"Determinants of Tourists' Length of Stay: A Hurdle Count Data Approach"

David Boto-García*[a], José F. Baños-Pino* and Antonio Alvarez*

*Department of Economics, University of Oviedo. Campus del Cristo, 33006, Oviedo (Spain)

[a]Corresponding autor: e-mail: botodavid@uniovi.es

Abstract:

This article analyses tourists' length of stay in a particular destination using a Hurdle Count Data model that allows us to firstly identify the determinants of the decision to be a same-day visitor or a tourist, and then explain the length of stay of those who actually stay for more than a day. Apart from sociodemographic characteristics, we are interested in the effects of distance, mode of transport and some relevant destination attributes of the destination such as tranquillity, natural environment or climate. Another feature this paper addresses is how advertising, recommendations and previous experience at the destination affect both the probability of an overnight stay and the length of the stay. The results indicate that the determinants of the decision to stay overnight and how long to stay are not the same. Besides, a positive previous experience and having seen advertising of the destination positively affect the decision to sleep at the destination and the number of days.

Keywords: *length of stay, tourist´s decision-making, conditional demand, Hurdle Negative Binomial model*

JEL codes: C35, D12, D81

Declaration of interest: none

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

1. INTRODUCTION

Length of stay at a tourist destination is one of the most relevant issues in the tourist decision-making process (Decrop and Snelders 2004). The economic impact of tourism basically depends on the number of days the tourist stays at the destination. In this sense, many studies have found evidence of a strong correlation between length of stay and total expenditure (Leones *et al.* 1998; Laesser and Crouch 2006). Because of this, uncovering the determinants of length of stay is critical for the proper design of marketing policies oriented to increase the revenues generated by tourism.

There are several studies in the literature that have examined the effects of sociodemographic characteristics, such as age, income, and nationality on length of stay (e.g. Barros and Machado 2010). Additionally, other scholars analyze the relationship between the number of days the visitor spends at a destination and the mode of transport used, the type of accommodation selected or the purpose of the trip (e.g. Alén *et al.* 2014). Another issue that deserves attention is the visitor's choice between being a same-day visitor or a tourist. In this sense, the determinants of the visitor's decision to stay overnight or not in a destination have been less studied in the literature, with the exception of Rodríguez et al. (2018). Given that tourism products are in essence experiential, tourists normally face a high level of uncertainty when deciding whether or not to stay overnight at a particular destination and how long to stay. Therefore, their knowledge about the characteristics of the destination will be a critical factor and justifies the interest of our analysis.

The aim of this research is twofold. Firstly, we analyze the determinants of the decision to stay overnight, differentiating between same-day visitors and tourists. Secondly, we model the length of the stay, focusing on the role that tourists' knowledge about the destination and the attributes they most value play in tourist decision-making. Specifically, we examine how a positive previous experience at the destination, looking for the natural environment or for tranquility and recommendations from friends or relatives (*word of mouth* effect) affect the length of the stay. Another issue of interest is how stay duration is connected with distance to origin and the chosen mode of transport.

This paper employs a pooled cross-sectional dataset of tourists visiting Asturias, a region located in northern Spain, during the period 2010-2016. For analyzing the effects of different sources of information on the length of stay, we estimate a hurdle count data model (Mullahy 1986). This methodology allows us to both identify the factors that determine the decision to stay overnight and how long to stay for those who spend at least one night in Asturias. From a methodological point of view, we consider two competing specifications for modelling the positive outcomes, namely, a Zero-Truncated Negative Binomial P (ZTNBP) and a Zero-Truncated Poisson-lognormal (ZTPN) model and compare them to determine which of them best fits the data. To the best of our knowledge, this is the first empirical paper that employs a hurdle count data model in tourism research that compares two alternative specifications of the unobserved heterogeneity for the positive outcomes.

Our results show that a recommendation from friends or relatives increases the probability of an overnight stay but has no effect on the length of the stay. Having seen some type of advertising about the destination and having had a positive previous experience there positively affect both the decision to stay overnight and the number of days. Furthermore, the climate and the natural environment are found to be the main destination attributes that increase the length of the stay. Booking the trip through travel agencies and lodging at hotels leads to the longest stays. Foreign visitors tend to stay longer than Spaniards, whereas education is not significant for explaining the number of days a tourist stays.

The paper is structured as follows. After this introductory section, we review the literature on this topic. We then present the theoretical model. The fourth section describes the database and the variables employed. In the fifth section we provide a brief methodological discussion and present the empirical model and the estimation procedure. The sixth section reports the results and discusses their implications. Finally, the last section outlines the main conclusions.

2. LITERATURE REVIEW

The economic relevance of tourism has sparked an increasing interest in analyzing its determinants. With regard to length of stay (hereafter LOS), in the last decade several studies have employed microeconometric regression models to analyze the effects of several explanatory variables. In what follows we discuss the main findings on the effects of tourists' sociodemographic characteristics, knowledge of destination, supply-based factors and trip-related features on LOS.

Sociodemographic characteristics

The empirical evidence about the effect of gender on LOS is mixed. Several studies find that female tourists stay longer (e.g. Rodríguez *et al.* 2018) whereas Barros and Machado (2010) and Machado (2010) find evidence of just the opposite. Moreover, other scholars do not find significant differences (e.g. Martínez-García and Raya 2008). As for the effect of age, several studies have found that LOS is positively associated with the tourist's age (e.g. Brida *et al.* 2013). Regarding labor status, the evidence is also inconclusive. Alegre and Pou (2006) show that highly-qualified workers display lower LOS whereas Martínez-García and Raya (2008) note that self-employed and low-level employees are the ones who tend to stay for a shorter time. Likewise, there is no consensus on the effect of education on LOS. While Barros and Machado (2010), Barros *et al.* (2010), Machado (2010) and Ferrer-Rosell *et al*. (2014) indicate that they are positively related, Gokovali *et al.* (2007), Gomes de Menezes *et al.* (2008), Martínez-García and Raya (2008) and Rodríguez *et al.* (2018) provide evidence of the contrary. Gomes de Menezes and Moniz (2011) and Brida *et al.* (2013) do not find significant effects and Oliveira-Santos *et al.* (2015) argues that the relationship between the level of education and the tourist's LOS seems to exhibit a complex pattern. When considering tourist nationality, most studies focus on LOS at a specific destination, so the effect of country's origin basically depends on the area being analyzed. In general, the literature agrees that tourists from further-away origins tend to stay longer (e.g. Oliveira-Santos *et al.* 2015). Regarding income, in general terms tourism is a normal good so that higher income leads to more extended stays. However, Rodríguez *et al.* (2018) find non-significant effects.

Supply-based factors

Another group of variables that seem to be relevant for explaining LOS are supply-based factors such as destination attributes and prices. According to Gokovali *et al.* (2007) and Gomes de Menezes *et al.* (2008), tourists who attach high importance to natural environment, landscape and beautiful surroundings display longer stays. In this sense, climate is one of the attributes that encourages tourists to stay for more extended periods (e.g Barros *et al.* 2008). In addition, some studies include tourist expenditures per day as a proxy of the price per stay (e.g. Alegre *et al.* 2011). As expected, they obtain a negative relationship with length of stay.

Authors (year)	Population, tourist destination and period	Methodology	Main results
Alegre and Pou (2006)	British and German tourists in the Balearic Islands (Spain) during the high seasons from 1993 to 2003.	Discrete logit model (0 if the tourist stayed for less than 7 days; 1 if he/she spent over a week).	Older people, travelling with a couple, mid-to-high accommodation, the number of yearly trips and the percentage of tourists that have previously visited the destination are the main factors that increase the probability of staying for more than 1 week.
Govokali *et al.* (2007)	Tourists who travelled to Bodrum (Turkey) by plane in the summer of 2005.	Duration models	The probability of staying increases with income, previous experience and party size but decreases with late accommodation, daily expenditures and high education.
Martínez-García and Raya (2008)	Low-cost travelers visiting Catalonia (Spain) in 2005.	Duration models	Type of accommodation, travelling in the high season and the level of education are quantitatively the most important factors when determining LOS.
Barros *et al.* (2008)	Portuguese tourists travelling to South America on charter flights.	Duration models.	The time spam a tourist stays at a destination is positively related to having booked in advance, having seen advertisements, previous visits and the frequency of travel.
Gomes de Menezes *et al.* (2008)	Tourists departing from the Azores (Portugal) in the summer of 2003.	Duration model	Repeat visitors and those who choose Azores due to its weather and remoteness stay for longer periods. Tourists who live far away from there (Nordic or German people) experience shorter stays.
Barros and Machado (2010)	Foreign tourists departing from Funchal Airport (Madeira Island).	Survival sample selection model proposed by Boehmke *et al.* (2006).	Age, gender, education and hotel quality increase LOS but expenditure reduces it. Besides, Germans stay longer than British, Dutch and French tourists.
Barros *et al.* (2010)	Golfers who visit the Algarve (Portugal) in the spring of 2004.	Duration model	Golfer´s LOS merely depends on nationality, education, age, the type of hotel where the individual stays, climate and the hospitality experience.
Machado (2010)	Homeward-bound foreign individuals departing from Madeira's Funchal Airport (Portugal) in 2008.	Duration model	LOS is positively related to age, gender, education, being German, and previous visits and negatively related to expenditure.
Gomes de Menezes and Moniz (2011)	Tourists departing from the Azores (Portugal) in the summer of 2003.	Duration models.	Educational level is not significant for explaining LOS. Besides, repeat visitors, taking charter flights and those who visit friends or relatives tend to exhibit longer stays.
Alegre *et al.* (2011)	British and German tourists in the Balearic Islands (Spain) during the high seasons from 1993 to 2003.	Latent class count data model with two groups based on the preference for short (a week) or long (2 weeks) stays.	For both segments, the price per day´s stay has a negative effect on the length of the stay, being the magnitude higher for the shorter-stay segment. The number of tourist trips per year also has a negative effect on LOS.

