Econometric modeling of long-distance travel

Pelayo Arbués José F. Baños Matías Mayor
Patricia Suárez

July 3, 2014

Abstract

The main goal of this study is to determine the effects of socio-economic, demographic and land-use factors on long distance travel. Long distance trips are not very frequent activities causing that a large share of the survey respondents reports zero kilometers travelled. Double Hurdle Model, Infrequent Purchase Model and Heckman selection Models are estimated employing data from the Spanish National Travel Survey. The dependent variable, distance travelled in trips longer than 50 km., is a function of gender, age, income, size of municipality, region of residence, mode of transport and purpose of the trip. The results indicate that zeroes reported in the survey are the result of corner solutions, which allow us making some policy recommendation about general designing of long-distance surveys.

1 Introduction

Long distance travel is attracting the attention of scholars and policy makers due to its rapid increase over recent decades. Technological innovations in transportation modes and recent economic growth are among the main factors contributing to make intercity travel more frequent. The development of high speed railway infrastructure, popularity of car ownership and greater accessibility to well connected airports allow the human activity to spread on larger spatial scales. In this sense, long distance travel figures are important indicators of economic and social integration making stronger links between distant urban regions. These interregional interactions shape the future of urban regions through higher mobility of the population, productive resources and companies.

Although intercity trips are less frequent activities than commuting and only represent a small share of all trips, a large share of total mileage travelled is long distance. Increases in long distance mobility have economical, social and environmental implications. Long distance travel increases the opportunities of regions to be better connected to trade and economic networks making clusters of industries more likely to occur between distant areas. Rural areas benefit from private and public facilities located far away making accessibility to services higher. Remote locations also seem more attractive in terms of recreational activities raising demand for tourism in these areas. On the other hand, higher long distance mobility also imposes negative externalities to society such as congestion or higher risks of traffic accidents. From an environmental perspective, long distance travel is through their high demand for fuel consumption and contribution to global greenhouse emissions.

These implications make long distance travel a greater concern to policy makers and companies in all industries, but specially in tourism and transportation sectors. It is of interest to have a better understanding of determinants of travel behaviour for several reasons. From a policy perspective, debate on proposed increases on fuel taxes is usually shaped by distributional issues, both regarding income groups and regions. Knowledge on the profile of long distance travellers can help designing price discrimination schemes of transport companies. This study contributes the scarce existing literature on long distance travel by increasing the empirical evidence on the factors behind the raise of long distance travel demand. In particular, the objective of this study is to examine the determinants of long distance travel in Spain using a large micro data set, the 2007 Spanish Mobility Survey (Movilia).

A final motivation for this study is to present an alternative econometric approach to analyse long distance travel determinants. As expected, a large share of the survey respondents reports zero kilometers travelled. In order to reduce the number of zeroes, survey designers have usually increase the duration of reporting periods. However, this solution also implies that respondents have more difficulties recalling events that occurred in the previous weeks and require a larger amount of resources to control and collect the data. Thus, reporting period is a fundamental characteristic of long-distance travel survey. In this study, we estimate a set of models including Double Hurdle Model, the Heckman Selection Model and the Infrequent Purchase Model. The application of these different techniques solves a econometric problem, but also allows us to gain some insights in the origin of the zeroes. According to the statistical tests, in this case Double Hurdle model is superior to Infrequent Purchase model, indicating that zeroes reported in the survey are the result of corner solutions and not of infrequent events. This result shows that the reporting period considered in this survey (3 months) is adequate to capture long-distance trips.

The structure of the remainder of the paper is as follows. In the following section, the relevant literature is briefly reviewed to place the paper within the context of existing studies. Section 3 describes information on long distance travel in Spain contained in Movilia 2007 database. The methodological part of the paper is presented in Section 4 where the econometric model and the results are discussed. Finally, some conclusions are drawn and directions for future research are sketched.

2 Previous studies on long-distance travel

National Mobility Surveys have usually collected data only on short distance trips, with an emphasis on commuting to work or study places, ignoring intercity transportation of passengers. This is probably caused by difficulties found in measuring long distance behavior as noted by Axhausen (1999). As it has been commented in Section 1 intercity trips are low frequent events which affects the design of long distance surveys in the duration of reporting periods. Moreover, long distance travel is defined as those trips longer than a distance threshold. Even threshold definition is a common problem and there is a lack of consensus in the literature with this lower limit ranging from 50 to 100 kilometers according to Dargay and Clark (2012). These are some of the reasons that have made short distance travel literature much more widespread than scientific works on long distance travel.

Among the studies dealing with intercity trips, it is worth noting that a great share of scientific work has paid special attention to a traditional topic on travel demand literature: mode choice decisions [1]. Other works have developed complete travel demand models applied to intercity trips including generation, distribution, mode choice and assignment (Erhardt et al., 2007; Baik et al., 2008).

In this literature review we focus on those relatively few studies concentrated on the determinants of long distance travel. We follow the work of Dargay and Clark (2012) where it is provided an up-to-date review of literature concerning long distance travel demand. Some of these papers concentrate in different groups of the population differentiating by demographics or socio-economic status.

Relying on data about trips longer than 100 miles coming from 1995 American Travel Survey, Mallett (1999) studies gender differences in long-distance travel. Women made fewer intercity trips than men and this was mostly caused by the fact that female respondents made less business travel, and also because of womens lower income and lower employment rates. Men also reported more trip making for outdoor recreation, which is more limited in the case of women as they are more likely to be caregivers of children. Georggi and Pendyala (2001) provide an analysis of long distance travel behaviour of two key segments of market: the elderly and the low income also using data of 1995 American Travel Survey. As expected, the authors found a loss in long distance mobility in terms of travel frequency in those individuals aged above 65 years, and specially in those above 75 years. Surprisingly, average trip length seem to increase with age. The comparison between income groups show that there are clear differences in long distance trip making regarding the frequency and the length of the trip. On average, lower income groups make half the trips (4 or 5) of those made by higher income households. Trips of higher income population are, on average, an 80 % longer than those of the low income group. These differences are explained in part by the fact that low income group makes less business and, also, less recreational trips and because they are less likely to use air transportation than people with higher disposable income. Similar results were found in a parallel analysis of the same survey conducted by Mallett (2001).

