# Habit Formation in Tourism Travelling

David Boto-García
botodavid@uniovi.es

University of Oviedo
Department of Economics, Faculty of Business and Economics
Avenida del Cristo s/n 33006, Oviedo (Spain)

**Abstract:**

This paper conducts a microeconometric analysis of individual participation in tourism activities. We examine the existence of habit formation in the form of state dependence by which past trips increase the taste for travelling. We also study the role of regional unemployment rates, regional price indexes and climate conditions at origin on the likelihood of tourism participation. Individual and household characteristics are also controlled for. We use monthly longitudinal microdata for Spain between 2015-2018 involving more than 92,000 individuals. We estimate static and dynamic random effects Probit models finding evidence of habit formation. The initial conditions and participation in the previous month raise the propensity to make a tourist trip in the following period by 28 percentage points. Habit formation is found to be strongly associated with income and education.

---

**JEL codes: C23, D12, L83**
**Keywords:** *tourism participation, habit formation, state dependence, dynamic panel data*

**Acknowledgements:**

I acknowledge financial support from the Spanish Ministry of Education, Culture and Sport (FPU 16/00031) and the Project PAPI-18-GR-2011-0026 (University of Oviedo).

I also acknowledge the Spanish National Statistics Institute for the provision of the primary data. All potential errors and mistakes derived from the analysis are mine.

## 1. Introduction

The tourism industry is nowadays an important driver of economic growth, both in developed (e.g. Capó-Parrilla et al., 2007) and developing countries (e.g. Faber and Gaubert, 2019). Despite its economic relevance for regional development, the role of personal and regional factors in sustaining tourism participation (i.e. the decision to make a tourist trip away from home) are still not properly understood. Specifically, there is a need for studying the dynamics of participation in tourism activities at the individual level considering both household characteristics and economic conditions.

In line with the cultivation of taste model developed by McCain (1981), taste for tourism travelling can be developed through consumption. That is, participation in tourism might exhibit habit formation by which past trips increase the taste for travelling (Pollak, 1970). This is what is generally called *state dependence*. This emerges because of individuals being forward-looking and maximizing their utility over time, with their preferences over tourism goods changing with past consumption. Although there are several studies that have analysed the dynamics of tourism flows using aggregate data (Nordström, 2005; Santana-Gallego et al., 2011; Lorde et al., 2016; Li et al., 2017), there is little research on tourism state dependence using microdata. We are only aware of the studies by Alegre et al. (2009) and Wu et al. (2013). For the case of Spain, Alegre et al. (2009) find that participation in tourism in the previous year positively impacts subsequent participation. Using Japanese panel data, Wu et al. (2013) show by contrast that tourism participation in month *t-1* negatively influences participation in month *t*. We aim to contribute to this scarce evidence.

There are two sources of observed state dependence. One stems from genuine state dependence, implying that previous behaviour influences current behaviour (Heckman, 1981a; 1981b). A second source emanates from unobserved heterogeneity that induces correlation between past and current behaviour, producing spurious state dependence. Therefore, for the purposes of appropriate identification of habit formation, it is necessary to control for unobserved heterogeneity in the analysis (Naik and Moore, 1996; Wooldridge, 2002). As such, studies based on time series or aggregate data might provide an incomplete picture of habit formation due to problems of separability of mutual trends, simultaneity, and aggregation (Heien and Durham, 1991). We, instead, make use of longitudinal microdata for this purpose.

We estimate both static and dynamic panel data models for tourism participation. Specifically, we estimate i) Mundlak-Chamberlain correlated random effects Probit model in the static context (Mundlak, 1978; Chamberlain, 1984), and ii) the proposal by Rabe-Hesketh and Skrondal (2013) and Skrondal and Rabe-Hesketh (2014) to deal with the initial conditions problem in a dynamic framework. In both cases, we study the role of time-varying regional characteristics on the individual decision to travel. In particular, we examine the effect of prices, unemployment rates and two indexes of climate conditions at the place of residence. The model also controls for individual sociodemographic features, characteristics of the household, time effects and regional fixed effects. Apart from their already relevant effect on tourism participation, the inclusion of these variables in the analysis allows us to identify habit persistence in a cleaner way, once observable sources of heterogeneity are accounted for.

We use monthly microdata from a representative sample of individuals living in Spain during the period February 2015 and December 2018[1]. This 47-month window allows us to assess the dynamics of tourism participation considering different climatic and economic conditions. Our database involves more than 488,000 observations corresponding to more than 92,000 individuals. The data refers to any kind of leisure trips, either within the region, to another Spanish region or abroad. In this way, we are considering short breaks, which are becoming increasingly popular and have been less studied. We merge this dataset with monthly regional data (NUTS 2) on consumer price indexes, unemployment rates and two climate indicators: heating degree days and cooling degree days. These two indexes measure the demand for heating and cooling during a given period based on outside temperature, with the additional advantage of being non-linear. Therefore, they allow us to examine how engagement in tourism activities relates to climate conditions at the place of residence in a non-linear fashion.

We find clear evidence of habit formation in tourism travelling. Previous month participation in tourism raises the probability of participation by 28%. Entry probabilities are around 20%, with engagement into tourism being largely determined by the initial conditions at the first observation period. Habit formation is strongly connected with household income and educational level. Furthermore, we also document that participation decreases with regional consumer price indexes but increases with warm temperatures. Interestingly, we find that the likelihood of leisure travelling is negatively associated with the mean level of regional unemployment but increases with deviations over that benchmark. Also relevant, our estimates show that participation decreases with household size, unemployment status and having a temporary job contract, but increases with the population density of the municipality of residence.

We contribute to the literature by examining the dynamics of tourism participation at the individual level, distinguishing the role of personal characteristics, time-variant and time-invariant regional factors, time effects, unobserved heterogeneity and state dependence. Unlike most related studies that use either aggregate flows or cross-sectional microdata, by exploiting a large longitudinal dataset we are able to identify habit formation in a cleaner way net of personal and regional characteristics. As such, we separate genuine habit formation from spurious habit formation stemming from unobserved heterogeneity.

[1] The use of monthly data is a novel aspect of the paper. Most studies on state dependence use annual data due to the lack of data for shorter observation intervals. The study by Bhuller et al. (2017) compares the estimates from monthly and annual data for the case of state dependence in labor market dynamics. They show that the use of annual data grossly overestimates the degree of state dependence predicted by a model at the monthly level. Because of this, we consider monthly data to be better suited for the identification of state dependence.

The analysis of consumers' habit formation has raised concern among economists in different domains (Naik and Moore, 1996; Erdem and Sun, 2001; Williams, 2004; Nolan, 2010; Vesterberg, 2018). To the best of our knowledge, this is the first study that formally disentangles habit formation in tourism travelling.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 describes the database and the variables employed. Section 4 outlines the empirical modelling. Section 5 presents and discusses the results. Finally, Section 6 summarizes the findings and concludes with some policy implications.

## 2. Literature Review

Microeconomic models of tourism participation assume individuals maximize their utility subject to budget and time constraints. Their utility depends on the consumption of tourism goods and a composite Hicksian non-tourism good. If preferences are weakly separable (Deaton and Muellbauer, 1980), the decision to make a tourist trip can be explained by economic factors, personal constraints and taste for travelling.

*Economic factors*

Tourism is a normal good with a positive income elasticity of demand so that tourism participation is higher among high income people (Reece, 2004; Nicolau and Más, 2005; Eugenio-Martín and Campos-Soria, 2010; 2011; Alegre et al., 2013)[2]. For the Spanish case, Alegre et al. (2010) document that budget constraints are the main reason that restricts households to participate in tourism. About 48% of households in their sample could not afford a one-week holiday away. They also argue that tourism participation does not only depend on household income but also on the general economic situation. In this sense, participation rates change over the business cycle (Smeral, 2012; 2014; Wong et al., 2016).

The income elasticity of tourism demand has been shown to vary across the business cycle due to liquidity constraints and precautionary savings related to fears about the future (Gunter and Smeral, 2016; Smeral, 2017). It is well-documented in the economic literature that labour market uncertainty raises the propensity to save (Carroll et al., 2003). In this way, macroeconomic conditions impact individual decisions about engagement in tourism activities. Unemployment and fears about a job loss have been shown to be negatively related to tourism participation (Alegre et al., 2013; Nicolau and Más, 2005). For instance, Bernini et al. (2017) and Alegre et al. (2019) use the unemployment rate at the place of residence to explore how the decision to travel is affected by unfavourable

[2] Nonetheless, the income elasticity of demand significantly differs depending on the origin and the destination being analysed (Peng et al., 2015).

economic conditions. They find unemployment to be negatively associated with both domestic and abroad travelling.

Nevertheless, economic downturns do not necessarily deter tourism participation (Bronner and de Hoog, 2014). Some authors argue that tourist trips are necessary goods so that consumers economize but do not give up going on holidays (Bronner and de Hoog, 2016). Under bad economic circumstances, consumers cutback their tourism expenditure by travelling to closer destinations (Cafiso et al., 2016), staying for shorter periods or lodging at cheaper accommodations (Campos-Soria et al., 2015). Indeed, the tourism industry has been more resilient to the economic crisis than other sectors (Cellini and Cuccia, 2015).

*Personal constraints*

Together with economic factors, individuals normally face time and personal constraints. Economic models of time allocation show that the demand for recreation is strongly linked to labour market supply. In this regard, studies on tourism participation also find significant differences in the likelihood of tourism travelling across job occupations (Eugenio-Martín and Campos-Soria, 2010; 2011; Bernini and Cracolici, 2015). In the context of a two-member household, McConnell (1999) shows that full-time dual-earner households face tighter time constraints for joint recreation, among other things due to the difficulties in synchronizing their leisure times.

Another factor that constraints tourism participation is age. Although it is generally found that participation increases with age due to both higher available leisure time and wealth (Alegre et al., 2010; Eugenio-Martín and Campos-Soria, 2014; Bernini and Cracolici, 2016; Bernini et al., 2017; 2019), the relationship is nonlinear and has an inverted U shape. Seniors are less likely to travel, mainly due to mobility and health related problems (Fleischer and Pizam, 2002).

Regarding the effect of household size, the evidence is mixed. Some studies find that family size is a barrier that limits tourism participation (Alegre and Pou, 2004; Nicolau and Más, 2005; Alegre et al., 2013; Eugenio-Martín and Campos-Soria, 2011; Wu et al., 2013; Kim et al., 2019) while others find the opposite (Bernini and Cracolici, 2015, 2016; Bernini et al., 2017; 2019). This inconclusive evidence can be explained by the presence of children in the household, which seems to increase the likelihood of domestic travelling (Eugenio-Martín and Campos-Soria, 2010; Alegre et al., 2013).

