) \tag{4.1}$$

In a log-linear Cobb-Douglas specification:

$$\ln Y = \alpha_0 + \alpha_1 \ln X + \alpha_2 \ln K + \alpha_3 \ln G + \upsilon \tag{4.2}$$

where $\upsilon \sim \mathcal{N}(0, \sigma_\upsilon^2 I_n)$.

Ideally, in (4.2), inputs should be measured in terms of service flows. When inputs such as capital, K, enter the production function as a stock, unbiased comparative-static effects are computed on the assumption that changes in input services are proportional to changes in input stocks. However, in the presence of positive adjustment costs, this assumption may not hold for capital, K. The non-proportional changes in private capital stock, K, and its flow of services, K^*,

---

that information on factor prices is required. Estimating a profit function is an alternative that permits the estimation of unconditional demand effects, but it is even more extensive than the cost function approach in terms of data requirements. The information required by the cost and profit function approaches is not available at the NUTS-3 level in Spain.

are represented as variations in the capacity utilization rate, (CU). In particular, we consider the following expression: $CU = \dfrac{K^*}{K}$ [4].

The lack of regional and provincial statistics for CU makes this variable an unobservable factor [5], and as a consequence the same happens to K^*. At this juncture, we suggest that CU depends on economic environment (Gajanan & Malhotra, 2007) or business cycles and besides we posit that this relationship would exhibit spatial dependence.

From a theoretical point of view, shocks in the production of *neighbor* units might increase the demand for products in the region under study. In international macroeconomics, when an economic boom produces an increase in the output of a country such as the United States of America, simultaneous increases in outputs in other countries are observed. Open economy models frequently have difficulty in explaining why business cycles are so closely related among countries. According to Baxter & Farr (2005), this fact frequently requires implausibly high cross-country correlations of productivity shocks. These authors show that variable capital utilization explains these events. Consequently, an alternative explanation would posit that the economic agents of one region might accommodate the utilization rate of capital to meet output increases in other regions (Burnside & Eichenbaum, 1996).

Thus, we can formally express this using a spatial process as follows:

$$\ln CU \;=\; \lambda + \phi W \ln Y + \nu \qquad (4.3)$$

---

[4]Note that CU can also be expressed as the deviation of actual output from optimal output.
[5]Empirical regional measures of capacity utilization are described in Garofalo & Malhotra (1997)

where λ is a constant term and ν is distributed as a $\mathcal{N}(0, \sigma_\nu^2 I_n)$.

In (4.3), the n by n spatial weight matrix, W, reflects the connectivity of the provinces, and the scalar parameter, ϕ, reflects the strength of spatial dependence in Y. If the scalar dependence parameter, ϕ, is positive, then the CU rate in region i will be positively associated with the output of neighboring regions.

Substituting the spatial specification (4.3) in (4.2),

$$
\begin{aligned}
\ln Y &= \alpha_0 + \alpha_1 \ln X + \alpha_2 (\ln CU + \ln K) + \alpha_3 \ln G + \upsilon \\
&= \alpha_0 + \alpha_2 \lambda + \alpha_1 \ln X + \alpha_2 \phi W \ln Y + \alpha_2 \ln K + \alpha_3 \ln G + \upsilon + \alpha_2 \nu \\
&= \mu + \alpha_1 \ln X + \beta W \ln Y + \alpha_2 \ln K + \alpha_3 \ln G + \epsilon \qquad (4.4)
\end{aligned}
$$

where the intercept $\mu = \alpha_0 + \alpha_2 \lambda$ and $\upsilon + \alpha_2 \nu = \epsilon \sim \mathcal{N}(0, \sigma_\nu^2 I_n)$

According to Manski (1993), the $W \ln Y$ variable in (4.4) denotes the endogenous interaction effects and $\beta = \alpha_2 \phi$ is called the spatial autoregressive coefficient.

As discussed above, the effect of the different inputs on productivity should be measured in terms of service flows. Considered in terms of productivity changes, the significance of the role that the transportation infrastructure plays in the economy of a region is determined by the infrastructure services that it provides. Improvements in these services are expected to reduce generalized transportation costs as a result of shorter distances, less congestion and higher speeds that reduce fuel, capital and labor costs (Forkenbrock & Foster, 1990). However, transportation projects create other significant spatial location services in addition to reducing travel and logistics costs. They may enlarge the market potential of

businesses by enabling them to serve broader markets more economically. Furthermore, improvements in the transportation system can provide firms with a greater variety of specialized labor skills and input products, making them more productive. (Rietveld, 1994) offers a description of the spatial development effects resulting from transportation infrastructure supply as a complete theoretical framework.

Measuring infrastructure as a stock fails to account for the actual supply of the services that determine its contributions to productivity (Oosterhaven & Knaap, 2003). Because the main purpose of this study is to compute the effect of road infrastructure projects on provincial output, public transportation infrastructure projects (G) were divided into two different variables, one for roads and the other for all other modes of transport. Despite the heavy dependence of Spanish companies on road transportation [6], investment in ports, airports and railways are also included to test for their possible effects. However, better information about road transportation infrastructure projects and vehicles is available, which allows a measurement model to be built that accounts for road services.

Following Fernald (1999), we suppose that road services, (RS), depend upon the flow of services provided by the aggregate stock of government roads ($ROAD$) and the stock of vehicles (VEH), as shown in (4.5),

$$RS_{it} \quad = \quad f(ROAD_{it} * VEH_{it}) \tag{4.5}$$

---

[6]According to the National Statistics Institute, road transportation was chosen in more than 77% of freight movements in Spain in year 2007.

Based on Fernalds idea we have built a measure that tries to accommodate the road stock to its degree of utilization. Following his proposal, if road infrastructure made the companies more productive, sectors of the economy making a heavy use of roads would benefit more from their quality improvements. In the road services variable, we intend to adapt this idea applying it to provinces instead of sectors.

4.2.1 Treatment of spillovers

Public capital is likely to produce spillovers in other provinces. Positive and negative spillovers have been detected and explained in the literature. To explain possible negative spillovers, we follow Boarnet (1998); if there were an increase in public capital in region A, there would be a rise in the price of labor and capital in the region, inducing resources to move from other regions to region A. This migration would yield a new output in region A, reducing the output in the rest of the regions. Therefore, total output in one region would depend positively on its infrastructure stock and negatively on the infrastructure stock of other regions as a result of negative output spillovers. These negative spillovers, called distributive effects by Rietveld (1994), might not arise in an analysis at a low spatial level. For instance, if we focus on an urban area, we might observe the building of offices or industrial facilities near a new highway; because these would have been built elsewhere, they would remain outside of our study.

Conversely, the foundations of the existence of positive spillovers rely on the network characteristics of transportation infrastructure in which every piece is subordinate to the entire system (Moreno & Lopez-Bazo, 2007)[7]. Road network

---

[7]We are aware of Braess's paradox, which states that an increase in the capacity of a

improvements in neighboring provinces might lead to a decrease in the transportation costs of moving inputs and final products for the economy of a particular province, which might translate into an increase in the demand for manufacturing goods and services. Congestion might also play a significant role when explaining positive spillovers; new transportation infrastructures in regions in which bottlenecks exist might improve the performance of the entire network.

In (4.6), the provincial Cobb Douglas production function is augmented by including spillover effects using the spatial lag of the variable that contains information about transportation infrastructure projects, (G),

$$
\begin{aligned}
\ln Y \;=\;& \mu + \beta W \ln Y + \alpha_1 \ln X + \alpha_2 \ln K + \alpha_3 \ln G + \\
& +\; \theta_1 W \ln X + \theta_2 W \ln K + \theta_3 W \ln G + \epsilon
\end{aligned}
\tag{4.6}
$$

where Y is the output of province, X is a matrix containing variable inputs, K contains quasifixed input private capital, and G contains public transportation infrastructure variables; α, β and θ are the parameters to be estimated. W is the row standardized N-by-N spatial weight matrix with $W_{ij} > 0$ when observation j is a spatial neighbor to observation i. To test for the consistency of the results, models are estimated using two different weighting matrices, W, that will be explained in the next section.

The specification of the Equation (4.6) leads to what has been labeled the

transportation network might reduce its overall performance. However, we assume that this phenomenon is less likely to be felt on an aggregate level than are the positive effects of network improvement.

Spatial Durbin Model (SDM) that includes both the lagged dependent variable and lagged independent variables [8]. The SDM can be simplified to the spatial lag model and the spatial error model because these models are special cases of SDM. Our approach approximates a general to specific selection strategy after the recent contributions about model specifications in spatial econometrics (Elhorst, 2010b; Lesage & Pace, 2009). The most general model may include three different types of spatial interactions, which were identified by Manski (1993) as the following: endogenous interaction effects, exogenous interaction effects and correlated effects. Elhorst (2010b) found that the parameter estimates of the endogenous and exogenous interaction effects are biased when all interaction types are considered. To solve this problem, Lesage & Pace (2009) proposes the exclusion of the spatially autocorrelated error term, taking SDM as the departure from the general model. Ignoring spatial dependence in the disturbances will only cause a loss of efficiency while omitting relevant variables will cause a more severe problem, a biased and inconsistent estimator of the parameters. Furthermore, the spatial Durbin model produces unbiased coefficient estimates when the true data-generation process is any spatial regression specification other than the Manski model.