Thrane (2012)	Undergraduate students attending an Scandinavian University colleague in September-October 2007	OLS and duration models.	Tourists who booked the trip on the Internet, travelled in July and planned the trip further in advance stay longer. By contrast, as the daily expenditure per person increases, trips become shorter.
Thrane and Farstad (2012)	International visitors to Norway during the summer 2007.	OLS and duration models.	Tourists from neighboring countries to Norway appear to stay for shorter periods while on holiday in Norway than tourists from elsewhere in Europe.
Brida *et al.* (2013)	Visitors of the Archaeological Museum of Bolzano (Italy) from June to August 2010.	Count data (Zero Truncated Negative Binomial model).	Visitors under 30 tend to have a shorter vacation than other age categories. Hosting the Ötzi museum is the most valuable attribute for visiting the city. Bad weather and travel costs negatively influence LOS.
Alén *et al.* (2014)	Spanish residents over 55 in 2012.	Count data (Zero Truncated Negative Binomial model).	The variables that increase LOS are age, visiting friends or relatives, the climate attribute, accommodation in a holiday apartment or in a second residence, travelling alone and the IMSERSO type of holiday.
Grigolon *et al.* (2014)	Dutch tourists in the period 2002-2009.	Dynamic mixed multinomial logit model for panel data.	The effect of a particular vacation length made in the past affects travelers' choice of a future vacation with the same length.
Ferrer-Rosell et al. (2014)	Foreign visitors arriving by air to Spain in 2010.	Ordered logit model.	Low cost airlines users have slightly longer than legacy airline travelers. Tourists from the Benelux stay longer when they travel on package trips and behave similarly to UK visitors when they book the trip themselves.
Oliveira-Santos *et al.* (2015)	Visitors to Brazilian destinations between 2004-2010.	Shared heterogeneity duration model	Income does not have a significant effect on LOS; Asians and Oceanians are the ones who stay longer; tourists visiting two destinations stay shorter than those who visit only one, and the effect of party size is negative following a non-monotic path.
Nicolau *et al.* (2016)	Visitors to an Atlantic Coast destination of the United States.	Count data (Zero Truncated Negative Binomial Model).	As distance increases, LOS increases too in order to compensate for the effort made in the journey and to spread the fixed costs. First-time visitation has a significant positive effect on LOS, maybe due to the willing to widely explore a new destination.
Rodríguez *et al.* (2018)	Visitors to Santiago de Compostela (Spain)	Heckman selection model and separate Probit and Zero-Truncated OLS regression.	Young and retired people who visit Santiago for leisure purposes display a higher probability of being a same-day visitor, whereas labor-related visitors are the ones with the longest stays.

Table 1.- Studies on tourists' length of stay.

Knowledge of destination

When deciding how long to stay, tourists, and especially first-timers, face a substantial risk of making a bad decision as the specific characteristics of a destination are unknown until the individual arrives there (i.e., intangibility). This "experience good" nature of tourism (Mill and Morrison 2009) induces travelers to carry out extensive information search strategies (Roehl and Fesenmaier 1992). Consequently, some authors have included informational-type variables when explaining tourists' LOS. One of the most important ones is advertising, which reduces the consumer's search costs as it provides critical information to potential and current consumers. Woodside and Dubelaar (2002) indicate that advertising helps the individual to gain positive perceptions of the destination. Brochures and advertising of the destination seem to positively affect LOS (Rodríguez *et al.* 2018), being considered nowadays the most influential information source for prospective and current visitors (Kim *et al.* 2005; Park and Nicolau 2015). However, some scholars point out that individuals tend to rely more on recommendations from friends and relatives, as they perceive them as trustworthy (Bieger and Laesser 2004). In this sense, the well-known "word-of-mouth" effect is well-documented in the tourism industry (e.g. Luo and Zhong 2015) as a key element in tourist decision-making. In spite of this, Govokali et al. (2007) do not find a significant relationship with LOS.

Less information search is needed when the individual has previously been at the destination and has first-hand information. In this situation, the tourist has more confidence in the decisions made and the perceived risk is substantially lower (Kerstetter and Cho 2004). Nonetheless, the effects of previous experience at the destination has been widely analyzed by scholars without a clear conclusion. On the one hand, some studies have shown that first-time visitors stay longer (e.g. Nicolau *et al.* 2016). On the other hand, Gomes de Menezes and Moniz (2011) and Machado (2010) provide evidence suggesting that repeaters tend to stay more days, while Alegre *et al.* (2011), Oliveira-Santos *et al.* (2015) and Rodríguez *et al.* (2018) find the opposite. When researchers take into account not only whether or not the tourist has been to the destination before but also the number of times, a clearer picture emerges, with the number of previous visits to the destination being positively associated with LOS (e.g. Thrane and Farstad 2012).

Trip-related characteristics

The distance between the tourist's origin and the destination is another key factor in tourism demand (Bell and Leeworthy 1990). As Nicolau *et al.* (2016) state, "the literature shows little consensus about the effects of distance on length of stay at the destination". On the one hand, Taylor and Knudson (1976) argue that distance reduces utility as it entails physical, temporal and financial effort. Moreover, self-drivers or train riders may prefer to stop at various places along the way (e.g. Zillinger 2007). On the other hand, as travel costs are fixed, longer stays allow tourists to spread the costs over a longer period. When the mode of transport is taken into account, it appears that tourists who travel by public modes of transport tend to have longer stays (Rodríguez *et al.,* 2018).

Concerning the purpose of travel, some researchers find that tourists visiting friends or relatives stay for the longest periods (e.g. Oliveira-Santos *et al*. 2015). By contrast, others such as Rodríguez *et al.* (2018) argue that those who travel for business purposes are the ones who stay for the greatest number of days, whereas non-significant effects for travel purpose are found by Martínez-García and Raya (2008). Party size appears to exert a negative influence in LOS (e.g. Alén *et al.* 2014), with tourists who travel with friends staying fewer days than those who travel with a partner (e.g. Gomes de Menezes *et al.* 2008). As far as accommodation is concerned, Alegre and Pou (2006) indicate that tourists who lodge at higher-quality hotels stay longer than their lower-quality counterparts do, whereas Ferrer-Rosell *et al.* (2014) find just the opposite. A general conclusion is that those staying at hotels remain at the destination for the shortest period, with the longest stays associated with those dwelling at private accommodations (e.g. Oliveira-Santos *et al.* 2015).

Booking a package holiday is associated with longer stays according to Ferrer-Rosell *et al.* (2014) but is not found to be significant in Alegre and Pou (2006) and Martínez-García and Raya (2008). Finally, tourists visiting more than one destination stay for shorter periods at each one than those who spend their whole trip period at a single destination (Gomes de Menezes *et al.* 2008). In line with this, some researchers have found significant differences in tourists' LOS depending on the geographical area where they stay when visiting a certain region (Oliveira-Santos *et al.* 2015). As for seasonal differences along the years, LOS is longer during the high season (e.g. Grigolon *et al.* 2014).

As for the methodologies employed, different econometric strategies can be identified in tourists' length of stay the literature: OLS regression (Thrane and Farstad 2012), heckman model (Rodríguez *et al.* 2018), duration models (Gokovali *et al.* 2007; Martinez-Garcia and Raya 2008; Gomes de Menezes *et al.* 2008; Barros *et al.* 2008; Barros *et al.* 2010; Barros and Machado 2010; Machado 2010; Gomes de Menezes and Moniz 2011; Oliveira-Santos *et al.* 2015), binomial logit (Alegre and Pou 2006), ordered logit (Ferrer-Rosell *et al.* 2014), multinomial logit (Grigolon *et al.* 2014), nested logit (Nicolau and Más 2009), latent class (Alegre *et al.* 2011) and count data models (Brida *et al.* 2013; Alén *et al.* 2014; Nicolau *et al.* 2016). We believe this last methodology is the most suitable for modelling LOS and we discuss and justify it in Section 5. Whereas previous studies that used this approach for studying LOS specified a zero-truncated count data model for explaining the strictly positive stays, we extend it by including a previous hurdle that models the probability of being a tourist in comparison to be a same-day visitor.

In Table 1, a summary of some recent studies about tourists' LOS is presented. This table provides a description of the geographic area and the population under analysis, the methodology employed and the main findings.

3. EMPIRICAL MODEL

We build our empirical model on the random utility model framework (McFadden 1974; Manski 1977) – firstly applied to length of stay by Alegre and Pou (2006), by assuming that each individual chooses a destination j among a choice set S under a utility maximization criterion; thus, the utility of each destination j for each individual is given by:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1)$$

where V_{ij} is a deterministic component that can be explained by observable characteristics of both the individual and the destination, and ε_{ij} is a random error term for non-observable factors.

As the number of possible destinations the consumer considers is unknown, our analysis is conditional on the election of the observable destination j. If the individual's utility function is weakly separable, the conditional demand function for the length of stay at destination j given the chosen characteristics of the trip can be expressed as follows:

$$LOS = f(Price, Income, Pref, \omega) \quad (2)$$

This conditional demand function allows us to estimate the length of stay taking pre-fixed values of the selected destination and trip characteristics so that LOS explicitly depends on daily prices, income, consumer preferences and a random error term for non-observable characteristics (ω). Following this approach, we can decompose tourist's preferences (Pref) among sociodemographic characteristics (Soc), destination attributes (Attrib), destination knowledge (Knowledge) and trip-related characteristics (Trip).

