Gutiérrez, A., Miravet, D., Saladié, Ò., Anton Clavé, S. (2019). High-speed rail, tourists' destination choice and length of stay: A survival model analysis. *Tourism Economics*. https://doi.org/10.1177/1354816619855286

High-speed rail, tourists' destination choice, and length of stay: A survival model analysis

Aaron Gutiérrez [1*], Daniel Miravet [2 3], Òscar Saladié [1], Salvador Anton Clavé [1]

[1] Department of Geography, Universitat Rovira i Virgili, 43480 Vila-seca, Spain

[2] Consortium of Public Transport of Camp de Tarragona, 43004 Tarragona, Spain

[3] Department of Economics, Universitat Rovira i Virgili, 43204 Reus, Spain

* corresponding author: aaron.gutierrez@urv.cat

Abstract

We analysed the determinants of the length of stay for tourists arriving in a Mediterranean coastal destination by means of high-speed rail (HSR) service. This study is based on data obtained from a survey completed by HSR passengers returning from holiday in Costa Daurada (Catalonia). The empirical analysis is based on estimations made using a survival model. The influence of the availability of HSR service on tourists' destination choices together with the tourists' profiles, party structure or accommodation characteristics were used as explanatory variables. Results revealed that the existence of HSR services played a minor role in tourists' decision of whether to visit the Costa Daurada. Also, evidence suggests that the existence of the HSR station would only affect the length of stay of those tourists who stay overnight in second residences.

Keywords: High-speed rail; tourists' destination choice; length of stay; survival model; Costa Daurada

Introduction

The extension of a high-speed rail (HSR) network has introduced new facilities for medium-distance trips. It has shortened travel times and made trips more comfortable between the regions connected by this network. Within this context, there is growing literature analysing the contribution of HSR to tourism development (mostly in China, Spain, France and Italy).

The analysis of the effects of travel time reduction and the improvement in accessibility to tourists is a preeminent research field within theoretical and ex-ante studies (Campa et al., 2016; Masson and Petiot, 2009). In the case of ex-post evaluation studies, the main topics are its influence in tourists' destination choice (Delaplace et al. 2014; Gutiérrez et al., 2018; Pagliara et al., 2015) and in the arrival of new tourists (Albalate and Fageda, 2016; Pagliara et al., 2017). These ex-post studies based on empirical data have obtained results that tend to differ. Some studies have demonstrated clear positive effects of HSR on tourism development (Chen and Hayes, 2015; Pagliara et al., 2017), while others noted no single direct contribution of HSR to tourist outcomes (Albalate and Fageda, 2016). This shows that it is a complex field of study and that a wide range of factors related to the territorial context and destination characteristics need to be taken into account (Gutiérrez and Ortuño, 2017).

Despite the increasing number of analyses of tourists' length of stay and its worldwide emergence as a key demand variable, to the best of our knowledge, none of these studies have focused on HSR passengers. This is a relevant issue because it is a common concern that rapid transport modes, such as HSR, increase short-term visits. Moreover, the availability of HSR services could facilitate opening new markets and diversifying the profiles of tourists. Thus, the effects of the availability of HSR services on destination choice is a relevant research line. This scope has been addressed previously by Delaplace et al. (2014) in the cases of Paris and Rome and by Pagliara et al. (2015) in the case of Madrid. The study by Delaplace et al. (2014) concluded that 28% and 49% of tourists who arrived in Rome and Paris, respectively, by means of HSR were positively influenced in their choice by the possibility of reaching their destination by HSR. In contrast, the study by Pagliara et al. (2015) based on tourist arrival by any transport mode (12.8% by HSR) highlighted that only 0.8% of tourists arriving in Madrid indicated that the availability of HSR services was the primary reason for their choice of destination.

Taking this into account, the relationship between destination choice and length of stay among tourists traveling by means of HSR has been identified as a research challenge. Consequently, the research objective of this paper in identifying the determinants of length of stay at destination for tourists arriving by means of HSR is to analyse the role of destination choice as a determinant of the length of stay. In particular, whether tourists who are more strongly influenced by the availability of HSR in their choice of destination tend to have different lengths of stay than other tourists is analysed.

The study is based on two initial research hypotheses. The first hypothesis was that, in addition to the tourist profiles, party structure and accommodation characteristics, the influence of the availability of HSR services on tourists' destination choices is a key variable that determines their length of stay. The second research hypothesis was that those tourists for whom the existence of HSR services played a significant role in their decision to visit are also those who tend to have a shorter stay. This assumption is in line with previous studies demonstrating that travel time is a key determinant in shorter visits.

The study is based on a survey of tourists visiting Costa Daurada (a coastal destination in Southern Catalonia) who arrived at the Camp de Tarragona HSR Station. The influence of the availability of HSR service on tourists' destination choices was addressed by means of a specific indicator put forward by Young et al. (2010) which was used as explanatory variable, together with tourists' profiles, party structure or accommodation characteristics for the empirical analysis of the determinants of length of stay among tourists arriving by HSR. The estimations were made using a survival model.

The rest of this article is organized as follows. Following the introduction, the second section reviews existing literature pertaining to three main topics: 1.) HSR and tourism, 2.) length of stay at a given destination and 3.) the approaches used for modelling the determinants of tourists' lengths of stay. The third section introduces the area of study. The fourth section explains how the data have been obtained and how the causal indicator (the influence of HSR services on destination choice) has been developed. The fifth section presents the descriptive statistics. In the sixth section, the econometric model specification is introduced and justified. The results of the model are shown and discussed in the seventh section. Finally, we present our key conclusions in the eighth section.

Background

High speed rail and tourism

HSR services improve accessibility and, in some cases, shorten the travel time between the origin and destination of a trip (Masson and Petiot, 2009; Wang et al., 2018). Moreover, several studies have highlighted the trend towards a diversification of HSR passenger profiles and trip purposes that correlates with the increasing use of HSR services for tourism and leisure (Santos et al., 2006; Bazin et al., 2013; Delaplace et al., 2014). Therefore, HSR services are creating new opportunities for tourism, especially national tourism (Campa et al., 2016; Sun and Lin, 2017). This has been shown to be the case for urban destinations by Delaplace et al. (2014) and Pagliara et al. (2015) (for Paris and Rome, along with Madrid, respectively). Furthermore, Gutiérrez and Ortuño (2017) have demonstrated this for coastal destinations in Spain, as have Delaplace et al. (2016) for the Disneyland Paris and Futuroscope theme parks. At the same time, HSR services also extend tourist markets (Wang et al., 2014).

For some destinations, the primary impact of the availability of HSR services is a change in the mode of transport (Dobruszkes et al. 2014; Saladié et al., 2016), and a growing number of studies have investigated the competition between HSR and air transportation, underscoring the substitution effects of the expansion of HSR networks (Givoni and Dobruszkes, 2013). Nevertheless, the arrival of new tourists to a destination could be less than expected by stakeholders and policy-makers (Bellet et al., 2012). As a consequence, the effects of HSR services on tourism development at a local level appear controversial in recent studies. Thus, new studies are required to enrich the debate.

Length of stay

There is a wide range of factors influencing tourists' destination choices, which are mainly linked to the individual tourist (Um and Crompton, 1990) and to the destinations (Crouch and Ritchie, 2005). Transportation, which is the way that allows the tourists move from their place of residence to the chosen destination, plays a key role in the positioning and competitiveness of tourist destinations (Lumson and Page, 2004). Consequently, this becomes one of the more important determinants of a tourist's destination choice (Decrop and Snelders, 2004). Moreover, improvements in transportation are one of the key reasons that contribute to the fact that people are taking more trips each year. This ultimately can result in a decline in the length of stay. For example, see Alegre and Pou (2006) for a study of the Balearic Islands; Gokovali et al. (2007) in Turkey; Gomes de Menezes et al. (2008) in the Azores; Martínez-García and Raya (2008) on tourists arriving at Girona airport (North-eastern Spain); Barros and Machado (2010) on golf tourism in the Algarve (Portugal); Salmasi et al. (2012) in Italy and Ferrer-Rosell et al. (2014) on international tourists arriving by air in Spain. This is also one of the reasons why the tourist industry has begun to show interest in analysing the determinants of length of stay.