Another advantage of the SDM is that it does not impose prior restrictions on the magnitude of indirect effects, e.g., the spatial spillovers; thus, this model is more appropriate for the aim of this study [9].

---

[8]See Tong *et al.* (2013) and Yu *et al.* (2013) for two recent applications of SDM to the analysis of transport infrastructure

[9]These spatial spillovers are set to zero in a non-spatial model and in the spatial error model. In the spatial lag model, the spatial spillover effects in relation to the direct effects are identical for each explanatory variable.

4.3   Econometric Model and Estimation Issues

4.3.1   Model specification

The empirical model we estimate is based on the log-linear Cobb-Douglas production function. Following the previous discussion about different spatial econometric models, we estimate a Spatial Durbin Model:

$$
\begin{aligned}
y_{it} \;=\;&\; \mu_i + \beta W y_{it} + \alpha_1 l_{it} + \alpha_2 hk_{it} + \alpha_3 k_{it} \\
+\;&\; \alpha_4 rs_{it} + \alpha_5 trans_{it} + \theta_1 W l_{it} + \theta_2 W kh_{it} \\
+\;&\; \theta_3 W k_{it} + \theta_4 W rs_{it} + \theta_5 W trans_{it} + \epsilon
\end{aligned}
\tag{4.7}
$$

where variables on both sides of the equations are in logarithms, ϵ is a well-behaved error term, and subscripts i and t denote provinces and time periods, respectively. Compared to Equation (4.6), this equation also includes human capital, hk, and public capital, G, separated into two variables, road services, rs, and other transportation modes infrastructure stock, $trans$. Finally, spatial fixed effects, μ_i, are introduced into the model to control for all time-invariant variables.

Moreover, as discussed above, these equations include the spatial lag of the dependent variable and the spatial lag of the explanatory variables. Two different criteria have been used to build W [10]. Wn stands for a physical contiguity matrix,

---

[10]The weighting matrices have been row normalized following standard practice in the spatial econometrics literature. After transformation, the sum of all elements in each row equals one. Note that the row elements of a spatial weighting matrix show the effect on a particular unit of all other units.

in which its values would be 1 for two bordering provinces and 0 for all others. $Wd150$ is another binary weighting matrix with elements valued at 1 for those provinces within a radius of 150 kilometers from the centroid of the province of reference and 0 for provinces beyond that distance. These matrices treat physical proximity as the main driver for the presence of spillovers.

4.4   Description of data and variables

In this section, we discuss the data employed in the estimation of the model. Spain is a decentralized country made up of 2 autonomous cities (Ceuta and Melilla) and 17 autonomous communities, each with its own heritage and government. These autonomous communities correspond to NUTS-2 in the European territorial unit classification and are composed of 47 mainland provinces (NUTS-3). Both Autonomous Communities and provinces may be considered regional economies nested within a national system. The main property of this system is interdependence among the Spanish provinces because the evolution of each region depends on the behavior of neighboring regions [11].

We use a balanced panel dataset of 47 Spanish peninsular provinces covering the period from 1986 to 2006 that results in 987 observations, as shown in Table 4.1. The dependent variable, Gross Added Value, measured in thousands of year 2000 constant Euros, came from the National Statistics Institute (INE). The source for the explicative variables, labor force, as measured in thousands

---

[11]Márquez & Hewings (2003), analyze regional competition between Spanish regions (NUTS-2).

Table 4.1: Summary statistics of variables in logs

	Mean	Std. Dev.	Min.	Max.
y	15.61	0.9	13.76	18.63
k	16.3	0.85	14.44	19.31
l	5.21	0.86	3.39	8.00
kh	-1.96	0.34	-3.14	-1.14
rs	27.57	1.49	24.15	32.88
trans	13.26	1.09	10.5	16.93

of workers, and human capital, as measured by the share of total employment
with higher level education (secondary school, technical college and university
degrees), are INE and BBVA Foundation-Ivie, respectively.

Figure 4.1: Stock of transportation infrastructure aggregated at national level
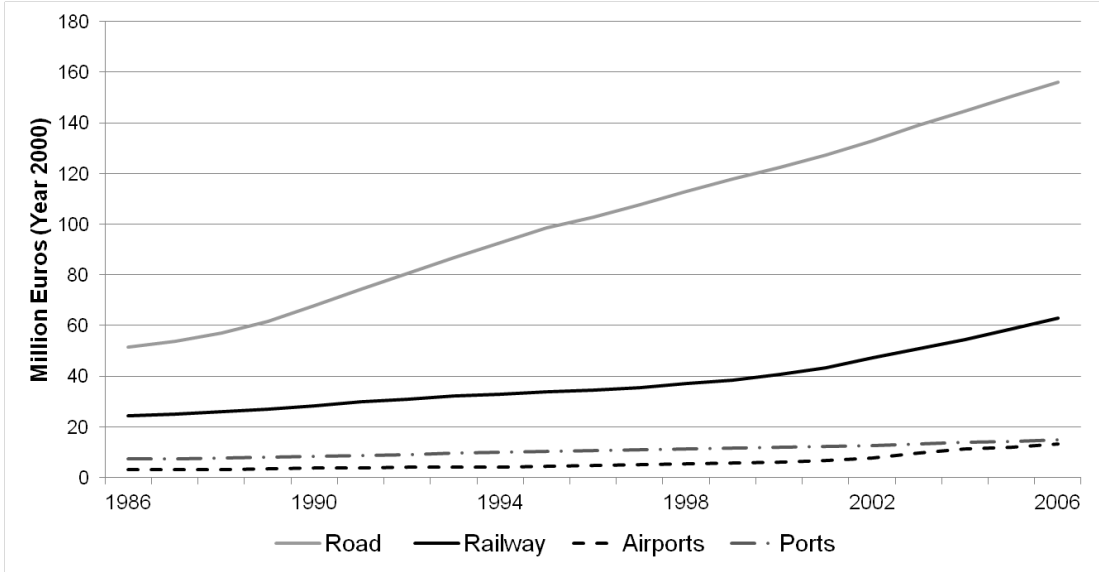
The latest series of capital stock for the Spanish economy were also obtained
from BBVA Foundation-Ivie (Más *et al.*, 2011), where net wealth and productive
capital stock data are available for both public and private capital. Productive

capital stock at constant pricing is a quantity factor that takes into account loss of efficiency as assets age and is the relevant component for productivity analysis [12]. Transportation infrastructure projects, such as ports, airports and railways, have been collapsed into one single variable excluding the stock of roads because these were previously included in the road services variable. As explained in Section 4.2, this variable is the result of the product of the stock of road infrastructure and the stock of private vehicles in each province. The source for the required information to build road services is BBVA Foundation-Ivie.

Since the 1970s, there has been substantial development of road transportation infrastructure in Spain; during the 1990s, in particular, the implementation of the Infrastructure and Transport Strategic Plan caused a significant boost in investment into High Capacity Networks . Figure 4.1 represents the national stock of infrastructure for the different transportation modes for the period of 1986-2006. Figure 4.2 shows the spatial distribution of road infrastructure stock growth rates at the provincial level during that same period. Neighbor provinces share similar growth rates for this variable that display an uneven distribution far from a random spatial process.
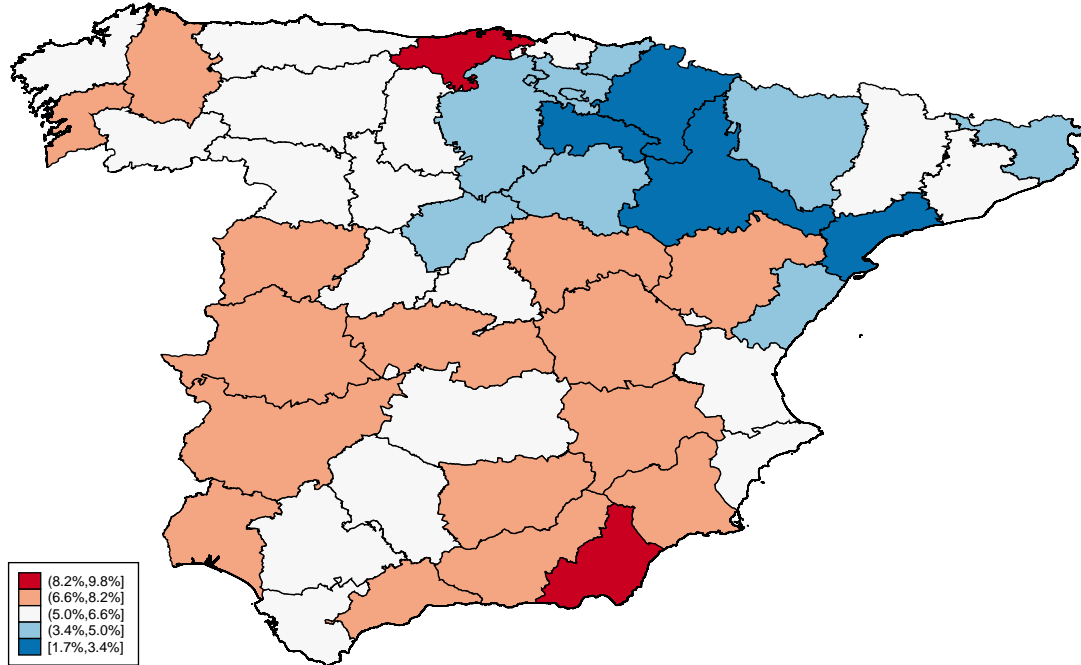
---

[12]The computations of the productivity of capital stock are obtained using a new methodology applied to Spanish capital stock estimates that is based on two OECD manuals (Schreyer, 2001; Schreyer *et al.*, 2003).