Collia et al. (2003) analyze trips with a length over 50 miles using data

[1] See for instance Bhat (1997) Mallett (1999) Georggi and Pendyala (2001) Limtanakool et al. (2006b) and Arbués et al. (2013).

from 2001 National Household Travel Survey. More than two thirds of the total intercity trips were taken by adults under 65, while only an 8 % of long distance trips were made by the elder. Cross tabulation of the data also indicates gender differences in trip making for all age groups, with men enjoying more frequently long distance trips than women. When comparing trip length by age groups, these authors found that elderly tend to take longer trips.

Using multivariate analysis, Mokhtarian et al. (2001) explore factors apart from demographics or socioeconomic background to explain distance travel. Some characteristics of the population such as personality, lifestyle or travel-related attitudes are measured through different variables which are found to have an impact on distance travelled. Authors conclude that some segments of the population have a desire to travel for its own sake, instead of the traditional approach of travel as a derived demand. Income is found to have a positive and significant impact, while increasing age seems to reduce the demand of airplane demand. As usual, differences in travel distance are found by gender, with women travelling less than men for work-related purposes but farther for long-distance entertainment and airplane.

In Limtanakool et al. (2006a) the authors study the propensity of participate in medium and long-distance travel In United Kingdom and the Netherlands. Relying on binary logit models applied to different segments of the sample, the overall results show how in leisure travel gender and household composition hardly affect the decision of travelling long-distance as males and females within a household usually undertake this trips jointly. In business travel and commute, gender and income differences have been found, where males, full time workers and higher income are more likely to travel long. Land use related factors also play a significant role, with citizens of metropolitan and high density populated areas travel shorter distances as they can fin a wider range of services within a relatively shorter distance from home that their counterparts residing in low density rural areas. Frändberg and Vilhelmson (2003) offer an account of long-distance travelling for Swedish with a focus on international mobility, estimating regression models on trips frequency to other countries as a function of sex, region of origin, income and other personal variables.

Dargay and Clark (2012) study the socio-economic, demographic and geographic determinants influencing long distance travel in Great Britain in the period 1995-2006. Econometric results of models estimated by weighted least squares indicate that distance travelled is lower for women, while it is greater for those under 60 years old than elders. Labour situation also affects distance travelled with students and workers travelling longer than other groups. Authors also include length of residence, and the estimates show that the longer living the in same address the less distance travelled. Household composition also affects distance, and seems to support the idea that adults living on their own travel longer distances than those adults sharing with other adults and children. Regarding geographic factors, Londoners travel least in terms of distance than residents in other british regions, while people living in rural areas travel further distances.

In Holz-Rau et al. (2014), the research question is focused on spatial differences in travel behaviour and a comparison of factors affecting distance travelled in daily and long-distance trips. Socio-economic and demographic variables are introduced in the mode using interactions between age, labor situation and gender. Using a Heckman model explanatory variables are found to similarly affect distance travelled in both sort of trips although income and level of education affects more the distance decision in long distance than in daily trips. Concerning geographic factors, participation and distance travelled increase with municipality size and population density in long distance trips, while opposite signs were found for daily commuting.

Overall, this literature review has shown that socio-economic, demographic and geographic factors have been previously found to have significant impacts on distance travelled in intercity trips. In regard to socio-economic and demographic factors, some consensus have been found in all studies reviewed. Opposite results have appeared when studying geographic determinants. While in Dargay and Clark (2012) and Limtanakool et al. (2006b) residents in high-density and high-populated metropolitan areas travel less distance, opposite results are found in Holz-Rau et al. (2014).

3 Long-distance travel in Spain: Movilia 2007

The main objective of this work is to study the determinants of long distance travel in Spain. To this end, we use a mobility survey (MOVILIA 2007) that comes from the Statistical Office of the Ministry of Public Works [2]. This survey provides information that increases knowledge about the mobility patterns of the resident population in private households in Spain, their characteristics and determinants. The basis for the selection of households has been the Municipal Register and long distance trips are defined as those with a length over 50 kms. It is important to note that some groups are not included in this survey because they are not living in private households in Spain. It is the case of tourists, illegal immigrants and people whose residence is established in institutions or collective households. Information on long distance trips is collected by telephone interview questionnaire for a 4 week period between February 2007 and January 2008. The size of the sample and its distribution has been designed to guarantee the representative of the data at provincial level. Consequently, the figures about the number of trips and the distribution by mode and purpose are statistically representative at provincial and metropolitan level.

Table 1 shows different information on the composition of the database and cross tabulation between the decision to travel and variables included in the survey. We can observe how there is a similar share of the total 50,730 observations of male respondents (48.6%) and females (51.4%). Low and medium income individuals represent almost an 83% of total observations. We can also observe how the majority of the surveyed population live in not metropolitan