*Taste for travelling*

Heterogeneity in tastes mainly emanate from heterogeneity in sociodemographic characteristics (Pollak and Wales, 1981). A stylized finding is that the probability of travelling increases with education level (Reece, 2004; Bernini and Cracolici, 2015; 2016; Bernini et al., 2017; Alegre et al., 2010; 2013; Wu et al., 2013; Eugenio-Martín and

Campos-Soria, 2010; 2011;2014; Li et al., 2020). This is argued to be due to easier access to information and a better knowledge of foreign languages. Participation is also related with marital status. Eugenio-Martín and Campos-Soria (2010; 2011) find that married people are significantly more likely to travel domestically than separated or divorced individuals. Since tourism is a social activity, this is explained by married or cohabiting people having more opportunities for joint travelling. In this sense, Rashidi and Koo (2016) find that the travel party choice is largely influenced by household size and the presence of children at home.

Empirical research also shows the existence of important geographical differences in tourism participation. The studies by Alegre and Pou (2004), Nicolau and Más (2005), Bernini et al. (2017), Eugenio-Martín and Campos-Soria (2014), Alegre et al. (2013), Kim et al. (2019) and Li et al. (2020) show that tourism participation is positively associated with the population density of the place of residence. This could be due to a higher need to scape for relaxation (Eymann and Ronning, 1997). Furthermore, the observed higher participation rates among residents in densely populated areas can be associated with a low air quality, which exerts a pushing effect on outbound tourism (Wang et al., 2018). Another source of heterogeneity are regional differences in transport infrastructure and accessibility. Albalate and Fageda (2016) show that the development of high-speed rail services has positively impacted tourism flows.

Additionally, tourism participation has been found to depend on climate conditions at the place of residence. Eugenio-Martín and Campos-Soria (2010; 2011) find that the warmer the origin, the higher the probability of domestic travelling but the lesser the probability of an international trip. Similar findings are reported in Wu et al. (2013), who show that people are more likely to travel when temperatures rise. Li et al. (2017) document that home climate is a significant predictor of aggregate flows from Hong Kong to Mainland China.

As it happens with other experience goods like culture (Castiglione and Infante, 2016), taste for travelling might be developed though consumption. Tourism is an experience good whose utility might be contingent on the accumulation of travelling capital (experience). By investing time in tourism activities, consumers accumulate knowledge and skills, thereby finding it easier and more appealing to travel in the future[3]. This implies that preferences for tourism participation depend on previous trip experiences (Adamowicz, 1994). This is the cultivation of taste framework developed by McCain (1981). Accordingly, the necessary skills and travelling capital for tourism appreciation might be acquired with exposure in a ‘learning-by-doing’ process à la Stigler and Becker (1977). In other words, the marginal utility of tourism participation becomes dynamic and increases with past consumption, leading to habit formation in tourism travelling (Pollak, 1970).

[3] Information asymmetries, the cost of information and all other uncertainties associated to travelling are reduced when this activity becomes a habit.

There is a large body of research on the dynamics of tourism demand. Whereas some conduct time series analysis for a single country (Nordström, 2005), others adopt panel data specifications for modelling aggregate flows considering different countries or regions (Song et al., 2010; Seetaram, 2010; Santana-Gallego et al., 2011; Lorde et al., 206; Habibi, 2017; Dogru et al., 2017; Li et al., 2017). These studies adopt autoregressive specifications in which current aggregate flows are regressed on past demand. A positive coefficient of the lagged dependent variable is usually interpreted as capturing habit persistence, word-of-mouth effect, reputation or interdependent preferences (Dogru et al., 2017). However, a proper identification of habit persistence requires to consider the behavior of the same individual over time[4]. That is, aggregate flows might inform of persistence in demand to a destination or between an origin and a destination, but not about habit formation at the individual level.

There are few studies that study the effect of habit formation in tourism travelling at the microlevel, being this evidence mixed. Alegre et al. (2009) find a positive effect of previous year's participation on subsequent participation whereas Wu et al. (2013) provide evidence of the opposite. We aim to contribute to this inconclusive body of research.

## 3. Data

### *3.1. Microdata on tourism participation*

Our database is drawn from the Spanish Domestic Tourist Survey (ETR/Familitur). This survey is conducted on a monthly basis by the Spanish National Statistics Institute to a representative sample of the Spanish population. The sample is obtained by multistage sampling, stratified by conglomerations with proportional section of primary (cities) and secondary units (census sections). Each month, around 8,000 individuals are interviewed at home (by telephone and in some cases personally) about tourist and same-day trips made two months before, and then the data is assigned to the corresponding period. A tourist trip is defined in the survey as any trip that implies at least one overnight stay away from home. These trips involve travelling outside the municipality where the respondent usually lives. Therefore, they include both within-region, within-country and international trips. Households are selected as a subsample of the Household Continuous Survey. One person in the household is randomly selected with equal probability, with the only requirement of being older than 15. Respondents are followed over time, so the database has a panel structure. However, the sample is updated each month, so it is a rotated panel.

[4] As shown by Heien and Durham (1991), habit effects using time series or aggregate data are overstated and substantially differ from the estimates using microdata.

We have monthly data for the period February 2015-December 2018 (47 months), involving a total of 219,675 different individuals and 736,147 observations. Since the panel is unbalanced, with some individuals being observed for short periods and others remaining in the sample for longer, we exclude from our analysis those with less than 3 and more than 20 observed periods. After further excluding some other respondents with missing values in the variables of interest, our final sample comprises 488,265 observations for 92,472 individuals living in any of the 17 Spanish regions (Ceuta and Melilla are excluded)[5].

The survey provides information about the number of tourist trips the respondent made in each month (*numtrips*). This variable ranges from 0 to 25 and exhibits a large inflation of zeroes (75%). Detailed information on the specific features of each trip is only gathered about up to three trips. Since this indicator of intensity might be mixing different trip purposes, we focus only on the participation decision. To this end, we construct a binary indicator for whether respondent $i$ travelled in month $t$ (denoted by $y_{it}$). This variable will act as our dependent variable.

The survey collects information about individual and household characteristics. Specifically, individual data includes age, gender, education level, nationality, civil status and labour market situation. Household features comprise household monthly income, household size, household composition, the number of employed and unemployed household members and the number of members under 15. Additionally, we have data about the number of inhabitants in the municipality, its population density, and the region (NUTS2) where the respondent lives. Based on this information, we define the following variables to be used in the analysis:

- Sociodemographic characteristics: age (in levels and in a squared form, denoted by *age* and *agesq* respectively), education level (*seceduc* and *higheduc*, being *primeduc* the base category), nationality (*foreign*) and civil status (*single* and *married*, being the rest of possibilities the excluded category).
- Labour market situation: *unemployed*, *retired*, *selfemployed*, *businessman* and *employee*, with other situations like student, housekeeper or disable acting as the excluded category. To control for the type of contract and labour stability of employees, the latter variable is replaced in the specification by *permemployed* (i.e. employee with an indefinite contract) and *tempemployed* (i.e. employee with temporary contract).
- Household characteristics: household income (*inc2*, *inc3*, *inc4*, *inc5* and *inc6*, with *inc1* acting as the base category), household size (*housesize*), household composition (*singleparentkids*, *couplenokids*, *couplewithkids*, being *alone* the excluded category), number of unemployed members (*numunempl*) and number of children in the household (*numless15*).

[5] See Table A1 in Appendix for more details.

- Characteristics of the place of residence: population density (*mediumdensity* and *highdensity*, taking *lowdensity* as the reference category) and municipality size (*mun2* and *mun3*, being *mun1* the reference category).

Table 1 presents summary statistics of these variable along with a full description and notation. These descriptive statistics are obtained based on the total number of observations. Since some individuals remain for longer than others due to the rotated nature of the sampling, the figures does not exactly correspond to individuals but to individual-time units.