Figure 4.2: Rate growth of road transportation infrastructure stock (1986-2006)

(8.2%,9.8%]
(6.6%,8.2%]
(5.0%,6.6%]
(3.4%,5.0%]
[1.7%,3.4%]

4.5   Results

4.5.1   Spatial Durbin Model interpretation

Before discussing the results, it is worth noting that the coefficient estimates must be interpreted carefully because they are dependent on model specifications. For example, if the estimated model had the form of the Spatial Error Model (*SEM*), the coefficient estimates in log-form can be directly interpreted as elasticities. However, the effect of the independent variables on the dependent variable in the *SDM* has no straightforward interpretation, and direct and indirect effects must be computed. Lesage & Pace (2009) shows that the partial derivates take the form of an N-by-N matrix for each k regressor and comments on their fundamental properties . For instance, the partial derivates matrix corresponding to the road

services regressor (rs) from Equation (4.7) would have the following form,

$$\frac{\delta y_t}{\delta rs_t} = (I_N - \beta W)^{-1}(\alpha_4 I_N + \theta_4 W) \tag{4.8}$$

These authors propose scalar summary averages to increase the ease of reporting the effects associated with the regressors; thus, direct effects measure what effect changing an independent variable has on the dependent variable of a province. Direct effects, which appear in the main diagonal of the matrix shown in Equation (4.8), are their own partial derivatives and are summarized using the average of these elements of the matrix. This measure includes feedback effects, i.e., those effects passing through neighboring units and back to the unit that instigated the change. The cross-partial derivatives are named indirect effects, and they measure the effect of changing an independent variable in a province on the dependent variable of all the other provinces. Indirect effects appear as off-diagonal elements and are summarized as row sum averages. Finally, total effects are computed as the sum of direct and indirect effects.

4.5.2   Comments on the results

The results obtained through the estimation process are shown in Table 4.2, which contains the point estimates of the production function model using two alternative spatial weight matrices, as discussed above. In Table 4.3, direct, indirect and total effects computations are reported for the SDM.

Table 4.2: Spatial Durbin Model with Spatial Fixed Effects

	Wn		**Wd150**	
	Coef.	t-stat	Coef.	t-stat
K	0.157***	7.98	0.178***	8.98
L	0.272***	15.19	0.268***	14.99
HK	0.016	1.63	0.020*	1.84
RS	0.060***	7.26	0.061***	7.38
Trans	-0.002	-0.47	-0.001	-0.13
W*K	0.039	1.12	-0.001	-0.04
W*L	-0.080***	-2.97	-0.058***	-2.24
W*HK	0.001	0.08	0.005	0.29
W*RS	0.001	0.06	0.011	0.87
W*Trans	-0.018**	-2.00	-0.017*	-1.96
W*Y	0.297***	7.34	0.251***	6.18
Corrected R^2	0.967		0.966	
Log-likelihood	2,010.00		1,995.10	
Wald Test Spatial Lag	20.117	$p = 0.001$	15.25	$p = 0.009$
LR Spatial Lag	20.586	$p = 0.000$	15.276	$p = 0.009$
Wald Test Spatial Error	45.135	$p = 0.000$	32.63	$p = 0.000$
LR Spatial Error	50.256	$p = 0.000$	35.307	$p = 0.000$
Observations	987		987	

Significance code: *p<.1, **p<.05, ***p<.01

Spatial fixed effects are not displayed, but are available under request.

85

Overall, the results are consistent with other production function studies and indicate the existence of transportation infrastructure spillovers. As discussed below, there are certain results shared by all the estimated models. All the specifications of the model yield similar results regarding the output point estimates of the coefficients accompanying the regressors. It is worth underlining the positive and highly significant effects of the spatial lag of the dependent variable that shows values of 0.297 and 0.251, depending on the W specification adopted. This result indicates that the weighted average of the output of neighbor provinces positively affects production in the geographic unit under analysis. According to the theoretical model that we developed in Section 4.2, changes in the business cycle in other provinces would significantly affect productivity in a particular province. In this case, because the sign of the parameter accompanying the spatial lag of the dependent variable is positive, economic agents in that province would increase the capacity utilization of quasifixed inputs when output in other provinces grows and would decrease the usage when output falls.

The Wald and Likelihood Ratio (LR) tests permit the determination of the validity of the hypothesis positing that the Spatial Durbin Model can be simplified to the Spatial Lag Model. The results reported using the Wald test (20.12, $p. = 0.001$ for the contiguity matrix and 15.25, $p. = 0.009$ for the W matrix, using neighbors within a distance of 150 km, respectively) or using the LR test (20.58, $p. = 0.000$ and 15.28, $p. = 0.000$) indicate that the hypothesis must be rejected. Similarly, the hypothesis that the SDM can be simplified to the Spatial Error model must be rejected, according to the Wald tests (45.135, $p. = 0.000$ and 32.63, $p. = 0.000$) and the LR tests (50.256, $p. = 0.000$ and 35.307, $p. = 0.000$). To investigate the null hypothesis that the spatial fixed effects are jointly

insignificant, an LR test may be conducted. The results (3,248.95, $p. < 0.01$ and 3,233.27, $p. < 0.01$ both with 47 degrees of freedom) indicate that this hypothesis must be rejected and justify the extension of the model with spatial fixed effects.

However, as discussed above, inferences must be made about the effect of independent variables on the productivity of a province with regard to the direct, indirect and total effects displayed in Table 4.3. According to these results, the direct effects of labor and private capital on the aggregated output of a particular province are positive and significant. Moreover, these elasticities are stable. The elasticities of labor (approximately 0.27) and private capital (between 0.16 and 0.18) are positive and significant in all the models [13]. These coefficients are similar to those obtained in some of the latest applied studies in Spain (Márquez *et al.*, 2010). Estimations of the direct effects of other modes of transportation capital are also not significant, regardless of the empirical specification.

The estimated coefficients accompanying the variable of interest in this work, the road services variable, are positive and highly significant, and their sizes show little variation. On average, a 100% increase of the road services of a certain province causes a 6.1% increase in its productivity.

In the SDM, the indirect effects reveal the existence and size of effects across boundaries. We find evidence of positive spatial spillovers for the road services

---

[13]Elhorst (2010a) emphasize that empirical studies usually find significant differences among the coefficient estimates from models with and without spatial fixed effects. Models that include spatial fixed effects use time-series variations of the data, whereas models without controlling for spatial fixed effects utilize cross-sectional components of the data. Models of the first type tend to give short-term estimates, and models without controls for spatial fixed effects tend to give long-term estimates (Baltagi, 2008).

Table 4.3: Direct, Indirect and Total Effects

		Wn		Wd150	
		Coef.	t-stat	Coef.	t-stat
Direct Effects					
	K	0.162***	8.47	0.180***	9.43
	L	0.272***	15.33	0.270***	16.01
	HK	0.016*	1.66	0.019*	1.92
	RS	0.061***	7.76	0.062***	7.60
	Trans	-0.003	-0.69	-0.002	-0.42
Indirect Effects					
	K	0.114***	2.79	0.057	1.35
	L	0.002	0.05	0.011	0.41
	HK	0.009	0.38	0.013	0.59
	RS	0.025*	1.73	0.034**	2.28
	Trans	-0.025**	-2.09	-0.022**	-2.10
Total Effects					
	K	0.276***	6.67	0.234***	5.59
	L	0.274***	10.13	0.281***	11.77
	HK	0.025	0.95	0.032	1.35
	RS	0.086***	6.44	0.096***	6.78
	Trans	-0.028**	-2.19	-0.024**	-2.14

Significance code: *p<.1, **p<.05, ***p<.01

variable with estimates of 0.025 and 0.034, which depend on the W matrix employed to define neighbors. According to these results, a 100% increase in the services provided by the road infrastructure of a province would yield positive effects for the productivity of its neighbor up to 3.4%. For the remaining modes of transportation, we found clear evidence of negative spillovers. The indirect effect coefficient is close to -0.02 in different settings. Following the interpretations presented in section 2.2, improvements in the services provided by the road infrastructure in one province would cause positive spillovers to other provinces by raising the quality of the road transportation network as a whole. Conversely, increased investment into ports, airports and railway infrastructure projects in a province would produce negative spillovers through the migration of productive factors to those regions with the larger levels of public capital investment.

These results offer evidence of spatial spillovers for the different types of transportation infrastructure projects consistent with most of the literature using Spanish provincial data. For instance, using a stochastic frontier approach, Delgado & Álvarez (2007) found positive and negative spillovers depending on the sector of the economy under review and the definition of the weighting matrix. Utilizing a production function, Moreno & Lopez-Bazo (2007) found the existence of negative spatial spillovers for transportation infrastructure development. By contrast, using Spanish provincial data, Álvarez et al. (2006) replicated the models used by Holtz-Eakin & Schwartz (1995) and Mas et al. (1996) and did not find either positive or negative spillovers.

Finally, we obtain the total effects of the variables in the productivity of a province by adding direct and indirect effects together. We find that all the variables included in the model are significant with the expected signs, except

that the human capital variable does not appear to be statistically significant. As discussed above, the average total effect of road transportation services is positive and significant (ranging from 0.086 to 0.096). Conversely, the average total effect of investment into other modes of transport infrastructure appear to be dominated by negative spillovers, lowering the productivity of other provinces [14].