Length of stay is an interesting research topic for several reasons. First, it influences overall spending, a sense of affiliation and satisfaction with a given destination (Davies and Mangan, 1992). It is also a key factor in research of sustainable tourism in general because tourism activities stress local resources (Saarinen, 2006). Moreover, it is fundamental to tourists' decision-making process (Decrop and Snelders, 2004), which may be strongly affected by socio-demographic factors (Alegre and Pou, 2006). Economic impact studies also show that the length of stay strongly correlates with expenditures at a given destination (Alegre and Pou, 2006; García-Sánchez et al., 2013; Thrane, 2015).

From a destination management perspective, length of stay is essential for effective planning and management (Martínez-García and Raya, 2008). How tourists decide on the duration of their stay and which factors determine this are key questions for public and private practitioners. Since the pioneering studies of Archer and Shea (1975), research has shown that longer-stay tourists visit a wider range of attractions, explore more peripheral regions and generate more diverse economic, social and environmental impacts. Analysing the determinants of length of stay in different contexts and for different markets and types of tourist destinations has become an issue of the utmost theoretical and practical importance. Ultimately, it is critical for designing marketing policies and achieving higher occupancy rates and revenue streams (Alegre and Pou, 2006).

Modelling the length of stay

Several studies have addressed the determinants of tourists' length of stay. Researchers have approached this issue using a variety of estimation strategies that range from the simplest ordinary least squares (OLS) to the more complex and frequently used duration models. Even though OLS is the most restrictive estimation procedure, it is advocated by Thrane (2012) as the most suitable technique in this area of research, provided that the data are cross-sectional, there is no right-censoring and there are no time-invariant dependent variables. In fact, OLS was the first approach used for analysing tourists' length of stay (Mak and Nishimura, 1979; Paul and Rimmawi 1992), and more recently, it has been favoured due to its lack of complexity (Thrane, 2016; Mortazavi and Cialani, 2016). Another option includes discrete-choice models, which treat the dependent variable as a decision made by the tourist based on a set of independent variables. For instance, multinomial logit has been applied by Nicolau and Más (2009) and Grigolon et al. (2014), whereas ordered logit has been used by Ferrer-Rosell et al. (2014). As demonstrated by Cameron and Trivedi (1998), estimators lose efficiency as the number of alternatives grows. This problem can be mitigated by grouping alternatives (Ferrer-Rosell et al., 2014), although this would result in a loss of information. Other authors have favoured count models (Rodríguez et al, 2003; Alegre et al, 2011; Brida et al, 2013; Adongo et al, 2017). Thrane (2015) argues that, aside from their complexity, these kinds of models are not meant to deal with a time dimension. Finally, since Gokovali et al. (2007) used survival models (also known as duration models) for the first time, this sort of model has become the most popular estimation technique in this field (see Santos et al., 2015 for a list of contributions using this type of model).

Concern for how to deal with the potential bias caused by unobserved heterogeneity has also been pointed out (Barros et al., 2008). Strategies to cope with unobserved heterogeneity are diverse. Researchers have opted for continuous survival models that allow unobserved heterogeneity (Barros et al., 2008; Thrane 2012; Santos et al, 2015), the random-logit model (Nicolau and Más, 2009), the mixed-logit model (Grigolon et al., 2014) or the utilization of latent class models (Alegre et al., 2011; Yang and Zhang, 2015).

Finally, another econometrical concern is the potential sample selection bias. This has been addressed by Barros and Machado (2010) and Santos et al. (2015). However, as Thrane (2012) argues, data gathered through surveys at tourist destinations are always selected samples because non-tourists are excluded.

Study area

The Costa Daurada (Figure 1) is a coastal tourist destination in Catalonia. The main attractions in this area include its 14 km of beach, the city of Tarragona (once the capital of the Roman province of Hispania and now a UNESCO World Heritage site) and Port Aventura (one of the largest theme parks in Europe). According to data from regular surveys conducted by the Science and Technology Park for Tourism and Leisure of Catalonia almost 5 million tourists visited Costa Daurada in 2014, 57% of whom were Spanish. Those tourists were accommodated in hotels (41%), second residences (33%), rented apartments (18%) and campsites (8%).

The destination is well-connected to Madrid and western Spain as well as with the Spanish Mediterranean regions and the French border via toll motorways. Reus airport is close to the main destinations, and Barcelona airport is roughly 100 km north-east of Salou. According to the same source, the most common modes of transport for tourists arriving at the Costa Daurada are private vehicles (59%) and planes (26%), followed by buses (7%), trains (5%) and other types of transportation (3%). This distribution varies when only Spanish tourists are considered: private vehicles (75%), trains (9%, which include both conventional and HSR services), buses (8%), planes (7%) and other modes of transportation (1%) (Gutiérrez and Miravet, 2016a and 2016b). Only 4% of Spanish tourists arrived on the Costa Daurada via HSR services in 2014. Nevertheless, this figure has increased year on year. It was 1.5% in 2010. Thus, the previsions is that this figure will continue growing in the forthcoming years.