| Label | Description | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| ***Dependent variable*** | | | | | |
| *y* | =1 if makes a tourist trip in month t | 0.246 | 0.431 | 0 | 1 |
| *numtrips* | Number of tourist trips in month t | 0.392 | 0.871 | 0 | 25 |
| ***Individual characteristics*** | | | | | |
| *age* | Age (in years) | 52.97 | 17.15 | 15 | 85 |
| *female* | =1 if female | 0.520 | 0.499 | 0 | 1 |
| *primeduc* | =1 if primary education | 0.146 | 0.353 | 0 | 1 |
| *seceduc* | =1 if secondary education | 0.471 | 0.499 | 0 | 1 |
| *higheduc* | =1 if high education | 0.378 | 0.485 | 0 | 1 |
| *foreign* | =1 if foreign | 0.057 | 0.233 | 0 | 1 |
| *single* | =1 if single | 0.279 | 0.448 | 0 | 1 |
| *married* | =1 if married | 0.542 | 0.498 | 0 | 1 |
| *widow* | =1 if widow/widower | 0.102 | 0.303 | 0 | 1 |
| *sepdiv* | =1 if separated or divorced | 0.077 | 0.267 | 0 | 1 |
| *unemployed* | =1 if currently unemployed | 0.092 | 0.290 | 0 | 1 |
| *retired* | =1 if retired | 0.232 | 0.422 | 0 | 1 |
| *inactive* | =1 if inactive (student, housekeeper, disabled) | 0.173 | 0.379 | 0 | 1 |
| *businessman* | =1 if businessman/businesswoman | 0.032 | 0.177 | 0 | 1 |
| *selfemployed* | =1 if self-employed | 0.059 | 0.237 | 0 | 1 |
| *permemployed* | =1 if employee and has a permanent labour contract | 0.326 | 0.469 | 0 | 1 |
| *tempemployed* | =1 if employee and has a temporary labour contract | 0.079 | 0.270 | 0 | 1 |
| ***Household characteristics*** | | | | | |
| *inc1* | =1 if monthly household income is less than €999 | 0.231 | 0.421 | 0 | 1 |
| *inc2* | =1 if monthly household income is between €1,000 and €1,499 | 0.247 | 0.431 | 0 | 1 |
| *inc3* | =1 if monthly household income is between €1,500 and €2,499 | 0.300 | 0.458 | 0 | 1 |
| *inc4* | =1 if monthly household income is between €2,500 and €3,499 | 0.137 | 0.344 | 0 | 1 |
| *inc5* | =1 if monthly household income is between €3,500 and €4,999 | 0.063 | 0.244 | 0 | 1 |
| *inc6* | =1 if monthly household income is higher than €5,000 | 0.021 | 0.146 | 0 | 1 |
| *housesize* | Number of people living in the household | 2.526 | 1.197 | 1 | 14 |
| *oneperson* | =1 if lives alone | 0.221 | 0.415 | 0 | 1 |
| *singleparentkids* | =1 if lives with children without more adults | 0.085 | 0.280 | 0 | 1 |
| *couplenokids* | =1 if lives with partner without children | 0.253 | 0.435 | 0 | 1 |
| *couplewithkids* | =1 if lives with partner and children | 0.381 | 0.486 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| *otherh* | =1 other type of household | 0.059 | 0.237 | 0 | 1 |
| *emplhouse* | Number of employed people in the household | 1.016 | 0.896 | 0 | 8 |
| *unemplhouse* | Number of unemployed people in the household | 0.594 | 0.828 | 0 | 7 |
| *less15house* | Number of children under 15 living in the household | 0.391 | 0.738 | 0 | 7 |
| ***Time periods*** | | | | | |
| *y2015* | =1 if year 2015 | 0.174 | 0.379 | 0 | 1 |
| *y2016* | =1 if year 2016 | 0.275 | 0.446 | 0 | 1 |
| *y2017* | =1 if year 2017 | 0.298 | 0.457 | 0 | 1 |
| *y2018* | =1 if year 2018 | 0.254 | 0.435 | 0 | 1 |
| *q1* | =1 for first quarter | 0.233 | 0.423 | 0 | 1 |
| *q2* | =1 for second quarter | 0.249 | 0.432 | 0 | 1 |
| *q3* | =1 for third quarter | 0.256 | 0.437 | 0 | 1 |
| *q4* | =1 for fourth quarter | 0.262 | 0.440 | 0 | 1 |
| ***Place of residence*** | | | | | |
| *lowdensity* | =1 if lives in a sparsely populated area | 0.286 | 0.452 | 0 | 1 |
| *mediumdensity* | =1 if lives in a medium populated area | 0.254 | 0.435 | 0 | 1 |
| *highdensity* | =1 if lives in a highly populated area | 0.460 | 0.498 | 0 | 1 |
| *mun1* | =1 if lives in a municipality with more than 100,000 inhabitants | 0.380 | 0.485 | 0 | 1 |
| *mun2* | =1 if lives in a municipality between 20,000 and 100,000 inhabitants | 0.270 | 0.444 | 0 | 1 |
| *mun3* | =1 if lives in a municipality with less than 20,000 inhabitants | 0.350 | 0.477 | 0 | 1 |
| *Andalusia* | =1 if lives in Andalusia | 0.120 | 0.325 | 0 | 1 |
| *Aragon* | =1 if lives in Aragon | 0.043 | 0.203 | 0 | 1 |
| *Asturias* | =1 if lives in Asturias | 0.042 | 0.201 | 0 | 1 |
| *BalearicIslands* | =1 if lives in the Balearic Islands | 0.036 | 0.187 | 0 | 1 |
| *CanaryIslands* | =1 if lives in the Canary Islands | 0.043 | 0.203 | 0 | 1 |
| *Cantabria* | =1 if lives in Cantabria | 0.033 | 0.180 | 0 | 1 |
| *CastileLeon* | =1 if lives in Castile and Leon | 0.061 | 0.240 | 0 | 1 |
| *CastillaMancha* | =1 if lives in Castilla-LaMancha | 0.050 | 0.220 | 0 | 1 |
| *Catalonia* | =1 if lives in Catalonia | 0.100 | 0.300 | 0 | 0 |
| *ValencianCom* | =1 if lives in Valencian Community | 0.090 | 0.287 | 0 | 1 |
| *Extremadura* | =1 if lives in Extremadura | 0.040 | 0.198 | 0 | 1 |
| *Galicia* | =1 if lives in Galicia | 0.065 | 0.248 | 0 | 1 |
| *Madrid* | =1 if lives in Madrid | 0.097 | 0.296 | 0 | 1 |
| *Murcia* | =1 if lives in Murcia | 0.042 | 0.202 | 0 | 1 |
| *Navarre* | =1 if lives in Navarre | 0.040 | 0.198 | 0 | 1 |
| *BasqueCountry* | =1 if lives in the Basque Country | 0.057 | 0.233 | 0 | 1 |
| *Rioja* | =1 if lives in La Rioja | 0.033 | 0.180 | 0 | 1 |
| Individuals | 92,472 | | | | |
| Observations | 488,265 | | | | |

Table 1.- Descriptive statistics of the sample

Around 25% of the individual-time units make at least one monthly tourist trip during the period considered. Average age is 53 years, with 52% of females and only 6% of foreign people. More than a half are married (54%) and 47% of the sample attains secondary education. About 9% are unemployed, 23% are retired and 17% are inactive. Concerning participants in the labour market, 3% are businessman, 6% are self-employed and around 40% are employees (32% with a permanent contract and 6% with a temporary one).

Regarding household characteristics, 30% monthly earn between €1,500 and €2,500, with a non-negligible 23% earning less than €1,000 per month. The average number of household members is 2.5. The most prevalent household structure is a couple with children (38%), followed by a couple without children (25%). However, 22% of the sample lives alone. On average, 1 person in the household is currently employed and 0.6 unemployed. The average number of children under 15 living in the household is 0.4. Finally, almost half of the sample lives in a highly populated area (46%), with 38% living in a municipality with more than 100,000 inhabitants.

### *3.2. Regional characteristics*

Tourism demand is a partial demand model based on a multistage budgeting process that assumes consumer demand to be weakly separable in tourism and non-tourism goods (Deaton and Muellbauer, 1980). However, disposable income for travelling is likely to be affected by the general level of prices of non-tourism goods. To control for this, we collected monthly regional consumer price indexes (*rCPI*) in base 2016 from the Spanish National Statistics Institute. This variable thus allows us to analyse how tourism travelling depends on the evolution of price levels in the area where the respondent lives. We expect it to be negatively associated with the likelihood of tourism participation, ceteris paribus. Our empirical model includes regional fixed effects in the form of dummy variables (see Section 4). As such, the price levels are captured by these dummies and *rCPI* thus measures the effect of price deviations over time relative to the base level.

As discussed in Section 2, participation in tourism depends not only on personal characteristics but also on the phase of the business cycle and labour market prospects. We collect quarterly unemployment rates (*unemp_rate*) for each of the 17 Spanish regions for this purpose. This data is also provided by the Spanish National Statistics Institute. We expect the probability of tourism travelling to decrease with *unemp_rate*, in line with Bernini et al. (2017) and Alegre et al. (2019). We do not consider regional Gross Domestic Product for two reasons. First, it is expected to be highly correlated with unemployment rates. Second, regional GDP is only provided on an annual basis, which does not provide enough variation for appropriate identification if the specification includes regional fixed effects.

Consistent with the findings by Eugenio-Martín and Campos-Soria (2010; 2011), tourism participation seems to be affected by climate conditions at the place of origin. We measure it using two different but interrelated variables: i) heating degree days (HDD) and cooling degree days (CDD). These are two well-known variables used to quantify the demand for energy needed to heat or cool a building based on outside temperature. They are defined as the number of degrees that a day’s average temperature is below (above) a given threshold. Therefore, they capture how warm or cool an area is. Specifically, these two measures for a given month *t* are calculated as follows:

$$HDD_t = \begin{cases} \sum_{d=1}^{30} (18ºC - T_d) \; if \; T_d \leq \tau_{HDD} \\ 0 \; otherwise \end{cases}$$

$$CDD_t = \begin{cases} \sum_{d=1}^{30} (T_d - 21ºC) \; if \; T_d \geq \tau_{CDD} \\ 0 \; otherwise \end{cases} \tag{1}$$

where *d* refers to a given day in month *t*, $T_d$ is the air temperature at day *d*, and $\tau_{HDD}$ and $\tau_{CDD}$ are threshold values for each measure.

These two indexes are derived from meteorological observations of air temperature in more than 3,000 weather stations in Europe, interpolated to regular grids at 25-kilometre resolution. The data is collected daily and added up to a calendar month. The primary data is published by the Joint Research Centre (AGRI4CAST Resources Portal) and reproduced on a monthly basis by Eurostat. We use the corresponding data for each of the 17 Spanish regions during the 47-month window. Since the indexes are defined for all the regions using the same thresholds, they are directly comparable. Specifically, HDD uses 15ºC as the threshold point whereas CDD considers 24ºC.

The higher the values of HDD, the cooler the region is. Similarly, the higher the values of CDD, the warmer the region is. Since these indexes take value zero for mild temperatures, they offer the great advantage of capturing potential non-linear effects of temperature on tourism participation. In this sense, it is acknowledged that there is an inverted U-shaped relationship between temperatures and tourism demand (e.g. Rosselló-Nadal, 2014). Another advantage is that these indexes capture the within month variability in temperatures by summing the deviations from the benchmark level over days.

Table 2 presents descriptive statistics of these four variables.

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| *rCPI (index, base 2016)* | 101.55 | 1.68 | 97.95 | 105.62 |
| *Unemplrate (%)* | 17.51 | 5.65 | 7.16 | 33.62 |
| *HDD* | 127.22 | 127.42 | 0 | 443.36 |
| *CDD* | 21.72 | 45.35 | 0 | 228.68 |

Table 2.- Descriptive statistics of the regional characteristics

# 4. Empirical Modelling

### *4.1. Static model*

For the purpose of modelling participation in tourism, we propose the following panel Probit specification with individual random effects:

$$U_{it}^{*} = \alpha + \lambda Z_{jt} + \beta X_i + \delta T_t + \theta R_j + \varepsilon_{it} + \mu_i$$

$$y_{it} = \begin{cases} 1 & if\ U_{it}^{*} \geq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

where subscript *i* indexes individuals, *t* refers to the time period and *j* to the region where the individual lives, $U_{it}^{*}$ is the unobserved latent utility of travelling, $y_{it}$ is the observed binary indicator for whether individual *i* travels in period *t*, $\alpha$ is the intercept, $Z_{it}$ is the vector of regional time-varying characteristics, $X_i$ is a set of time-invariant individual characteristics, $T_t$ is a vector of time effects, $R_j$ is a set of regional fixed-effects, $\lambda, \beta,\ \delta$ and $\theta$ are vectors of parameters to be estimated, $\varepsilon_{it}$ is an idiosyncratic error term that is iid normally distributed with mean zero and variance $\sigma_e$, and $\mu_i$ is a mean-zero normally distributed individual-specific term with variance $\sigma_u$.

The idiosyncratic errors $\varepsilon_{it}$ are assumed to be serially independent once we condition on the individual effects $\mu_i$. Nevertheless, the composite error $\epsilon_{it} = \varepsilon_{it} + \mu_i$ is correlated over time due to the individual effects. This correlation is supposed to be constant and given by:

$$\rho = Corr(\epsilon_{it}, \epsilon_{is}) = \frac{\sigma_u}{\sigma_u + \sigma_e} \qquad \forall t \neq s \tag{3}$$

where $\sigma_e$ is normalized to 1 for identification.