4.6   Conclusions

In this paper, we attempt to find the correct specifications for an aggregated production function to measure the effects of road infrastructure public investments on the economy of Spain. The main contribution of the work is twofold. First, we *revisit* Fernald (1999) using a variable that combines the stock of vehicles in a certain province with information about the road network to assess the effects of road transport infrastructure on productivity. Second, the empirical models include spatial lags of the independent variables and of the dependent variable, which is not common in the literature. We empirically and theoretically justify the inclusion of the spatial lag as an explanatory variable. Primarily, we accommodate the private capital grade of utilization in the business cycles contained in the geographical units. In this fashion, we attempt to avoid shortcomings caused by the use of stock indicators of private and public capital while capturing the underlying spatial processes at work.

As our main empirical result, we find strong evidence of the positive effects

---

[14]We must remind ourselves that the estimated effects in Table 7 are computed as national average effects using the whole set of geographical units. Thus, the estimated effects for a particular province might be different from those results reported in 4.3.

of road infrastructure projects on the private economy of a province. Spillovers caused by investment in transportation infrastructure (i.e., the effects on one province of changes in the flows of road services in other provinces) are approximately half the size of the direct effects of such investment. Improvements in the road infrastructure of one spatial unit increases productivity in neighboring units by approximately half of the amount of the improvement in the spatial unit in which the infrastructure is located. According to the outcome of this model, specially the importance of spillover effects, seems to support the idea that road transport infrastructure investment effects are not confined within the territory where the infrastructure project is located. In the Spanish political context, this conclusion can have major consequences because both regional and provincial governments share with the national government the decision making about where to make infrastructure investments.

Chapter 5

Conclusions and extensions

The essays that compose this dissertation address different topics on the economics of transportation. The first two essays analise the demand side of transportation services. In the first essay, we study the phenomena related to the first two steps in the traditional four-step model for traffic assignment: trip generation and trip distribution. The trip generation step determines the frequency or volume of trips in each zone, and the trip distribution step matches the origin with the destination of corresponding flows. The second essay, which focuses on transport mode choice, tackles the third step of the process in an attempt to gain knowledge about transportation customers. The final essay aims to evaluate a fundamental aspect of the supply side of the transportation sector by computing the services that the transportation infrastructure contributes to the regional economy.

In the first essay, we analyse the existence of a dynamic component of road freight flows. The main contribution of this portion of the thesis is to specify a model that applies a dynamic version of the gravity equation. Two lags of the dependent variable, the log of freight goods transported from one region to another, were included as explanatory variables. It was then statistically tested

whether systematic changes in the past values of road freight flows positively affect the current values.

In the beginning section of the essay, we reviewed the literature related to gravity models and focused on their dynamic version. Although this type of inertia has not been extensively addressed in the transportation literature, scholars have investigated its effects in international trade. According to these authors, businesses may make investments in developing their commercial networks to increase their presence in certain markets. These investments (which can be thought of as sunk costs) signal long-term planning that might partly influence the patterns of freight movements to maintain their current trends.

In the empirical portion of the essay, we explore a dynamic specification of the gravity equation with some peculiarities. The model includes common regressors, such as GDP measures and distance, as drivers and deterrents of transportation flows, respectively. In addition, a measure of the quality of transportation infrastructure was used as an explanatory variable. Its inclusion attempted to test whether better infrastructure favored the transportation of goods in an aggregate demand framework.

The data used to calibrate the model consist of a balanced panel set of road freight flows among Spanish regions over the 1999-2009 period. The results obtained confirm the theory of the gravity model and find a significant and positive effect for the scale measures of the regions, measured as GDP, and a negative effect for distance. The constructed proxy of the quality of infrastructure also indicates that providing more road infrastructure benefits road transportation flows. The key result shows that there is a positive effect of lagged freight variables in the explanation of current freight variables. Although its magnitude is

not large, this result indicates that the inertia of flows must be taken into account to avoid model misspecification.

As a complementary exercise, CO_2 emissions emanating from road freight transportation are calculated in a straightforward fashion. An equation relates the total CO_2 emissions per year through the interaction of data with respect to transport operations between each pair of regions and some average values on diesel consumption and distance travelled. Using the aggregate transport operations collected in the database and, alternatively, those predicted by our model, we provide yearly data on CO_2 emissions caused by the most used mode of transportation in Spain.

Based on the above discussion, it appears that the volume of CO_2 emissions in recent years is quite stable. This outcome is a bit surprising, as we expected a larger decrease in CO_2 emissions resulting from the reduction of traffic movements. Nevertheless, it seems that the economic slowdown may have had additional effects apart from reducing the pace of growth of road freight flows. The analysis of the data indicates that the vehicles renewal rate is much lower than in previous years, which is a product of the economic environment. Considering that older vehicle models are less energy efficient and pollute more than newer versions, policy-makers must take these results into account if a reduction in greenhouse gases is among their primary goals.

This work is already being extended in my current study that involves performing an analysis of microdata on vehicles shipping freight goods. The main objective is to measure the impact of distance on decisions about vehicle shipment size. Many transportation models assume that the vehicles used in transportation activities are of optimal size. Therefore, it is also assumed that this optimal

outcome is the result of minimizing total logistics cost in the Economic Order Quantity (EOQ) model. This function typically includes a trade-off between transport costs and inventory costs; as the shipment size increases, transport costs decrease (although inventory costs increase). In such an optimization process, economies of distance and economies of scale are likely to be important. Economies of distance occur when transport rates taper as haulage distance increases, whereas decreasing transport rates with respect to increasing haulage quantity for a given distance is typically known as economies of scale. For many years, both effects have been assumed in many transportation models without being properly tested. Currently, I am working on an empirical specification that allows us to test the latest theoretical developments related to this issue.

Another potential extension of this work is the application of spatial econometric techniques to the gravity model primarily by estimating a specification of the model that accounts for spatial dependence among the regions. Conventionally, it has been accepted that distance functions in origin-destination models were enough to capture the possible spatial dependence of flows. However, recent discussions have shown that residuals of aspatial specifications of gravity models exhibit spatial dependence. The latest developments of spatial econometric models using interaction data point to the need to test the inclusion of the following three different spatial effects: "origin-based", "destination-based" and "origin-destination-based" spatial dependences. The first, "origin-based" spatial dependence, would include a spatial lag of origin flows, i.e., an average of the flows from neighbors to the origin to the destination region. The second type of spatial dependence, "destination-based" spatial dependence, is intuitively explained as the forces leading to flows from an origin state to a destination state

that create similar flows to nearby or neighboring destinations. Finally, "origin-destination-based" spatial dependence is a combination of these two types of spatial dependence and reflects an average of flows from neighbors of the origin state to neighbors of the destination state. The model allows test whether these effects are significant and finds the appropriate empirical specification.

The second essay aims to study the determinants of the choice of long-distance travel modes. In practice, this chapter focuses on how socio-economic characteristics of travelers, land use features and trip attributes shape choices among a set of ground transportation modes consisting of private car, public bus and train. Mode-choice models have been widely studied in the transport economics literature in the context of daily commuting or short-distance trips. However, data on long-distance or inter-city movements are not as common, which makes the analysis of this topic less usual.

Although commuting is a daily activity for almost everyone, long-distance travel is less frequent for all except a few job positions, such as salespersons or diplomats. The set of differences between short- and long-distance trips includes differences in travel times and costs incurred. Generally, more time and out-of-pocket costs are required for inter-city trips, which make the traveler more likely to study the supply of services more closely. In addition, travel purposes and mode availability might differ. In the literature review section, we provide up-to-date empirical evidence about the effect of socio-demographic, land use and trip attribute variables on the selection of the mode of transportation regarding inter-city trips.

The data description of mobility survey (Movilia 2007) plays a fundamental role in this essay. The objective of this survey is to study the main features of long-

distance trips and travelers using telephone interviews as the main data collection method. More than 19,000 microdata observations were analyzed and described, and these provided the main substance of the database. This preliminary analysis combined with a literature review help properly address the research question in two ways. First, it helped check the consistency of the results to be obtained by the multivariate analysis. Second, it allowed us to identify the drawbacks of the database.

In general, the data are rather powerful and rich, although some caveats from the spatial point of view must be flagged. An ideal database would include data on time of the trip, in addition to accurate geographical information on the points of origin and destination. With this information in hand, it would be possible to calculate interesting variables, such as accessibility measures to public transportation services, origin and destination land-use indices and ingress and egress times to transport stations.

The empirical approach adopted attempts to overcome these drawbacks by exploiting the hierarchical structure of the data. The multinomial model allows the probabilities of choosing one specific mode of transportation over another that was previously set as the base category to be computed. The multilevel version of the multinomial model permits estimating random intercepts associated with higher levels that nest individual observations. In this essay, these intercepts are related to the departure provinces collecting information about unmeasured characteristics that are shared by all travelers with origins in a certain province. In this way, we are able to calibrate the model and take the spatial heterogeneity of trip-makers into account.

Among the avenues for future research, I would like to explain the variation

in these random intercepts to find the drivers that cause spatial heterogeneity in the choice of long-distance transport modes. In particular, future arrangements might include studying the effects of accessibility to transport stations and the impact of the degree of land-use characteristics. Controlling for these variables, almost all other spatial information except preferences would be erased from the provincial intercepts.

