PARTICIPATION DECISION IN TOURISM DEMAND: THE SPANISH CASE.

Juan Luis Eugenio Martín*
Departamento de Análisis Económico Aplicado
Universidad de Las Palmas de Gran Canaria
e-mail: jluis@empresariales.ulpgc.es

Environment Department
University of York
e-mail: jlem100@york.ac.uk

Abstract

The main purpose of this paper is to study the determinants of the decision of participation in tourism demand. We consider three different levels of participation. First, the decision of have ever participated; second, the decision of participation either taking holidays in their own country or abroad, within a period of a year; and third, the decision of participation travelling abroad within a period of a year. Previously to proceed with the analysis we predict income of some missing observations of the sample. For this purpose we consider sample selection bias and double censored regression with grouped data. Finally, we model the participation decisions with the traditional binary choice approach.

Keywords: Tourism Demand, Participation Decision, Sample Selection Bias, Grouped Data.

* *Acknowledge*: Author wishes to thank the help obtained from Sara González and Noelia Martín as well as UK Data Archive for facilitating part of the data for the analysis.

1. INTRODUCTION

In the last decades, the economy of some regions has turned to develop their tourism sector. Provided tourists consider the region as an attractive place to visit, tourism sector may offer a great opportunity for economic growth. Generally speaking, tourists need accommodation, food, transportation and entertainment services. Most of these services are labour intensive and therefore it provides growth in terms of PIB and employment. Consequently, some authorities have seen in tourism activities an interesting alternative to the traditional economies. Thus, current and potential regions based on tourism activities must look at the market in order to understand tourists' needs and practice an adequate tourism policy. For this purpose, they have to figure out which are the determinants of the tourists for choosing their tourist destination and try to adapt their attributes more conveniently. However, the tourism policy cannot be based on the preferences of current tourists only but on those people who are not travelling as well. Understanding the reasons why those people are not travelling is relevant for current tourist regions but even more for new or potential tourist regions because they can try to attract a latent demand. The main purpose of this paper is to model this participation decision in holidays tourism demand.

During a period of time, say a year, any individual, family or household usually face the decision of whether travelling for taking holidays or not. This decision is what we refer to as the participation decision in holidays tourism[1] demand. The main purpose of this paper is to study the determinants of this decision. We consider three different levels of participation. First, the decision of have ever participated; second, the decision of participation either taking holidays in their own country or abroad, within a period of a year; and third, the decision of participation travelling abroad within a period of a year. One of the objectives of the paper is to estimate the main determinants of these different decisions of participation.

Depending on the objective, the analysis can be carried out from a regional or national perspective, i.e. we can study the determinants for participation of either all the regions

[1] The traditional definition of tourism includes stays due to business trips. However, these kind of trips are quite often imposed by the labour conditions of the traveller and the analysis of this kind of tourism is not as relevant as the holidays tourism.

of a country or a set of countries. Results may vary depending on how heterogeneous a country is in terms of the attributes that define each region, e.g. weather conditions and the existence of tourist resorts or outdoor recreation facilities. For instance, the better weather conditions and recreation facilities of a region are, the less likely that a household of that region wishes to travel. In order to deal with this heterogeneity, we test the significance of the differences among regions. For this purpose, we propose two different models. One consists on splitting up the analysis into a regional level and consider differences on the parameters associated with the main determinants. An alternative model consists on including a set of attributes of the place of residence as exogenous variables. This kind of approach is absolutely novel and it offers unknown aspects to the determinants of the participation decision of tourists.

2. MODELLING PARTICIPATION DECISION

2.1 Sample

We will base our analysis on a stratified weighted sample of 16.186 households from all EU members in 1997. Amongst various issues, the sample covers information concerning holidays taking and socio-economic variables. A critical objective in this kind of samples is capturing the information about income. In order to increase the probabilities of obtaining an answer, income is asked as a coded question in 12 intervals. Unfortunately, as expected, not all the interviewees answer the question. Indeed, about 75 per cent answered the question. Nevertheless, we can try to estimate the income of the 25 per cent building an income function from the socio-economic variables available.