As opposed to a pooled Probit, the specification of random effects captures unobserved individual-specific factors that affect the likelihood of tourism participation not captured in the regressors, such as health status or intrinsic taste for travelling[6]. This is relevant, since it allows us to obtain more accurate estimates of the effect of personal and regional characteristics on participation once unobserved heterogeneity is controlled for (Nolan, 2010).

Importantly, the random effects in (2) are assumed to be uncorrelated with the explanatory variables. However, this assumption might be restrictive, since it might happen that these

[6] In the context of a panel Probit specification, it is not advisable to specify the individual effects (unobserved heterogeneity at the individual level) as ‘fixed effects’ (parameters to be estimated). As discussed in Wooldridge (2002, p. 484), the use of a fixed effect estimator for the panel Probit produces an incidental parameter bias problem that leads to inconsistent estimates for T fixed and N large.

unobserved factors correlate with individual characteristics. In such case, the parameter estimates are biased. A common way to relax this is to adopt the Mundlak-Chamberlain approach (Mundlak, 1978; Chamberlain, 1984), generally known as the correlated random effects[7]. This consists of specifying the individual effects as a function of the time means of the time-varying variables as follows:

$$\mu_i = \pi_1 \bar{Z}_t + \pi_2 \bar{T}_t + \varsigma_i \tag{4}$$

where $\varsigma_i$ are iid normally distributed errors independent from $\bar{Z}_t$, $\bar{T}$ and $\varepsilon_{it}$ for all i, t. This formulation thus allows unobserved heterogeneity to be associated with the mean of observed regional and time factors. An example could be that the individual-specific taste for travelling depends on average climate conditions, the season within the year or the economic situation. As a result, the model specification of the correlated random effects Probit becomes:

$$U_{it}^{*} = \alpha + \lambda Z_{jt} + \beta X_i + \delta T_t + \theta R_j + \varepsilon_{it} + \pi_1 \bar{Z}_t + \pi_2 \bar{T}_t + \varsigma_i \tag{5}$$

A test comparing the standard random effects Probit with the correlated random effects Probit implies testing whether $\pi_1 = 0$ and $\pi_2 = 0$.

### *4.2. Dynamic model*

To explore the potential existence of state dependence, we expand the specification in (2) by introducing a lag of the dependent variable[8]. Therefore, the latent equation for the Dynamic Random Effects Probit model becomes:

$$U_{it}^{*} = \alpha + \gamma y_{it-1} + \lambda Z_{jt} + \beta X_i + \delta T_t + \theta R_j + \varepsilon_{it} + \mu_i \tag{6}$$

where $y_{it-1}$ is a binary indicator for whether the individual *i* travelled in period *t-1* that captures state dependence. The rest of variables are the same as before.

The estimation of the dynamic model in (6) faces the well-known 'initial conditions problem'. The initial period that the researcher observes whether each individual travels

---

[7] Wooldridge (2019) develops in detail the correlated random effects strategy for unbalanced panels and nonlinear models.

[8] When using aggregate data (e.g. tourism flows), the adjustment process to a long-run steady state is usually argued as the reason for the need of a dynamic specification. When working with microdata on binary outcomes, we model latent utility instead of aggregate demand. As such, the dynamic specification could be seen as the modelling of a continuum of probabilities of the steady-state tourism participation (see Section 5.2).

($y_{i1}$) does not correspond to the beginning of the stochastic process that drives the outcome. That is, individuals do not start their travelling history at the period they are observed for the first time. A naïve model that treats the initial response to be exogeneous produces inconsistent estimates (Skrondal and Rabe-Hasketh, 2014).

The econometrics literature has proposed mainly two ways to deal with the initial conditions problem. Heckman (1981a;1981b) suggests the joint modelling of the initial period and the subsequent ones. We disregard this approach because it requires a balanced panel. Wooldridge (2005) develops a simpler solution that conditions on the response at the initial period. Basically, Wooldridge proposes to model $y_{it}$ for $t = 2, \dots T$ assuming that the time-varying explanatory variables ($Z_{jt}$ and $T_t$ here) are strictly exogenous conditional on the individual-specific unobserved effects ($\mu_i$). He proposes the following auxiliary model:

$$\mu_i = \alpha_0 + \alpha_2 y_{i1} + \alpha_3 Z_{jt} + \alpha_4 T_t + \xi_i \qquad \text{for } t = 2, \dots, T \tag{7}$$

where $y_{i1}$ are the values of the dependent variable at the first observed period for each individual, and $\xi_i$ is a normally distributed individual effect with zero mean and variance $\sigma_\xi$ independent from $y_{i1}$, $Z_{jt}$, $T_t$ and the rest of regressors in (6) including $\varepsilon_{it}$.

Akay (2012) introduced a variant of this model that specifies within-individual means of $Z_{jt}$ and $T_t$ based on all the periods including the first. This way of proceeding has become popular for handling the initial conditions problem. However, Rabe-Hesketh and Skrondal (2013) have shown that Akay's procedure can be severely biased[9]. These authors propose a variant of Wooldridge's solution that specifies the individual-specific effect as follows:

$$\mu_i = \alpha_0 + \alpha_2 y_{i1} + \alpha_3 Z_{j1} + \alpha_4 \overline{Z_{jt}} + \alpha_5 T_1 + \alpha_6 \overline{T_t} + \xi_i \tag{8}$$

where $y_{i1}, Z_{j1}$ and $T_1$ are the first observed period of the dependent and the time-varying explanatory variables for each individual. As a result, the dynamic specification is given by:

$$U_{it}^* = \alpha + \gamma y_{it-1} + \lambda Z_{jt} + \beta X_i + \delta T_t + \theta R_j + \alpha_2 y_{i1} + \alpha_3 Z_{j1} + \alpha_4 \overline{Z_{jt}} + \alpha_5 T_1 + \alpha_6 \overline{T_t} + \varepsilon_{it} + \xi_i \tag{9}$$

[9] The reason is that Akay's model restricts the coefficients of the initial period of the explanatory variables ($Z_{j1}$ and $T_1$) to be equal to the coefficients for the means of the subsequent periods (Rabe-Hesketh and Skrondal, 2013), and this affects compliance with consistency requirements developed by Wooldridge (2005).

The models in (2), (5) and (9) are estimated in Stata 16 using the *xtprobit*, *xthybrid* (Schunck and Perales, 2017) and *xtpdyn* (Grotti and Cutuli, 2018) modules.

## 5. Results

### *5.1. Static model*

Column 1 in Table 3 presents the parameter estimates for the Random Effects Probit (RE Probit) specified in (2) together with their robust standard errors[10]. The model is estimated by Maximum Likelihood using Gauss-Hermite quadrature with 12 integration points[11]. To facilitate interpretation, Column 2 reports the Average Marginal Effects (AME)[12]. In column 3 we report the coefficient estimates for the Correlated Random Effects Probit (CRE Probit) specified in (5) and their robust standard errors. Column 4 presents the AME, which are derived following the formulas outlined in Wooldridge (2002, ch.15) and Wooldridge (2005). As seen, the AME from the two models are similar.

To capture time-invariant differences across spatial units arising from differences in transport infrastructures or the attractiveness of the surrounding area, all the regressions include a set of regional dummies (NUTS2) for the place of residence[13]. *Andalusia* is taken as the base category. Similarly, we also control for time effects, being *y2015* and *q1* the reference categories. The estimates are not reported to save space but are available upon request.

The means of *y2016*, *y2018*, *q2*, *q4*, *unemplrate*, *HDD* and *CDD* are statistically significant at conventional levels in the CRE Probit. Under the Mundlak-Chamberlain formulation, this implies that part of the unobservable heterogeneity is explained by the mean levels of the time effects and regional characteristics. Consequently, the assumption of independence between the explanatory variables and the individual-effect is quite restrictive and therefore the CRE Probit is preferable. Please note that these means are calculated *for each individual*. Since the panel is not balanced and individuals enter and exit the panel at different periods, the means of the time-varying regional variables exhibit variability across units and therefore are identified in addition to the regional fixed effects.

---

[10] We do not test for unit roots for two reasons. First, the number of individuals (N) is large relative to the number of periods (T). Indeed, although we use a 47-month window the maximum number of observed periods is 20. Second, most of the variables used in the analysis are binary, including the dependent variable.
[11] The likelihood function involves an integral that needs to be approximated by numerical analysis. The results are robust to the number of integration points.
[12] In line with Bland and Cook (2019), in the computation of the AME we integrate out the individual effects.
[13] Everything else being equal, those who live close to sightseeing spots, natural areas or the beach might have more incentives to make a tourist trip, either within the region or to a nearby one.