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

Table 1: Cross-tabulation of travel decision

	Indicates if the person travels or not				
	No travel		**Travel**		**Total**
	Col %	row %	Col %	row %	Col %
Gender					
Female	56.3	51.5	47.0	48.5	51.4
Male	43.7	42.2	53.0	57.8	48.6
Age					
15 to 29	15.4	34.7	25.6	65.3	20.8
30 to 39	13.8	38.2	19.8	61.8	17.0
40 to 49	15.5	38.4	22.0	61.6	18.9
50 to 64	21.9	46.3	22.5	53.7	22.2
65 or more	33.4	74.6	10.1	25.4	21.0
Income					
Low	62.6	58.8	38.8	41.2	50.0
Medium	28.7	41.1	36.4	58.9	32.8
High	8.7	23.7	24.8	76.3	17.2
Area					
Not metropolitan	68.8	46.1	71.2	53.9	70.1
Metropolitan area	31.2	49.0	28.8	51.0	29.9
Municipality size					
Less than 10000	36.8	49.0	33.9	51.0	35.3
10000 to 50000	25.2	48.3	23.9	51.7	24.5
50000 to 500000	32.5	44.5	36.0	55.5	34.3
More than 500000	5.5	44.0	6.2	56.0	5.9
Region of residence					
Madrid	3.3	48.2	3.1	51.8	3.2
Galicia	9.7	51.9	8.0	48.1	8.8
Asturias	2.9	55.6	2.1	44.4	2.5
Cantabria	2.8	53.1	2.2	46.9	2.4
Euskadi	7.4	53.1	5.8	46.9	6.5
Rioja	1.7	46.7	1.7	53.3	1.7
Navarra	2.1	39.0	2.9	61.0	2.5
Aragon	5.7	42.7	6.8	57.3	6.3
Catalonia	8.7	46.4	8.9	53.6	8.8
Com Valenciana	8.2	53.8	6.2	46.2	7.2
Murcia	2.7	56.5	1.8	43.5	2.2
Castilla y Leon	15.9	42.1	19.4	57.9	17.7
Extremadura	3.1	39.1	4.2	60.9	3.7
Castilla-La Mancha	7.7	38.4	10.9	61.6	9.4
Andalucia	18.2	50.2	16.0	49.8	17.0
Sample size	23,833		26,897		50,730
Total	100.0	47.0	100.0	53.0	100.0

areas (70%) and in municipalities with standing population below 500000 inhabitants.

In the analysis of the travel decision it is important to note that the percentage of individuals who make long-distance trips (53%) is slightly higher than the individuals who did not make any trip with a length over 50 kilometers in the previous three months (47%). When these global figures are analyzed by gender, we observe that the percentage of men who do long distance travel is greater than the womens percentage (57.8% and 51.5% respectively).

Exploratory analysis also indicates important differences by age; we find that the 74.6% of individuals with 65 or older do not travel whereas in the rest of the age groups, percentage of non-travel individuals is lower than 50%. For instance, this percentage of non-participation is minimum for individuals between 15-29. These values suggest an inverse relation between age and the travel decision.

We constructed a proxy for a personal income variable relying on educational level and labor information. 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 we expected, individuals with low income mainly do not travel (58.8%) while in the rest of income categories (medium and high income) the proportion of traveler individuals is higher, especially individuals with high income (76.3%).

Regarding the geographical variables, cross tabulation indicates that the percentage of individuals who decide to travel is slightly higher than 50% and the size of the municipality exerts positive effects in terms of participation. It is also possible to analyze the distribution of individuals who travel and do not travel by region of residence of the respondent. In some regions like Castilla-La Mancha, Navarra or Extremadura, the percentage of individuals who travel is equal to 61.6%, 61% and 60.9%, respectively. Note, however, that in Murcia on average 56.5% of the individuals do not travel in the reference period.

Distance travelled is analyzed in Table 2 which includes different descriptive statistics for the sample restricted to observations representing a long distance trip. The mean distance is slightly higher for female travellers than men, while the median of the distance travelled is higher for male trip makers although these differences are very small. An analysis of distance travelled by age groups show that individuals in the categories representing younger ages travel shorter distances than elders. According to this statistics, the maximum is reached for those individual in the group between 50 and 64 years, while distance declines for seniors above 65. The residence of traveller also seem to affect distance travelled. In particular, length of the trip is higher on average for those individuals living in metropolitan areas. Residents of the Madrid Autonomous Community travel further than habitants of other regions, with a mean distance value of more than 200 kilometers, followed by travellers from Euskadi and La Rioja. On average, people living in Aragn are the travellers making shorter long distance trips (130.1 km).

This section has provided some insights in long distance travel in Spain relying on an exploratory analysis. The estimation of econometric models presented in next section, allows us to make inference of the effects of socio-economic, demographic and geographical variables on long-distance travel.

Table 2: Summary of distance variable in kms for travellers (N= 26,879)

	Mean distance	**Sd** distance	**Median** distance
Gender			
Female	151.2	191.2	89.0
Male	150.7	182.2	90.0
Age			
15 to 29	144.0	172.6	89.0
30 to 39	148.6	180.0	89.0
40 to 49	153.0	183.5	91.0
50 to 64	157.3	195.2	91.0
65 or more	154.6	216.7	82.0
Income			
Low	148.8	195.4	86.0
Medium	147.4	177.4	90.0
High	159.5	184.9	96.0
Area			
Not metropolitan	143.3	175.4	87.0
Metropolitan area	169.7	210.2	96.0
Municipality size			
Less than 10000	133.5	170.1	83.0
10000 to 50000	150.1	181.2	89.0
50000 to 500000	164.0	203.3	93.0
More than 500000	173.6	181.5	104.0
Region of residence			
Madrid	201.8	205.4	123.0
Galicia	141.9	203.7	78.0
Asturias	171.0	281.1	82.0
Cantabria	166.7	225.3	90.5
Euskadi	187.5	243.8	90.0
Rioja	182.9	258.9	102.0
Navarra	151.7	215.9	78.0
Aragon	130.1	126.2	93.0
Catalonia	153.6	223.4	90.0
Com Valenciana	150.3	158.9	86.0
Murcia	145.7	168.9	73.0
Castilla y Leon	150.6	177.3	90.0
Extremadura	138.4	136.6	89.0
Castilla-La Mancha	138.3	147.8	90.0
Andalucia	144.5	165.0	90.0
Mode of transport			
Car	127.9	119.7	86.0
Bus	157.3	149.3	97.0
Train	200.8	172.9	126.0
Plane	875.7	570.4	661.0
Other	149.7	154.9	84.0
Trip purpose			
Pleasure	171.6	212.8	100.0
Business	152.2	158.1	90.0
Second residency	106.9	119.2	69.0
Other purpose	97.0	86.3	74.0
Total	150.9	186.5	89.0

4 Econometric analysis

4.1 Model specification

In the empirical part of the study we aim to estimate the contribution of different factors to the long-distance travelled by Spanish population. The dependent variable, length of the trip measured in kilometers, is censored since there are zero for a large fraction of the observations as reported in the previous section. Hence, a basic ordinary least square (OLS) regression would yield biased paramenter estimates. In this situation, different methodologies are usually applied to deal with this problem.