2.2 Income prediction

First of all, we test if those interviewees who decided not to state their income are significantly different from those who stated, i.e. we need to test the existence of sample selection bias. We follow Heckman (1979), who proposes a two-step procedure. Such that a first step is a binary decision which estimates if those respondents who did not state their income are significantly different from those who did. The second step proceeds with the regression but taking into account, if appropriate, the sample selection bias.

Hence, if we consider a random sample of N observations, the model proposed is as follows:

$I_i = SI_i\beta_I + \varepsilon_{Ii}, (i = 1,..., S)$

$S_i = SS_i\beta_S + \varepsilon_{Si}, (i = 1,..., N)$, with $S < N$, because N corresponds to the full sample, while S is limited to the subsample of those respondents who stated their income.

In the first equation, I_i denotes income of individual or household i, SI_i represents socioeconomic variables which influence on income, β_I is the vector of parameters associated with SI_i and ε_{Ii} corresponds to error term associated with income regression. We assume $\varepsilon_{Ii} \sim N(0,\sigma_I)$. In the second equation, S_i denotes the sample selection rule, i.e. if respondent has stated his or her income. SS_i indicates socioeconomic variables related with stating or not the income, with its associated vector of parameters β_S and its error term ε_{Si}, where $\varepsilon_{Si} \sim N(0,\sigma_S)$.

If respondents state their income according to any specific rule but pure randomly, then we can assume $E(\varepsilon_{Ii}\varepsilon_{Si}) = 0$. We need to test this last assumption. If the assumption is true then sample selection bias is not significant, otherwise we need to take it into account for the income regression.

For this purpose, we assume that the joint density of ε_{Ii} and ε_{Si} is a bivariate normal density. Consequently, as Heckman shows, we can obtain the inverse of Mill's ratio as $\lambda_i = \frac{\phi(Z_i)}{\Phi(-Z_i)}$, where Z_i is a standard normal variable, defined as $Z_i = -\frac{SS_i\beta_S}{(\sigma_S)^{\frac{1}{2}}}$. Once the inverse of Mill's ratio has been estimated, we may use it as a regressor in the income regression and test its significance. The results of this estimation are presented in table 1.

In order to estimate income, we have to note, as previously commented, the income variable is grouped in intervals, where the first interval lies between zero and the first cut-off point and the last interval lies between the last cut-off point and infinity. Consequently, we have censored information in the tails and thus it is convenient to estimate it with a double censored regression model. Furthermore, Stewart (1983) showed that taking mid-values of the interval is not an efficient procedure for the estimation and he suggested using socio-economic variables to estimate the most likely value which may correspond to each observation within the interval. Stewart's suggestion concerned OLS regression. The case for double censored models has been dealt by Bhatt (1994). The results of this model are shown below in table 2.

Table 1. Estimation of sample selection bias. First step of the Heckman procedure

Variable	Coefficient	Standard error	z	P>z	[95% Confidence interval]	
Size of Community	.0641773	.0074259	8.64	0.000	.0496227	.0787319
Education	.0077466	.0024818	3.12	0.002	.0028824	.0126108
Married	-.0847869	.0310032	-2.73	0.006	-.1455522	-.0240217
Divorced	.1146637	.0615396	1.86	0.062	-.0059517	.2352791
Home	.3648421	.0832853	4.38	0.000	.2016058	.5280783
Student	.7780305	.1072385	7.26	0.000	.5678469	.9882141
Unemployed	.4566631	.0769331	5.94	0.000	.3058768	.6074493
Retired	.5244438	.0473034	11.09	0.000	.4317309	.6171568
Independent professional	.2367499	.1103503	2.15	0.032	.0204673	.4530325
Own business	.120095	.0992451	1.21	0.226	-.0744218	.3146119
Employed professional	.3693362	.1159991	3.18	0.001	.1419822	.5966902
General manager	.3260642	.1145951	2.85	0.004	.1014619	.5506665
Middle manager	.6158252	.0732629	8.41	0.000	.4722326	.7594178
Desk job	.5370718	.0721185	7.45	0.000	.3957221	.6784216
Travelling job	.4361894	.0941859	4.63	0.000	.2515884	.6207904
Service	.6676019	.0774932	8.61	0.000	.5157181	.8194857
Supervisor	.4580355	.1252063	3.66	0.000	.2126357	.7034352
Skilled worker	.5820272	.0606128	9.60	0.000	.4632283	.7008261
Unskilled worker	.48905	.0787948	6.21	0.000	.334615	.6434849