| | RE Probit | | | CRE Probit | | |
|---|---|---|---|---|---|---|
| | (1) | | (2) | (3) | | (4) |
| Variables | Coefficient | St. Error | AME (%) | Coefficient | St. Error | AME (%) |
| *rCPI* | -0.029*** | (0.006) | -0.304*** | -0.030*** | (0.006) | -0.305*** |
| *unemplrate* | 0.011*** | (0.003) | 0.112*** | 0.019*** | (0.004) | 0.199*** |
| *HDD* | -1.15e-04*** | (3.7e-05) | -1.19e-03*** | -2.0e-04*** | (0.000) | -1.7e-03*** |
| *CDD* | 0.001*** | (7.4e-06) | 0.013*** | 0.001*** | (0.000) | 0.014*** |
| Mean *rCPI* | | | | 0.014 | (0.014) | 0.148 |
| Mean *unemplrate* | | | | -0.025*** | (0.006) | -0.256*** |
| Mean *HDD* | | | | 3.0e-04** | (0.000) | 3e-03** |
| Mean *CDD* | | | | -0.001*** | (0.000) | -0.012*** |
| *age* | 0.008*** | (0.002) | -0.160*** | 0.007*** | (0.002) | -0.159*** |
| *agesq* | -2.13e-04*** | (1.65e-06) | | -2.0e-04*** | (0.000) | |
| *seceduc* | 0.233*** | (0.014) | 2.411*** | 0.235*** | (0.014) | 2.441*** |
| *higheduc* | 0.538*** | (0.016) | 5.569*** | 0.540*** | (0.016) | 5.602*** |
| *foreign* | -0.216*** | (0.016) | -2.233*** | -0.215*** | (0.016) | -2.231*** |
| *single* | -0.085*** | (0.014) | -0.876*** | -0.083*** | (0.014) | -0.874*** |
| *married* | 0.005 | (0.014) | 0.055 | 0.007 | (0.015) | 0.063 |
| *unemployed* | -0.147*** | (0.017) | -1.520*** | -0.149*** | (0.017) | -1.536*** |
| *retired* | 0.062*** | (0.015) | 0.640*** | 0.063*** | (0.015) | 0.633*** |
| *selfemployed* | -0.023 | (0.019) | -0.235 | -0.027 | (0.040) | -0.239 |
| *businessman* | 0.011 | (0.022) | 0.112 | -0.008 | (0.047) | 0.102 |
| *permemployed* | -0.019 | (0.014) | -0.201 | -0.038 | (0.031) | -0.202 |
| *tempempoyed* | -0.070*** | (0.017) | -0.721*** | -0.068*** | (0.019) | -0.719*** |
| *inc2* | 0.326*** | (0.012) | 3.371*** | 0.326*** | (0.012) | 3.365*** |
| *inc3* | 0.641*** | (0.013) | 6.628*** | 0.643*** | (0.013) | 6.636*** |
| *inc4* | 0.894*** | (0.016) | 9.249*** | 0.897*** | (0.016) | 9.261*** |
| *inc5* | 1.171*** | (0.019) | 12.108*** | 1.175*** | (0.019) | 12.128*** |
| *inc6* | 1.329*** | (0.027) | 13.744*** | 1.331*** | (0.027) | 13.746*** |
| *housesize* | -0.162*** | (0.007) | -1.673*** | -0.167*** | (0.007) | -1.679*** |
| *singleparentkids* | -0.151*** | (0.015) | -1.557*** | -0.150*** | (0.015) | -1.553*** |
| *couplenokids* | 0.012 | (0.014) | 0.120 | 0.013 | (0.014) | 0.121 |
| *couplewithkids* | -0.086*** | (0.015) | -0.891*** | -0.086*** | (0.015) | -0.887*** |
| *numunempl* | 0.036*** | (0.006) | 0.373*** | 0.004 | (0.014) | 0.374*** |
| *less15* | 0.062*** | (0.008) | 0.638*** | 0.067*** | (0.008) | 0.642*** |
| *highdensity* | 0.132*** | (0.018) | 1.364*** | 0.139*** | (0.018) | 1.449*** |
| *mediumdensity* | 0.052*** | (0.012) | 0.542*** | 0.057*** | (0.012) | 0.597*** |
| *mun2* | -0.072*** | (0.015) | -0.748*** | -0.070*** | (0.015) | -0.723*** |
| *mun3* | -0.126*** | (0.019) | -1.307*** | -0.121*** | (0.019) | -1.248*** |
| *q2* | 0.186*** | (0.013) | 1.927*** | 0.205*** | (0.014) | 2.117*** |
| *q3* | 0.414*** | (0.015) | 4.277*** | 0.445*** | (0.017) | 4.607*** |
| *q4* | 0.130*** | (0.016) | 1.346*** | 0.186*** | (0.018) | 1.921*** |
| σ | 0.810*** | (0.004) | | 0.655*** | (0.007) | |
| ρ | 0.396*** | (0.002) | | | | |
| Constant | 1.278** | (0.576) | | 0.481 | (1.243) | |
| Year dummies | | YES | | | YES | |
| Means of time varying regressors | | NO | | | YES | |
| Regional dummies | | YES | | | YES | |
| Log Likelihood | | -228,304 | | | -228,200 | |
| Observations | | 488,265 | | | 488,265 | |
| N individuals | | 92,472 | | | 92,472 | |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.- Coefficient estimates and AME for RE Probit and CRE Probit

Note: the omitted categories are $primeduc$, $sepdiv$, $widow$, $inactive$, $inc1$, $oneperson$, $otherh$, $lowdensity$, $mun1$, $y2015$, $q1$. The first period of time varying regressors include $unemplrate_1$, $rCPI_1$, $HDD_1$, $CDD_1$, $y2016_1$, $y2017_1$, $y2018_1$, $q2_1$, $q3_1$, $q4_1$. The means of time varying regressors include $\overline{unemplrate}$, $\overline{rCPI}$, $\overline{HDD}$, $\overline{CDD}$, $\overline{y2016}$, $\overline{y2017}$, $\overline{y2018}$, $\overline{q2}$, $\overline{q3}$, $\overline{q4}$.

Consistent with expectations, the probability of tourism participation is negatively associated with rises in regional price indexes (*rCPI*). As the general level of prices in the region increases, individuals are less likely to make a tourist trip because, for the same budget constraint, goods are relatively more expensive. This is in line with Bernini et al. (2017). Strikingly, tourism participation is negatively related to the mean level of unemployment at the region of residence but increases with temporary deviations from that level. Specifically, a marginal increase in the unemployment rate conditional on the mean level during the observed period for a given individual translates into a 0.20% higher probability of making a tourist trip. This result is counterintuitive and therefore deserves further discussion. This effect appears to be heterogeneous across regions. We interacted *unemplrate* with the regional dummies (available upon request) and we still document positive effects for Andalusia, Asturias, the Balearic Islands, Cantabria, Valencian Community, Extremadura, Galicia, Madrid, Murcia and the Basque Country but negative effects for Aragon, the Canary Islands, Castile and Leon, Castilla-LaMancha, Catalonia, Navarre and La Rioja. Nevertheless, bear in mind that the model already controls for individual and household labour status. As such, this variable is intended to capture the effect of variations in the economic situation of the region on individual decisions.

The positive effect of *unemplrate* is conditional on controls for year, quarter, and regional differences in the form of dummy variables. This effect is also positive in the baseline RE Probit that does not consider the within-unit mean. We alternatively estimated a Linear Probability Model (both considering fixed and random effects) with all the controls and the effect remains positive and significant. Since the regional unemployment rates have quarterly frequency rather than monthly, one might think the positive effect could be due to conflation with the quarter dummies. We repeated the estimation using dummies for months instead and the effect of unemployment remains unchanged, positive and significant.

A possible explanation for the positive relationship could be the following. In line with Eugenio-Martín and Campos-Soria (2014), cutback decisions in tourism are more likely among individuals living in areas with lower GDP growth and higher unemployment rates. If the change in tourism consumption involves substituting long-haul trips during vacation periods involving long stays by several short breaks to nearby locations, this could explain the positive association between regional unemployment and tourism participation. To explore this, we estimated a Zero Inflated Ordered Probit (ZIOP) model (Harris and Zhao, 2007) using *numtrips* as the dependent variable. The model distinguishes genuine zeroes (non-participants) from apparent zeroes (potential participants). The dependent variable (*numtrips*) is redefined as an ordered indicator with

5 categories (0,1,2,3,4 or more). Although the model does not account for the panel structure of the data, standard errors are clustered at the individual level. Results are presented in Tables A2 and A3 in Appendix. The marginal effects show that the probabilities of making two, three and four (or more) tourist trips per month increase with regional unemployment, *ceteris paribus*. By contrast, the likelihood of making just one decreases. Therefore, this auxiliary regression supports the notion that regional unemployment is associated with more trips, everything else being equal.

Concerning the effect of climate conditions at origin, we find that the probability of tourism participation increases with warm temperatures (CDD) and decreases with cold temperatures (HDD). Specifically, following the formulas in (1), the probability of tourism participation increases by 0.56% in a given month if during 10 days the temperature is 25ºC (0.014*10*(25-21)=0.56). Similarly, the probability of tourism participation in month *t* decreases by 0.07% if during 10 days the temperature is 10ºC (0.0017*10*(18-14)=0.068). Although these climate indexes are not comparable with the one used by Eugenio-Martín and Campos-Soria (2010, 2011), our results fall in line with theirs, showing that participation increases with good climate conditions at origin.

Consistent with previous findings, participation in tourism exhibits an inverted U-shaped relationship with age and increases with educational level. The likelihood of participation is lower among foreign and single individuals, whereas married status is not significant. As for labour status, unemployed people are significantly less likely to participate in tourism relative to inactive individuals. The opposite holds for retired people. Although elderly people are less likely to participate according to the negative coefficient of *agesq*, this result might account for the great success in Spain of IMSERSO, a subsidize program for travelling that target the young senior population (Losada et al., 2016). Interestingly, temporary employees have a lower propensity of making a tourism trip. This might reflect their higher labour instability.

Regarding household characteristics, participation increases with monthly income. For example, compared to a household that earns less than €1,000 per month, an individual earning between €1,500 and €2,500 has a 6.6% higher probability of making a tourist trip. Household size appears to be a factor that limits tourism participation, in line with Alegre et al. (2013) and Eugenio-Martín and Campos-Soria (2011). Compared to those living alone, couples with children and single parents with children exhibit a lower probability of making a tourist trip. Surprisingly, participation increases with the number of unemployed people in the household. This might account for more joint time availability. Additionally, households with minors participate more in tourism. As regards the characteristics of the municipality of residence, we see that participation is notably larger in big cities that are densely populated. This is consistent with earlier findings pointing to those living in metropolitan areas being in more need to relax (Eymann and Ronning, 1997; Nicolau and Más, 2005; Bernini et al., 2017). Higher accessibility to transportation hubs may be also behind this pattern. Finally, participation is larger in the summer period (*q3*).

*5.2 Dynamic model*

In Table 4 we report the coefficient estimates and robust standard errors for the Dynamic Random Effects Probit specified in (9) following Rabe-Hesketh and Skrondal (2013). Before discussing state dependence in detail, we document that the direction and significance of the rest of explanatory variables remains largely unchanged compared to Table 3, so we abstract from commenting them again.