Zero observations may appear in any microeconomic data as a consecuence of different reasons Gibson and Kim (2011): an individual is a genuine non-consumer of the transport services, due to either abstention (for any sort of preferences) or to a corner solution where the individual is a potential consumer of the service but cannot afford it at the current level of income and prices. The case of abstention can be modeled using Craggs model, which is compounded by a Probit equation that accounts for the participation decision in the market and a Tobit model that explains the quantity of the service consumed or the expenditures. Conventionally, a Tobit model is applied to corner solutions where the demand for the service is censored from below. Besides, even when the individual usually consumes the analyzed (goods or) services, the existence of zeroes may be due to infrequently purchased goods if the survey period is shorter than the consumer of the good. In this context, the design of the survey becomes an important theme since several waves of the survey may reduce the second source of zeroes. A more detailed explanation can be found in Newman et al. (2003) and in the references given therein. In this situation the appropriate specification would be the Infrequency of Purchase Model.

The standard Tobit model (Tobin, 1958) has been frequently used to estimate demand relationships with limited dependent variables. However, the Tobit model can be very restrictive because, apart from econometric reasons, it assumes that zero consumption observations alone arise from corner solutions generated by a constrained budget. To overcome this restrictive assumption, different double-hurdle models, as used in the present analysis, have been proposed in the economic literature.

An alternative commonly known as the Double-Hurdle Model (DHM), originally formulated by Cragg (1971), supposes that individuals make their consumption decisions in two steps, each of which is determined by a different set of explanatory variables. Accordingly, two separate hurdles (the decision about whether or not to consume and the quantity consumed) must be passed before, to observe a positive level of consumption. Thus, the model contains a participation equation,

$$d^* = \alpha' z + u \qquad u \sim \mathbf{N}(0,1) \tag{1}$$

where the latent variable d^* represents, in our case, the decision to make long-distance trips as a linear function of first-hurdle regressors (z), and a consumption equation,

$$y^* = \beta' x + v \qquad v \sim \mathbf{N}(0,1) \tag{2}$$

characterized by the latent variable y^*, the distance travelled, as a linear function of second-hurdle regressors (x). The observed distance travel, y, is such that

$$y = \begin{cases} y^*, & \text{if } d^* > 0 \text{ and } y^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Assuming that the error terms u and v can be correlated allows for the possibility that travel decision and distance travel are simultaneously taken. Then, it is considered that u and v are distributed as a bivariate normal,

$$(u,v) \sim \mathbf{BVN}(0,\Sigma), \qquad \mathbf{\Sigma} = \begin{pmatrix} 0 & \sigma\rho \\ \sigma\rho & \sigma^2 \end{pmatrix} \tag{4}$$

where ρ is the correlation coefficient.

Estimating this bivariate model requires the maximization of the following likelihood equation,

$$L = \prod_0 \left[1 - \Phi\left(\alpha' z, \frac{\beta' x}{\sigma}, \rho\right)\right] \prod_+ \left[\Phi\left(\frac{\alpha' z + \frac{\rho}{\sigma}(y - \beta' x)}{\sqrt{1-\rho^2}}\right) \frac{1}{\sigma}\phi\left(\frac{y - \beta' x}{\sigma}\right)\right] \tag{5}$$

where Φ and ϕ are the standard normal cumulative and density functions, respectively.

If considered the independent double-hurdle model (independence between u and v), the likelihood reduces to

$$L = \prod_0 \left[1 - \Phi\left(\alpha' z, \frac{\beta' x}{\sigma}\right)\right] \prod_+ \left[\Phi\left(\alpha' z\right) \frac{1}{\sigma}\phi\left(\frac{y - \beta' x}{\sigma}\right)\right] \tag{6}$$

It should be noted that the standard Tobit model is a nested version of this Cragg model, when $\Phi(\alpha' z) = 1$.

An alternative to the double-hurdle model is to considerer that the participation decision dominates the quantity decision, as in the Heckman selection model (Heckman, 1979). In this case, the corresponding likelihood function can be written as

$$L = \prod_0 \left[1 - \Phi\left(\alpha' z\right)\right] \prod_+ \left[\Phi\left(\frac{\alpha' z + \frac{\rho}{\sigma}(y - \beta' x)}{\sqrt{1-\rho^2}}\right) \frac{1}{\sigma}\phi\left(\frac{y - \beta' x}{\sigma}\right)\right] \tag{7}$$

The Heckman specification reduces to a probit for participation and an ordinary least squares model for the consumption if both equations are independent, that is $\rho = 0$.

Besides, we have considered the infrequency of purchases model (Deaton and Irish, 1984; Blundell and Meghir, 1987), also based on a bivariate structure. In this model, zero observations may be due either to the survey period being too short for positive purchase to be observed, or to a standard corner solution. In fact, the likelihood of the infrequency of purchases model with independent error terms is similar to the double-hurdle model and can be expressed as

$$L = \prod_{0} \left[1 - \Phi \left(\alpha' z, \frac{\beta' x}{\sigma} \right) \right] \prod_{+} \left[\left[\Phi \left(\alpha' z \right) \right]^{2} \frac{1}{\sigma} \phi \left(\frac{\Phi \left(\alpha' z \right) y - \beta' x}{\sigma} \right) \right] \tag{8}$$

although now the purchase decision is related to the actual costs through the expression $\Phi \left(\alpha' z \right) y$. Indeed, the Tobit model is nested within the above model when $\Phi \left(\alpha' z \right) = 1$.