Table 2. Income regression with a double censored model with grouped data. Modified second step of the Heckman procedure

Variable	Coefficient	Standard error	t statistic	P>t
Inverse of Mill's ratio	-208474.7	59980.45	-3.48	0.001
Education	2543.601	1114.766	2.28	0.023
Potential experience	1868.342	851.8789	2.19	0.029
Potential experience squared	-27.01026	11.30141	-2.39	0.017
Male	10013.57	7494.742	1.34	0.182
Size of the community	-6574.633	2271.273	-2.89	0.004
No. of adults	12569.62	2607.538	4.82	0.000
Married	14918.44	7611.222	1.96	0.051
As married	36416.52	15946.17	2.28	0.023
Home	-55333.46	20565.22	-2.69	0.007
Unemployed	-78986.14	13686.44	-5.77	0.000
Retired	-66042.46	17513.46	-3.77	0.000
Fisher	-54253.24	16608.52	-3.27	0.001
Employed professional	83449.17	28976.03	2.88	0.004
Desk	-17911.87	15216.53	-1.18	0.240
Service sector	-32169.19	23649.3	-1.36	0.174
Skilled worker	-46809.26	13071.32	-3.58	0.000
Unskilled worker	-49455.32	11732.25	-4.22	0.000
Andalucía	174105.4	47372.87	3.68	0.000
Aragón	172067.2	50802.38	3.39	0.001
Asturias	178823.9	51384.02	3.48	0.001
Baleares	226650.6	54321.85	4.17	0.000
Canarias	156233.1	48442.89	3.23	0.001
Cantabria	155804.8	54217.88	2.87	0.004
Castilla-León	165273.2	49771.11	3.32	0.001
Castilla – La Mancha	189362.1	50818.16	3.73	0.000
Cataluña	200404.6	48793.15	4.11	0.000
Extremadura	173449.8	49628.69	3.49	0.001
Galicia	164497.8	50335.52	3.27	0.001
Madrid	175457.3	49783.56	3.52	0.000
Murcia	169719.5	50442.46	3.36	0.001
Navarra	173028	46955.86	3.68	0.000
La Rioja	143995.6	52199.72	2.76	0.006
Com.Valenciana	189630.8	46880.17	4.05	0.000
País Vasco	205829.1	50359.81	4.09	0.000
sigma	47625.23	2448.407	19.45	0.000

Mergoupis y Steuer (2003) face a similar problem and decided to predict income employing an ordered probit model. However, since it is a discrete model, predictions are discrete and they need to use mid-values to transform the discrete variable into continuous. Moreover, they do not test for sample selection bias.

2.3 Participation decision

Following the methodology proposed by Eugenio-Martin (2003) we model participation decision, we consider it is a binary choice, denoted by T_i, such that, $T_i = 1$ if household or individual decides to travel and $T_i = 0$ otherwise. We want to model probability that $T_i = 1$, i.e. $\Pr(T_i = 1)$. We assume $\Pr(T_i = 1)$ is linked to a set of exogenous variables, which may be those already shown above. More precisely, for some appropriate function $g(\cdot)$, $\Pr(T_i = 1) = g\left(\alpha + \sum_{j=1}^{k} \beta_j SE_{ji}\right)$, where $0 \leq g(\cdot) \leq 1$, α denotes a constant, SE_{ji} denotes j^th^ socioeconomic variable of household or individual *i* and β_j denotes associated parameter to j^th^ socioeconomic variable.

Traditional linear probability model is not recommended to be used to estimate the probability function because it would present non normal errors, heteroskedasticity and logical inconsistency, since prediction of probabilities may lie out of range (0,1). It is well-known that the suggested model for binary choice estimations is latent variable model. This model considers the existence of a latent variable T_i^*. Since this latent variable is unobserved by the researcher we can consider it is composed by two parts: one observed by the researcher, which includes all the socioeconomic variables and another part that it is unobserved by the researcher and that corresponds to heterogeneity reasons among tourists. Thus the model can be represented as: $T_i^* = \alpha + \sum_{j=1}^{k} \beta_j SE_{ji} + \varepsilon_i$, where ε_i denotes unobserved part or error term. For our purposes, the latent variable will work as an index function, such that we will set $T_i^* = 1$ if $T_i^* > 0$ and $T_i^* = 0$ if $T_i^* \leq 0$.