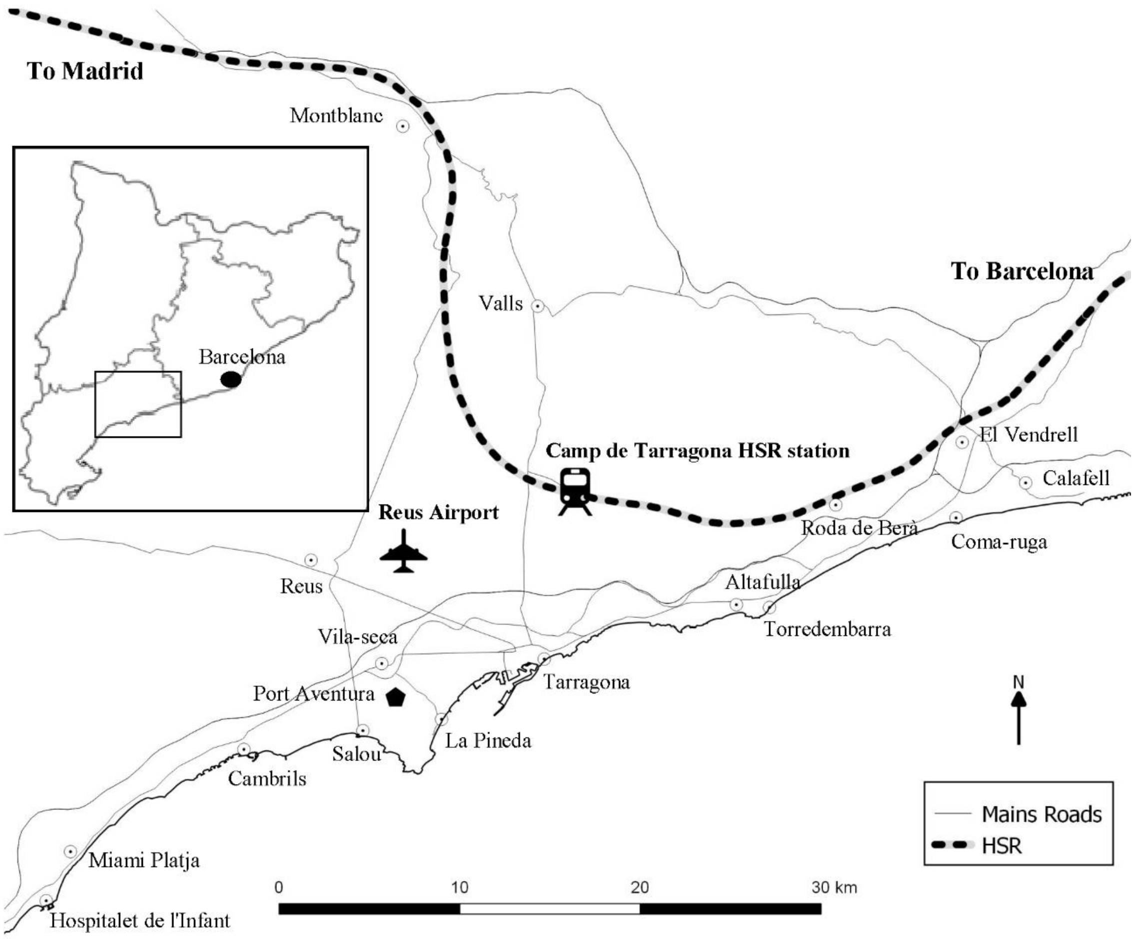


Figure 1. Study area

HSR services have been available in Costa Daurada since 2007. The HSR station, called Camp de Tarragona, is an intermediate node of the Madrid-Barcelona-French border high speed line, and it is located at a notable peripheral location. A taxi service is available, and the station is connected to the Central Costa Daurada and the cities of Tarragona and Reus by bus. According to RENFE (a Spanish rail company), the highest volume of passengers is reported in the summer. It is the eighth-largest Spanish HSR station in terms of volume of passengers, but it is the largest among the peripheral stations.

Data

Data collection

Between July and August 2014, 1,225 passengers were interviewed at the departure lounge of the Camp de Tarragona HSR station. For this specific study, we disregarded passengers who (a) were not on holiday in the Costa Daurada and (b) were not from Spain. Thus, our results are based on 574 Spanish passengers who were waiting to take the train home after their holiday at the Costa Daurada. These passengers were asked about their socio-demographic characteristics, including their gender, age and region of origin. They were also asked about their trip's characteristics, including their length of stay, the number of days in advance that they had decided on their trip before departure, whether this was their first visit to the destination, whether they had used the Camp de Tarragona HSR station on previous occasions, how they booked their trip, their party structure, how they transferred to their accommodations, their type of accommodations and their final destination at the Costa Daurada. As one of the main goals of the present work focusses on assessing to what extent the capacity of the HSR station to attract tourists influences their length of stay, tourists were also asked about the availability of HSR services in relation to their decision to visit the Costa Daurada for their holidays. To this end, a causal analysis aiming to identify the probability of HSR passengers visiting the Costa Daurada due to the existence of the Camp de Tarragona HSR Station was applied. This approach has been applied previously using the same sample by Gutiérrez et al. (2018).

Causal Analysis

Young et al. (2010) applied a causal chain approach in order to examine empirically the impact that cultural events exert on tourists' probability of visiting a town or village, which can easily be extrapolated to the influence of the HSR station. More recently, this approach has also been applied in the context of HSR services by Saladié et al. (2016) and Gutiérrez et al. (2018). In addition, it has been used in cases involving other modes of transportation, such as low-cost flight routes (Anton Clavé et al., 2015). This approach was used to identify the upward trend in the probability of visiting the destination.

This can be expressed as:

$$\Delta P = P(V|S) - P(V|{\sim}S) \quad (1)$$

where $P(V|S)$ refers to the probability that a tourist will actually visit a destination (the Costa Daurada) given that the HSR station exists, and $P(V|{\sim}S)$ denotes the probability that the visit would occur without the presence of the HSR station.

These conditional probabilities are very difficult to observe or measure. As a result, ΔP cannot be calculated directly by means of expression (1). Nonetheless, Young et al. (2010) put forward the following alternative expression:

$$\Delta P = ck \quad (2)$$

It is assumed that there are two (and only two) circumstances that would prevent the probability of visiting the Costa Daurada being increased by the existence of the HSR station:

1. The existence of the HSR station causes no attraction at all to the tourist to visit the Costa Daurada.
2. The tourist would have visited the Costa Daurada anyway, regardless of the existence of the HSR station.

The occurrence of either of these two circumstances will imply that $\Delta P = 0$. As a consequence, the probability that neither of these circumstances occurs yields an increase in the probability of visiting the Costa Daurada. k is then defined as the complementary probability attached to circumstance 1 and c is defined as the complementary probability associated to circumstance 2. Taking this into account it can be concluded that the expression $\Delta P = ck$ is appropriate to obtain a suitable indicator that measures to what extent the HSR station is critical to the attraction of tourists to the Costa Daurada. From here on, we will refer to this indicator as the causal indicator (CI).

Young et al. (2010) incorporated a feasible measurement for the probability that circumstance 1 would not occur (k), as well as a feasible measurement for the probability that circumstance 2 would occur $(1-c)$, from which its complementary probability, c, could easily be derived.

Adapted from their survey, the question that has allowed us the measurement for k is:

> Q1. On a scale from 0 to 10, to what extent do you agree with the following statement: 'I chose the Costa Daurada as a tourist destination because of the existence of the Camp de Tarragona HSR Station'. (0 means 'false', and 10 means 'absolutely true').

Similarly, the question that provided us with $(1-c)$ is:

> Q2. On a scale from 0 to 10, to what extent do you agree with the following statement: 'I would have chosen the Costa Daurada as a tourist destination even if there was no Camp de Tarragona HSR Station'. (0 means 'false', and 10 means 'absolutely true').

Descriptive statistics

Table 1 presents descriptive statistics. All variables (with the exceptions of the length of stay and the Causal Indicator) are dichotomous. Thus, each observation can only equal 1 or 0. According to these data, the Madrid area (36%) is the main origin among tourists arriving by the HSR, followed by the Aragon region (21%). Women are predominant (making up 58% of the sample), while the most important age group is the one ranging from 31 to 45 years old (38%). Up to 64% of the sample has a tertiary education. The most preferred accommodation option is staying with family or friends (42%), whereas hotel accommodations are chosen by 25% and second residences by 19%. The destination of almost half of the sample is the central coastal area of Costa Daurada (48%). Only 12% of the tourists have organized their trips using a travel agency. Most of them have already used the railway station (57%), and this is a repeat tourist destination for them (89%). Regarding the travel party, 49% have travelled alone[1], whereas 20% have had family trips with children and adult family trips.

The causal indicator (CI) of the HSR station is a continuous variable. Given the fact that it is calculated as the product of two probabilities, its values are circumscribed between 0 and 1, both inclusive. Its mean and standard deviation were 0.19 and 0.22,

[1] This is not equivalent to spending holidays alone.

respectively, indicating that the tourists' reported degree of attraction of the HSR station to the Costa Daurada was relatively limited.

The mean length of stay of tourists arriving at the HSR station is 8.2 nights with a standard deviation of 10.5 nights, which reflects the considerable fluctuation in the number of nights spent in the area. These fluctuations are reflected in the histogram presented in Figure 2. It becomes apparent that the density function is clearly left-skewed because 70% of the tourists interviewed reported stays of seven nights or less. In fact, the median number of nights is six, while the maximum number reported is 100. Despite the large number of two-night stays, the variable reaches the absolute mode at the seventh day. The number of stays then suddenly drops. Notwithstanding this reduction, several modes can be identified. Local maximums reached at the 10th, 15th and the 30th nights are particularly remarkable. Beyond the 31st night, spells tend to be isolated.