The lag of the dependent variable is positive and significant, suggesting that there is positive genuine state dependence. Therefore, tourism participation in *t-1* increases the likelihood of participation in *t*. Importantly, this effect is net of all other sources of observable and unobservable heterogeneity across individuals. Interestingly, the initial condition ($y_{i1}$) is also positive and significant. This means that, as hypothesized, the initial conditions of the stochastic process are not exogenous to the individual effect and need to be explicitly accounted for in the model for consistency. A such, if we assume there is habit formation by which past trips increase the taste for travelling, the history of travel experience accumulated until the first observation period matters to a great deal for explaining the likelihood of participation.

| | Dynamic RE Probit | |
|---|---|---|
| Variables | Coefficient | St. Error |
| $y_{t-1}$ | 0.334*** | (0.009) |
| $y_1$ | 0.579*** | (0.009) |
| *rCPI* | 6.5e-06 | (0.008) |
| *unemplrate* | 0.008* | (0.005) |
| *HDD* | -7.8e-05* | (4.7e-05) |
| *CDD* | 0.002*** | (0.000) |
| $\text{rCPI}_1$ | 0.020** | (0.009) |
| $unemplrate_1$ | -0.017*** | (0.005) |
| $\text{HDD}_1$ | -1.6e-05 | (6.2e-05) |
| $\text{CDD}_1$ | -1.0e-04 | (1.1e-04) |
| Mean *rCPI* | -0.018 | (0.018) |
| Mean *unemplrate* | 0.004 | (0.008) |
| Mean *HDD* | 4.4e-05 | (1.4e-04) |
| Mean *CDD* | -0.001*** | (2-7e-04) |
| *age* | 0.009*** | (0.001) |
| *agesq* | -1.7e-04*** | (1.4e-05) |
| *seceduc* | 0.173*** | (0.013) |
| *higheduc* | 0.380*** | (0.014) |
| *foreign* | -0.143*** | (0.015) |
| *single* | -0.061*** | (0.013) |
| *married* | -0.012 | (0.014) |
| *unemployed* | -0.123*** | (0.015) |
| *retired* | 0.046*** | (0.013) |
| *selfemployed* | -0.032* | (0.017) |
| *businessman* | -0.001 | (0.020) |
| *permemployed* | -0.034*** | (0.013) |
| *tempempoyed* | -0.051*** | (0.016) |
| *inc2* | 0.250*** | (0.011) |
| *inc3* | 0.464*** | (0.012) |
| *inc4* | 0.627*** | (0.015) |
| *inc5* | 0.815*** | (0.017) |
| *inc6* | 0.916*** | (0.024) |
| *housesize* | -0.111*** | (0.006) |
| *singleparentkids* | -0.105*** | (0.014) |
| *couplenokids* | 0.022* | (0.012) |
| *couplewithkids* | -0.057*** | (0.013) |
| *numunempl* | 0.038*** | (0.006) |
| *less15* | 0.054*** | (0.007) |
| *highdensity* | 0.097*** | (0.016) |
| *mediumdensity* | 0.041*** | (0.010) |
| *mun2* | -0.042*** | (0.013) |
| *mun3* | -0.080*** | (0.017) |
| *q2* | 0.134*** | (0.017) |
| *q3* | 0.324*** | (0.020) |
| *q4* | 0.091*** | (0.023) |
| σ | 0.255*** | (0.007) |
| Constant | -1.381 | (1.230) |
| Year dummies | YES | |
| First period of time varying regressors | YES | |

| | |
|---|---|
| Means of time varying regressors | YES |
| Regional dummies | YES |
| Log Likelihood | -159,466.28 |
| Observations | 343,260 |
| N individuals | 92,472 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.- Coefficient estimates for dynamic RE Probit

Note: the omitted categories are $primeduc$, $sepdiv$, $widow$, $inactive$, $inc1$, $oneperson$, $otherh$, $lowdensity$, $mun1$, $y2015$, $q1$. The first period of time varying regressors include $unemplrate_1$, $rCPI_1$, $HDD_1$, $CDD_1$, $y2016_1$, $y2017_1$, $y2018_1$, $q2_1$, $q3_1$, $q4_1$. The means of time varying regressors include $\overline{unemplrate}$, $\overline{rCPI}$, $\overline{HDD}$, $\overline{CDD}$, $\overline{y2016}$, $\overline{y2017}$, $\overline{y2018}$, $\overline{q2}$, $\overline{q3}$, $\overline{q4}$.

To explore in more detail the dynamics of tourism participation, we now examine i) participation persistence, ii) the transitions into and out of tourism participation, iii) the expected spell duration and iv) the long-term steady state probability.

Participation persistence is defined as the probability of participation in period *t* conditional on having also participated in period *t-1* (i.e. $Prob(y_{it} = 1 | y_{it-1} = 1, X_{it})$, where $X_{it}$ gathers all the remaining regressors). The transition into participation (entry probability) is obtained as the probability of participating in period *t* conditional on not having participated in period *t-1* (i.e. $Prob(y_{it} = 1 | y_{it-1} = 0, X_{it})$). Contrariwise, the transition out of participation (exit probability) is the probability of non-participation in period *t* conditional on having participated in period *t-1* (i.e. $Prob(y_{it} = 0 | y_{it-1} = 1, X_{it})$)[14]. The expected duration of the spell (mean duration) is given by the inverse of the exit probability (i.e. $1/Prob(y_{it} = 0 | y_{it-1} = 1, X_{it})$). Finally, the proportion of time in which unit *i* continues participating (also referred to as steady state probability) is given by:

$$SE = \frac{Prob(y_{it} = 1 | y_{it-1} = 0, X_{it})}{(Prob(y_{it} = 1 | y_{it-1} = 0, X_{it}) + Prob(y_{it} = 0 | y_{it-1} = 1, X_{it}))} \tag{10}$$

Consistent with Biewen (2009), entry and exit probabilities are derived from the model estimates and then averaged across individuals. In doing so, unobserved heterogeneity is considered together with observed characteristics. However, as highlighted in Grotti and Cutuli (2018), these statistics are net from the inertia effects arising from unobserved heterogeneity, thereby capturing *genuine* state dependence. This constitutes a key point of the analysis. Table 5 presents the entry and exit probabilities, expected spell duration and steady state probabilities for the whole sample and also based on income and educational level.

[14] This is derived as $Prob(y_{it} = 0 | y_{it-1} = 1, X_{it}) = 1 - Prob(y_{it} = 1 | y_{it-1} = 1, X_{it})$

| | Persistence Prob. | Entry Prob. | Exit Prob. | Mean duration | Steady State Prob. |
|---|---|---|---|---|---|
| All | 0.289*** | 0.199*** | 0.710*** | 0.219 | 1.407 |
| *inc2* | 0.348*** | 0.250*** | 0.651*** | 1.534 | 0.277 |
| *inc3* | 0.388*** | 0.284*** | 0.611*** | 1.634 | 0.317 |
| *inc4* | 0.462*** | 0.350*** | 0.537*** | 1.860 | 0.394 |
| *inc5* | 0.540*** | 0.424*** | 0.459*** | 2.177 | 0.480 |
| *inc6* | 0.586*** | 0.470*** | 0.413*** | 2.415 | 0.532 |
| *seceduc* | 0.317*** | 0.224*** | 0.682*** | 1.465 | 0.247 |
| *higheduc* | 0.357*** | 0.253*** | 0.642*** | 1.556 | 0.283 |

*** p<0.01, ** p<0.05, * p<0.1

Table 5.- State dependence probabilities by income and education

As seen, the overall persistence probability (the average marginal effect of the lagged dependent variable) is 0.28. This means that those who travelled in month *t-1* have a 28% higher probability of also travelling in month *t*. Also interesting, the entry probability is 0.20, implying that those who did not make a tourist trip past month have a 20% probability of engaging in tourism participation in the current month. Net of other observable and unobservable factors, the dynamics of tourism participation are largely associated with household income and educational level. That is, it is not only that these two factors explain participation in a given period but participation spells over time. As shown, persistence probabilities monotonically increase with income and education.

In line with our discussion in Section 4.2, the part of unobserved heterogeneity in the dynamic model that is correlated with the explanatory variables is measured by the initial period of the dependent variable ($y_{i1}$), the initial period of the time-varying explanatory variables ($Z_{i1}$, $T_{i1}$) and the within-means of the time-varying explanatory variables ($\overline{Z_{\iota 1}}, \overline{T_{\iota 1}}$). For the purposes of evaluating how genuine state dependence (habit formation) varies with individual-specific heterogeneity, as a final exercise we evaluate exit and persistence probabilities at the quintiles of its distribution and at different initial conditions ($y_{i1} = 1$ and $y_{i1} = 0$). Hence, we are comparing habit formation in tourism among individuals with the same unobservables.

Figure 1 depicts the predicted probabilities over the quintiles of the distribution of unobserved heterogeneity[15]. We see that participants in the previous month exhibit always a larger the likelihood of participation in the current period than non-participants. Furthermore, there is a substantial gap in participation probabilities based on the initial conditions ($y_{i1}$). Since the initial conditions collapse the history of tourism traveling until the first observation period (i.e. travelling experience), this provides clear evidence of habit formation. Consistent with the cultivation of taste framework, individuals develop a taste for tourism travelling through consumption (i.e. past participation significantly raises the propensity to continue consuming). Whereas an individual that did not

[15] See Table A4 in Appendix for the specific figures.

participate neither in the first observation period nor in the previous one exhibits a 15% probability of participation, one that did participate both in the first period and in the previous month has an average likelihood of participation of 50%.

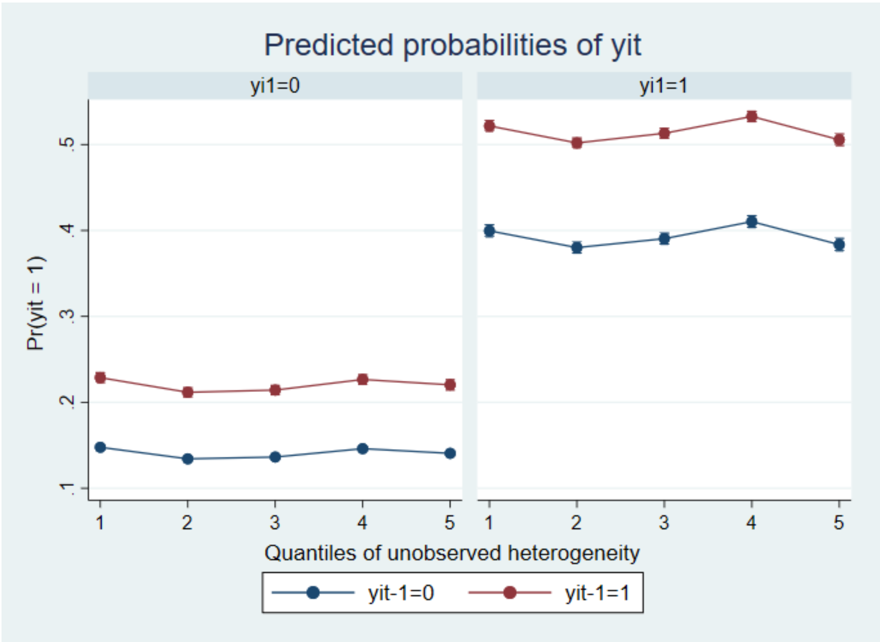


Figure 1.- Predicted probabilities of participation over the quintiles of the distribution of unobserved heterogeneity

## 6. Conclusions

### *6.1. Summary of findings*

This paper has studied the dynamics of tourism participation at the micro level. Given the experience nature of tourism activities, we have examined the existence of habit formation (state dependence) by which current engagement in tourist trips depends on past travelling. For this purpose, we have used a large longitudinal dataset involving more than 92,000 individuals in Spain. We have combined monthly microdata about their participation in tourism, individual and household characteristics with regional data on prices, unemployment rates and climate conditions.