Let $S_i = \alpha + \sum_{j=1}^{k} \beta_j SE_{ji}$, such that $T_i^* = S_i + \varepsilon_i$.

Then, $\Pr(T_i = 1) = \Pr(S_i + \varepsilon_i > 0) = \Pr(\varepsilon_i > S_i) = 1 - \Pr(\varepsilon_i \leq S_i) = 1 - F_\varepsilon(-S_i)$, where F_ε denotes cumulative density function of unobserved part. Due to a problem of identification of location and scale of T_i^*, researcher needs to choose a distribution and a value for the variance of ε_i. The most common approaches assume ε_i is independently and identically distributed, either following a normal distribution with zero mean and variance of one, or following a logistic distribution with zero mean and variance of $\frac{\pi^2}{3}$. If we assume that ε_i follows the former distribution we are employing the well-known probit model, and if we assume the latter distribution we are employing the also well-known logit model. Any of these distributions can be employed for the participation decision and both present similar results. Finally, maximum likelihood estimation is applied to the model in order to estimate parameters of interest. Under correct specification, these estimates are consistent and asymptotically normal[2].

3. MAIN RESULTS

The main results of the participation decision are shown below in tables 3, 4 and 5. In all the models the income variable seems to be highly significant, as expected. Moreover, income is the only variable which is significant in the three decisions and justifies the especial care we have considered previously in its prediction.

The case of have ever participated, as we can see in table 3, shows how relevant is the fact of having children or a large family in order to be able to travel for tourism purposes. Besides, if household lives in a large town they will be more likely to travel than those households who live in rural surroundings.

Table 3. The determinants of ever have participated in tourism demand

	Coefficient	**Standard error**	**z**	**P>z**	**[95% confidence interval]**	
Income	.0000174	2.27e-06	7.65	0.000	.0000129	.0000218
Age	-.0056109	.0035847	-1.57	0.118	-.0126557	.0014338
Number of children	-.2176923	.1372301	-1.59	0.113	-.4873836	.051999
Number of adults	-.3931319	.0801545	-4.90	0.000	-.5506554	-.2356084
Size of community	.1035657	.0459335	2.25	0.025	.013295	.1938364

[2] For a complete exposition of the methodology see Greene (2003).

For the participation decision, it is interesting to point out that the more the place of residence is visited the more likely is that the residents of that place go out for tourism, either national or international. Since most of the tourism participation is local or national this remark shows that areas highly visited are usually well equipped in terms of transportation infrastructure which results in an advantage for the tourism purposes of the residents. This idea is also reinforced by the fact that population density is also significant. The age is seen as a negative determinant for the decision of travelling. This is due to the link that age has with health status, which may condition the probabilities of travelling. As in the previous case, the number of children and the number of adults in the household are constraining the chances for taking tourism trips. As it was also shown by Mergoupis and Steuer (2003), women are more likely to travel than men. Finally, as expected, labour conditions are affecting the chances for travelling. It is the case of unemployed people, unskilled workers and those people who are the owners of a shop.

Table 4. The determinants of participation in tourism demand (domestic and abroad)

Variable	Coefficient	Standard error	z	P>z	[95% Confidence interval]	
Income	6.95e-06	1.36e-06	5.13	0.000	4.29e-06	9.61e-06
Attractiveness of the place of residence (No.visitors/population)	.6419998	.1989475	3.23	0.001	.2520699	1.03193
Population per km2	.0015141	.0003815	3.97	0.000	.0007664	.0022619
Age	-.0189922	.0032988	-5.76	0.000	-.0254578	-.0125266
Male	-.3606121	.1490899	-2.42	0.016	-.6528229	-.0684014
Number of children	-.212125	.0899872	-2.36	0.018	-.3884967	-.0357533
Number of adults	-.1630899	.0618846	-2.64	0.008	-.2843815	-.0417984
Separate	-.859349	.4887586	-1.76	0.079	-1.817298	.0986002
Unemployed	-.6452606	.3136969	-2.06	0.040	-1.260095	-.030426
Shop	-.4120333	.2069645	-1.99	0.046	-.8176762	-.0063905
Unskilled worker	-.979449	.271214	-3.61	0.000	-1.511019	-.4478794