Table 1. Descriptive statistics

		Mean	**St. Dev**
Length Stay	Length of stay	8.24	10.48
Origin-Cat	Origin: Catalonia	0.10	
Origin-Aragon	Origin: Aragon	0.21	
Origin-BC	Origin: Basque Country, Navarra and la Rioja	0.16	
Origin-Mad	Origin: Madrid	0.36	
Origin-Rest	Origin: Rest of Spain	0.16	
Sex-man	Sex: Man	0.42	
Age-30	Age: <31	0.28	
Age-40	Age: >30 and < 46	0.38	
Age-60	Age: >45 and < 61	0.23	
Age-older	Age: >60	0.12	
Educ-Primary	Education: Primary education	0.13	
Educ-Secundary	Education: Secundary education	0.19	
Educ-Tertiary	Education: Tertiary education	0.64	
Educ-Unknown	Education: Not reported education	0.03	
CI	Causal indicator	0.19	0.22
CI^2	Squared Causal indicator	0.08	0.16
Accom-Hotel	Accommodation: Hotel	0.25	
Accom-2resid	Accommodation: Second residence	0.19	
Accom-Famfriends	Accommodation: Staying with family or friends	0.42	
Accom-Rent	Accommodation: Renting an ppartament	0.08	
Accom-Other	Accommodation: Other	0.06	
Trav-Agency	Trip contracted by means of a travel agency	0.12	
Dest-CD	Destination: Touristic destinations in Central Costa Daurada	0.48	

Dest-Tarragona	Destination: Tarragona	0.14
Dest-Rest	Destination: Other destinations	0.38
Who-Alone	Accompanied by: Visiting alone	0.49
Who-Friends	Accompanied by: Friends	0.10
Who-Adultfam	Accompanied by: Adult family members	0.20
Who-Chlidfam	Accompanied by: family trip with children	0.20
Repeater	Not the first time in the touristic destination	0.89
Sample size: N=588		

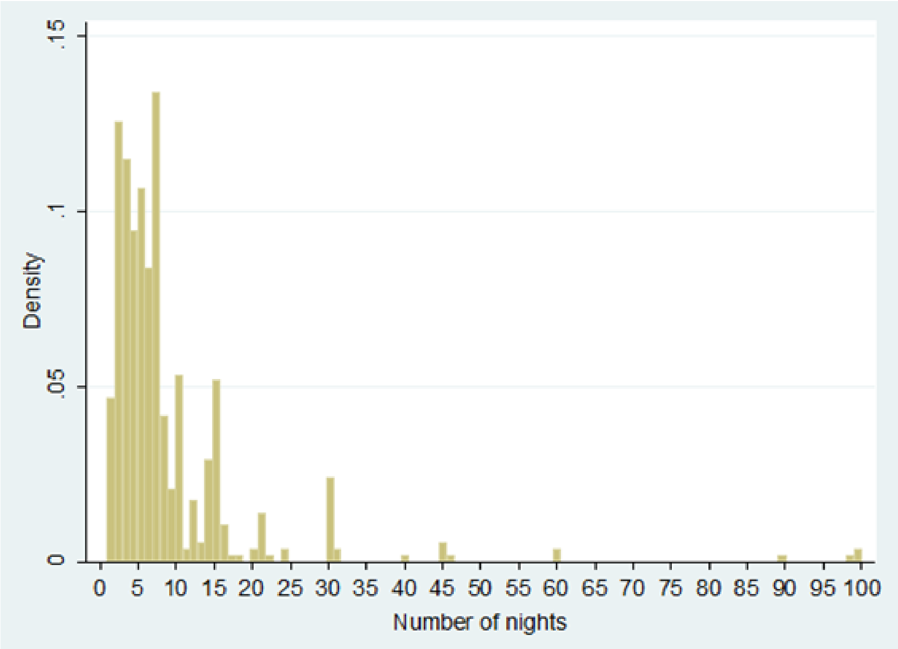


Figure 2. Density function of length of stay

Method

Time is a continuous phenomenon. By contrast, the length of stay is measured as the number of nights and has a discrete nature. Following Chalita et al. (2002), the number of nights spent at the tourist destination can be regarded as a continuous variable if the number of ties among event times is small, or in other words, if the number of failures (tourists leaving) coinciding in each of the intervals is small enough. They suggested a ratio to discern whether the number of ties is low enough to consider the variable of interest as continuous:

$$pt = (nf - r)/n \tag{3}$$

where pt denotes the proportion of ties, nf refers to the number of failures, r is the number of distinct failures (distinct intervals when tourists leave) and n denotes sample size. The rule suggests the use of discrete models when the proportion of ties exceeds 25%. The proportion of ties in our dataset reaches 95%, which verifies the discrete nature of our variable.

From this perspective, OLS and traditional specifications of duration models where the variable of interest is continuous are discarded because they do not meet the discrete nature of our dependent variable. Discrete choice models are also dismissed due to the fact that their efficiency is compromised as the number of distinct spells increases. Finally, the decision between count data models and continuous duration models is not straightforward. The first issue to consider is the fact that count data models are meant to be applied to 'how-many-times-something-happened' phenomena instead of time phenomena (Thrane, 2015), whereas survival models were specifically designed to deal with time phenomena that occur at a maximum of once to each of the individuals, as failure implies that the individual is no longer observed in the data set. From a conceptual perspective, the examination of the determinants of tourists' length of stay is much closer to the latter definition than the former. From a more technical point of view, Thrane (2015) also argues that count-data distributions are intended to suit data whose minimum values are zero. As an example, a study that analyses the number of car accidents in different areas requires a distribution whose minimum value is zero. Given that the minimum number of nights spent at a tourism destination will never equal zero, survival models provide a more accurate fit than do count-data models. This relative advantage of duration models is also signalled by Santos et al. (2015). Taking these elements into consideration, discrete duration models align better with both the characteristics of our data and the object of our research in comparison to count-data models. These advantages do not imply that duration models are free of drawbacks, for instance, their complexity (Thrane, 2012). To our knowledge, none of the previous contributions to the analysis of the determinants of length of stay have adopted a discrete duration model.

Duration models were developed to consider time as continuous. A discrete perspective is adopted by assuming that each night constitutes an interval j, which comprises time within$(j-1, j]$. Our aim is to estimate a proportional hazard model in which the probability that a tourist stays until the end of interval j is written as:

$$S_i(j) = \Pr(T_i > j) = \prod_{k=1}^{j}(1 - h_{ik}) \tag{4}$$

Where h_{ik} is the discrete probability that a tourist ends his/her stay in interval *k*, conditional to the probability of having stayed until the beginning of this interval:

$$h_{ik} = \Pr(k - 1 < T_i \leq k | T_i > k - 1) \tag{5}$$

Thus, the probability that the tourist ends the stay at interval j is represented as follows:

$$\Pr(j - 1 < T_i \leq j) = S_i(j - 1) - S_i(j) = \frac{h_{ij}}{1-h_{ij}} \prod_{k=1}^{j}(1 - h_{ik}) \tag{6}$$

Following Prentice and Gloeckler (1978), it is assumed that h_{ik} is distributed as a complementary log-log (cloglog). Under these circumstances, the discrete time equation of an underlying continuous time proportional hazard can be written as:

$$c\ log\ log\left[1 - h_j(x_{ij})\right] \equiv log\left(-log\left[1 - h_j\left(x_{ij}\right)\right]\right) = \beta_0 + \beta x_{ij} + \gamma_j$$

$$\Rightarrow h_j\left(x_{ij}\right) = 1 - exp\left[-exp\ \ (\beta_0 + \beta x_{ij} + \gamma_j)\right] \tag{7}$$

The baseline hazard is defined non-parametrically by γ_j, which provides a higher degree of flexibility when specifying duration dependence, instead of forcing a specific form. Furthermore, x_{ij} stands for the non-time varying explanatory variables of the model.

The next step involves introducing unobserved heterogeneity into the expression, which is also known as frailty. At this stage, assumptions for how frailty enters the equation are required. We suggest a first specification, which assumes an individual heterogeneity term $u_i \equiv \ln(v_i)$, where v_i follows a Gamma distribution. The gamma distribution has been a popular choice in empirical practice involving unobserved heterogeneity within discrete duration models.

$$h_j\left(x_{ij}\right) = 1 - exp\left[-exp\ \ (\beta_0 + \beta x_{ij} + \gamma_j + u_i)\right] \tag{8}$$

An alternative and more flexible approach implies non-parametrical unobserved heterogeneity. Following Heckman and Singer (1984), the existence of a diversity of mass points or types is assumed. The membership of each class is unobserved, and each individual has a probability to be associated with each of the mass points.

$$h_{j,type}(x_{ij}) = 1 - exp\left[-exp\ \ (m_{type} + \beta_0 + \beta x_{ij} + \gamma_j)\right] \quad ..... (9)$$

For instance, an individual belonging to type 2, thus, would be facing the following hazard function:

$$h_{j,2}(x_{ij}) = 1 - exp\left[-exp\ \ (m_2 + \beta_0 + \beta x_{ij} + \gamma_j)\right] \quad (10)$$

According to this parametrization, mass point 2 equals $m_2 + \beta_0$. Since m_1 is normalized to 0, mass point 1 becomes β_0.