4. DATABASE

Our analysis of tourist's LOS employs a pooled time series cross-section database of individuals visiting Asturias in the period 2010-2016. The *Tourist Information System of Asturias* conducts a detailed survey throughout the entire year to a representative sample of all visitors over 18 to the Principality of Asturias, a region in northern Spain of 10,604 km^2. Data were collected through personal interviews for a total of 33,461 individuals using a mixture of i) a quota random sampling procedure[1] based on type of visitor, type of accommodation, geographical area, day of the week and month and ii) a pure random sampling. The sample size was determined according to a 95% confidence level with a 5% error. The questionnaires were completed both on the street and in collective establishments all over the Asturian geography at different tourist sites. They were available in Spanish, German, English and French. The survey gathers microdata regarding the respondent's sociodemographic characteristics, travel motivation, places visited, total number of nights spent, mode of transport, place of origin, expenditure and type of accommodation, etc.

Visitors who stayed for more than 30 days were removed from the sample, as they should not be considered tourists (Hellström and Nordström 2008; Greene 2009). Moreover, local tourists (those who live in the region) were not considered in our analysis, as their behavior is quite different from those coming from other places (Bell and Leeworthy 1990). Since we are interested in the role of information about the destination on the tourist's length of stay due to uncertainty, it is not appropriate to consider residents in the sample. Therefore, our final sample consists of 19,111 individuals.

Asturias is a region characterized by its natural surroundings, its beautiful landscape and mild weather. It has experienced a notable increase in the number of visitors during the last decade, from six million in 2006 to more than seven in 2016. The tourist sector is currently one of the most important sectors for this region, representing 10 per cent of its Gross Domestic Product and 12 per cent of its total employment. The average length of stay continuously fell between 2010-2014, decreasing from 4.6 average days in 2010 to 4.26 in 2014. However, in 2015 and 2016 it increased to 4.61 and 4.52, respectively. Only 8.5 percent of the total visitors are same-day visitors, 34.7 percent have seen some type of advertisement regarding Asturias, 38.7 percent declare they have come for the first time and 35 percent consider novelty seeking as the main reason for coming. The principal trip

[1] In contrast to random sampling, quota sampling allows the sample to be properly representative of the total population under study, overcoming the possible selection bias that may arise with random sampling, as respondents are self-selected (See Santos-Silva (1997) for a discussion of this issue). In this sense, quota random sampling guarantees that each type of tourist is proportionally represented in the sample.

purpose is holiday/leisure (86 percent) and they mainly travel by car (82 percent) and as a couple (51 percent). Most visit Asturias in the second trimester (49 percent) and have organized the trip themselves (90 percent). The distance to their origin is, on average, 675 kilometers, although only 8.1 percent come from foreign countries. The average expenditure per person per day is 72 €, and the most chosen accommodation for those who spend at least one night is hotels (56.4 percent).

According to our theoretical model, we consider the following groups of explanatory variables:

- *Sociodemographic characteristics (Soc):* gender, age (both in levels and in a quadratic form), labor status (distinguishing among employed, self-employed, student, housewife/househusband, unemployed and retired), education level (primary, secondary and higher education) and nationality (Spaniard versus foreign). It is important to note that when we refer to nationality we mean the individual's country of residence. Therefore, the dummy variable *foreign* takes value one if the person does not live in Spain. Unfortunately, we lack data on income in our dataset. We are aware of its critical importance from an economic point of view, as the economic budget is a basic determinant of LOS. Given that we have information about age, education level and labor status, and according to the "Mincer earnings function" (Mincer, 1974), we proxy income with these three variables.
- *Supply-based factors (Attrib):* in the survey tourists are asked about the main reason for having chosen Asturias. They can choose among the following alternatives: novelty seeking, natural environment, heritage, tranquility, gastronomy and climate. All of them are defined as dummies. As for the daily prices, we consider the daily price paid for accommodation per person (denoted as *accom_price*) in euros. Lodging expenditures per day are a good proxy of the minimum cost of each day spent (Gokovali *et al.,* 2007).
- *Knowledge of destination (Knowledge):* in the survey, individuals are asked whether they have seen any type of advertising about Asturias, and whether they have previously visited this region. If so, they also report the total number of visits. Furthermore, they indicate their main reason for choosing Asturias, with two of the possibilities being recommendations from friends or relatives or a positive previous experience. We define the following variables: *advert* (which takes the value 1 if the tourist has seen any type of advertisement, regardless of whether it was via the internet, a brochure or a TV spot), *recommend* (if the individual declares that he/she has visited Asturias based on a recommendation), *first* (if it is the first time the

individual has visited Asturias), *num_vis* (which accounts for the number of visits made during the year, in order to control the frequency of visits as a normal habit) and *experience* (if the individual states that a positive previous visit is one of the reasons for returning).

- *Trip-related characteristics (Trip):* distance to origin (measured as the total number of kilometers from the tourist's residence to Oviedo (the centroid), also considered in a squared form to allow for further flexibility), mode of transport to reach the destination (by car, bus, train, or plane), the purpose of the trip (leisure, labor-related, visiting relatives or other options, such as for sport events, doctor visits, religious peregrination or making purchases), party size (number of members in the travel group), trip companions (alone, as a couple or in a group), type of accommodation (hotel, rural house, hostel, campsite or private accommodation), how the trip was organized (the individual did it himself, through a travel agency or the company where he works/a club to whom he belongs organized it for him), if the individual only visits Asturias in this trip or not, and if the tourist conducts active tourism activities or not. We also consider temporal factors (Temp) that may influence the decision regarding how long to stay. Specifically, we control for the year and the trimester when the visit takes place. Additionally, we also control for the regional area (Area) where the tourist stays (distinguishing five different areas: west area, central area, capital city area, east coast and east inner).

Dummy variables were created for each categorical variable. Annex 1 presents the descriptive statistics of all the variables employed in the analysis, their acronym and definition.

5. METHODOLOGY

Most of the literature about tourists' length of stay has employed duration models (Gokovali *et al.* 2007; Martínez-García and Raya 2008; Barros and Machado 2010; Barros *et al.* 2010; Gomes de Menezes and Moniz 2011; Oliveira-Santos *et al.* 2015). The relevant issue in this type of models when applied to tourism demand is not the duration of the trip but the probability of ending the stay at period t, conditional on having stayed in the destination until that moment (Kiefer, 1988, p.651). However, the use of duration models does not seem to be the best way to model tourists' length of stay. As Thrane (2012) criticizes, tourist´s LOS can hardly be understood as a process by which, in each period, visitors face a real "risk" of leaving the destination. In fact, most tourists decide their trip duration in advance and, consequently, they have booked accommodation and transport for specific dates. The arrival and departure dates have “de facto” been previously decided. Therefore, it makes little sense to apply duration models for studying tourists’ LOS.

The variable of interest (total number of overnight stays at a given destination) is assumed to be discrete and non-negative, so that LOS $\in N = \{0,1,2,\ldots\}$. We believe that its modeling could be better characterized by count data models (Hellerstein and Mendelsohn 1993), which have been previously used in the related literature (Alegre *et al.* 2011; Brida *et al.* 2013; Alén *et al.* 2014; Nicolau *et al.* 2016).

One of the basic assumptions of the standard count data models is that both zeros and positive values of the dependent variable come from the same Data Generating Process. However, in our study case, it makes sense to assume that there are two types of visitors: those who spend the night at the destination (tourists) and those who do not (same-day visitors). In this sense, Mullahy (1986) suggested that the effect of the different covariates on the probability of participation and on the intensity (number of positive counts) should not be restricted to being equal. To do so, it seems necessary to firstly separate participants from non-participants through a binary model, and then, in a second step, model the number of days they stay conditional on participation using a count data model. This model is known as the hurdle model and has been widely applied in the economic literature, especially in health (e.g. Sarma and Simpson 2006) and environmental economics (e.g. Bilgic and Florkowski 2007). However, it has not been employed in the tourism context to date.

The hurdle model can be constructed as follows:

a) Participation equation: we define a latent participation variable (d_i^*), which is given by a set of explanatory variables Z_i.

$$d_i^* = Z_i\gamma + u_i \tag{3}$$

where $i = 1,\ldots,N$ indexes the N observations in the sample and u_i is a random error term. We assume it follows a logistic distribution, which results in the Logit model, although Probit is a common alternative. The observation mechanism assigns d_i=1 if d_i^*>0, and di=0 otherwise. The probabilities for each alternative are given by:

$$P(d_i=1|Zi) = P(d_i^*>0) = \frac{1}{1+e^{-Z_i\gamma}}$$
$$P(d_i=0|Zi) = P(d_i^*\leq 0) = 1 - \frac{1}{1+e^{-Z_i\gamma}} \tag{4}$$

b) Intensity equation: the positive values of the dependent variable come from a zero-truncated count data model.

Therefore, maximizing the hurdle log L is equivalent of maximizing both log L functions separately. For modelling the number of nights spent at the destination in the intensity equation, we start from the benchmark Poisson model. One of its main limitations is that it imposes the conditional mean and variance to be equal (equidispersion property). This assumption is quite restrictive and is commonly violated in applied work, generating the *over*dispersion problem, by which the conditional variance exceeds the conditional mean. Hence, researchers normally seek better alternatives to the Poisson model. The *over*dispersion problem seems to be present in our data since the mean of the LOS is 4.31, whereas its variance is 14.66. We will then test this formally (see Section 6).

Cameron and Trivedi (2009) indicate that the *over*dispersion problem arises due to the presence of unobserved heterogeneity, suggesting the need for a new specification in which the error term adequately represents unobservable or omitted variables. The econometric literature has proposed several alternatives. The most common one is to introduce multiplicative randomness (v) in the Poisson model. We now proceed to introduce two alternative models depending on the assumption about the distribution of the unobserved heterogeneity.