Outbound tourism is influenced similarly as the previous case, with income, age, gender and the number of children significant and with a similar interpretation as before. However, in this case, independent professionals are those workers more likely to travel abroad with respect to other kind of occupation. Nevertheless, we have to point out that for this kind of tourism it is relevant to consider how attractive for tourism purposes it is the place of residence. In this sense, we show that those people who live in a place of tourist interest are less likely to travel than those who live in not so attractive places. As a complement to this determinant, if the place of residence is relatively congested, then residents are more likely to travel. Finally and opposite to the previous cases where the bigger the community is the more chances to travel at the local, national or international level; in the case of only international tourism, those who do not live in large communities are more likely to travel than those who live in large cities. This last point needs further research; however, it seems that lower cost of transportation and better transportation infrastructures for national tourism in favour of inhabitants of large cities are affecting their preferences for national tourism rather than international tourism.

Table 5. The determinants of participation in outbound tourism demand (abroad only)

Variable	Coefficient	Standard error	z	P>z	[95% Confidence interval]	
Income	8.15e-06	1.81e-06	4.50	0.000	4.60e-06	.0000117
Age	-.0351744	.005974	-5.89	0.000	-.0468832	-.0234656
Male	-.5782049	.2335874	-2.48	0.013	-1.036028	-.1203821
Number of children	-.5042481	.2193643	-2.30	0.022	-.9341944	-.0743019
Independent professional	1.355236	.6739983	2.01	0.044	.0342238	2.676249
Attractiveness of the place of residence (No.tourists/Population)	-.376096	.1078011	-3.49	0.000	-.5873823	-.1648097
Tourist congestion of the place of residence (No.tourists/km2)	.0015567	.0005052	3.08	0.002	.0005665	.002547
Size of community	-.1646511	.0519341	-3.17	0.002	-.26644	-.0628621

4. CONCLUSIONS

We have seen, as expected, that the most relevant determinant for participating in tourism demand is the income level and for not participating is the number of children.
Sample employed presents some income missing observations. Therefore and due to its relevance in the model, we need to predict income with special care. For this purpose, we predict income missing observations employing two steps. In a first step we consider sample selection bias and in the second step we predict income with a double censored

regression with grouped data. Once income is predicted for those missing observations we proceed with modelling participation decision.

For the case of Spain, the decision of have ever participated is positively linked to expected variables as income and size of the community, while negatively to the age, the number of children and adults. It confirms the important role that financial, familiar and health conditions have in the chances that a family has in travelling. Nevertheless, we have seen particularities in the other two kinds of participation. On the one hand, tourism participation decision, mainly local or national tourism is positively related with the number of national visitors that they receive and consequently with the level of transport infrastructures. Moreover, we show how labour conditions constrain the chances for travelling. It is the case of the unemployed and unskilled workers and the owners of a shop. Finally, for the case of outbound tourism, we showed that if the place of residence is a tourist attraction, then it is more likely that those residents will stay there rather than travelling abroad. However, this last point is negatively linked to the size of the community and the level of tourist congestion in the place of residence.

References

Bhatt, C.R. (1994). "Imputing a continuous income variable from grouped and income missing observations." *Economics Letters* 46(4): 311-9.

Eugenio-Martin, J.L. (2003). "Modelling the determinants of tourism demand as a 5-stage process. A discrete choice methodological approach." *Tourism Research and Hospitality. Surrey Quarterly Review*(forthcoming).

Heckman, J.J. (1979). "Sample selection bias as a specification error." *Econometrica* 47(1): 153-62.

Mergoupis, T. y M. Steuer (2003). "Holiday taking and income." *Applied Economics* 35(3): 269-84.

Stewart, M.B. (1983). "On least squares estimation when the dependent variable is grouped." *Review of Economic Studies* 50(4): 737-53.