Results and discussion

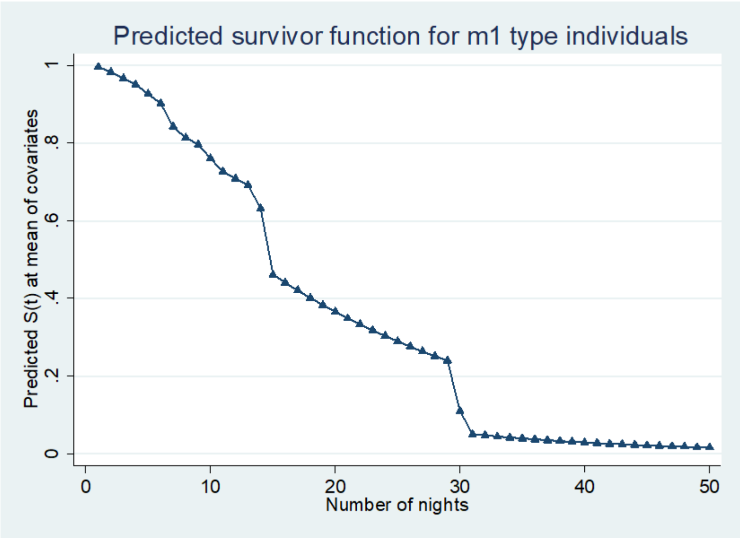
Estimation results are shown in Table 2, which contains hazard rates arising from the cloglog model without frailty (1), the cloglog model with Gamma distributed frailty (2) and the cloglog model with non-parametric frailty (3). As previously set forth, the main difference between the specifications of models (2) and (3) lies in the fact that model (2) imposes a specific distribution for the unobserved heterogeneity (gamma distribution), whereas frailty enters model (3) non-parametrically. As a result, model (3) enables a higher degree of flexibility. Discrepancies between the three models in terms of the magnitude of the hazard rates are, in some particular cases, not negligible. Changes in the significance of the variables are reduced. A first issue to disclosure relates to whether the inclusion of frailty in the models is appropriate. Departing from model (2), for the one considering Gamma-distributed frailty, whether the unobserved heterogeneity variance component (σ^2) is equal to zero is tested. The null hypothesis is rejected because Chibar2 (01) is equal to 14.066 with a p-value of 0.0001. Regarding the model with non-parametric frailty (model 3), the constant associated with the mass-point 2 and the probabilities belonging to both mass-points are also significant, which also sustains the inclusion of unobserved heterogeneity within the survival models. Thus, models (2) and (3) are much more adequate than model (1) as it has been demonstrated that unobserved heterogeneity is a critical issue. In model (3), frailty is added through the consideration of two different mass points or tourist types. The probability of belonging to mass point 1 is 0.15 and 0.85 for mass point 2. Figure 3 presents the predicted survivor functions for both types of individuals at the mean of covariates. It becomes apparent that each type of tourist behaves differently. Indeed, individuals within mass type 1 are more prone to enjoying long stays, as their predicted mean stay reaches a total of 23.9 nights. Visitors within type 2, in contrast, have a predicted average stay of 5.5 nights. According to these results, substantial differences in terms of the length of stay stem from the impact of unobserved heterogeneity.

Table 2. Maximum likelihood estimates for the discrete time proportional hazard models

	Model (1)		Model (2)		Model (3)	
	cloglog model without frailty		cloglog model with Gamma distributed frailty		cloglog model with non-parametric frailty	
	Hazard rate	St. errors	Hazard rate	St. errors	Hazard rate	St. errors
Duration dependence						
d2	3.0501	(0.6876)***	4.5743	(1.3238)***	3.2187	(0.7281)***
d3	3.3622	(0.7677)***	7.4337	(2.8914)***	3.7996	(0.8781)***
d4	3.325	(0.7833)***	10.1486	(4.9057)***	4.0156	(0.9676)***
d5	4.6158	(1.0684)***	19.5575	(11.4143)***	5.9628	(1.4345)***
d6	4.5467	(1.0948)***	26.7752	(18.6226)***	6.3386	(1.6116)***
d7	10.1678	(2.2958)***	95.3956	(81.4693)***	16.3213	(4.0813)***
d8	4.379	(1.2402)***	60.6460	(60.8018)***	8.0718	(2.5133)***
d9	2.5283	(0.8835)***	41.5434	(45.0648)***	4.9754	(1.8772)***
d1011	4.5625	(1.2028)***	112.1915	(135.4197)***	10.849	(3.4961)***
d1213	2.1188	(0.7224)**	71.9909	(97.2377)***	5.883	(2.4075)***
d14	6.398	(2.0062)***	285.6859	(412.878)***	21.1659	(8.7508)***
d15	16.3758	(4.4588)***	1,309.6050	(2198.641)***	74.1547	(32.6505)***
d1629	1.6341	(0.4854)*	332.3192	(668.7125)***	10.9825	(5.6061)***
d3031	20.9212	(6.8795)***	16,430.6300	(42203.52)***	183.8796	(95.0299)***
d32100	1.2644	(0.4731)	2,842.9180	(8230.71)***	13.9011	(7.2541)***
Explanatory variables						
Origin-Cat	Reference category		Reference category		Reference category	
Origin-Aragon	0.8628	(0.1426)	0.37	(0.1618)**	0.593	(0.1333)**
Origin-BC	0.499	(0.0885)***	0.0999	(0.0582)***	0.2458	(0.0619)***
Origin-Mad	0.6043	(0.0949)***	0.1915	(0.0853)***	0.3634	(0.0876)***
Origin-Rest	0.4359	(0.078)***	0.0982	(0.0584)***	0.2446	(0.0591)***
Sex-man	1.0355	(0.0911)	1.1543	(0.2227)	0.9371	(0.1064)
Age-30	0.9877	(0.114)	0.8191	(0.2039)	0.8065	(0.1168)
Age-40	Reference category		Reference category		Reference category	
Age-60	0.9685	(0.1106)	0.7029	(0.182)	0.7937	(0.1249)
Age-older	0.631	(0.0954)***	0.2339	(0.1014)***	0.3377	(0.0913)***
Educ-Primary	Reference category		Reference category		Reference category	

Educ-Secundary	1.142	(0.1784)	1.0843	(0.3627)	1.2086	(0.2364)
Educ-Tertiary	1.2598	(0.1731)*	1.2137	(0.3611)	1.2306	(0.2113)
Educ-Unknown	1.7459	(0.4549)**	1.8793	(0.9863)	1.8116	(0.8083)
CI	4.1475	(2.28)**	2.7602	(3.0945)	3.4587	(2.3687)*
CI^2	0.1456	(0.1114)**	0.2203	(0.3272)	0.2367	(0.2099)
Accom-Hotel	Reference category		Reference category		Reference category	
Accom-2resid	0.3838	(0.0599)***	0.2584	(0.0849)***	0.3286	(0.0736)***
Accom-Famfriends	0.6224	(0.0846)***	0.4941	(0.1363)**	0.5478	(0.1088)***
Accom-Rent	0.5675	(0.1041)***	0.4681	(0.1828)*	0.5241	(0.1273)***
Accom-Other	0.9268	(0.1891)	1.6252	(0.7817)	1.0447	(0.2792)
Trav-Agency	1.1356	(0.154)	1.7281	(0.5447)*	1.2835	(0.2734)
Dest-CD	Reference category		Reference category		Reference category	
Dest-Tarragona	0.8809	(0.0882)	0.8419	(0.1835)	0.8763	(0.1151)
Dest-Rest	1.135	(0.1851)	0.6407	(0.2261)	0.8352	(0.1731)
Who-Alone	Reference category		Reference category		Reference category	
Who-Friends	0.7475	(0.0898)**	0.5178	(0.143)**	0.6878	(0.124)**
Who-Adultfam	0.7405	(0.0953)**	0.4135	(0.1305)***	0.5708	(0.1034)***
Who-Chlidfam	1.0276	(0.1559)	1.0168	(0.3107)	1.0048	(0.1847)
Repeat	0.886	(0.1303)	0.9448	(0.2646)	1.024	(0.1745)
Constant	0.1012	(0.0319)***	0.386	(0.248)	0.0213	(0.0104)***
m2 Constant					10.1434	(3.374)***
Log likelihood	-1597.522882		-1590.4897		-1588.0774	
Number of observations		4,844		Number of spells	588	

Standard errors within parentheses. * Significant at 10%, ** significant at 5%, *** significant at 1%.