We have estimated static and dynamic random effects Probit models that control for unobserved heterogeneity in the form of individual effects. In this way, our modelling framework allows us to disentangle genuine state dependence from spurious dependence emanating from omitted unobserved heterogeneity. We have adopted Mundlak-Chamberlain correlated random effects approach and Rabe-Hesketh and Skrondal (2013) procedure to deal with the initial conditions problem, respectively. Together with sociodemographic, household and regional characteristics, our model specification controls for time and regional effects.

We have found evidence of significant positive state dependence in tourism participation. Having made a tourist trip in month *t-1* increases the probability of also travelling in month *t* by 28 percentage points. Interestingly, we have found that the initial conditions (i.e. the accumulated trip experience at the first observation period) matter to a large extent in raising the propensity to engage in tourism activities. Although the entry probabilities from non-participants are around 20%, the persistence probabilities for current participants are larger over the whole distribution of unobserved heterogeneity. Also relevant, the likelihood of continuing participating in tourism is strongly linked to household income and educational level.

Our results also show that participation decreases with regional prices and the mean level of unemployment. However, deviations from the mean unemployment rate raise the propensity to make a tourist trip. We show that this effect accounts for a substitution pattern by which under economic downturns individuals substitute long-haul trips by a larger number of short breaks. We also find that participation increases with warm temperatures and decreases with bad weather conditions. Additionally, we show that participation i) exhibits an inverted U-shape relationship with age, ii) decreases with household size, unemployment status and holding a temporary job, and iii) is positively associated with educational level, household income, municipality size and the population density of the place of residence.

### *6.2. Policy implications*

Our results have some policy implications. Understanding the role of personal and regional characteristics in sustaining tourism participation together with potential habit formation is particularly important in the current context produced by COVID-19. Due to the pandemic outbreak and the uncertainties surrounding it, there has been a drop in the number of tourist trips everywhere, which is expected to continue in the short run. Policy makers and regional authorities are starting to launch campaigns aimed at encouraging people not to cancel or change their travelling plans. Identifying the profile of those for whom tourism travelling is a habit and who exhibit a high persistence in tourism participation might help in the development of these promotional strategies. Those who develop a taste for travelling in the form of participation persistence might be less deterred to travelling. Targeting this segment might thus of great relevance for the recovery of the sector.

### *6.3. Limitations and future research*

The study possesses some limitations. First, our analysis only considers the decision to make at least one tourist trip in the corresponding month, without distinguishing whether it is domestic or abroad or the specific trip purpose. Future research should explore whether the effect of individual and regional characteristics varies by type of destination and travel purpose. Second, our data does not provide us information on participation in other leisure activities. Future research should analyse whether tourism participation exhibits complementarities with other activities. Finally, although the study considers a 47-month window, it would be interesting to expand the analysis to a larger period to explore in greater detail how participation changes over the business cycle.

## REFERENCES


Adamowicz, W.L. (1994). Habit formation and variety seeking in a discrete choice model of recreational demand. *Journal of Agricultural and Resource Economics* 19(1): 19-31.

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics* 27(7): 1189–1204.

Albalate, D. and Fageda, X. (2016). High speed rail and tourism: empirical evidence from Spain. *Transportation Research Part A* 85: 174-185.

Alegre, J., Mateo, S. and Pou, L. (2009). Participation in tourism consumption and the intensity of participation: an analysis of their socio-demographic and economic determinants. *Tourism Economics* 15(3): 531-546.

Alegre, J., Mateo, S. and Pou, L. (2010). An analysis of households' appraisal of their budget constraints for potential participation in tourism. *Tourism Management* 31: 45-56.

Alegre, J., Mateo, S. and Pou, L. (2013). Tourism Participation and expenditure by Spanish households: the effects of the economic crisis and unemployment. *Tourism Management* 39: 37-49.

Alegre, J. and Pou, L. (2004). Micro-economic determinants of the probability of tourism consumption. *Tourism Economics* 10(2): 125-144.

Alegre, J., Pou, L. and Sard, M. (2019). High unemployment and tourism participation. *Current Issues in Tourism*, 22(10), 1138-1149.

Bernini, C. and Cracolici, M.F. (2015). Demographic change, tourism expenditure and life cycle behavior. *Tourism Management* 47: 191-205.

Bernini, C. and Cracolici, M.F. (2016). Is participation in the tourism market an opportunity for everyone? Some evidence from Italy. *Tourism Economics* 22: 57-79.

Bernini, C., Cracolici, M.F. and Nijkamp, P. (2017). Place-based attributes and spatial expenditures behavior in tourism. *Journal of Regional Science* 57(2): 218-244.

Bhuller, M., Brinch, C.N. and Königs, S. (2017). Time aggregation and state dependence in welfare receipt. *The Economic Journal* 127(604): 1833-1873.

Biewen, M. (2009). Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics* 24: 1095-1116.

Bland, J.R. and Cook, A.C. (2019). Random effects probit and logit: understanding predictions and marginal effects. *Applied Economics Letters* 26(2): 116-123.

Bronner, F. and de Hoog, R. (2014). Vacationers and the economic "double dip" in Europe. *Tourism Management* 40: 330-337.

Bronner, F. and de Hoog, R. (2016). Crisis resistance of tourist demand: the importance of quality of life. *Journal of Travel Research* 55(2): 190-204.

Cafiso, G., Cellini, R. and Cuccia, T. (2016). Do economic crises lead tourists to closer destinations? Italy at the time of the Great Depression. *Papers in Regional Science* 97(2): 369-386.

Campos-Soria, J.A., Inchausti-Sintes, F. and Eugenio-Martín, J.L. (2015). Understanding tourists' economizing strategies during the global economic crisis. *Tourism Management* 48: 164-173.

Capó-Parrilla, J., Riera, A. and Rosselló-Nadal, J. (2007). Tourism and long-term growth: a Spanish perspective. *Annals of Tourism Research* 34(3): 709-726.

Castiglione, C. and Infante, D. (2016). Rational addiction and cultural goods: the case of the Italian theatregoer. *Journal of Cultural Economics* 40: 163-190.

Carroll, C.D., Dynan, K.E. and Krane, S.D. (2003). Unemployment risk and precautionary wealth: evidence from households’ balance sheets. *The Review of Economics and Statistics* 85(3): 586-604.

Cellini, R. and Cuccia, T. (2015). The economic resilience of tourism industry in Italy: What the 'great recession' data show. *Tourism Management Perspectives* 16: 346-356.

Chamberlain, G. (1984). *Panel data*. In Handbook of Econometrics, ed. Z. Griliches and M. Intrilligator, vol. 2, 1247–1318. Amsterdam: North Holland.

Deaton, A.S. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review* 70(3): 312-326.

Dogru, T., Sirakaya-Turk, E. and Crouch, G.I. (2017). Remodeling international tourism demand: old theory and new evidence. *Tourism Management* 60: 47-55.

Erdem, T. and Sun, B. (2001). Testing for choice dynamics in Panel Data. *Journal of Business & Economic Statistics* 19(2): 142-152.

Eugenio-Martín, J.L. and Campos-Soria, J.A. (2010). Climate in the region of origin and destination choice in outbound tourism demand. *Tourism Management* 31: 744-753.

Eugenio-Martin, J.L. and Campos-Soria, J.A. (2011). Income and the substitution pattern between domestic and international tourism demand. *Applied Economics* 43(20): 2519-2531.

Eugenio-Martín, J.L. and Campos-Soria, J.A. (2014). Economic crisis and tourism expenditure cutback decision. *Annals of Tourism Research* 44: 53-73.

Eymann, A. and Ronning, G. (1997). Microeconometric models of tourists' destination choice. *Regional Science and Urban Economics* 27: 735-761.

Faber, B. and Gaubert, C. (2019). Tourism and Economic Development: Evidence from Mexico's Coastline. *American Economic Review* 109(6): 2245-2293.

Fleischer, A. and Pizam, A. (2002). Tourism constraints among Israeli seniors. *Annals of Tourism Research* 29(1): 106-123.

Grotti, R. and Cutuli, G. (2018). xtpdyn: A community-contributed command for fitting dynamic random-effects probit models with unobserved heterogeneity. *The Stata Journal* 18(4): 844-862.

Gunter, U. and Smeral, E. (2016). The decline of tourism income elasticities in a global context. *Tourism Economics* 22(3): 466–483.

Habibi, F. (2017). The determinants of inbound tourism to Malaysia: a panel data analysis. *Current Issues in Tourism* 20(9): 909-930.

Harris, M.N. and Zhao, X. (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics* 141: 1073–1099.

Heckman, J.J. (1981a). *The incidental parameters problem and the problem of initial conditions in estimating a discrete time — discrete data stochastic process*. In: Manski, C.F., McFadden, D.L. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 179–195.

Heckman, J.J. (1981b). *Heterogeneity and state dependence*. In Studies in labor markets (pp. 91-140). University of Chicago Press.

Heien, D. and Durham, C. (1991). A test of the habit formation hypothesis using household data. *The Review of Economics and Statistics* 73(2): 189-199.

Kim, H.R., Yi, C. and Jang, Y. (2019). Relationships among overseas travel, domestic travel, and day trips for latent tourists using longitudinal data. *Tourism Management* 72: 159-169.

Losada, N., Alén, E., Domínguez, T., and Nicolau, J.L. (2016). Travel frequency of senior tourists. *Tourism Management* 53: 88-95.

Li, H., Song, H., and Li, L. (2017). A Dynamic Panel Data Analysis of Climate and Tourism Demand: Additional Evidence. *Journal of Travel Research* 56(2): 158–171.

Li, Z., Shu, H., Tan, T., Huang, S.S. and Zha, J. (2020). Does the demographic structure affect outbound tourism demand? A panel smooth transition regression approach. *Journal of Travel Research* 59(5): 893-908.

Lorde, T., Li, G. and Airey, D. (2016). Modeling Caribbean tourism demand: an augmented gravity approach. *Journal of Travel Research* 55(7): 946-956.

McCain, R.A. (1981). Cultivation of taste, catastrophe theory, and the demand for works of art. *American Economic Review* 71(2): 332-334.

McConnell, K.E. (1999). Household Labor Market Choices and the Demand for Recreation. *Land Economics* 75(3): 466-477.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* 46: 69–85.

Naik, N.Y. and Moore, M.J. (1996). Habit formation and intertemporal substitution in individual food consumption. *The Review of Economics and Statistics* 78(2): 321-328.

Nicolau, J.L. and Más, F. J. (2005). Stochastic Modeling: a three-stage tourist choice process. *Annals of Tourism Research* 32(1): 49-69.

