



THE LENGTH OF STAY IN TOURISM

Carlos Pestana Barros

Technical University of Lisbon, Portugal

Luís Pinto Machado

Escola Profissional de Hotelaria e Turismo, Portugal

Abstract: Economic theory generally treats the duration of a vacation as a constraint on demand imposed by available time. In contrast, in this paper, it is shown that the length of stay is a determinant of destination demand rather than a demand characteristic. The length of stay is largely explained by the socio-demographic profile of the tourist, and moderated by the perceived characteristics of the destination. The length of stay is also found to have sample selection. Moreover, previous research that does not take sample selection into account is inadequate. Policy implications of the research findings, particularly those to increase the length of stay, are discussed. **Keywords:** duration models, survival models, sample selection, Madeira Island. © 2010 Elsevier Ltd. All rights reserved.

INTRODUCTION

In the present context of emerging new tourist markets, a declining average stay is observed in tourism length of stay, coupled with a trend toward an increasing frequency of trips throughout the year, supported by business travelers and low-cost flights (Martínez-García & Raya, 2008; Yang & Liu, 2003). Length of stay is therefore a key demand variable which has previously been neglected, but merits much greater research attention (Gokovali, Bahar, & Kozak, 2006). The research question of this paper is to examine the determinants of length of stay in a tourism destination—Madeira Island—using a sample of foreign tourists waiting to depart from the island at the end of their vacations and adopting a sample selection survival model. Survival models in tourism have been adopted by Barros, Correia, and Crouch (2008), Menezes, Moniz, and Vieira (2008), Gokovali et al. (2006) and Hong and Jang (2005). However, none of these authors took sample selection into consideration when analyzing the length of stay.

The motivations for the present research are the following. First, vacation duration is of paramount importance in tourism management. Hotels, for example, aim to maximize yield and G.O.P (gross operational profit), reduce fixed costs and maintain high rates of

Carlos Pestana Barros is an assistant professor of microeconomics (Email <cbarros@iseg.utl.pt>). His research interests are productivity analysis, survival analysis and categorical variables. **Luís Pinto Machado** is a teacher of tourism management and his main research interest is island tourism.

occupancy. Part of the strategy for doing so is to attract tourists who wish to stay longer, including repeaters, since the length of stay is influenced by previous visits, particularly during off-peak seasons. Therefore, it is important to ascertain which covariates best explain the length-of-stay decision (Gokovali et al., 2006), since the covariates of the length of stay are specific to the destination (Barros et al., 2008).

Second, survival models have proven to be particularly suited to the modeling and analysis of duration events. Although the usefulness and reliability of survival modeling for predicting duration events has been recognized (Menezes et al., 2008), their application to tourism has not attracted much interest research.

Third, it is important for policy purposes to investigate how tourists decide the duration of their stay. Tourists who visit only for short periods tend to stay centrally and visit only the major tourist attractions. Longer-stay tourists, by comparison, visit a greater range of attractions, explore more peripheral regions, and generate more diverse economic, social and environmental impacts. Families with children also tend to stay longer, in order to save money in the overall package.

Fourth, the research on length of stay has to be parsimonious if it aims to present relevant policy implications. Despite this aim, all published papers use many variables that restrict the utility of the result for policy prescription purposes, since having too many relevant variables confounds the policy. Therefore, one aim of the present research is to adopt a parsimonious procedure that restricts the variables to be used in order for them to have relevant policy implications.

Finally, the consequence of sample selection is taken into account in the present research. Ignoring it in duration models will bias parameter estimates in an unknown direction, rendering conclusions drawn from them at best tenuous (Boehmke, Morey, & Shannon, 2006). Sample selection may involve selecting the samples that will preferably include or exclude certain types of results. Typically, this causes measures of statistical significance to appear much stronger than they in fact are. Selection bias can be the result of scientific fraud which manipulates data directly, but more often is either unconscious, or due to biases in the instruments used for observation. To determine in a particular setting whether there is selection bias or not, it is not sufficient to establish that there has been selection. Instead, one must establish that the quantity of interest (length of stay, for example) is systematically different in the sample than in the entire population of interest, as the selection procedure may simultaneously lead to bias in the length of stay, but not in another variables.

The paper contributes to the relevant literature, first, by analyzing the determinants of the length of stay, in this instance, of foreign tourists in Madeira Island. A long-established tourist destination that has attracted little research so far, and second, by adopting a survival model with sample selection, which is a novelty in the survival models context. We are not aware of any research paper previously adopting survival models with sample selection in tourism.