My interests now proceed into the details of long-distance mobility of low-income populations and particularly the elderly and the unemployed. A large portion of these population groups make no trips at all because of economical reason (in all likelihood), but I am also interested in analyzing the importance of other explanatory variables. Using a double-hurdle type of model, I aim to investigate the socio-demographic and spatial factors that affect the decision whether to travel or not and to model the choice of transportation mode in a second stage.

The third issue attempts to answer a question that has been at the center of the economic debate for a long time. In particular, we are concerned with measuring the economic effects of public infrastructure on the productivity of regions. Much emphasis is put on capital related to transportation infrastructure, particularly in the Spanish road transportation network. The contribution to the vast literature is twofold. First, we apply the latest advances in spatial econometrics modeling that allows us to test whether spatial dependence of the dependent and the independent variables must be included in the model specification, which has been labeled the Spatial Durbin Model in the literature. The second contribution attempts to overcome another drawback in the literature, i.e., the use of stock variables instead of flows. Measuring private and public capital as a stock adds information about the quantitative properties of the infrastructure but not the

services it provides. In this sense, stock measures are combined with information about the vehicles potentially using the infrastructure to obtain a proxy of the transportation services provided by this type of public capital.

Although this is not the first time this Spatial Durbin Model has been applied to regional production functions, previous authors have typically focused on its capabilities to account for the spillover effects of independent variables by using spatial lags of these variables. For studying transportation infrastructure spillovers, this feature might be particularly important because it allows the researcher to capture those effects of the infrastructure that go through the network and affect neighboring regions. The specification with the lag of the dependent variable has been supported by econometric reasons but has remained unexplained from the economic point of view. In this chapter, we provide theoretical justification for its inclusion by showing how the use of private capital in a region might be influenced by the business cycles of its neighboring regions.

The resulting model specification is applied to Spanish provinces (NUTS-3) during the 1986-2006 period, which was a period that witnessed heavy investments to increase the quality of Spain's road transportation network. Our main empirical result appears to support the idea that better transportation services increase regional productivity. However, this result might be a straightforward justification to maintain investments in upgrading the infrastructure, and we suggest it be taken with caution. From our perspective, this outcome of the model is important in explaining the productivity increases in that period; however, we are aware that the marginal contribution of the early rounds of upgrading certain corridors might be much more productive than subsequent contributions.

Additional results are related to the significant spillover effects caused by

transportation infrastructure. According to the estimations, an increase in transportation services provided by road infrastructure in one province raises productivity in neighboring units by approximately one-half the amount of the improvement in the spatial unit in which the infrastructure is located. This result supports the existence of spillover effects and shows that the potential impacts of some public infrastructure projects are not confined to the geographical area in which they are located. Thus, the network characteristics of this public capital are confirmed and might suggest the need for coordination among administrators in charge of transportation infrastructure planning. In the case of Spain, national, regional and provincial governments share this task. As a policy matter, we suggest testing the inclusion of spillover effects even in the analysis of those projects that are not particularly inter-regionally oriented.

An alternative version of this essay, which has recently been recently chosen to be part of a book edited by Edward Elgar Publishing, uses recent findings on the role of accessibility to study its consequences on regional economy. In addition, we are undertaking the development of a formal background of the Spatial Durbin Model based on New Economic Geography foundations. The objective is to provide other scholars with a robust theoretical framework when using this technique. Other potential work derived from this last chapter might involve the estimation of a spatial stochastic frontier to study the efficiency of provinces. Finally, we would also like to calculate the usage level of public and private capital in the Spanish provinces.

Resumen y conclusiones en castellano

Los ensayos que constituyen esta tesis analizan diferentes aspectos de la economía de transporte. Los dos primeros estudian aspectos de la demanda de servicios de transporte. En el primero de ellos, estudiamos fenómenos relacionados con las dos primeras etapas del Modelo clásico de Cuatro Etapas de Asignación de Tráficos: la generación de viajes y la distribución de viajes. La generación de viajes determina la frecuencia o el volumen de viajes en cada zona, mientras que la distribución de viajes empareja origen y destino de los flujos. El segundo capítulo, que analiza la elección de modo de transporte, se centra en la tercera etapa en un intento de mejorar el conocimiento acerca de los usuarios de transporte. El tercer ensayo evalúa un aspecto fundamental por el lado de la oferta del sector transporte, calculando los impactos que la infraestructura de transporte provoca en la economía regional.

En el primer capítulo, se analiza la existencia de un componente dinámico de los flujos de transporte por carretera. La contribución principal de esta parte de la tesis es la especificación dinámica de un modelo basado en la ecuación gravitacional. Para ello, se incluyen como variables explicativas dos rezagos de la

variable dependiente, el logaritmo de las mercancías transportadas de una región a otra. El objetivo es contrastar empíricamente si cambios sistemáticos en los valores pasados de los flujos de mercancías afectan positivamente a los valores actuales.

Al comienzo del capítulo se revisa la literatura referente a los modelos gravitacionales, y especialmente aquellos trabajos centrados en su especificación dinámica. Aunque este tipo de efecto persistencia no ha sido tratado de forma extensiva en la literatura de transporte, sí que existen trabajos aplicados al campo del comercio internacional. Según estos autores, las empresas pueden realizar inversiones para mejorar sus redes comerciales y así aumentar su presencia en determinados mercados. Dichas inversiones (las cuales pueden considerarse costes hundidos) informan acerca de la existencia de una estrategia a largo plazo que podría determinar parcialmente los patrones de movimientos de mercancías afectando de esta manera las tendencias actuales.

En la parte empírica del ensayo, se explora una especificación dinámica de la ecuación gravitacional que incluye ciertas particularidades. El modelo incluye las variables habituales, como Producto Interior Bruto y distancia, que actúan como estímulo y freno de los flujos, respectivamente. Adicionalmente, como variable explicativa se incorpora una medida de la calidad de la infraestructura. Con su inclusión se intenta contrastar si una mejor dotación en infraestructuras potencia el transporte agregado de mercancías.

El modelo se calibra utilizando datos de un panel equilibrado de los flujos de mercancías por carretera entre las regiones españolas en el periodo comprendido entre 1999 y 2009. Los resultados obtenidos confirman la teoría del modelo gravitacional encontrando un efecto significativo positivo de las medidas de escala de

las regiones, recogidas por el Producto Interior Bruto, y un efecto negativo para la distancia entre las mismas. La variable que recoge información acerca de la calidad de las infraestructuras señala que una mayor provisión de las infraestructuras de alta capacidad beneficia a los flujos de mercancías por carretera. El resultado clave muestra que existe un efecto positivo de los flujos de mercancías pasados sobre la variación de los flujos actuales. A pesar de que su magnitud no es muy elevada, este resultado indica que existe un efecto persistencia en los flujos de transporte que ha de tomarse en consideración para evitar una especificación errónea del modelo.

Como un ejercicio complementario, se incluye el cálculo de las emisiones de CO_2 provocadas por el transporte de mercancías en camiones. Para ello se utiliza una ecuación que relaciona el total de emisiones anuales de CO_2 emitidas a la atmósfera con una interacción de variables referentes al número de operaciones de transporte entre regiones, los valores medios de consumo de combustible y la distancia recorrida entre dichas regiones. Utilizando el agregado de operaciones de transporte procedentes de la base de datos y, alternativamente, aquellas predichas utilizando el modelo estimado, se proporciona una serie anual de emisiones de CO_2 provocadas por el modo de transporte más utilizado en el transporte de mercancías en España.

Según los datos obtenidos en dicha estimación, el volumen de emisiones de CO_2 en los últimos aos es bastante estable. Este resultado es un poco sorprendente, dado que cabría esperar una caída de las emisiones de gases contaminantes debido la importante reducción de tráficos en nuestro país causada por la coyuntura económica. Sin embargo, es posible que la desaceleración económica provoque efectos adicionales además de la reducción del número y volumen de

operaciones de transporte. Un análisis más detallado de los datos parece indicar que la tasa de renovación de los vehículos es mucho menor que en los años precedentes. En este sentido es importante considerar que los vehículos más antiguos son menos eficientes desde el punto de vista energético y, por tanto, contaminan más que los nuevos modelos. Desde el punto de vista de política económica se podría recomendar un sistema de incentivos para renovar la flota de vehículos en el caso de que la reducción de gases contaminantes se encuentre entre los objetivos principales de la sociedad.

El análisis de flujos de mercancías por carretera está siendo extendido en un estudio de los microdatos de los vehículos que transportan las mercancías. El principal objetivo es medir la influencia de la distancia sobre la elección del tamaño del vehículo. La mayoría de los modelos de transporte asume que los vehículos que realizan las actividades de transporte tienen el tamaño óptimo. De esta forma, también se asume que este óptimo se alcanza como resultado de minimizar los costes logísticos totales en el modelo de Cantidad Económica de Pedido. Éste, se caracteriza por una función que incluye un *tradeoff* entre los costes de transporte y los costes de inventario; al aumentar el tamaño del vehículo disminuirían los costes de transporte, aumentando los costes de inventario. En este proceso optimizador, es probable que las economías de distancia y escala sean importantes. Las economías de distancia ocurren cuando las tasas de transporte descienden a medida que la distancia del envío aumenta, mientras que el fenómenos por el cual las tasas de transporte se reducen con respecto a la cantidad transportada se conoce como economías de escala. Durante muchos años, ambos efectos se han considerado dados en muchos modelos de transporte sin ser contrastados de forma adecuada. Actualmente, me encuentro trabajando

en una especificación empírica que permita contrastar los desarrollos teóricos más recientes en este tema.