4.2 Model selection and estimation Results

Before discussing these results in detail, we make some comments about the consideration of alternative models. In this exercise, we consider some exclusion restrictions to identify the participation effect although there is not a common strategy in the literature about their need and selection. Moreover, first stage participation hurdle is likely the result of psychological and cultural factors aside from the levels of prices and income which should be included with any other economic factors in the second hurdle of the model according to the basic assumptions of the models (Newman and Matthews, 2001) . In the second equation, we have added the income variables to explain the length of the trip allowing the existence of corner solutions. From an econometric point of view, the use of different sets of regressors that is imposing some exclusion restrictions in each hurdle can help parameter identification (Jones, 1992). Similarly to Humphreys et al. (2011) the convergence is not achieved without these restrictions when the double-hurdle model is estimated allowing for correlation between the errors of the participation equation and quantity (distance trip) equation.

In order to select the most suitable model for this particular application we estimated different models and ran appropiate statistical tests The first step is the comparison between the Tobit model and the independent double-hurdle model using a likelihood ratio test (LR) because Tobit model is nested in Cragg's Double Hurdle Model (1971) under the hypothesis that the parameters on both equations are identical. This is the expression of the LR test

$$LR = -2 \left[L(\hat{\beta_1}, \hat{\sigma_1}) - L(\hat{\beta_2}, \hat{\sigma_2}) \right] \sim \chi^2 \tag{9}$$

being m the number of exclusion restrictions, $L(\hat{\beta_1}, \hat{\sigma_1})$ is the maximum value of the double hurdle log-likelihood function and $L(\hat{\beta_2}, \hat{\sigma_2})$ is the maximum value of the Tobit log-likelihood function. The null hypothesis establishes the equivalence between the double hurdle model and the Tobit model, which implies that the more parsimonious Tobit model fits the data better. Consequently, the rejection of the null hypothesis means that the double hurdle model is a better alternative to fit the data. In this application, a likelihood ratio test with value of 56486.18 clearly rejects the null hypothesis of model equivalence between the Tobit model and Double Hurdle Model, favoring the selection of the latter. Similarly, instead of assuming independence of error terms between the two parts of the model, we estimated a Dependent Double Hurdle Model finding a ρ parameter statistically not different from zero so we stick with Independent Double Hurdle Model.

In the second step, we also estimated another common model used to deal with zeroes in the dependent variable, the Heckman Selection Model (HSM), where the parameter ρ that indicates dependence between the participation equation and the second hurdle was not significant. Then, we have to choose between the independent double-hurdle model and HSM which is not nested in the double hurdle model. Then, the LR test is not feasible in this case and the best model specification is determinated using the Voung test for non-nested models Vuong (1989).

Suppose that $f(y_i|X_i, \theta)$ and $g(y_i|X_i, \gamma)$ are two alternative specifications for the density of the random variable y, the test requires the computation of both sets of predicted probabilities to obtain $m_i = \log \dfrac{f(y_i|X_i, \theta)}{g(y_i|X_i, \gamma)}$. This value is the difference for each observation between their contributions to the likelihood function. The Voungs statistics is

$$v = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}} \tag{10}$$

The distribution of this standard statistic is a limiting normal standard distribution and it is bidirectional [3]. If the absolute value of the standard statistic is less than two, the two competing models are considered equivalents. In our results, the value of the standard statistic is 4.747 so DHM is preferred instead of the HSM.

Finally, as it was described above the IPM model is estimated and evaluated in comparison with the independent double-hurdle model. In this case, the Voung test statistic is 11.45 so the hypothesis of model equivalence is not rejected and both models are not statistically different. The large value of this statistic indicates that the DHM is the preferred specification alternative.

Overall, test results indicate that Craggs Double Hurdle Model is superior to the rest of specifications. This result is consistent with our previous hypothesis

[3] We follow the description of the test described in Greene (2003).

since DHM allow us to introduce flexibility in the model contemplating the possibility of budget restrictions of individuals, and therefore, choose not to travel. Consequently, zeroes found in the survey seem to respond to corner solutions instead of the infrequency of long distance trips [4].

Table 3: Model selection tests

Models	Test	Value	Conclusion
IDHM vs. Tobit	LR test (m=2)	56486.180***	Reject H_0
IDHM vs. HSM	Non-nested Vuong	4.513***	Reject H_0
HSM vs. IPM	Non-nested Vuong	11.016***	Reject H_0
IDHM vs. IPM	Non-nested Vuong	11.051***	Reject H_0

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

Table 6 displays Cragg's Double Hurdle model (DHM), revealing interesting patterns in long-distance decisions [5]. In the main, our results are consistent with those reported from studies on long-distance travel reviewed in Section 2. The participation equation indicates the effect of the explanatory variables on the probability of being involved in long-distance travel, while the second equation tries to assess the impact of the covariates on distance travelled. This second equation follows a semilog linear form by employing the logarithms of the dependent variable. The magnitude of the ML estimates in these models cannot be interpreted in the same fashion as say OLS estimates as they are based on latent equations. However, the sign of the parameters can be interpreted presenting an intuitive interpretation of the factors determining long-distance travel.

Females are less likely than men to participate in long-distance travel and, given they participate at all, are involved in shorter trips than men. Long-distance travel participation clearly declines with age, however the effect of age on distance travelled is not so clear. According to DHM estimates, distance travelled increases for ageing adults, taking as a reference the base category 15 to 29 years old. Although it seems that being a senior is not different, in terms of distance travelled, from being in the omitted category. These results are similar to those obtained in Dargay and Clark (2012) but contrary to the findings in Collia et al. (2003) and Georggi and Pendyala (2001), although the latter two studies only relied on cross-tabulation analysis.

Income variables only appear in the second equation. Moreover, first stage participation hurdle is likely the result of psychological and cultural factors

[4]Estimation of the DHM is by William Burkes Craggit routine available from www.stata-journal.com/software/sj9-4/ for Stata. IPM is estimated using code based on the work provided by Gibson and Kim (2011).