5.1. The Negative binomial model: a Poisson-gamma mixture.

Suppose we specify v such that E(v)=1 and Var(v)= σ^2. Then, the first two moments of the dependent variable are given by:

$$E(LOS_i|X_i)= e^{Xi\beta+v}= \lambda_i h_i \tag{5}$$

$$Var\ (LOS_i|X_i)= \lambda_i\ (1+ \lambda_i\, \sigma^2) \tag{6}$$

where h_i= exp(v) and X_i is a vector of covariates that explain the length of stay. We assume a constant term in the model.

In the particular case that v ~ Gamma (1, α), we obtain the Negative Binomial (NB) model (also known as Poisson-gamma mixture model), which is regarded as more flexible and suitable for empirical research (Gurmu and Trivedi 1992; Winkelmann and Zimmermann 1995). The probability mass distribution of the NB is given by:

$$\text{Prob}\ (Y=y_i|X_i)=\frac{\Gamma\ (\theta+y_i)r_i^{\theta}(1-r_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)}, \tag{7}$$

where $\Gamma(.)$ denotes the gamma integral that specializes to a factorial for an integer argument, $y_i= 0,1,\dots$, $\theta>0$ so that $\theta=1/\alpha$, and $r_i=\theta/(\theta + \lambda_i)$.

The introduction of latent heterogeneity induces *over*dispersion while preserving the conditional mean as $E(v)=1$. Therefore, the conditional variance is expressed as follows:

$$\text{Var}\ (LOS|X_i)= \lambda_i\,(1+\alpha\lambda_i^{P-1}) \tag{8}$$

When P takes the values 1 and 2 we obtain the well-known NB1 and NB2 models (Cameron and Trivedi 1986; Gurmu and Trivedi 1996). The former specifies a linear variance function, whereas the latter considers a quadratic variance function. Cameron and Trivedi (1998) also note that other exponents apart from 1 and 2 in the conditional variance are possible (p.73). By replacing θ with $\theta\ \lambda_i^{2-P}$ in the probability mass function, we obtain the NBP model, whose probability mass function is then given by:

$$\text{Prob}\ (Y=y_i|X_i)=\frac{\Gamma\left(\theta\,\lambda_i^{2-P}+y_i\right)s_i^{\theta\,\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta\,\lambda_i^{2-P})}, \tag{9}$$

being $y_i= 0,1,\dots$; and $s_i = \frac{\lambda_i}{\lambda_i+\theta\,\lambda_i^{2-P}}$

Greene (2008) suggests that as the NBP model estimates the parameter P endogenously, this model is likely to be the preferable alternative among the negative binomial family. Although a quadratic conditional variance (NB2) often works well in empirical research, it may be badly specified in case the true P is higher than 2. For these reasons, we estimate the general NBP model to explain the positive outcomes. In case the estimated value of P is 1 or 2, the NBP model reduces to the classical NB1 and NB2 variants. As LOS is necessarily a positive variable, it is necessary to truncate the distribution of the dependent variable. Therefore, we model the intensity equation in terms of a Zero Truncated Negative Binomial P Model (in the following ZTNBP)[2].

$$LOS^*|X_i \sim \text{ZTNBP} \tag{10}$$

Its truncated probability mass function will be given by dividing the probability function by Prob (y_i>0|X_i):

[2] The reason why we truncate the distribution after having introduced the latent heterogeneity (v) as gamma distributed is not innocuous. In order to have a closed-form solution of the truncated models based on the NB distribution, it is required to perform the mixing first.

$$\text{Prob (Y=y}_i|\text{ y}_i > 0,\text{ X}_i) = \frac{\text{Prob (Y=}y_i|\text{X}_i)}{\text{Prob (}y_i > 0|\text{X}_i)} = \frac{\frac{\Gamma\left(\theta\,\lambda_i^{2-P}+y_i\right)s_i^{\theta\,\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta\,\lambda_i^{2-P})}}{1-(1+\alpha\lambda\text{i})^{-\alpha^{-1}}} \tag{11}$$

Being $y_i = 0,1,\ldots$, $s_i = \frac{\lambda_i}{\lambda_i+\theta\,\lambda_i^{2-P}}$, and Prob ($y_i$ >0|X_i)= $1-(1+\alpha\lambda_i)^{-\alpha^{-1}}$

Estimation of the Zero Truncated Negative Binomial P model is conducted by maximum likelihood. As the log-likelihood function to maximize is not globally concave and there is no certainty of a unique maximum, the estimates of the truncated NB2 model were used as starting points. An application of this general Hurdle Negative Binomial model can be found in Farbmacher (2013).

5.2. The Poisson lognormal mixture model

Instead of assuming that the multiplicative randomness (v) follows a gamma distribution, another alternative is to suppose that it is normally distributed with a zero mean and σ standard deviation. The Zero-Truncated Poisson Log Normal (ZTPN) model conditioning on both X_i and v is given by the following:

$$\text{Prob (Y=y}_i|\text{y}_i>0,\text{ X}_i, v) = \frac{\exp(-h_i\lambda_i)(-h_i\lambda_i)^{y_i}}{\{1-\exp(-h_i\lambda_i)\}\,y_i!}, \tag{12}$$

where $h_i\lambda_i = \exp(X_i\beta+\sigma v)$, $v \sim N(0,1)$

The density of y_i conditioning on X_i is as follows:

$$\text{Prob (Y=y}_i|\text{ y}_i > 0,\text{ X}_i) = \int_{-\infty}^{\infty} \text{Prob (Y} = \text{y}_i|\text{y}_i\ 0, \text{X}_i, v)\ \phi(v)dv \tag{15}$$

The conditional variance in the log normal model is quadratic in the conditional mean, so it accounts for *over*dispersion in the same way as the commonly applied NB2 model does (Greene 2009). The integrals in the log likelihood function do not exist in closed form, so parameter estimation needs to be conducted by Gauss-Hermite quadrature using the BHHH estimator after having reparametrized the log likelihood following Butler and Moffitt (1982).

Greene (2009) argues that the log normal model seems to be a more natural specification than the Poisson-gamma mixture. The reason is that if v captures unobserved heterogeneity across the sample, then the normality of v can be established by central limit theorems (Winkelmann 2008). Several authors point out that the normal distribution would be a preferable alternative for the unobserved heterogeneity instead of the traditional gamma

(Riphahn *et al.* 2003; Winkelman 2004). In any case, which is the most suitable model needs to be tested empirically for each dataset.

6. RESULTS

Before discussing the estimated coefficients, we must first choose which of the two alternatives for the intensity equation fits the data best. Model choice plays a critical role in our research, as the marginal effects that we will present later and the policy implications that can be derived from them depend crucially on the estimated parameters and, consequently, on the empirical model.

The Vuong test (Vuong 1989) is the most commonly employed test for statistically discriminating between non-nested models. However, ZTNBP and ZTPN are not strictly non-nested, but overlapping, as they collapse to Zero Truncated Poisson when $\alpha=0$ and $\sigma=0$, respectively. Because of this, we employ the HPC test proposed by Santos-Silva *et al.* (2015). These authors develop a testing procedure based on Davidson and MacKinnon´s (1981) seminal work, which basically discriminates between two models by checking whether the conditional expectation of the dependent variable under the alternative outperforms the corresponding conditional mean under the null. If so, we reject the null as the alternative improves the prediction of the outcome.

Table 2 displays the results of the HPC test. As Santos-Silva *et al.* (2015) suggest, we reverse the roles of the null and the alternative so that model choice does not depend on which one you compare against the other. The test clearly indicates that the ZTNBP model fits our data best and it is thus the chosen one.

Model comparison	t-Statistic (p-value)	Selected model
ZTNBP vs ZTPN	-3.187 (0.99)	ZTNBP
ZTPN vs ZTNBP	4.603 (0.00)	ZTNBP

Table 2.- Santos-Silva et al. (2015) HPC test for choosing the proper specification.

Table 3 reports the estimation results of the hurdle count data model to explain LOS. The first column shows the estimates of the binary Logit model for the participation decision, whereas the second refers to the chosen alternative for modelling the intensity, namely, the Zero Truncated Negative Binomial P model (ZTNBP). The α parameter, which accounts for the *over*dispersion phenomenon in its corresponding conditional variance, is statistically significant at the 1 percent level, providing evidence of the necessity of accounting for unobserved heterogeneity when modelling the length of stay in our sample (Cameron and

Trivedi 1998). The estimated value of the parameter P in the ZTNBP model is 3.65 and is statistically significant. This value is quite far from the imposed 1 and 2 corresponding to the ZTNB1 and ZTNB2 alternatives, indicating the need for allowing the model to be flexible when estimating the structure of the conditional variance instead of exogenously imposing it.

Starting with the sociodemographic characteristics, gender is not significant in either the participation or the intensity equations, which is in line with most of the literature (Martínez-García and Raya 2008; Brida *et al.* 2013). In the same vein, age is not significant as an explanation of the overnight stay decision (it is significant only at the 10 percent level). It is, however, positively related with the number of days spent, though at a decreasing rate given the negative coefficient of the squared term. This is in line with Fleischer and Pizam (2002), who found a concave relationship between age and length of stay. Regarding labor status, we set *self-employed* as the reference category. Our estimations indicate that these individuals display the highest probability of staying overnight. Regarding the length of the stay, retired people (*retired*), students (*student*), unemployed people (*unemployed*) and housewives (*housewife*) stay longer than self-employed individuals. As for the education level, compared to primary education (reference category), visitors with secondary and higher-level studies (*secondary* and *high* respectively) have a higher probability of sleeping in Asturias, though neither of them is significant in explaining the intensity of the stay. It is important to highlight here that these last three variables (age, educational level and labor status) may also account for income differences among individuals. With reference to nationality, we differentiate between people who live in Spain and those who do not with the dummy variable *foreign*. It seems that foreign individuals do not display a statistically different probability of an overnight stay to Spaniards. However, conditional on having decided to stay, they tend to stay for longer.