Otra extensión potencial de esta parte de la tesis es la aplicación de técnicas de econometría espacial a la ecuación gravitacional permitiendo la dependencia espacial entre regiones. Convencionalmente, los modelos de datos de interacción han considerado que las funciones de distancia eran suficientes para capturar esas posibles dependencias espaciales entre flujos. Sin embargo, trabajos recientes han demostrado cómo los residuos de especificaciones no espaciales de los modelos gravitacionales presentan dependencia espacial. Los últimos desarrollos de modelos de econometría espacial señalan la necesidad de contrastar la inclusión de tres tipos de dependencias espaciales diferentes: "basadas en el origen", "basadas en el destino" y "basadas en origen-destino".

La primera, la dependencia espacial "basada en el origen", incluiría un rezago espacial de los flujos de origen, es decir, un promedio de los flujos provenientes de los vecinos a la región de origen a la región de destino. El segundo tipo de dependencia espacial, "basada en el destino", se puede explicar intuitivamente como el caso en el que las fuerzas que impulsan los flujos de una región de origen a una de destino, crean flujos similares a aquellas regiones cercanas al destino. Finalmente, la dependencia espacial "basada en origen-destino" es una combinacin de las dos anteriores y recoge el promedio de los flujos de los vecinos de una región de origen a los vecinos de una región de destino. El modelo permite contrastar si estos efectos son significativos y encontrar la especificación empírica apropiada.

El segundo capítulo tiene por objetivo el estudio de los determinantes de la elección de modo de transporte en viajes de larga distancia. En la práctica, este

ensayo se centra en cómo las características socioeconómicas de los viajeros, los aspectos geográficos y los atributos del viaje condicionan las elecciones entre un grupo de modos de transporte terrestre: vehículo privado, autobús y tren. Los modelos de elección modal han sido ampliamente analizados en la literatura de economía de transporte en el contexto de viajes cotidianos de corta distancia. Sin embargo, los datos de viajes entre ciudades o de larga distancia, no son tan habituales, haciendo de este análisis un tema menos recurrente.

Mientras los viajes por motivos de estudio o trabajo son una actividad diaria para una mayoría de la población, los viajes de larga distancia son menos frecuentes con la excepción de algunas profesiones específicas como los agentes comerciales o los diplomáticos. El conjunto de diferencias entre los viajes de corta y larga distancia incluye distintos tiempos y costes de viaje. Generalmente, se necesita más tiempo y dinero para viajes interurbanos, lo que hace que el viajero estudie más detenidamente la oferta de servicios de transporte. Adicionalmente, los motivos de viaje y la disponibilidad de modos también puede diferir. En la sección destinada a la revisión de la literatura, proporcionamos evidencia actualizada acerca de los efectos que las variables sociodemográficas, geográficas y atributos de viaje tienen sobre la elección de modo de transporte en los desplazamientos de pasajeros interurbanos.

La descripción de los datos provenientes de una encuesta de movilidad (Movilia 2007) juega un papel fundamental en este ensayo. El objetivo de la encuesta es estudiar las características principales de los viajes y viajeros de larga distancia a través de encuestas telefónicas. Más de 19,000 microdatos se analizan y describen en esta parte de la tesis. Este análisis preliminar combinado con la revisión de la literatura ayuda a contextualizar la investigación de dos formas diferentes.

Primero, ayuda a comprobar la consistencia de los resultados obtenidos con el análisis multivariantes. Segundo, nos permitió identificar los puntos débiles de la base de datos.

En general los datos son bastante ricos en información, si bien algunos aspectos espaciales de dicha encuesta son mejorables. Una base de datos ideal debería incluir información acerca de las coordenadas geográficas exactas del origen y destino del viaje. La disponibilidad de estos datos permitiría calcular variables de interés como, por ejemplo, medidas de accesibilidad a los modos de transporte públicos, índices de uso del suelo en origen y destino y los tiempos de acceso y egreso de las estaciones de transporte.

La aproximación empírica utilizada trata de superar estos problemas explotando la estructura jerárquica de los datos. El modelo multinomial permite calcular las probabilidades de escoger un modo de transporte sobre otro alternativo. Una versión multinivel de dicho modelo permite estimar efectos aleatorios asociados a niveles superiores que engloban observaciones individuales. En este caso dichos niveles son las provincias de origen de viaje, por lo tanto las constantes aleatorias recogen información acerca de características no medidas que son compartidas por todos los viajeros con un origen en común. De esta manera, somos capaces de calibrar el modelo teniendo en cuenta la heterogeneidad espacial de los viajeros.

Entre las posibles derivaciones de este trabajo, estoy interesado en estudiar la variación de estos efectos aleatorios con el objetivo de encontrar qué fuerzas causan la heterogeneidad espacial en la elección modal en desplazamientos interurbanos. En concreto, trataré de evaluar los efectos de la accesibilidad a las estaciones de transporte público y el impacto de los distintos usos del suelo. Al

controlar por estas variables, espero poder sacar conclusiones acerca de la heterogeneidad en las preferencias a nivel provincial.

Mis intereses abarcan también el estudio de la movilidad de larga distancia por parte de distintos extractos de población, en especial, de las personas mayores y los desempleados. Una amplia proporción de estos grupos de población apenas realiza viajes de larga distancia, presumiblemente por razones económicas, pero posiblemente por otras razones que permanecen ocultas y sobre las que es posible que se pueda actuar para facilitar la movilidad de estos grupos. A través de la utilización de modelos de doble valla, pretendo investigar acerca de los factores espaciales y sociodemográficos que afectan a la decisión de realizar un viaje o no, y posteriormente modelizar la elección de modo.

El tercer capítulo intenta hacer aportaciones en torno a una cuestión que ha sido ampliamente estudiada en la literatura. En concreto, tratamos de calcular los efectos económicos que las infraestructuras por carretera causan sobre la productividad de las regiones, con especial énfasis en las infraestructuras de transporte. Dos contribuciones fundamentales se aportan a esta extensa literatura. Primero, aplicamos los últimos avances en modelos de econometría espacial con una especificación que nos permite contrastar si existe dependencia espacial en las variables independientes y la variable dependiente, lo cual tradicionalmente se conoce como Modelo Durbin Espacial.

La segunda contribución intenta superar otro problema encontrado durante la revisión de la literatura, la utilización de variables stock en lugar de variables flujo. La medición del capital público y privado como un stock proporciona información acerca de las propiedades cuantitativas de la infraestructura pero no de los servicios que ésta provee. Para minimizar dicho problema, los indicadores

stock son combinados con información acerca de los vehículos que potencialmente utilizan dichas infraestructuras con el fín de obtener una aproximación de los servicios de transporte proporcionados por este tipo de capital pblico.

Aunque el Modelo Durbin Espacial ha sido aplicado a funciones de producción regional en trabajos anteriores, los autores se han centrado habitualmente en su capacidad para tener en cuenta los efectos desbordamiento de las variables independientes, incluyendo rezagos espaciales de dichas variables. En el estudio de los efectos desbordamiento de las infraestructuras de transporte, esta característica del modelo puede tener especial importancia debido a que permite capturar los efectos que se dispersan a través de la red y afectan a regiones vecinas. La especificación del modelo, que incluye la variable dependiente rezagada espacialmente, ha sido habitualmente justificada a través de razonamientos econométricos pero ha permanecido inexplicada desde el punto de vista económico. En este capítulo, tratamos de realizar una aportación en este sentido, sentando las bases teóricas para su inclusión demostrando cómo el uso del capital privado de una región puede estar influido por los ciclos económicos que atraviesan sus regiones vecinas.

El modelo resultante se aplica a una base de datos de las provincias españolas (NUTS-3) en el periodo 1986-2006, durante el cual se realizaron fuertes inversiones en capital público para mejorar la calidad de la red española de transporte por carretera. Como principal resultado empírico obtenemos que un nivel más elevado de servicios de transporte mejora la productividad regional. Sin embargo, y aunque ésta salida del modelo podría ser una razón importante para mantener el ritmo de inversiones para mejorar la red, sugerimos que se tome con precaución. Desde nuestro punto de vista, este resultado es importante al explicar los aumentos de productividad en ese periodo. No obstante, somos conscientes de que la

contribución marginal de siguientes rondas de inversión en ciertos corredores de transporte puede que sea más reducida que las primeras inversiones realizadas.

Resultados adicionales indican que los efectos desbordamiento son considerables. Las estimaciones obtenidas señalan que un incremento en los servicios de transporte, proporcionados por las infraestructuras de carretera en una provincia, aumentan la productividad en las regiones vecinas en una magnitud de aproximadamente la mitad de la mejora que disfruta la región donde las infraestructuras están localizadas. Este resultado proporciona evidencia a la existencia de efectos desbordamiento, que provocarían que los impactos de algunos proyectos de infraestructura públicos no están confinados únicamente en el área donde están construidos. Así, se confirman las características de red del capital público lo cual podría servir de justificación para la coordinación de planificación de infraestructuras de transporte a niveles nacional o supranacional. En el caso de España, los gobiernos provinciales, regionales y nacional comparten esta tarea. Desde el punto de vista metodológico, sugerimos que se contraste la inclusión de los efectos desbordamiento incluso en aquellos análisis de proyectos que no tengan una orientación decididamente inter-regional.