[5]Estimates of Heckman Selection Model and Infrequent Purchases Model can be found in the Appendix

aside from the levels of prices and income which should be included with any other economic factors in the second hurdle of the model according to the basic assumptions of the models (Newman and Matthews, 2001). In both models we find that individuals in the higher income categories travel further than low income citizens, that is, there is a positive relationship between income and distance travelled.

Geographical, or land-use variables, also play a significant role explaining the participation and distance travelled decisions. Residents of larger citites are more likely to be involved in a long-distance travel, and moreover, these segments of the population travel further than those living in smaller municipalities. As an intuition, we think that longer distances associated with higher populated areas might be affected by the accessibility to better transportation services, specially those designed for long-distance travel, as high capacity roads, well-connected airports and high-speed rail. Higher participation of residents of great cities might also be related to this sort of reasoning, although in some studies where have been previously found a negative relationship between city size and long-distance travel participation have motivated this result differently. More precisely, authors like Limtanakool et al. (2006a) stated that individuals who have a better access to services, i.e. those living in areas with high density of population, travel shorter than rural areas population. That might be the case of the estimates related to the variable that indicates if a person lives in a metropolitan area or not. Our results show that living in a metropolitan area reduces the chances of taking a long-distance travel and also lowers the trip distance for those who travel.

The omitted region is Autonomous Community of Madrid when we control for the region of residence of the respondent. Results of IPM show that residents in any other region are more, or at least equally likely to make a long-distance trip than those living in Madrid. This result does not hold for the DHM where people living Valencian Autonomous Comunity and Region of Murcia are less likely to travel more than 50 km. than Madrid population. DHM estimates of the second equation show that population living in any region different from Madrid Autonomous Comunity make shorter trips. More variability is found when estimating the IPM, where depending on the region impacts on distance travelled are positive, negative or null.

For purpose of the trip the omitted category is the one collecting holidays and leisure motivated trips, while private vehicle is the base category in the mode of transport used. Variables related to purpose and mode only appear in the second equation, as only the indivuals who actually travel report information on these. Results confirm that when the purpose of the trip is leisure and holidays, long-distance users travel much further than in business trips in those trips heading to second residency. Mode of transport used are also introduced in the model as control variables, and as expected, trips by car reduce the distance travelled compared to any other mode of transport, but specially when compared to train and plane.

Table 4: Independent Double Hurdle Model

	Participation Equation		ln(Distance)	
	Coeff.	S.E.	Coeff.	S.E.
Gender: Male	0.190***	(0.01)	0.039***	(0.01)
Age: 30 to 39	-0.0765***	(0.02)	0.076***	(0.02)
Age: 40 to 49	-0.0949***	(0.02)	0.121***	(0.01)
Age: 50 to 64	-0.298***	(0.02)	0.129***	(0.01)
Age: 65 or more	-1.059***	(0.02)	0.006	(0.02)
Income: Medium	-	-	0.046***	(0.01)
Income: High	-	-	0.101***	(0.01)
Municipality: 10,000-50,000	0.045***	(0.02)	0.070***	(0.01)
Municipality: 50,000-500,000	0.159***	(0.02)	0.119***	(0.01)
Municipality: >500,000	0.235***	(0.03)	0.200***	(0.03)
Area: Metropolitan	-0.0671***	(0.02)	0.027*	(0.01)
Region: Galicia	0.031	(0.04)	-0.319***	(0.03)
Region: Asturias	-0.058	(0.05)	-0.240***	(0.04)
Region: Cantabria	-0.059	(0.05)	-0.169***	(0.04)
Region: Euskadi	-0.060	(0.04)	-0.072**	(0.04)
Region: Rioja	0.054	(0.06)	-0.083*	(0.05)
Region: Navarra	0.349***	(0.05)	-0.192***	(0.04)
Region: Aragon	0.264***	(0.04)	-0.087***	(0.03))
Region: Catalonia	0.125***	(0.04)	-0.216***	(0.03)
Region: Com Valenciana	-0.113***	(0.04)	-0.204***	(0.03)
Region: Murcia	-0.184***	(0.05)	-0.236***	(0.05)
Region: Castilla y Leon	0.286***	(0.04)	-0.070**	(0.03)
Region: Extremadura	0.300***	(0.05)	-0.070*	(0.04)
Region: Castilla-La Mancha	0.302***	(0.04)	-0.091***	(0.03)
Region: Andalucia	-0.0457	(0.04)	-0.204***	(0.03)
Purpose: Business	-	-	-0.073***	(0.02)
Purpose: Second residency	-	-	-0.497***	(0.01)
Purpose: Other purpose	-	-	-0.273***	(0.02)
Mode: Bus	-	-	0.196***	(0.02)
Mode: Train	-	-	0.395***	(0.02)
Mode: Plane	-	-	1.976***	(0.03)
Mode: Other	-	-	0.155***	(0.04)
Constant	0.136***	(0.04)	4.570***	(0.03)
σ	-	-	0.790***	(0.00)
Log likelihood		-63994.6		
AIC		128105.1		
BIC		128617.5		

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

As previously indicated, the interpretation of the estimates must be taken with caution and in order to get sensible interpretation of the results, marginal effects are computed and presented in Table 5. For categorical explanatory variables, marginal effects are used to compute percentage change in probability and unconditional level when the value of the variable shifts from zero to one, holding the rest of the variables constant.

For instance the effect of 0.068 on the probability of travel implies that the probability of making a long-distance trip by a male who is average in all other respects increases by almost a 7 per cent of the average probability when the individual is a female with similar characteristics. Of most interest is the overall effect on the dependent variable for values of the explanatory variables, known as unconditional expectation. The unconditional effect of 0.337 implies that the average individual distance is expected to increase by 34% when the person is male. More interestingly, we analyse the unconditional effects of age. With respect to the base category we can observe how growing older sharply reduces distance travelled, with a 176 per cent reduction in the case of seniors. The size of the city of residence also affects distance travelled, raising it in a 50% for cities larger than 500,000 habitants in comparison with cities below 10,000 population.