The motivations to visit the selected destination are also crucial elements to consider when explaining LOS. Our empirical estimations find that the appealing attributes Asturias provides, such as tranquility (*tranquility*), the natural environment (*natural*), and its oceanic weather (*climate*), positively influence both the length of the stay and the probability of staying overnight. However, those who indicate that their main reason for travelling was either novelty seeking (*novelty_seeking*) or its gastronomy (*gastronomy*) display a higher probability of spending the night but do not stay for significantly longer.

Dependent variable: LOS	Participation	Intensity
Independent variables	Logit	ZTNBP
man	-0.0875	-0.0064
	(0.059)	(0.011)
age	0.0297*	0.0180***
	(0.017)	(0.003)
age^2	-0.0003	-0.0001***
	(0.000)	(4.57e-05)
housewife	-0.5183***	0.0776**
	(0.154)	(0.033)
retired	-0.3100*	0.0789**
	(0.179)	(0.039)
employed	-0.2051**	-0.0171
	(0.086)	(0.015)
student	-0.4648***	0.0609**
	(0.144)	(0.029)
unemployed	-0.6361***	0.1339***
	(0.191)	(0.046)
secondary	0.2311**	0.0329
	(0.106)	(0.025)
high	0.3983***	0.0108
	(0.106)	(0.024)
foreign	-0.1692	0.2099***
	(0.131)	(0.031)
natural	1.4149***	0.0547**
	(0.109)	(0.027)
novelty_seeking	1.3746***	-0.0050
	(0.095)	(0.025)
tranquility	1.2438***	0.0972*
	(0.313)	(0.058)
climate	1.4340***	0.1828***
	(0.418)	(0.054)
gastronomy	0.6610***	-0.0204
	(0.215)	(0.073)
accom_price		-0.0003
		(0.000)
advert	0.3482***	0.0297**
	(0.067)	(0.012)
recommend	1.0665***	-0.0044
	(0.116)	(0.028)
first	0.3501***	0.0779***
	(0.079)	(0.014)
num_year_vis	-0.0266***	0.0005
	(0.003)	(0.003)
experience	1.3585***	0.0533**

	(0.089)	(0.026)
distance	0.0008***	0.0002***
	(0.000)	(3.72e-05)
Dependent variable: LOS	Participation	Intensity
Independent variables	Logit	ZTNBP
distance^2	-7.74e-08**	-2.71e-08***
	(3.29e-08)	(3.91e-09)
bus	-1.0888***	-0.0572
	(0.243)	(0.040)
train	1.2759***	0.0412
	(0.412)	(0.039)
plane	0.5714***	-0.0053
	(0.196)	(0.027)
leisure	-0.8672***	0.1767***
	(0.182)	(0.045)
labor	-0.3875*	0.2510***
	(0.218)	(0.070)
family	-0.1984	0.1603***
	(0.210)	(0.050)
party_size	-0.0002	0.0002
	(0.005)	(0.001)
alone	0.6313***	-0.0055
	(0.174)	(0.037)
couple	0.2639***	-0.0365***
	(0.061)	(0.012)
hotel		-0.3322***
		(0.027)
hostel		-0.0858**
		(0.033)
rural		-0.1527***
		(0.028)
private		0.2716***
		(0.032)
travel-agency	0.8898***	0.1213***
	(0.270)	(0.026)
club_comp	0.6747***	-0.1543**
	(0.198)	(0.062)
only_ast		0.0907***
		(0.016)
act_tour	1.5424***	0.1481***
	(0.201)	(0.019)
y11	0.6240***	-0.1212***
	(0.120)	(0.020)
y12	0.1460	-0.1282***
	(0.107)	(0.021)

y13	0.1176	-0.1328***
	(0.108)	(0.023)
y14	0.2451**	-0.0848***
	(0.109)	(0.020)
y15	-0.1078	-0.0290
	(0.102)	(0.021)
Dependent variable: LOS	Participation	Intensity
Independent variables	Logit	ZTNBP
y16	0.1456	-0.0534**
	(0.112)	(0.024)
t2	0.3751***	0.4000***
	(0.073)	(0.014)
t3	0.3456***	0.1377***
	(0.076)	(0.017)
west	-1.6733***	0.0816***
	(0.083)	(0.016)
centralr	0.0021	-0.0027
	(0.183)	(0.027)
east_inner	-1.3624***	0.0591***
	(0.085)	(0.017)
east_coast	-0.9123***	0.1236***
	(0.090)	(0.015)
constant	1.0192**	0.4412***
	(0.443)	(0.102)
alpha		0.0107***
		(0.001)
P		3.6564***
		(0.118)
Log L	-4,654.679	-38,7764.976
Observations	19,111	17,478

Table 3.- Estimated coefficients of the hurdle model (robust standard errors in parentheses).

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

Contrary to our expectations, the daily price of accommodation per person (*accom_price*) is not found to be significant in the intensity equation. One possible explanation is that, for some tourists, high prices may be interpreted as signals of high quality (e.g. Keane, 1997). Those who plan to stay for several days may be willing to pay higher prices if this guarantees them a certain level of quality. Another reason why the price is not significant may be the fact that we control for the type of selected accommodation in the regression, which implicitly reflects price differences.

We now move to the effect of the tourist's knowledge about the destination on the probability of staying overnight and on LOS. First-visitors (*first*) have a positive coefficient in both equations, implying that those who have never been to Asturias stay for longer than repeat visitors. This positive relation between coming for the first time and the number of days can be explained in terms of the willingness to widely get acquainted with the destination (Nicolau *et al.* 2016). Nevertheless, as stated before, in our model those who declare novelty seeking as their main reason for visiting the destination do not stay significantly longer, which in turn implies that the explanation of first time visitors staying longer may be due to other reasons. Another possible explanation for this result is that repeat visitors have already explored the destination widely in previous visits. The number of visits during the year (*num_vis*) is negatively related with the likelihood of staying overnight, whereas it is not significant in the intensity equation. Conversely, those who declare that a positive previous experience at the destination is the main reason for returning (*experience*) exhibit longer stays and higher probability of an overnight stay. For some tourists, if their previous experience was satisfactory and provided them with high levels of utility, a good risk-reduction method is to return to the same destination and stay for longer periods.

As for the effect of advertising on LOS, those who state that they have seen some type of advertisement (*advert*) show a higher probability of spending a night in Asturias and stay longer, in line with some previous findings (e.g. Barros 2008). This is not surprising given that the experience nature of tourism induces people to build indirect experience from advertising contents such as texts, images or videos (Park and Nicolau, 2015). In fact, tourism advertising is considered one of the main external information and communication sources, as it both consciously and unconsciously affects consumer decision-making (Woodside and King 2001). In the same way, the recommendation of the destination from friends or relatives (*recommend*) also increases the probability of spending a night. This result seems to provide more evidence on the strong reliance that tourists have on friends' and relatives' opinions and advice regarding the characteristics of the destination (Fodness and Murray 1997).

As long as they receive trustworthy information, they perceive the destination as less risky and tend to stay longer. Surprisingly, recommendation is not significant for explaining the number of days the tourist stays. Although this is contrary to our expectations, the results match those of Gokovali *et al.* (2007).

Regarding distance to origin, this variable is significant in both the participation and intensity equations, revealing a positive relationship between distance and length of stay, although at a decreasing rate. When considering the chosen mode of transport, the longer the tourist spends on reaching the destination, the less time he/she can then allocate to staying there. As our analysis is *conditional* on trip characteristics, tourists travelling by plane will arrive sooner than those by car and, consequently, may stay for longer. However, faster modes of transport will be more expensive so, given budget constraints, the individual would have less money to spend at the destination and may stay for shorter periods. Setting *car* as the reference category, this double reasoning may justify why neither means of transport is statistically significant in the intensity equation. The trade-off between monetary and time savings might cancel out the differences across modes of transport, as both effects go in opposite directions. Nonetheless, travelling by train or by plane positively affects the likelihood of an overnight stay. Conversely, reaching the destination by bus reduces the probability of an overnight stay. This may account for the fact that most same-day visitors come to the destination by bus.

Regarding the purpose of the trip, visiting Asturias for leisure and entertainment (*leisure*) or due to job or study-related issues (*labor*) reduce the probability of an overnight stay. Although this may initially seem counterintuitive, the omitted category here (*other*) includes, among others, sports events or doctor visits, which normally require the visitor to spend at least one night at the destination. Nonetheless, when it comes to analyzing the intensity of the stay, leisure purposes positively contribute to lengthen the stay, in line with the literature (Gomes de Menezes *et al.* 2008). Visiting friends or relatives (*family*) and labor-related reasons are also associated with longer stays.

Party size and its composition also matter for explaining LOS, as tourism consumption is usually a social activity in which activities are mainly group-based (Thornton et al. 1997). The fact that only 6.3% of the tourists in our sample traveled alone indicates that they generally travel as a couple or group. Because of this, deciding how long to stay is a balance between the personal preferences of the members of the group. Consequently, *who* the tourist travels with needs to be controlled for in our regression analysis. Compared to travelling in a group (*group*), those who come to Asturias alone (*alone*) or with their couple (*couple)* exhibit a higher probability of spending at least a night there. Regarding the length of the stay, there are no statistical differences between travelling alone or in a group, whereas couples are linked with longer stays. As for party size, this variable is not statistically significant either in the participation or in the intensity equation.

With regard to accommodation type, staying at a private dwelling leads to the longest stays. Focusing on formal market-based accommodations, tourists who lodge at a rural house (*rural*), a *hostel* or a *hotel* stay for shorter periods than those who stay on *campsites* (the reference category). Another issue of interest is how the trip and the selected accommodation were booked. Letting the tourist himself be the omitted category, we find that hiring the trip through a travel agency (*travel_agency*) increases both the probability of spending the night at the destination as well as LOS. Booking the trip through a club or a company (*club_comp*) is also positively related to the decision of staying overnight, but it reduces the number of expected stays.