Una versión alternativa de este ensayo en la cual se valora el rol de la accesibilidad en la economía regional ha sido recientemente seleccionada para formar parte de un libro editado por Edward Elgar Publishing. Adicionalmente, seguimos tratando de mejorar el respaldo teórico al modelo Durbin espacial basándonos en los fundamentos de la Nueva Geografía Económica. El objetivo es proporcionar a otros académicos un marco teórico robusto al utilizar esta técnica en las funciones de producción regionales. Una derivación adicional de este trabajo incluye la estimación de una frontera estocástica espacial para estudiar la eficiencia de

las provincias. Finalmente, también esperamos calcular el nivel de utilización del capital público y privado en las provincias españolas.

References

ÁLVAREZ, A., OREA, L. & FERNÁNDEZ, J. (2003). La productividad de las infraestructuras en España. *Papeles de Economía Española*, **95**, 125–136. 67, 71

ÁLVAREZ, A., OREA, L. & ARIAS, C. (2006). Econometric testing of spatial productivity spillovers from public capital. *Hacienda pública española*, **178**, 9–22. 89

ANDERSON, J.E. (1979). Theoretical foundation of the gravity equation. *The American Economic Review*, **69**, 106–116. 11, 12

ANDERSON, J.E. & VAN WINCOOP, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, **93**, 170–192. 11, 12

ARELLANO, M. (1989). A note on the Anderson-Hsiao estimator for panel data. *Economics Letters*, **31**, 337–341. 21

ARELLANO, M. & BOND, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, **58**, 277–297. 21, 23

ASCHAUER, D.A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, **23**, 177–200. 67

AXHAUSEN, K., MADRE, J.L. & TOINT, P., eds. (2003). *Capturing Long-Distance Travel*. Research Studies Press Ltd., Hertfordshire, England, United Kingdom. 33

BALTAGI, B. (2008). *Econometric Analysis of Panel Data*. John Wiley and Sons, 4th edn. 19, 20, 87

BAXTER, M. & FARR, D.D. (2005). Variable capital utilization and international business cycles. *Journal of International Economics*, **65**, 335–347. 73

BEL, G. (1997). Changes in travel time across modes and its impact on the demand for inter-urban rail travel. *Transportation Research Part E: Logistics and Transportation Review*, **33**, 43–52. 36

BEN-AKIVA, M. & BIERLAIRE, M. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, 1985. 31, 50

BERGSTRAND, J. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The review of economics and statistics*, **67**, 474–481. 12

BERGSTRAND, J. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The review of economics and statistics*, **71**, 143–153. 12

BHAT, C.R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, **31**, 34–48. 34

BHAT, C.R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, **32**, 495–507. 31

BHAT, C.R. (2000). A multi-level cross-classified model for discrete response variables. *Transportation Research Part B: Methodological*, **34**, 567–582. 51, 52

BLUNDELL, R. & BOND, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**, 115–143. 21

BOARNET, M.G. (1998). Spillovers and the Locational Effects of Public Infrastructure. *Journal of Regional Science*, **38**, 381–400. 69, 70, 76

BRICKA, S. (2001). Variations in Long-Distance Travel. *Transportation research circular*, 197–206. 36

BUEHLER, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography*, **19**, 644–657. 33, 35

BUN, M.J. & KLAASSEN, F.J. (2002). The Importance of Dynamics in Panel Gravity Models of Trade. *SSRN Electronic Journal*. 13

BUN, M.J.G. & KLAASSEN, F.J.G.M. (2007). The Euro Effect on Trade is not as Large as Commonly Thought. *Oxford Bulletin of Economics and Statistics*, **69**, 473–496. 14, 20

BURGER, M., VAN OORT, F. & LINDERS, G.J. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis*, **4**, 167–190. 19

BURNSIDE, C. & EICHENBAUM, M. (1996). Factor hoarding and the propagation of business cycles shocks. *The American Economic Review*, **86**, 1154–1174. 73

CANTOS, P., GUMBAUIALBERT, M. & MAUDOS, J. (2005). Transport infrastructures, spillover effects and regional growth: evidence of the Spanish case. *Transport Reviews*, **25**, 25–50. 12

CHANDRA, A. & THOMPSON, E. (2000). Does public infrastructure affect economic activity? *Regional Science and Urban Economics*, **30**, 457–490. 68

CHOW, J.Y.J., YANG, C.H. & REGAN, A.C. (2010). State-of-the art of freight forecast modeling: lessons learned and the road ahead. *Transportation*, **37**, 1011–1030. 13

COHEN, J.P. & MORRISON PAUL, C.J. (2003). Airport infrastructure spillovers in a network system. *Journal of Urban Economics*, **54**, 459–473. 69

COHEN, J.P. & MORRISON PAUL, C.J. (2004). Public Infrastructure Investment, Interstate Spatial Spillovers, and Manufacturing Costs. *Review of Economics and Statistics*, **86**, 551–560. 68

DE BENEDICTIS, L. & VICARELLI, C. (2005). Trade Potentials in Gravity Panel Data Models. *Topics in Economic Analysis & Policy*, **5**. 13, 14, 21

DE WITTE, A., HOLLEVOET, J., DOBRUSZKES, F., HUBERT, M. & MACHARIS, C. (2013). Linking modal choice to motility: A comprehensive review. *Transportation Research Part A: Policy and Practice*, **49**, 329–341. 33, 35

DELGADO, M.J. & ÁLVAREZ, I. (2007). Network infrastructure spillover in private productive sectors: evidence from Spanish high capacity roads. *Applied Economics*, **39**, 1583–1597. 70, 71, 89

EGGER, P. (2001). European exports and outward foreign direct investment: A dynamic panel data approach. *Weltwirtschaftliches Archiv*, **3**, 427–449. 12

EGGER, P. & PFAFFERMAYR, M. (2003). The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects. *Empirical Economics*, **28**, 571–580. 9, 13

EGGER, P. & PFAFFERMAYR, M. (2005). The Determinants of Intrafirm Trade: In Search for Export-Import Magnification Effects. *Review of World Economics*, **141**, 648–669. 12

EICHENGREEN, B. & IRWIN, D. (1998). The role of history in bilateral trade flows. In J. A. Frankel, ed., *The regionalization of the world economy*, 33–57, University of Chicago Press, Chicago. 9, 13

ELHORST, J. (2010a). Spatial panel data models. In M.M. Fischer & A. Getis, eds., *Handbook of applied spatial analysis*, 377–407, Springer-Verlag, Berlin. 87

ELHORST, J.P. (2010b). Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis*, **5**, 9–28. 78

FERNALD, J.G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review*, **89**, 619–638. 68, 70, 75, 90

FIDRMUC, J. (2009). Gravity models in integrated panels. *Empirical Economics*, **37**, 435–446. 12, 13

FORKENBROCK, D.J. & FOSTER, N.S. (1990). Economic benefits of a corridor highway investment. *Transportation Research Part A: General*, **24**, 303–312. 74

GAJANAN, S. & MALHOTRA, D. (2007). Measures of capacity utilization and its determinants: a study of Indian manufacturing. *Applied Economics*, **39**, 765–776. 73

GARCIA-MILA, T., MCGUIRE, T.J. & PORTER, R.H. (1996). The Effect of Public Capital in State-Level Production Functions Reconsidered. *The Review of Economics and Statistics*, **78**, 177–180. 68

GARMENDIA, M., UREÑA, J. & CORONADO, J. (2011). Long-distance trips in a sparsely populated region: The impact of high-speed infrastructures. *Journal of Transport Geography*, **19**, 537–551. 35

GAROFALO, G.A. & MALHOTRA, D.M. (1997). Regional Measures of Capacity Utilization in the 1980s. *Review of Economics and Statistics*, **79**, 415–421. 73

GEORGGI, N.L. & PENDYALA, R.M. (2001). Analysis of long-distance travel behavior of the elderly and low income. *Transportation Research Board*, **E-C026**, 121–150. 34, 35, 53

GLICK, R. & ROSE, A.K. (2002). Does a currency union affect trade? The time-series evidence. *European Economic Review*, **46**, 1125–1151. 12

GRAMLICH, E. (1994). Infrastructure investment: A review essay. *Journal of economic literature*, **32**, 1176–1196. 67, 68

GRILLI, L. & RAMPICHINI, C. (2007). A multilevel multinomial logit model for the analysis of graduates skills. *Statistical Methods and Applications*, **16**, 381–393. 48, 59

HARKER, P.T. (1985). The state of the art in the predictive analysis of freight transport systems. *Transport Reviews*, **5**, 143–164. 10

HARKER, P.T. (1987). *Predicting intercity freight flows. Topics on Transportation*. VNU Science Press. 10

HARRIS, M. & MÁTYÁS, L. (1998). *The econometrics of gravity models*. Melbourne Institute, Working papers 03. 13

HAUSMAN, J. & MCFADDEN, D. (1984). Specification Tests for the Multinomial Logit Model. *Econometrica*, **52**, 1219. 50

HELPMAN, E. & KRUGMAN, P.R. (1985). *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Press, MIT, Cambridge, MA. 11