Table 5: Discrete effects for Double Hurdle Model

Variable	Prob.	Uncond.
Gender: Male	0.068***	0.337***
Age: 30 to 39	-0.027***	-0.088***
Age: 40 to 49	-0.034***	-0.094***
Age: 50 to 64	-0.107***	-0.428***
Age: 65 or more	-0.383***	-1.763***
City size: 10000 to 50000	0.016***	0.112***
City size: 50000 to 500000	0.057***	0.327***
City size: More than 500000	0.085***	0.498***
Area: Metropolitan	-0.024***	-0.097***

Significance code: *p<.1, **p<.05, ***p<.01
Bootstrapped standard errors.

5 Summary and conclusions

Long distance travel is increasing in frequency and, specially, in trip length. Although these sort of trips are still outnumbered by commuting daily short trips, the distance travelled in intercity trips represents a growing share over the total distance. These facts cause better integration between distant regions, higher accessibility of public and private services but also congestion and pollution, among other important social, economical and environmental implications. A current topic of interest is studying factors affecting different aspects of long distance travel demand such as mode choice, trip length and the travel-or-not decision.

In this study, we investigated the factors determining long-distance travel. Using 2007 Spanish Ministry of Public Works' Mobility Survey, and different econometric modeling techniques, we studied the importance of different socio-economic, demographic and land-use factors on the decision of making a long-distance trip and on the length of it. Other factors shown to be important for long distance travel are the main motivation to travel and the mode of transport used.

The main findings indicate that male travel further than females, younger segments of the population participate more in the market of long distance travel and also travel further. Individuals who enjoy higher income levels travel more and further. Also interesting findings related to geographical variables emerged. Citizens living in more populated cities travel further, but those whose residence is located in metropolitan areas travel less and make shorter trips. Airplane and railway passengers travel much further than those who use public bus services and specially those who drive private vehicles. The profile of the long-distance traveller is completed by the results showing that those who travel for leisure or holidays reasons travel much further than those travelling for business or who move to a second residency.

Additionally, testing different models specifications allows us to make some comments about long-distance survey design. According to our results Double Hurdle Model is the preferred method over Infrequent Purchase Model. The prevalence of DHM indicate that most zeroes in the survey are caused by corner solutions and the decision of abstention, that is, with the different prices of transport and personal income given, individuals choose not to travel either because of economic reasons or preferences. This means that zeroes do not seem to be the result of the low-frequency of a rare event as an intercity trip. In this regard, it seems that the three-month period used in this survey is adequate to capture travel behaviour. However, our intuition is that this result may hold only if the current definition of long distance travel is kept at 50 kilometers. Our guess is that in the case of increasing the threshold to 100 kilometers, all other things being equal, may require raising the duration of the reporting period if survey designers aim to avoid infrequent zeroes. It is also important to note that although three months seem to be adequate in statistical terms for long-distance travel surveys design, reducing the period in some weeks might increase the quality of survey responses. It must be taken into account the likely trade-off between reducing infrequency of purchase and the difficulty of recalling past events caused by long reporting periods. We encourage further research in these questions in order to find the optimal reporting period in intercity trips.

References

Arbués, P., Baños, J.F., Mayor, M., Suárez, P., 2013. Determinants of ground transport modal choice in long-distance trips in Spain. Efficiency Series Paper 7/2013, Departamento de Economía. Universidad de Oviedo .

Axhausen, K., 1999. Preparing for the long of it: Methodological research for an European survey of long distance travel. paper presented at TRB Conference: Personal Travel: The Long and Short of It, Washington DC, 28 June1 July .

Baik, H., Trani, A.a., Hinze, N., Swingle, H., Ashiabor, S., Seshadri, A., 2008. Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States. Transportation Research Record: Journal of the Transportation Research Board 2052, 9–20.

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

Blundell, R., Meghir, C., 1987. Bivariate alternatives to the tobit model. Journal of Econometrics 34, 179–200.

Collia, D.V., Sharp, J., Giesbrecht, L., 2003. The 2001 National Household Travel Survey: A look into the travel patterns of older Americans. Journal of Safety Research 34, 461–470.

Cragg, J., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica: Journal of the Econometric Society 39, 829–844.

Dargay, J.M., Clark, S., 2012. The determinants of long distance travel in Great Britain. Transportation Research Part A: Policy and Practice 46, 576–587.

Deaton, A., Irish, M., 1984. Statistical models for zero expenditures in household budgets. Journal of Public Economics 23, 59–80.

Erhardt, G., Freedman, J., Stryker, A., Fujioka, H., Anderson, R., 2007. Ohio Long-Distance Travel Model. Transportation Research Record 2003, 130–138.

Frändberg, L., Vilhelmson, B., 2003. Personal mobility: a corporeal dimension of transnationalisation. The case of long-distance travel from Sweden. Environment and Planning A 35, 1751–1768.

Georggi, N.L., Pendyala, R.M., 2001. Analysis of long-distance travel behavior of the elderly and low income. Transportation Research Circular E-C026, 121–150.

Gibson, J., Kim, B., 2011. Testing the Infrequent Purchases Model Using Direct Measurement of Hidden Consumption from Food Stocks. American Journal of Agricultural Economics 94, 257–270.

Greene, W.H., 2003. Econometric Analysis. volume 97. Prentice Hall, New Jersey. 5th edition.

Heckman, J., 1979. Sample Selection Bias as a Specification Error Author. Econometrica: Journal of the econometric society 47, 153–161.

Holz-Rau, C., Scheiner, J., Sicks, K., 2014. Travel distances in daily travel and long-distance travel: what role is played by urban form? Environment and Planning A 46, 488–507.

Humphreys, B., Lee, Y., Soebbing, B., 2011. Modeling consumers' participation in gambling markets and frequency of gambling. Journal of Gambling and Business and Economics 5, 8–14.

Jones, A.M., 1992. A note on computation of the double-hurdle model with dependence with an application to tobacco expenditure. Bulletin of Economic Research 44, 67–74.