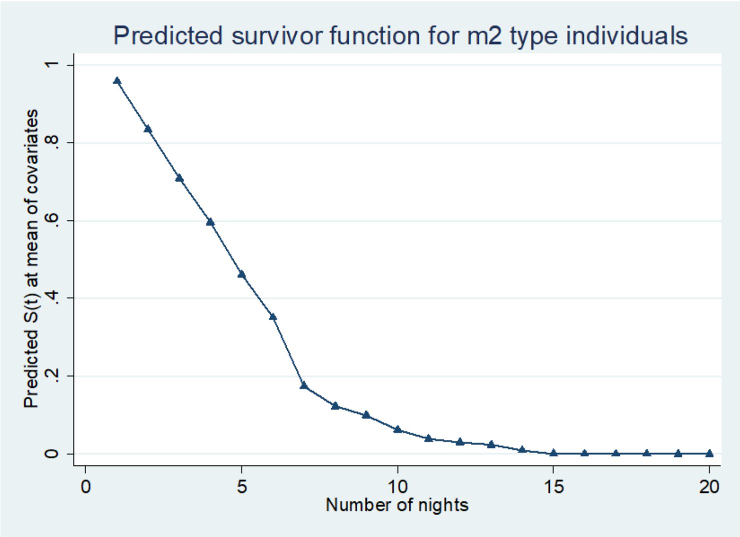


Figure 3. Predicted survivor functions for m1 and m2 types at mean of covariates

With respect to the impact of the CI, the variable has been introduced in the equations linearly and considering its quadratic value as well. The introduction of the squared value in the equations allows us to assess the existence of non-linear relationships. Both models (2) and (3) concur that there is no evidence of non-linear effects as the hazard rates associated with the squared value of CI are not significant. Focusing on the non-exponential effect of CI, we depart from the results obtained by Nicolau and Más (2009) and Martinez García and Raya (2008), who concluded that the availability of days is crucial for determining the length of stay. In accordance with their findings, it could be assumed that visitors attracted by the HSR station should have a preference for shorter stay periods, and hence, it is hypothesized that the hazard rate associated with the CI should be significant and larger than 1. Evidence regarding the impact of CI in models (2) and (3) is nonetheless mixed. While the hazard rate in model (2) is not significant, the hazard rate in model (3) is significant and signals, as expected, that higher CI values are linked to a higher risk of ending the stay at each interval and hence shorter stays. According to the definition of the CI, visitors with null CI values are those who firmly intend to spend a holiday in a specific tourist destination, and at the same time they should be those who do not have a preference for reducing the travel time to their destination by means of HSR services. In contrast, CI values larger than 0 are related to factors that push visitors towards the HSR station to some extent. In the context of HSR services, time constraints that limit visitors' length of stay should be the most decisive element among the restrictions. From this perspective, the higher the time constraints, the higher the CI. As a result, if the results of model (3) are accepted, the existence of an HSR station that shortens travel times would emerge as an incentive that would make the destination more attractive for short trips.

However, it must be taken into consideration that the results of models (2) and (3) oppose each other. This discrepancy could be simply the result of the more flexible, and as a consequence preferable, specification of the unobserved heterogeneity allowed in model (3) compared to model (2). It could happen as well that some underlying factors could be behind the mixed evidence obtained regarding the effect of the CI. In this sense, some attention should be placed on second residences, as the decision of visiting the Costa Daurada, in the case of this segment of visitors, is undoubtedly determined by the ownership of a property. In order to test to what extent the effect exerted by the CI is influenced by visitors staying in second residences, an interaction term between both variables was created and introduced in model (2) and model (3). The results regarding both variables and the interaction term are shown in table 3.

Table 3 Hazard rates estimates of the interaction term: CI x second residences

	Model (2)*		Model (3)*	
	cloglog model with Gamma distributed frailty		cloglog model with non-parametric frailty	
CI	2.4957	(2.7893)	2.1206	(1.47)
Accom-2resid	0.2114	(0.0812)	0.2808	(0.0745)***
CI x Accom-2resid	2.8437	(3.1932)	3.4224	(2.126)**
Number of observations: 4,844		Number of spells: 588		

Standard errors within parentheses. * Significant at 10%, ** significant at 5%, *** significant at 1%.

In the model where frailty follows a gamma distribution, neither the CI nor its interaction term with second residences are significant. Conversely, the hazard rate associated with second residences continues to be highly significant, and its effect (increasing the expected length of stay) is consistent with the previously estimated model without the interaction term (model 2 of table 2). On the other hand, the evidence obtained with the non-parametric frailty model depicts a different scenario in which the previous significance of the CI vanishes, while the interaction term is significant and its hazard rate acquires a similar magnitude to the hazard rate associated with the CI when the interaction term was not considered (model 3 of table 2). Second residences, similarly to the model with gamma-distributed frailty, remain highly significant, and its hazard rate is slightly smaller than in the model without the interaction term. The interpretation of the results obtained when non-parametric frailty is incorporated into the specification is interesting, as those results suggest that the effect of the CI only appears for the owners of second residences. In this sense, the availability of HSR in the area would foster additional short stays, although only for that segment that travels to the Costa Daurada with a certain degree of regularity as they own an apartment there. In contrast, the length of stay of visitors who stay overnight in alternative sorts of accommodations would not be impacted by the value of the CI; however, these conclusions should be taken with caution given the discrepancies between the models.

Back to table 2, the inclusion of unobserved heterogeneity slopes upwards with the duration-dependence dummies, which present a growing trend. The fact that all tourists in the sample finish their stay at some stage results in an increasingly positive relationship with the hazard rate. Furthermore, hazard rates associated with numbers of nights that refer to periods of time (such as a week [7 nights], fortnight [15 nights] or a whole month) are larger.

Longer travel times contribute to increasing the number of nights spent. In fact, those areas with the longest travel times (Madrid, Basque Country, and the rest of Spain) are associated with longer stays. Neither gender nor education level attained exert significant impacts, whereas age extends length of stay for the oldest population group given the fact that they face fewer time constraints[2]. This latter result is consistent with most of the research in this area (Santos et al, 2015), although some previous work suggests the opposite results (Barros et al, 2008; Yang and Zhang, 2015). The type of accommodation emerges as a decisive factor. Hotel stays reduce the number of nights, whereas the longest stays are attributable to second residences. The hazard rates associated with apartment rentals and stays with family and friends are also clearly below 1, although their impacts on the length of stay are more modest. Previous studies have obtained similar results (Mak and Nishimura, 1979; Martínez-García and Raya, 2008; Alegre et al, 2011; Santos et al. 2015). The daily price of each of the accommodation options is the underlying factor that accounts for this highly significant impact of the different types of accommodations (Fleischer and Byk, 2009). According to Menezes and Moniz (2011), Grigolon et al. (2014) or Yang and Zhang (2015), party size has been analysed from a party composition perspective. All three models signal that the greatest number of nights spent at the destination corresponds to trips accompanied by adult family members, followed by trips undertaken with friends. Thus, this result gives support to the vision that it is not only a matter of the party size, but also its composition. Finally, no significant impact is found to be associated with the location of the tourist destination, or whether the tourists have visited the destination previously. This latter result can be shocking because it opposes the conclusions reached by most of the research in the field. However, it must be taken into consideration there might be variables related to the accommodation options that could capture the effect of a repeating destination, mainly second residences. Some past studies have neither observed a significant influence of repeating a visit to a destination (Yang and Zhang, 2015, for one of their latent classes) or even found a negative impact on the length of stay (Thrane, 2012; and Santos et al, 2015 for subsequent visits after a second trip to Brazil).