HOLTZ-EAKIN, D. & SCHWARTZ, A.E. (1995). Spatial productivity spillovers from public infrastructure: Evidence from state highways. *International Tax and Public Finance*, **2**, 459–468. 68, 69, 89

HONG, J., SHEN, Q. & ZHANG, L. (2013). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation*. 38

HULTEN, C. (2004). Transportation infrastructure, productivity and externalities. In *132nd Round Table of the European Conference of Ministers of Transport*, OECD/EMCT Transport Research Center, Paris, France. 68

JONES, K. & DUNCAN, C. (1996). People and places: the multilevel model general framework for the quantitative analysis of geographical data. In P. Longley & M. Batty, eds., *Spatial Analysis: Modelling in a GIS Environment*, chap. 5, 79–104, GeoInformation International, Cambridge. 52

JONG, G., VIERTH, I., TAVASSZY, L. & BEN-AKIVA, M. (2012). Recent developments in national and international freight transport models within Europe. *Transportation*, **40**, 347–371. 12, 13

JUNG, B. (2009). Adjustment Dynamics of Bilateral Trade Flows: Theory and Evidence. *Swiss Journal of Economics and Statistics (SJES)*, **145**, 421–442. 21

REFERENCES

KOPPELMAN, F.S. & SETHI, V. (2005). Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior. *Transportation Research Part B: Methodological*, **39**, 825–853. 36

LAPPARENT, M.D., AXHAUSEN, K. & FREI, A. (2013). Long distance mode choice and distributions of values of travel time savings in three European countries. *European Transport \ Trasporti Europei*, 1–7. 36

LESAGE, J.P. & PACE, R.K. (2009). *Introduction to Spatial Econometrics*. Taylor and Francis Group, CRC Press, Boca Raton, FL. 78, 83

LIMTANAKOOL, N., DIJST, M. & SCHWANEN, T. (2006). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer-distance trips. *Journal of Transport Geography*, **14**, 327–341. 31, 34, 35, 53

LLANO, C., ESTEBAN, A., PEREZ, J. & PULIDO, A. (2010). Opening the Interregional Trade "Black Box": The C-Intereg Database for the Spanish Economy (1995–2005). *International Regional Science Review*, **33**, 302–337. 9

MALLETT, W. (1999). Long-Distance Travel by Women: Results from the 1995 American Travel Survey. *Transportation Research Board*, **1693**, 71–78. 34

MALLETT, W. (2001). Long-distance travel by low-income households. *Transportation Research Board*, **E-C026**, 169–177. 34

MANSKI, C. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, **60**, 531–542. 74, 78

MANSKI, C.F. (1977). The Structure of Random Utility Models. *Theory and Decision*, **8**, 229–254. 48

MÁRQUEZ, M.a. & HEWINGS, G.J. (2003). Geographical competition between regional economies: The case of Spain. *The Annals of Regional Science*, **37**, 559–580. 80

MÁRQUEZ, M.a., RAMAJO, J. & HEWINGS, G.J.D. (2010). A spatio-temporal econometric model of regional growth in Spain. *Journal of Geographical Systems*, **12**, 207–226. 87

MARTÍN, J.C. & NOMBELA, G. (2007). Microeconomic impacts of investments in high speed trains in Spain. *The Annals of Regional Science*, **41**, 715–733. 35, 36

MARTÍNEZ-ZARZOSO, I. & NOWAK-LEHMANN D., F. (2003). Augmented gravity model: An empirical application to Mercosur-European Union trade flows. *Journal of Applied Economics*, **VI**, 291–316. 13

MARTÍNEZ-ZARZOSO, I., FELICITAS, N.L.D. & HORSEWOOD, N. (2009a). Are regional trading agreements beneficial? *The North American Journal of Economics and Finance*, **20**, 46–65. 13

MARTÍNEZ-ZARZOSO, I., NOWAK-LEHMANN D., F., KLASEN, S. & LARCH, M. (2009b). Does German Development Aid Promote German Exports? *German Economic Review*, **10**, 317–338. 13, 14

MAS, M., MAUDOS, J., PÉREZ, F. & URIEL, E. (1996). Infrastructures and Productivity in the Spanish Regions. *Regional Studies*, **30**, 641–649. 89

MÁS, M., PÉREZ, F. & URIEL, E. (2011). *El stock y los servicios de capital en España y su distribución territorial y sectorial (1964-2010)*. Fundación BBVA, Working papers. 81

McFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka, ed., *Frontiers in Econometrics*, 105–142, Academic Press. 51

McKINNON, A. (2010). The potential of economic incentives to reduce CO2 emissions from goods transport. *DILF Orientering*, **1**, 28–30. 25

MERA, K. (1973). Regional production functions and social overhead capital: An analysis of the Japanese case. *Regional and Urban Economics*, **3**, 157–185. 67

MERCADO, R. & PÁEZ, A. (2009). Determinants of distance traveled with a focus on the elderly: a multilevel analysis in the Hamilton CMA, Canada. *Journal of Transport Geography*, **17**, 65–76. 59

MIKELBANK, B.A. & JACKSON, R.W. (2000). The Role of Space in Public Capital Research. *International Regional Science Review*, **23**, 235–258. 70

MOECKEL, R., FUSSELL, R. & DONNELLY, R. (2013). Mode Choice Modeling for Long-Distance Travel. *Transportation Research Board*, 1–16. 35

MORENO, R. & LOPEZ-BAZO, E. (2007). Returns to Local and Transport Infrastructure under Regional Spillovers. *International Regional Science Review*, **30**, 47–71. 69, 71, 76, 89

MUNNELL, A. (1990). Why has productivity growth declined? Productivity and public investment. *New England Economic Review*. 68

NARDIS, S.D. & VICARELLI, C. (2003). *The Impact of the Euro on Trade*. ENEPRI, Working Paper. 13

NICKELL, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, **49**, 1417–1426. 20

OF TRANSPORTATION, O. & QUALITY, A. (2006). *Greenhouse gas emissions from the U.S. transportation sector, 1990-2003*. U.S. Environmental Protection Agency, , Washington, D.C. 24

OOSTERHAVEN, J. & KNAAP, T. (2003). Spatial economic impacts of transport infrastructure investments. In A. Pearman, P. Mackie & J. Nellthorp, eds., *Transport Projects, Programmes and Policies: Evaluation Needs and Capabilities*, 87–97, Ashgate Publishing Company, Aldershot. 75

OUM, T. (1989). Alternative demand models and their elasticity estimates. *Journal of Transport Economics and policy*, **23**, 163–187. 10, 16

PÁEZ, A. (2006). Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. *Journal of Transport Geography*, **14**, 167–176. 36

PEREIRA, A. & ANDRAZ, J. (2010). On the economic effects of public infrastructure investment: A survey of the international evidence. 67, 68

PFAHLER, W., HOFMANN, U. & BONTE, W. (1996). Does extra public infrastructure capital matter? *Public Finance Analysis*, **53**, 68–112. 72

RABE-HESKETH, S. & SKRONDAL, A. (2012). *Multilevel and Longitudinal Modeling Using Stata*. Press, Stata, Lakeway Drive, TX, 3rd edn. 50, 59

RABE-HESKETH, S., SKRONDAL, A. & PICKLES, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, **128**, 301–323. 51

REPHANN, T. & ISSERMAN, A. (1994). New highways as economic development tools: An evaluation using quasi-experimental matching methods. *Regional Science and Urban Economics*, **24**, 723–751. 71

RIETVELD, P. (1994). Spatial economic impacts of transport infrastructure supply. *Transportation Research Part A: Policy and Practice*, **28**, 329–341. 75, 76

ROODMAN, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, **9**, 86–136. 21

ROSE, A.K. (2000). One money, one market: the effect of common currencies on trade. *Economic Policy*, **15**, 7–46. 12

SCHAFER, A. (1998). The global demand for motorized mobility. *Transportation Research Part A: Policy and Practice*, **32**, 455–477. 30

SCHREYER, P. (2001). The OECD productivity manual: a guide to the measurement of industry-level and aggregate productivity. *International Productivity Monitor*, **2**, 37–51. 82

SCHREYER, P., BIGNON, P. & DUPONT, J. (2003). *OECD capital services estimates: Methodology and a first set of results*. OECD Statistics Working Papers, OECD Publishing. 82

SKRONDAL, A. & RABE-HESKETH, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, **68**, 267–287. 48, 51

SPANISH MINISTRY OF ENVIRONMENT, R. & AREAS, S. (2009). *Inventario de emisiones de gases de efecto invernadero de España*. Spanish Government. 26

TOBLER, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234–240. 52

TONG, T., YU, T.H.E., CHO, S.H., JENSEN, K. & DE LA TORRE UGARTE, D. (2013). Evaluating the spatial spillover effects of transportation infrastructure on agricultural output across the United States. *Journal of Transport Geography*, **30**, 47–55. 78

WESTERLUND, J. & WILHELMSSON, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, **43**, 641–649. 19

WHALEN, K.E., PÁEZ, A. & CARRASCO, J.a. (2013). Mode choice of university students commuting to schooland the role of active travel. *Journal of Transport Geography*, **31**, 132–142. 39

WINSTON, C. (1983). The demand for freight transportation: models and applications. *Transportation Research Part A: General*, **17**, 419–427. 10

YU, N., DE JONG, M., STORM, S. & MI, J. (2013). Spatial spillover effects of transport infrastructure: evidence from Chinese regions. *Journal of Transport Geography*, **28**, 56–66. 78