Nolan, A. (2010). A dynamic analysis of household car ownership. *Transportation Research Part A* 44: 446-455.

Nordström, J. (2005). Dynamic and stochastic structures in tourism demand modelling. *Empirical Economics* 30: 379-392.

Peng, B., Song, H., Crouch, G.I. and Witt, S.F. (2015). A Meta-analysis of international tourism demand elasticities. *Journal of Travel Research* 54(5): 611-633.

Pollak, R.A. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy* 78(4): 745-763.

Pollak, R.A. and Wales, T.J. (1981). Demographic analysis in demand analysis. *Econometrica* 49(6): 1533-1551.

Rabe-Hesketh, S. and Skrondal, A. (2013). Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem. *Economics Letters* 120: 346-349.

Rashidi, T.H. and Koo, T.T.R. (2016). An analysis of travel party composition and expenditure: a discrete-continuous model. *Annals of Tourism Research* 56: 48-64.

Reece, W.S. (2004). Are senior leisure travellers different? *Journal of Travel Research* 43: 11-18.

Rosselló-Nadal, J. (2014). How to evaluate the effects of climate change on tourism. *Tourism Management* 42: 334-340.

Santana-Gallego, M., Ledesma-Rodríguez, F. and Pérez-Rodríguez, J.V. (2011). Tourism and trade in OECD countries. A dynamic heterogeneous panel data analysis. *Empirical Economics* 41: 533-554.

Schunck, R. and Perales, F. (2017). Within- and between-cluster effects in generalized linear mixed models: a discussion of approaches and the xthybrid command. *The Stata Journal* 17(1): 89-115.

Seetaram, N. (2010). Use of dynamic panel cointegration approach to model international arrivals to Australia. *Journal of Travel Research* 49(4): 414-422.

Skrondal, A. and Rabe-Hesketh, S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society C: Applied Statistics* 63(2): 211-237.

Smeral, E. (2012). International tourism demand and the business cycle. *Annals of Tourism Research* 39(1): 379-400.

Smeral, E. (2014). Forecasting international tourism with due regard to asymmetric income effects. *Tourism Economics* 20(1): 61-72.

Smeral, E. (2017). Tourism forecasting performance considering the instability of demand elasticities. *Journal of Travel Research* 56(7): 913-926.

Song, H., Li, G., Witt, S. and Fei, B. (2010). Tourism demand modelling and forecasting: how should demand be measured? *Tourism Economics* 16(1): 63-81.

Stigler, G.J. and Becker, G.S. (1977). De Gustibus Non Est Disputandum. *American Economic Review* 67(2): 76-90.

Vesterberg, M. (2018). The effect of price on electricity contract choice. *Energy Economics* 69: 59-70.

Wang, L., Fang, B. and Law, R. (2018). Effect of air quality in the place of origin on outbound tourism demand: disposable income as a moderator. *Tourism Management* 68: 152-161.

Williams, J. (2004). Habit formation and college students' demand for alcohol. *Health Economics* 14(2): 119-134.

Wong, I.A., Fong, L.H.N. and Law, R. (2016). A longitudinal multilevel model of tourist outbound travel behavior and the dual-cycle model. *Journal of Travel Research* 55(7): 957-970.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Chapter 15.

Wooldridge, J.M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20(1): 39–54.

Wooldridge, J.M. (2019). Correlated random effects with unbalanced panels. *Journal of Econometrics* 211: 137-150.

Wu, L., Zhang, J., and Chikaraishi, M. (2013). Representing the influence of multiple social interactions on monthly tourism participation behavior. *Tourism Management* 36: 480-489.

# APPENDIX

| Time periods | Number of individuals | Share (%) | Tourism participation rate |
|---|---|---|---|
| T=3 | 9,656 | 10.44 | 0.266 |
| T=4 | 37,154 | 40.17 | 0.230 |
| T=5 | 655 | 0.70 | 0.254 |
| T=6 | 24,213 | 26.18 | 0.236 |
| T=7 | 5,496 | 5.94 | 0.274 |
| T=8 | 4,099 | 4.43 | 0.275 |
| T=9 | 1,357 | 1.46 | 0.326 |
| T=10 | 1,375 | 1.48 | 0.262 |
| T=11 | 592 | 0.64 | 0.281 |
| T=12 | 759 | 0.82 | 0.244 |
| T=13 | 359 | 0.38 | 0.247 |
| T=14 | 332 | 0.35 | 0.213 |
| T=15 | 171 | 0.18 | 0.220 |
| T=16 | 171 | 0.18 | 0.242 |
| T=17 | 120 | 0.13 | 0.257 |
| T=18 | 86 | 0.09 | 0.178 |
| T=19 | 63 | 0.07 | 0.173 |
| T=20 | 54 | 0.06 | 0.258 |
| Total | 92,472 | 100.00 | 0.246 |

Table A1.- Structure of the panel data

| | Dependent variable: *numtrips* | | | |
|---|---|---|---|---|
| | Outcome equation | | Inflation equation | |
| Explanatory variables | Coefficient | St. Error | Coefficient | St. Error |
| *rCPI* | -0.005 | (0.009) | -0.021*** | (0.007) |
| *unemplrate* | 0.031*** | (0.006) | -0.021*** | (0.005) |
| *HDD* | 0.000 | (0.000) | -0.001*** | (0.000) |
| *CDD* | 0.000 | (0.000) | 0.002*** | (0.000) |
| *age* | 0.030*** | (0.007) | -0.012** | (0.006) |
| *agesq* | -0.000*** | (0.000) | -0.000 | (0.000) |
| *seceduc* | -0.031 | (0.042) | 0.247*** | (0.027) |
| *higheduc* | -0.054 | (0.045) | 0.602*** | (0.043) |
| *foreign* | -0.194*** | (0.036) | -0.096*** | (0.030) |
| *single* | 0.244*** | (0.033) | -0.251*** | (0.039) |
| *married* | 0.202*** | (0.031) | -0.120*** | (0.030) |
| *unemployed* | 0.070 | (0.048) | -0.200*** | (0.037) |
| *retired* | 0.080** | (0.037) | 0.014 | (0.026) |
| *selfemployed* | 0.121** | (0.051) | -0.108*** | (0.041) |
| *businessman* | 0.044 | (0.054) | -0.003 | (0.045) |
| *permemployed* | 0.080* | (0.045) | -0.063* | (0.034) |
| *tempempoyed* | 0.214*** | (0.049) | -0.217*** | (0.041) |
| *inc2* | 0.086** | (0.034) | 0.264*** | (0.020) |
| *inc3* | 0.145*** | (0.035) | 0.555*** | (0.022) |
| *inc4* | 0.204*** | (0.042) | 0.816*** | (0.029) |
| *inc5* | 0.343*** | (0.051) | 1.034*** | (0.036) |
| *inc6* | 0.470*** | (0.058) | 1.129*** | (0.051) |
| *housesize* | -0.042*** | (0.016) | -0.151*** | (0.011) |
| *singleparentkids* | -0.112*** | (0.033) | -0.102*** | (0.025) |
| *couplenokids* | -0.197*** | (0.029) | 0.140*** | (0.030) |
| *couplewithkids* | -0.223*** | (0.034) | 0.049 | (0.030) |
| *numunempl* | 0.054*** | (0.015) | 0.009 | (0.013) |
| *less15* | -0.026 | (0.019) | 0.085*** | (0.014) |
| *highdensity* | 0.073** | (0.034) | 0.109*** | (0.029) |
| *mediumdensity* | 0.026 | (0.024) | 0.044** | (0.019) |
| *mun2* | -0.102*** | (0.026) | 0.008 | (0.025) |
| *mun3* | -0.112*** | (0.035) | -0.038 | (0.031) |
| Constant | | | 0.136 | (2.226) |
| Year dummies | | YES | | |
| Quarter dummies | | YES | | |
| Mean of time varying regressors | | YES | | |
| Regional dummies | | YES | | |
| Log likelihood | | -358,421.2 | | |
| Observations | | 488,265 | | |

Clustered standard errors at individual level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A2.- Parameter estimates for Zero-Inflated Ordered Probit

Note: the omitted categories are $primeduc$, $sepdiv$, $widow$, $inactive$, $inc1$, $oneperson$, $otherh$, $lowdensity$, $mun1$, $y2015$, $q1$. The first period of time varying regressors include $unemplrate_1$, $rCPI_1$, $HDD_1$, $CDD_1$, $y2016_1$, $y2017_1$, $y2018_1$, $q2_1$, $q3_1$, $q4_1$. The means of time varying regressors include $\overline{unemplrate}$, $\overline{rCPI}$, $\overline{HDD}$, $\overline{CDD}$, $\overline{y2016}$, $\overline{y2017}$, $\overline{y2018}$, $\overline{q2}$, $\overline{q3}$, $\overline{q4}$.

The inflation equation estimates a Probit on the probability of being a participant. The outcome equation estimates an Ordered Probit for those classified as participants.

| | AME for *unemplrate* (%) |
|---|---|
| Prob(*numtrips*=0) | -0.088 |
| Prob(*numtrips*=1) | -0.130** |
| Prob(*numtrips*=2) | 0.075*** |
| Prob(*numtrips*=3) | 0.041*** |
| Prob(*numtrips*≥4) | 0.101*** |

*** p<0.01, ** p<0.05, * p<0.1

Table A3.- Average Marginal Effects (in percentage) for *unemplrate*

| $y_{it-1}$ | $y_{i1}$ | quintile | Prob. |
|---|---|---|---|
| 0 | 0 | 1 | 0.147*** |
| 0 | 0 | 2 | 0.134*** |
| 0 | 0 | 3 | 0.136*** |
| 0 | 0 | 4 | 0.146*** |
| 0 | 0 | 5 | 0.140*** |
| 0 | 1 | 1 | 0.399*** |
| 0 | 1 | 2 | 0.380*** |
| 0 | 1 | 3 | 0.390*** |
| 0 | 1 | 4 | 0.410*** |
| 0 | 1 | 5 | 0.383*** |
| 1 | 0 | 1 | 0.228*** |
| 1 | 0 | 2 | 0.211*** |
| 1 | 0 | 3 | 0.214*** |
| 1 | 0 | 4 | 0.226*** |
| 1 | 0 | 5 | 0.220*** |
| 1 | 1 | 1 | 0.521*** |
| 1 | 1 | 2 | 0.501*** |
| 1 | 1 | 3 | 0.513*** |
| 1 | 1 | 4 | 0.532*** |
| 1 | 1 | 5 | 0.505*** |

*** $p<0.01$

Table A4.- Marginal effects for the probability of participation based on past participation, participation at the first period and quintile of the distribution of unobserved heterogeneity