[2] In the non-frailty model (model 1), the effect of tertiary education is significant at 10%, and it diminishes the length of stay. Nevertheless, the significance vanishes when frailty is introduced in models (2) and (3).

Conclusions

This research article develops an analysis of the determinants of the length of the stay at a given destination for specific tourists arriving by means of an HSR. The survival model has used the usual explanatory variables associated with tourist profiles and trip and stay characteristics. Additionally, it has also used the uneven incidence of the availability of HSR services on destination choice as an explanatory variable. This incidence has been calculated using a causal indicator. This novel approach has allowed us to obtain new data on the relationship between the capacity of HSR services to attract tourists and the length of the stay of those tourists.

The study findings unveil, in the first place, that the capacity to attract tourists associated with the availability of HSR services in the area is relatively limited. In second place, we have obtained weak evidence of an impact of this capacity of attraction on the tourists' length of stay as the imposition of a more rigid specification for unobserved heterogeneity results in a non-significant effect of the incidence of HSR services at all. On the contrary, when unobserved heterogeneity enters the estimation equation in a non-parametric form, which allows a higher degree of flexibility, HSR tends to reduce the number of nights spent by tourists at the destination. According to these results, tourists who reported a higher incidence of the HSR station on their decision to visit the Costa Daurada tended to undertake shorter stays. However, it has also been shown that this effect is restricted to those tourists who are lodged in second residences. This profile, which comprises recurrent visitors due to the ownership of a property within the area, would benefit from the travel time reduction yielded by the HSR to increase their number of visits; additional visits would be nonetheless shorter. In third place, the results have also highlighted a large and significant impact of unobserved heterogeneity on the number of nights spent.

This study also highlights other characteristics of the profiles of tourists who arrive by HSR and their stay at the destination, which significantly affects their length of stay. Older tourists, those arriving from farther away, those travelling in groups and, especially, those hosted in second residences are the ones who tend to spend longer time in the destination. On the other hand, the education level or gender have no significant effects on the length of stay of tourists arriving by means of the HSR.

The findings of this study provide evidences of interest for tourism-destination management, marketing, strategic planning, and for the planning of transportation facilities for destinations with HSR services (or those that plan to be served by HSR).

They are also of interest for railway-transportation study and management. Although less than 5% of all tourists vacationing at the Costa Daurada arrived by means of HSR in 2014, this is enough for the Camp de Tarragona Station to achieve its highest level of passengers in during the summer months, while the opposite occurs in the average Spanish HSR network. Nevertheless, the study aligns with those developed in Spain (that coincide in highlighting the moderate effects of HSR services on attracting new tourists to destinations). This contrasts with the over-estimated expectations linked to this infrastructure (Bellet et al., 2012). Moreover, as Saladié et al. (2016) noted, the availability of HSR services contributes to the diversification of the modes of transport used to reach a destination, which is also a key issue for tourism destinations and for transport planning and management.

The study also opens a door to future research. There is common agreement that the improvement of transportation systems has contributed to increasing the number of visits to some destinations while reducing the length of stays (Alegre and Pou, 2006; Barros and Machado, 2010; Lumsdon and Page, 2004). The availably of HSR services could be aligned to this context particularly when analysing repeat tourism, which has been identified as specific subgroup of interest. Considering this, a study based on repeat tourists as a target group could provide new evidence about whether a reduction in the length of the stay is associated with an increasing of number of visits to this or other destinations.

Finally, as some authors noted, the profiles of users of HSR services experience an evolution after the first years of entry into service of the high-speed line. However, as Gutiérrez and Ortuño (2017) highlighted, there is no specific profile of traveller using HSR for tourism, not even in the case of a coastal destination. The most-expected trend is the diversification of user profiles and the growing numbers of travellers due to tourism and leisure (Bazin et al., 2013; Santos et al., 2006). Taking these factors into consideration, a new study in the area in the coming years would offer evidence on the evolution of the effects of HSR stations in regard to the visitors attracted, their profiles, and their lengths of stay at their destinations.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Spanish Ministry of Economy and Competitiveness (POLITUR – CSO2017-82156-R).

References

Adongo, C. A., Badu-Baiden, F., & Boakye, K. A. A. (2017). The tourism experience-led length of stay hypothesis. *Journal of Outdoor Recreation and Tourism*, *18*, 65-74.

Albalate, D., & Fageda, X. (2016). High speed rail and tourism: Empirical evidence from Spain. *Transportation Research Part A: Policy & Practice*, 85, 174–185.

Alegre, J., & Pou, L. (2006). The length of stay in the demand for tourism. *Tourism management*, *27*(6), 1343-1355.

Alegre, J., Mateo, S., & Pou, L. (2011). A latent class approach to tourists' length of stay. *Tourism Management*, *32*(3), 555-563.

Anton Clavé, S., Saladié, O., Cortés-Jiménez, I., Fernandez Young, A., & Young, R. (2015). How different are tourists who decide to travel to a mature destination because of the existence of a low-cost carrier route? *Journal of Air Transport Management*, 42, 213-218.

Archer, B.H., & Shea, S. (1975). Length of stay problems in tourist research. *Journal of Travel Research*, 13(3), 8-10.

Barros, C.P., Correia, A., & Crouch, G. (2008). Determinants of the length of stay in Latin American tourism destinations. *Tourism Analysis*, *13*(4), 329-340.

Barros, C.P., & Machado, L.P. (2010). The length of stay in tourism. *Annals of Tourism Research*, *37*(3), 692-706.

Bazin, S., Beckerich, C., & Delaplace, M. (2013). Desserte TGV et villes petites et moyennes. Une illustration par le cas du tourisme à Arras, Auray, Charleville Mézières et Saverne. *Les Cahiers Scientifiques des Transport*, 63, 33-62.

Bellet, C., Alonso, P., & Gutiérrez, A. (2012). The High Speed Rail in the Spanish cities: urban integration and strategies for socioeconomic development. In: J.M. Ureña (coord.). *Territorial implications of High Speed Rail in Spain*. Aldershot: Ashgate, 163-192.

Brida, J. G., Meleddu, M., & Pulina, M. (2013). Factors influencing length of stay of cultural tourists. *Tourism Economics*, *19*(6), 1273-1292.

Cameron, A.C., & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, Econometric Society Monograph No.30.

Campa, J.L., López-Lambas, M.E., & Guirao, B. (2016). High speed rail effects on tourism: Spanish empirical evidence derived from China's modelling experience. *Journal of Transport Geography*, 57, 44-54.

Chalita, L. V., Colosimo, E. A., & Demétrio, C. G. (2002). Likelihood approximations and discrete models for tied survival data. *Communications in Statistics-Theory and Methods*, *31*(7), 1215-1229.

Chen, Z., & Haynes, K. (2015). Impact of high-speed rail on international tourism demand in China. *Applied Economics Letters*, 22(1), 57–60.

Crouch, G.I., & Ritchie, R.B. (2005). Application of the analytic hierarchy process to tourism choice and decision making: a review and illustration applied to destination competitiveness. *Tourism Analysis*, 10, 17-25.

Davies, B., & Mangan, J. (1992). Family expenditure on hotels and holidays. *Annals of Tourism Research*, 19, 691-699.