As expected, those who are only visiting Asturias in the current trip (*only_ast*) stay for longer periods than those who are not. Engaging in active tourism (*act_tour*) activities exerts a positive impact on both the decision to stay overnight and on the number of days to remain at the destination. This suggests that this type of tourist is an important segment for the tourism market, as their interest in engaging in different outdoor activities requires time and therefore implies longer stays.

Finally, in our regression framework we also control for temporal and geographical variables. The year dummies mainly reflect income differences throughout the business cycle, while the geographical ones refer to differences in preferences across the territory. Everything else being equal, tourists had fewer stays from 2011 onwards. This may be associated with the fact that during the economic crisis people faced more economic constraints when travelling (Smeral and Song 2015). Even if they did not, uncertainties and fears about the near future and labor instability might have urged them to spend more on necessities and less on luxuries (tourism) to save money (Gunter and Smeral 2016). With regard to seasonal effects during the year, we included trimester dummies, the first trimester being the omitted one. The estimations indicate that people exhibit a higher likelihood of an overnight stay and also longer stays in the second (*t2*) and third trimesters (*t3*) than in the first (*t1*) one. As for territorial preferences, the probability of sleeping at any accommodation of the central area is higher than in the east or the west. However, the opposite pattern is observed when it comes to analyzing the number of overnight stays.

Given that the magnitude of the estimates are not easy to interpret, Table 4 reports the average marginal effects on the probability of an overnight stay[3] in the first column, and the relative marginal effects on the conditional expected number of nights spent at the destination for those who stay overnight[4] in the second one (both in percentage), for the informative-type variables (*first*, *num_vis*, *experience*, *advert*, *recommend*) and the destination attributes (*natural*, *novelty_seeking*, *tranquility*, *climate*, *gastronomy*).

Variable	Logit model	ZTNBP
first	2.292***	7.098***
num_year_vis	-0.180***	0.047
experience	7.724***	4.621**
advert	2.289***	2.706**
recommend	5.492***	-0.406
natural	7.032***	4.984**
novelty	8.636***	-0.462
tranquility	5.600***	8.857*
climate	6.058***	16.657***
gastronomy	3.609**	-1.858

Table 4. - Average marginal effects on the participation (1) and intensity (2) equations (in percentage).
*** p<0.01, ** p<0.05, * p<0.1

First visitors (*first*) display a 2.3 percent higher probability of staying overnight and are associated with a 7.1 percent longer stay than repeat visitors. A unitary change in the number of visits during the year (*num_vis*) reduces the probability of becoming a tourist by almost 0.2 percent. Previous experience (*experience*) and having seen some type of advertisement (*advert*) increases the expected number of stays by 4.6 and 2.7 percent, respectively. Moreover, those who had previously visited the destination and those who declare that advertising has persuaded them to visit it display 7.7 and 2.3 percent higher probability of an overnight stay. Finally, recommendation from friends or relatives (*recommend*) only influences the probability of being a tourist (5.5 percent higher) but does not affect the expected stay.

[3] $\sum_{i=1}^{N} \frac{1}{N} \frac{\partial P(LOS_i>0)}{\partial Z_{ik}} = \sum_{i=1}^{N} \frac{1}{N} \beta_k \frac{e^{Xi\beta}}{\left(1+e^{Xi\beta}\right)^2}$

[4] Following Farbmacher (2013), the relative marginal effects for a ZTNBP model are given by: $\sum_{i=1}^{N} \frac{1}{N} \frac{\frac{\partial E(LOS_i|LOS_i>0,\ X_i)}{\partial X_{ik}}}{E(LOS_i|LOS_i>0,\ X_i)} = \beta_k - \frac{r_i m_i (Q\beta_k \ln\left(\frac{(m_i+e^{Xi\beta})}{m_i}\right) + \frac{m_i Q\beta_k + \beta_k e^{Xi\beta}}{m_i+e^{Xi\beta}} - Q\beta_k)}{1-r_i}$

where Q= (2-P); $m_i = \frac{1}{\alpha}\lambda_i^{2-P} = e^{(2-P)Xi\beta - \ln\alpha}$; and $r_i = \left(\frac{m_i}{\lambda_i + m_i}\right)^{m_i}$

Regarding the destination attributes, novelty seeking (*novelty_seeking*) emerges as the key one for discriminating between same-day visitors and tourists as it increases the probability of an overnight stay by 8.6 percent. The natural environment (*natural*) is another relevant attribute that encourages visitors to stay overnight. With reference to the expected stay, the Asturian climate (*climate*) and its tranquility (*tranquility*) have a remarkable effect on the number of days tourists stay. Individuals attracted by mild temperatures and the peacefulness of the region display 16.6 and 8.8 percent higher lengths of stay, respectively.

From the average relative marginal effects over the sample reported above, we can conclude that previous experience (*experience*) is the information source which gives rise to the highest probability of an overnight stay and has the largest impact on LOS, everything else constant. This result implies that there is no better source of information than having a positive experience in the past. To a lesser extent, advertising also has a significant positive contribution to lengthen the stay.

In summary we find that the different explanatory variables considered do not have the same effect on the probability of staying overnight (participation equation) and on the number of days spent (intensity equation). This highlights the relevance of distinguishing between tourists and same-day visitors, which is one of the novel aspects of this paper. Furthermore, our regression framework controls for a wide range of sources of observable heterogeneity among tourists, allowing us to correctly isolate the effects of the different information sources about the destination on LOS.

7. CONCLUSIONS

Using a hurdle count data model, this study examines the determinants of both the decision to stay overnight at a destination and the length of the stay. The determinants of tourist LOS have been widely analyzed in the literature but, to the best of our knowledge, no studies have considered the different nature of same-day visitors and tourists using a hurdle count data approach. Moreover, in this research we are interested in isolating the effect of tourists' knowledge about the destination as well as the destination attributes on the length of stay, controlling for a large spectrum of sources of observable heterogeneity. Specifically, we focus on the effect of being a first-time visitor, having seen some type of advertising, visiting because of a recommendation or having had a pleasant experience in the past.

From a methodological point of view, we have proposed two alternative specifications for the intensity equation in the hurdle model. On the one hand, a Zero-Truncated Negative Binomial

P model, which seems to be the better alternative among the Negative Binomial family: not only does it handle the *over*dispersion problem, but it also endogenously estimates the structure of the conditional variance through the parameter P. On the other hand, we estimate a Zero-Truncated Poisson Log-Normal model, which specifies a normal distribution for the unobserved heterogeneity instead of the gamma distribution assumed for the ZTNBP model. The HPC test clearly indicates that the ZTNBP model provides a better fit to our data.

Regarding our study case, we have employed a pooled cross-sectional database of visitors to Asturias (Spain) for the period 2010-2016. Our empirical model is based on a conditional demand function for tourist time given individual characteristics. Our results show that there is a higher probability of staying overnight for first-time visitors, those who had seen some type of advertising about the destination and those who declare that they had a positive past experience at the destination or that someone recommended it to them. The same pattern holds for the expected number of stays, except for the fact that a recommendation is not significant for explaining LOS. Quantitatively, we find that a first-time visit leads to the highest increase in LOS. These results are consistent with individuals trying to minimize the risk of uncertainty through different information strategies. They seem to rely more on their personal past experiences rather than on recommendations from friends or relatives.

Apart from the effects of these informational-type variables, we also consider other relevant covariates in our empirical model. To summarize, we find that the sociodemographic profile appears to be less relevant for the length of stay in comparison to previous studies once you control for a high number of trip-related characteristics. In this sense, the Asturian climate and the aim of performing active tourism emerge as two relevant factors that increase the length of the stay. Regarding the chosen mode of transport, travelling by train or by plane positively increases the likelihood of an overnight stay in comparison to travelling by car. Private accommodation is found to lead to the longest stays, followed by campsites. The results also suggest that the effect that the party size exerts on LOS is not significant.

With regard to policy implications, the identification of the drivers of length of stay seem to be a quite relevant issue given that revenues from tourism are directly related to LOS. The results provided in this study about the effects of a wide set of factors on the length of stay can help policy makers to improve the promotional campaigns and to develop proper strategies to adapt the tourism products to the desires of tourists, focusing on those who stay for longer periods. In this sense, the estimates of the relative marginal effects reveal that those who highly value a mild climate and the natural environment of the destination display higher likelihoods of both an overnight stay and longer-period stays. Consequently, policy makers should reinforce these appealing features when promoting this destination, highlighting the “green tourism” brand. Accordingly, the conservation of the natural environment needs to be a priority for policy makers, not only because of ecological concerns but also because it is one of the main attractions for visitors. Moreover, a positive past experience seems to be another relevant information source, implying that providing tourists with an enjoyable stay is essential for encouraging him/her to come back in the near future.

The limitations of this study include the fact that our analysis of length of stay is conditional on having previously decided to visit a particular destination. Thus, our study is conditional on coming to the Principality of Asturias, a decision that we cannot model with the data we have. Having said that, our results can be generalized to other nature-based tourist destinations in Europe and elsewhere characterized by mild weather, an attractive natural environment and beautiful landscapes. Moreover, the methodology proposed here can be applied to any destination.

ACKNOWLEDGEMENTS

We are grateful to Joao Santos-Silva, Helmut Farbmacher, Eduardo del Valle, Luís Valdés and members of the Department of Economics at the University of Oviedo for helpful comments and suggestions.

We would also like to thank the *Tourist Information System of Asturias* for providing us the database and for their assistance.