Limtanakool, N., Dijst, M., Schwanen, T., 2006a. On the participation in medium- and long-distance travel: A decomposition analysis for the UK and the Netherlands. Tijdschrift voor Economische en Sociale Geografie 97, 389–404.

Limtanakool, N., Dijst, M., Schwanen, T., 2006b. 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.

Mallett, W., 1999. Long-Distance Travel by Women: Results from the 1995 American Travel Survey. Transportation Research Record 1693, 71–78.

Mallett, W., 2001. Long-distance travel by low-income households. Transportation Research Circular E-C026, 169–177.

Mokhtarian, P.L., Salomon, I., Redmond, L.S., 2001. Understanding the Demand for Travel: It's Not Purely 'Derived'. Innovation: The European Journal of Social Science Research 14, 355–380.

Newman, C., Henchion, M., Matthews, A., 2003. A double-hurdle model of Irish household expenditure on prepared meals. Applied Economics 35, 1053–1061.

Newman, C., Matthews, A., 2001. Infrequency of purchase and double-hurdle models of Irish households meat expenditure. European Review of Agricultural Economics 28, 393–412.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society 26, 24–36.

Vuong, Q., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica: Journal of the Econometric Society 57, 307–333.

Appendices

Table 6: Model estimations

	Heckman Selection Model				Infrequent Purchase Model			
	Participation Equation		ln(Distance)		Participation Equation		ln(Distance)	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Gender: Male	0.190***	(0.01)	0.038***	(0.01)	0.135***	(0.01)	0.258***	(0.02)
Age: 30 to 39	-0.077***	(0.02)	0.078***	(0.02)	0.037***	(0.01)	0.109***	(0.02)
Age: 40 to 49	-0.095***	(0.02)	0.123***	(0.01)	0.033**	(0.01)	0.127***	(0.02)
Age: 50 to 64	-0.298***	(0.02)	0.132***	(0.01)	-0.102***	(0.01)	-0.109***	(0.02)
Age: 65 or more	-1.059***	(0.02)	0.0118	(0.02)	-0.630***	(0.01)	-1.107***	(0.03)
Income: Medium	-	-	0.046***	(0.01)	-	-	0.026***	(0.01)
Income: High Income	-	-	0.100***	(0.01)	-	-	0.063***	(0.01)
Municipality size: 10000 to 50000	0.045***	(0.02)	0.069***	(0.01)	0.010	(0.01)	0.0566***	(0.02)
Municipality size: 50000 to 500000	0.159***	(0.02)	0.118***	(0.01)	0.080***	(0.01)	0.207***	(0.02)
Municipality size: More than 500000	0.235***	(0.03)	0.199***	(0.03)	0.200***	(0.02	0.475***	(0.04)
Area: Metropolitan	-0.067***	(0.02)	0.027*	(0.02)	-0.081***	(0.01)	-0.130***	(0.02)
Region: Galicia	0.031	(0.04)	-0.319***	(0.03)	0.091***	(0.03)	-0.001	(0.05)
Region: Asturias	-0.058	(0.05)	-0.240***	(0.04)	0.056	(0.04)	-0.025	(0.07)
Region: Cantabria	-0.059	(0.05)	-0.169***	(0.04)	-0.018	(0.03)	-0.121*	(0.07)
Region: Euskadi	-0.060	(0.04)	-0.074**	(0.04)	-0.018	(0.03)	-0.114**	(0.06)
Region: Rioja	0.054	(0.06)	-0.083*	(0.05)	0.0303	(0.04	0.017	(0.08)
Region: Navarra	0.349***	(0.05)	-0.199***	(0.04)	0.269***	(0.03)	0.378***	(0.07)
Region: Aragon	0.264***	(0.04)	-0.087***	(0.03))	0.347***	(0.03)	0.567***	(0.06)
Region: Catalonia	0.125***	(0.04)	-0.217***	(0.03)	0.135***	(0.03)	0.127**	(0.05)
Region: Com Valenciana	-0.113***	(0.04)	-0.204***	(0.03)	-0.043	(0.03)	-0.182***	(0.05)
Region: Murcia	-0.184***	(0.05)	-0.235***	(0.05)	-0.023	(0.04)	-0.160**	(0.07)
Region: Castilla y Leon	0.286***	(0.04)	-0.070**	(0.03)	0.268***	(0.03)	0.447***	(0.05)
Region: Extremadura	0.300***	(0.05)	-0.070*	(0.04)	0.244***	(0.03)	0.399***	(0.06)
Region: Castilla-La Mancha	0.302***	(0.04)	-0.093***	(0.03)	0.307***	(0.03)	0.498***	(0.05)
Region: Andalucia	-0.0457	(0.04)	-0.205***	(0.03)	0.047*	(0.02)	-0.019	(0.05)
Purpose: Business	-	-	-0.072***	(0.02)	-	-	-0.040***	(0.01)
Purpose: Second residency	-	-	-0.498***	(0.01)	-	-	-0.280***	(0.01)
Purpose: Other purpose	-	-	-0.273***	(0.02)	-	-	-0.152***	(0.01)
Mode: Bus	-	-	0.196***	(0.02)	-	-	0.108***	(0.01)
Mode: Train	-	-	0.396***	(0.02)	-	-	0.219***	(0.01)

Continued on next page

Table 6 – *Continued from previous page*

Mode: Plane	-	-	1.976***	(0.03)	-	-	1.062***	(0.02)
Mode: Other	-	-	0.155***	(0.04)	-	-	0.112***	(0.02)
Constant	0.136***	(0.04)	4.570***	(0.03)	0.022	(0.03)	2.322***	(0.05)
σ	-	-	0.790***	(0.00)	-	-	0.451***	(0.00)
Log likelihood	-64042.8				-64595.5			
AIC	128203.5				129306.9			
BIC	128724.7				129819.3			
Vuong Test (DHM vs. IPM)								
Value	0.375							
p. value	0.354							

*Robust standard errors between parenthesis. * Significant at 10%. ** Significant at 5%. *** Significant at 1%*