Decrop, A., & Snelders, D. (2004). Planning the summer vacation. An adaptable process. *Annals of Tourism Research*, 31, 1008-1030.

Delaplace, M., Pagliara, F., Perrin, J., & Mermet, S. (2014). Can High Speed Rail Foster the Choice of Destination for Tourism Purpose? *Procedia - Social and Behavioral Sciences*, 111, 166–175.

Delaplace, M., Pagliara, F., & La Pietra, A. (2016). Does high-speed rail affect destination choice for tourism purpose? *Belgeo*, 3/2016.

Dobruszkes, F., Dehon, C., & Givoni, M. (2014). Does European high-speed rail affect the current level of air services? An EU-wide analysis. *Transportation Research Part A: Policy and Practice*, 69, 461-475.

Ferrer-Rosell, B., Martínez-Garcia, E., & Coenders, G. (2014). Package and no-frills air carriers as moderators of length of stay. *Tourism Management*, *42*, 114-122.

Fleischer, A., & Byk, J. R. (2009). Quality, quantity and duration decisions in household demand for vacations. *Tourism Economics*, *15*(3), 513-530.

García-Sánchez, A., Fernández-Rubio, E., & Collado, M.D. (2013): Daily expenses of foreign tourists, length of stay and activities: evidences from Spain. *Tourism Economics*, 19, 613-630.

Givoni, M., & Dobruszkes, F. (2013). A review of ex-post evidence for mode substitution and induced demand following the introduction of high-speed rail. *Transport Reviews*, 33(6), 720-742.

Gokovali, U., Bahar, O., & Kozak, M. (2007). Determinants of length of stay: A practical use of survival analysis. *Tourism Management*, *28*(3), 736-746.

Gomes de Menezes, A., Moniz, A., & Vieira, J.C. (2008). The determinants of length of stay in the Azores. *Tourism Economics*, 14, 205-222.

Grigolon, A. B., Borgers, A. W., Kemperman, A. D., & Timmermans, H. J. (2014). Vacation length choice: a dynamic mixed multinomial logit model. *Tourism Management*, *41*, 158-167.

Gutiérrez, A., & Miravet, D. (2016a). The Determinants of Tourist Use of Public Transport at the Destination. *Sustainability*, 8(9), 908.

Gutiérrez, A., & Miravet, D. (2016b). Estacionalidad turística y dinámicas metropolitanas: un análisis a partir de la movilidad en transporte público en el Camp de Tarragona. *Revista de Geografía Norte Grande*, 65, 65-89.

Gutiérrez, A., & Ortuño, A. (2017). High speed rail and coastal tourism: Identifying passenger profiles and travel behaviour. *PloS one*, 12 (6), e0179682.

Gutiérrez, A., Saladié, Ò, & Anton Clavé, S. (2018). High-speed rail and tourism destination choice: the role and significance of the Camp de Tarragona station for passengers visiting the Costa Daurada. *Boletín de la Asociación de Geógrafos Españoles*, 76, 479-503.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271-320.

Lumsdon, L., & Page, S.J. (2004). *Tourism and transport: issues and agenda for the new millennium*. Oxford: Elsevier.

Martinez-Garcia, E., & Raya, J. M. (2008). Length of stay for low-cost tourism. *Tourism Management*, *29*(6), 1064-1075.

Mak, J., & Nishimura, E. (1979). The economics of a hotel room tax. *Journal of Travel Research*, *17*(4), 2-6.

Masson, S. and Petiot, R. (2009). Can the high speed rail reinforce tourism attractiveness? The case of the high speed rail between Perpignan (France) and Barcelona (Spain). *Technovation*, 29, 611-617.

Menezes, A. G., & Moniz, A. (2011). Determinants of length of stay: A parametric survival analysis. *Tourism Analysis*, 16(5), 509-524.

Mortazavi, R., & Cialani, C. (2016). International tourists' length of overnight stay in Venice. *Tourism Economics*, te-2016.

Nicolau, J.L., & Más, F.J. (2009). Simultaneous analysis of whether and how long to go on holidays. *The Service Industries Journal*, *29*(8), 1077-1092.

Pagliara, F., La Pietra, A., Gomez, J., & Vassallo, J. M. (2015). High Speed Rail and the tourism market: evidence from the Madrid case study. *Transport Policy*, 37, 187-194.

Pagliara, F., Mauriello, F., & Garofalo, A. (2017). Exploring the interdependences between High Speed Rail systems and tourism: Some evidence from Italy. *Transportation Research Part A: Policy and Practice*, 106, 300-308.

Paul, B. K., & Rimmawi, H. S. (1992). Tourism in Saudi Arabia: Asir National Park. *Annals of Tourism Research*, *19*(3), 501-515.

Prentice, R. L., & Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 57-67.

Raya-Vilchez, J. M., & Martínez-Garcia, E. (2011). Nationality and low-cost trip duration. A microeconometric analysis. *Journal of Air Transport Management*, *17*(3), 168-174.

Rodríguez, S., Dávila, D., Rodríguez, A., & Tejera, M. (2003). Determinantes de la demanda de turismo vacacional en las Islas Canarias. *Investigaciones regionales*, (3), 81.

Saarinen, J. (2006). Traditions of sustainability in tourism studies. *Annals of Tourism Research*, 33(4), 1121-1140.

Saladié, O., Anton Clavé, S., & Gutiérrez, A. (2016). Measuring the influence of the Camp de Tarragona highspeed rail station on first-time and repeat tourists visiting a coastal destination. *Belgeo*, 3/2016.

Salmasi, L., Celidoni, M., & Procidano, I. (2012). Length of Stay: Price and Income Semi-Elasticities at Different Destinations in Italy. *International Journal of Tourism Research*, *14*(6), 515-530.

Santos, J. M., Aguilera, J., Borderías, M.P., & González, M. P. (2006). La movilidad interurbana en la Línea de Alta Velocidad Madrid-Sevilla: Rasgos definitorios a los 10 años de su implantación. *Anales de Geografía,* 26, 147-165.

Santos, G.E.D.O., Ramos, V., & Rey-Maquieira, J. (2015). Length of stay at multiple destinations of tourism trips in Brazil. *Journal of Travel Research*, *54*(6), 788-800.

Sun, Y.Y., & Lin, Z.W. (2018). Move fast, travel slow: the influence of high-speed rail on tourism in Taiwan. *Journal of Sustainable Tourism*, 26(3), 433-450.

Thrane, C. (2012). Analyzing tourists’ length of stay at destinations with survival models: A constructive critique based on a case study. *Tourism Management*, *33*(1), 126-132.

Thrane, C. (2015). Research note: The determinants of tourists' length of stay: some further modelling issues. *Tourism Economics*, *21*(5), 1087-1093.

Thrane, C. (2016). Modelling tourists’ length of stay: A call for a ‘back-to-basic’approach. *Tourism Economics*, *22*(6), 1352-1366.

Um, S., & Crompton, J.L. (1991). Attitude determinants in tourism destination choice. *Annals of Tourism Research*, 17, 432-448.

Wang, D., Niu, Y., & Qian J. (2018). Evolution and optimization of China's urban tourism spatial structure: A high speed rail perspective. *Tourism Management*, 64, 218-232.

Wang, D., Qian, J., Chem ,T., Zhao, M., & Zhang, Y. (2014). Influence of the High-Speed Rail on the Spatial Pattern of Regional Tourism – Taken Beijing–Shanghai High-Speed Rail of China as Example. *Asia Pacific Journal of Tourism Research*, 19 (8), 890-912.

Yang, Y., & Zhang, H. L. (2015). Modeling tourists' length of stay: does one Model fit all?. *Tourism Analysis*, *20*(1), 13-23.

Young, R., Fernandez Young, A., Parkin, J. and Diamond, A. (2010). Assessing the economic impact of culture in English market towns: a causal chain approach. *Tourism Economics*, 16, 925-951.