REFERENCES

Alegre, J., and Pou, L. 2006. "The length of stay in the demand for tourism". *Tourism Management*, 27(6): 1343–1355.
https://doi.org/10.1016/j.tourman.2005.06.012

Alegre, J., Mateo, S., and Pou, L. L. 2011. "A latent class approach to tourists' length of stay". *Tourism Management*, 32(3): 555–563.
https://doi.org/10.1016/j.tourman.2010.05.003

Alén, E., Nicolau, J.L, Losada, N. y Domínguez, T. 2014. "Determinant factors of senior tourist´s length of stay". *Annals of Tourism Research*, 49: 19-32.
https://doi.org/10.1016/j.annals.2014.08.002

Barros, C. P., Correia, A., and Crouch, G. 2008. "Determinants of the length of stay in Latin American tourism destinations". *Tourism Analysis*, 13: 329-340.

Barros, C. P., and Machado, L. P. 2010. "The length of stay in tourism". *Annals of Tourism Research*, 37(3): 692-706.
https://doi.org/10.1016/j.annals.2009.12.005

Barros, C. P., Butler, R., and Correia, A. 2010. "The length of stay of golf tourism: a survival analysis". *Tourism Management*, 31: 13-21.
https://doi.org/10.1016/j.tourman.2009.02.010

Bell, F.W., and Leeworthy, V.R. 1990. "Recreational demand by tourists for saltwater beach days". *Journal of Environmental Economics and Management*, 18: 189-205.
https://doi.org/10.1016/0095-0696(90)90001-F

Bieger, T. and Laesser, C. 2004. "Information sources for travel decisions: toward a source process model". *Journal of Travel Research,* 42: 357–371.
https://doi.org/10.1177/0047287504263030

Bilgic, A. and Florkowkski, W.J. 2007. "Application of a hurdle negative binomial count data model to demand for bass fishing in the southeastern United States". *Journal of Environmental Management*, 83: 478-490.
https://doi.org/10.1016/j.jenvman.2006.10.009

Brida, J. G., Meleddu, M., and Pulina, M. 2013. "Factors influencing length of stay of cultural tourists". *Tourism Economics*, 19: 1273-1292.
https://doi.org/10.5367/te.2013.0248

Butler, J. and Moffitt, R. 1982. "A computationally efficient quadrature procedure for the one factor multinomial probit model". *Econometrica*, 50: 761-764.
http://www.jstor.org/stable/1912613

Cameron, A. C., and Trivedi, P. K. 1986. "Econometric models based on count data. Comparisons and applications of some estimators and tests". *Journal of Applied Econometrics*, 1(1): 29–53.
DOI:10.1002/jae.3950010104

Cameron, A. C., and Trivedi, P. K. 1998. *Regression analysis of count data*. Nueva York: Cambridge University Press.

Cameron, A. C., and Trivedi, P. K. 2009. *Microeconometrics using Stata.* Stata Press, College Station, Texas, USA.

Davidson, R. and MacKinnon, J.G. 1981. “Several tests for model specification in the presence of alternative hypotheses”. *Econometrica*, 49: 781-793.
DOI: 10.2307/1911522

Decrop, A., and Snelders, D. 2004. “Planning the summer vacation. An adaptable process”. *Annals of Tourism Research*, 31(4): 1008–1030.
https://doi.org/10.1016/j.annals.2004.03.004

Farbmacher, H. 2013. “Extensions of hurdle models for overdispersed count data”. *Health Economics*, 22: 1398-1404.
DOI:10.1002/hec.2892

Ferrer-Rosell, B., Martínez-Garcia, E., and Coenders, G. 2014. “Package and no-frills air carriers as moderators of length of stay”. *Tourism Management*, 42: 114-122.
https://doi.org/10.1016/j.tourman.2013.11.002

Fleischer, A., and Pizam, A. 2002. “Tourism constraints among Israeli seniors”. *Annals of Tourism Research*, 29(1): 106–123.
https://doi.org/10.1016/S0160-7383(01)00026-3

Fodness, D., and Murray, B. 1997. “Tourist Information Search”. *Annals of Tourism Research*, 24: 503–523.
https://doi.org/10.1016/S0160-7383(97)00009-1

Gokovali, U., Bahar, O., and Kozak, M. 2007. “Determinants of length of stay: A practical use of survival analysis”. *Tourism Management*, 28(3): 736–746.
https://doi.org/10.1016/j.tourman.2006.05.004

Gomes de Menezes, A., Moniz, A., and Cabral, J. 2008. “The determinants of length of stay of tourists in the Azores”. *Tourism Economics*, 14(1): 205-222.
https://doi.org/10.5367/000000008783554866

Gomes de Menezes, A. G., and Moniz, A. 2011. “Determinants of length of stay. A parametric survival analysis”. *Tourism Analysis*, 16: 509-524.
https://doi.org/10.1007/978-3-7908-2124-6_6

Greene, W. 2008. “Functional forms for the negative binomial model for count data”. *Economic Letters*, 99: 585-590.
https://doi.org/10.1016/j.econlet.2007.10.015

Greene, W. 2009. "Models for count data with endogenous participation". *Empirical Economics*, 36: 133-173.
DOI 10.1007/s00181-008-0190-5

Grigolon, A.B., Borgers, A.W.J., Kemperman, A.D.A.M. and Timmermans, H.J.P. 2014. "Vacation length choice: a dynamic mixed multinomial logit model". *Tourism Management*, 41: 158-167.
https://doi.org/10.1016/j.tourman.2013.09.002

Gunter, U., and E. Smeral. 2016. “The Decline of Tourism Income Elasticities in a Global context”. *Tourism Economics,* 22 (3): 466–83.

https://doi.org/10.5367/te.2014.0431

Gurmu, S., and Trivedi, P. K. 1992. “Overdispersion tests for truncated Poisson regression models”. *Journal of Econometrics*, 54(1): 347–370.
https://doi.org/10.1016/0304-4076(92)90113-6

Gurmu, S., and Trivedi, P. K. 1996. “Excess Zeros in Count Models for Recreational Trips”. *Journal of Business and Economic Statistics*, 14(4): 469–477.
http://www.jstor.org/stable/1392255

Hellerstein, D., and Mendelsohn, R. 1993. “A theoretical foundation for count data models”. *American Journal of Agricultural Economics*, 75(3): 604-611.
http://www.jstor.org/stable/1243567

Hellström, J. and Nordström, J. 2008. "A count data model with endogenous household specific censoring: the number of nights to stay". *Empirical Economics*, 35: 179-192.
DOI 10.1007/s00181-007-0155-0

Keane, M. 1997. “Quality and Pricing in Tourism Destinations”. *Annals of Tourism Research*, 24: 117–130.
https://doi.org/10.1016/S0160-7383(96)00034-5

Kerstetter, D. and Cho, M.H. 2004. “Prior knowledge, credibility and information search”. *Annals of Tourism Research*, 31: 961–985.
https://doi.org/10.1016/j.annals.2004.04.002

Kiefer, N. 1988. “Economic duration data and hazard functions”. *Journal of Economic Literature*, 26: 646–667.
http://www.jstor.org/stable/2726365

Kim D. Y., Hwang Y. H. and Fesenmaier D. R. 2005. “Modeling Tourism Advertising Effectiveness.” *Journal of Travel Research*, 44 (1): 42-49.
https://doi.org/10.1177/0047287505276590

Laesser, C., and Crouch, G. I. 2006. “Segmenting markets by travel expenditure patterns: the case of international visitors to Australia”. *Journal of Travel Research*, 44(4): 397–406.
https://doi.org/10.1177/0047287505282952

Lehto, X., O’Leary, J., and Morrison, A. 2004. “The effect of prior experience on vacation behavior”. *Annals of Tourism Research*, 31: 801–818.
https://doi.org/10.1016/j.annals.2004.02.006

Leones, J., Colby, B., and Crandall, K. 1998. “Tracking expenditures of the elusive nature tourists of Southeastern Arizona”. *Journal of Travel Research*, 36(3): 56–64.
https://doi.org/10.1177/004728759803600306

Luo, Q. and Zhong, D. 2015. “Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social network sites”. *Tourism Management*, 46: 274-282.
https://doi.org/10.1016/j.tourman.2014.07.007

Machado, L. P. 2010. “Does Destination Image Influence the Length of Stay in a Tourism Destination?” *Tourism Economics*, 16 (2): 443-56.

https://doi.org/10.5367/000000010791305554

Manski, C. 1977. “The structure of random utility models”. *Theory and Decision*, 8: 229–254. DOI: 10.1007/BF00133443

Martínez-García, E., and Raya, J. M. 2008. “Length of stay for low-cost tourism”. *Tourism Management*, 29(6): 1064–1075. https://doi.org/10.1016/j.tourman.2008.02.011

McFadden, D. 1974. “Conditional logit analysis of qualitative choice Behavior”. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Mill, R. C., and Morrison, A. M. 2009. *The tourism system* (4th ed.). Dubuque, IA: Kendall/Hunt Publishing Company.

Mincer, J. 1974. *Schooling, Experience and Earnings.* New York: National Bureau of Economic Research.

Mullahy, J., 1986. “Specification and testing of some modified count data models”. *Journal of Econometrics*, 33: 341–365. https://doi.org/10.1016/0304-4076(86)90002-3

Nicolau, J.L., Zach, F. and Tussyadiah, I. 2016. "Effects of distance and first-time visitation on tourists’ length of stay". *Journal of Hospitality and Tourism Research*, 1096348016654972. https://doi.org/10.1177/1096348016654972

Oliveira-Santos, G.E., Ramos, V. and Rey-Maquieira, J. 2015. “Length of stay at multiple destinations of tourism trips in Brazil”. *Journal of Travel Research*, 54(6): 788-800. https://doi.org/10.1177/0047287514532370

Papadopoulos, G. and Santos-Silva, J.M.C 2012. "Identification issues in some double-index models for non-negative data". *Economic Letters*, 117: 365-367. https://doi.org/10.1016/j.econlet.2012.06.001

Park, S. and Nicolau, J.L. 2015. "Differentiated effect of advertising: joint vs separate consumption". *Tourism Management*, 47: 107-114. https://doi.org/10.1016/j.tourman.2014.09.018

Riphahn, R., Wambach, A. and Million, A. 2003. “Incentive effects in the demand for health care: a bivariate panel count data estimation”. *Journal of Applied Econometrics*, 18(4): 387-405. DOI: 10.1002/jae.680

Rodríguez, X.A., Martínez-Roget, F. and González-Murias, P. 2018. "Length of stay: evidence from Santiago de Compostela". *Annals of Tourism Research*, 68: 9-19. https://doi.org/10.1016/j.annals.2017.11.001

Roehl, W., and Fesenmaier, D. R. 1992. “Risk perceptions and pleasure travel: an exploratory analysis”. *Journal of Travel Research*, 30: 17-26. https://doi.org/10.1177/004728759203000403

Santos-Silva J.M.C. 1997. “Unobservables in count data models for on-site samples”. *Economics Letters*, 54: 217–220. https://doi.org/10.1016/S0165-1765(97)00032-3

Santos-Silva, J.M.C., Tenreyro, S., and Windmeier, F. 2015. "Testing Competing Models for Non-Negative Data with Many Zeros". *Journal of Econometric Methods*, 4: 29-46.
https://doi.org/10.1515/jem-2013-0005

Sarma, S. and Simpson, W. 2006. “A microeconometric analysis of Canadian health care utilization”. *Health Economics*, 15: 219-239.
DOI: 10.1002/hec.1057

Smeral, E., and Song, H. 2015. “Varying Elasticities and Forecasting Performance”. *International Journal of Tourism Research*, 17(2): 140-150.
DOI: 10.1002/jtr.1972

Taylor, C.E., and Knudson, D.M. 1976. “Area preferences of Midwestern campers”. *Journal of Leisure Research*, 7: 39-48.

Thornton, P., Shaw, G. and Williams, A.M. 1997. "Tourist group holiday decision-making and behavior: the influence of children". *Tourism Management*, 18(5): 287-297.
https://doi.org/10.1016/S0261-5177(97)00017-4

Thrane, C., and Farstad, E. 2012. “Tourists' length of stay: the case of international summer visitors to Norway”. *Tourism Economics*, 18(5): 1069-1082.
https://doi.org/10.5367/te.2012.0158

Thrane, C. 2012. “Analyzing tourists´ length of stay at destinations with survival models: a constructive critique based on a case study”. *Tourism Management*, 33: 126-132.
https://doi.org/10.1016/j.tourman.2011.02.011

Vuong, Q.H. 1989. “Likelihood ratio tests for model selection and non-nested hypotheses”. *Econometrica*, 57(2): 307–333.
http://www.jstor.org/stable/1912557

Winkelmann, R. 2004. “Health reform and the number of doctor visits – an econometric analysis”. *Journal of Applied Econometrics*, 19(4): 455-472.
DOI: 10.1002/jae.764

Winkelmann, R. 2008. *Econometric Analysis of Count Data*. 5th ed. Berlin, Springer-Verlag.

Winkelmann, R. and Zimmermann, K.F. 1995. “Recent developments in count data modelling: theory and applications”. *Journal of Economic Surveys*, 9(1): 1-24.
DOI: 10.1111/j.1467-6419. 1995.tb00108.x

Woodside, A. G., and Dubelaar, C. 2002. “A general theory of tourism consumption systems: A conceptual framework and an empirical exploration”. *Journal of Travel Research*, 41(2): 120–132.
https://doi.org/10.1177/004728702237412

Zillinger, M. 2007. “Tourist routes: A time-geographical approach on German car-tourists in Sweden”. *Tourism Geographies*, 9(1): 64-83.
https://doi.org/10.1080/14616680601092915

ANNEX 1. - Descriptive statistics

Type of variable	Variables	N	Mean	Standard deviation	Min	Max	Definition
TYPE OF TRAVELER	tourist	19111	.9145	.2795	0	1	The individual sleeps in the destination at least 1 night
TYPE OF TRAVELER	same-day	19111	.08544	.2795	0	1	The individual does not spend the night at the destination
DEPENDENT VARIABLE	LOS	19111	4.313	3.833	0	30	Number of nights spent at the destination
SOC	man	19111	.5437	.4980	0	1	Man
SOC	age	19111	40.315	12.110	18	91	Age
SOC	housewife	19111	.0360	.1862	0	1	Housewife/ househusband
SOC	retired	19111	.0566	.2311	0	1	Retired
SOC	employed	19111	.6521	.4762	0	1	Employed
SOC	student	19111	.0762	.2653	0	1	Student
SOC	unempl	19111	.0189	.1363	0	1	Unemployed
SOC	self-empl	19111	.1559	.3628	0	1	Self-employed
SOC	primary	19111	.0726	.2596	0	1	Primary studies
SOC	secondary	19111	.3097	.4623	0	1	Secondary studies
SOC	high	19111	.6176	.4859	0	1	Higher education
SOC	foreign	19111	.0819	.2743	0	1	The individual lives in another country
ATTRIB	natural	19111	.1133	.3170	0	1	The individual visits Asturias due to its natural environment
ATTRIB	novelty_seeking	19111	.3530	.4779	0	1	The individual visits Asturias due to novelty seeking
ATTRIB	tranquility	19111	.0077	.0879	0	1	The individual visits Asturias looking for tranquility
ATTRIB	climate	19111	.0079	.0885	0	1	The individual visits Asturias due to its climate
ATTRIB	heritage	19111	.0070	.0837	0	1	The individual visits Asturias due to its heritage
ATTRIB	gastronomy	19111	.0107	.1030	0	1	The individual visits Asturias due to its gastronomy
PRICE	accom_price	19111	28.602	22.536	0	575	Daily expenditure per person (€) on accommodation
KNOWLEDGE	recommend	19111	.0923	.2894	0	1	The individual visits Asturias due to recommendation
KNOWLEDGE	experience	19111	.2157	.4113	0	1	The individual visits Asturias due to previous experience
KNOWLEDGE	advert	19111	.3470	.4760	0	1	The individual has seen advertising.
KNOWLEDGE	first	19111	.3877	.4872	0	1	First time the individual visits Asturias
KNOWLEDGE	num_vis	19111	.9288	6.295	0	100	Number of visits during the year.
TRIP	distance	19111	675.12	1171.70	0	17713	Distance between origin and Oviedo (km)
TRIP	car	19111	.8247	.3802	0	1	Car
TRIP	bus	19111	.0266	.1610	0	1	Bus
TRIP	train	19111	.0310	.1735	0	1	Train
TRIP	plane	19111	.0754	.2641	0	1	Plane
TRIP	alone	19111	.0636	.2440	0	1	The individual comes alone
TRIP	couple	19111	.5175	.4997	0	1	The individual comes in a couple
TRIP	group	19111	.4188	.4933	0	1	The individual comes with his/her family, friends or work-mates (in a group)
TRIP	party_size	19111	3.704	7.110	1	250	Party size
TRIP	leisure	19111	.8302	.3754	0	1	The individual comes for leisure or on holidays.
TRIP	labor	19111	.0630	.2429	0	1	The individual comes because of his/her studies or job issues.
TRIP	family	19111	.0751	.2637	0	1	The individual comes for visiting relatives.
TRIP	other	19111	.0315	.1748	0	1	The individual comes for a doctor visit, making purchases, a religious peregrination or a sport competition.
TRIP	hotel	19111	.5648	.4957	0	1	The individual stays at a hotel
TRIP	hostel	19111	.0385	.1924	0	1	The individual stays at a hostel
TRIP	rural	19111	.1125	.3160	0	1	The individual stays at a rural house

TRIP	campsite	19111	.0541	.2263	0	1	The individual stays at a campsite
TRIP	private	19111	.1444	.3515	0	1	The individual stays at a private accommodation
TRIP	himself	19111	.9074	.2898	0	1	The individual organized the trip himself
TRIP	travel_agency	19111	.0446	.2066	0	1	The trip was organized by a travel agency
TRIP	club_comp	19111	.0478	.2135	0	1	The trip was organized by a club or the company the individual works for.
TRIP	act_tour	19111	.0747	.2629	0	1	The individual makes active tourism
TRIP	only_ast	17478	.8329	.3729	0	1	The individual only visits Asturias.
AREA	west	19111	.1614	.3679	0	1	West area
AREA	centraly	19111	.4586	.4983	0	1	Central area (Oviedo-Gijon-Avilés)
AREA	centralr	19111	.0463	.2102	0	1	The rest of the central area
AREA	east_coast	19111	.1725	.3778	0	1	East coast
AREA	east_inner	19111	.1610	.3675	0	1	East inner area
TEMP	t1	19111	.1942	.3956	0	1	First trimester (January-February-March-April)
TEMP	t2	19111	.4924	.4999	0	1	Second trimester (May-June-July-August)
TEMP	t3	19111	.3045	.4602	0	1	Third trimester (September-October-November-December)
TEMP	y10	19111	.1350	.3417	0	1	Year 2010
TEMP	y11	19111	.1297	.3359	0	1	Year 2011
TEMP	y12	19111	.1378	.3447	0	1	Year 2012
TEMP	y13	19111	.1133	.3170	0	1	Year 2013
TEMP	y14	19111	.1481	.3552	0	1	Year 2014
TEMP	y15	19111	.1798	.3841	0	1	Year 2015
TEMP	y16	19111	.1558	.3427	0	1	Year 2016