Madeira Island

Madeira Island is the largest in a archipelago of 4 islands, situated in the Atlantic Ocean 360 miles west of the African coast and some 600 miles from Lisbon. The uninhabited archipelago was discovered by the Portuguese in 1420 and its value immediately perceived for its timber, in view of the vast rain forests found there (madeira is the Portuguese word for wood).

Tourism to Madeira Island dates from the late nineteenth century, in the form of wealthy British visitors who were passengers on transatlantic liners and cargo ships. It remained the exclusive reserve of the well-to-do until the opening of the airport in the capital, Funchal, in 1963, which was to lead to the development and expansion of infrastructure and the construction of more hotels.

While most tourists today arrive by air, Funchal is a port of call throughout the year on the itineraries of many European and transatlantic cruise-ships. Blessed by a climate that benefits from the geographical location and its protective mountains, Madeira offers all-year-round benign temperatures and low rainfall. This nourishes the lush, semi-tropical vegetation and flora for which Madeira is famous, in the form of an abundance of exotic flowers, sub-tropical and tropical fruits and Madeira wine. Modern tourism in Madeira offers great opportunities to the many dedicated walkers who visit for this purpose, in addition to mountaineering, deep-sea fishing, sailing and surfing, [Oliveira and Pereira \(2008\)](#).

The destination is particularly popular with tourists from the UK, Germany and Scandinavia between October and the spring, the peaks being in March and April. The summer season tends to attract more visitors from southern Europe, who seek to escape to Madeira's less intense heat and more peaceful atmosphere. New Year's Eve is also a peak period, famously marked by a firework display over Funchal city that is claimed to be the largest in the world.

[Figure 1](#) presents Madeira Island.

However, the average duration of the vacations has tended to decline in recent years, raising concerns over the long-term competitive position of Madeira as a sun-and-sea tourism destination ([Aguiló, Alegre, & Sard, 2005](#)).

The Length of Stay in Tourism Research

Studies using duration models in tourism are rare. Recently, [Kaniovski, Peneder, and Smeral \(2008\)](#) have analyzed firm survival in the Austrian accommodation sector with a gamma duration model. They conclude that larger size and market share increase survival, while turnover and vintage decreases the survival.

[Barros et al. \(2008\)](#) have analyzed the length of stay of Portuguese tourists in South America with alternative survival models. The Cox model, the Weibull model, the logistic model and the Weibull with heterogeneity were adopted. The covariates used were budget, destination attributes, socio-demographic characteristics, previous visits, temporal

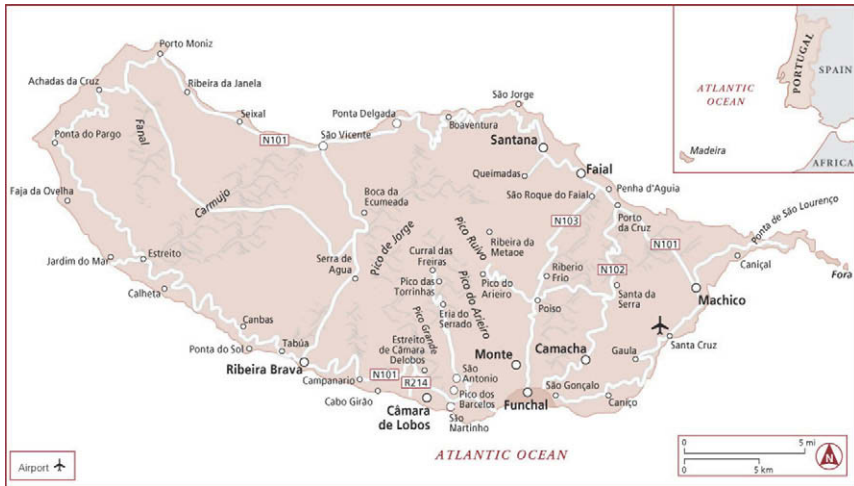


Figure 1. Map of Madeira Island

constraints and the frequency of travel. The main conclusion is that the length of stay depends on multiple determinants and is specific to each tourism destination.

Menezes et al. (2008) analyzed the length of stay of tourists in the Azores islands with a Cox proportional hazard model and using questionnaire data. The covariates used were socio-demographic profiles, trip attributes, sustainability practices and destination image attributes. The main conclusion is that the covariates explain the length of stay and the repeat visitor is particularly important in this context.

Martín-García and Raya (2008) have analyzed low-cost tourism length of stay in Spain with log-logistics and Cox survival models. Questionnaire data was used. The covariates used were nationality, socio-economic characteristics, occupation, reason for travel, type of accommodation, organized travel, travelling with family, high season, and geographical area. The main conclusion was that the hypothesis of hazard ratio is proportionality over time of the Cox survival model was not fulfilled and therefore, an accelerated survival model was adopted. Moreover, only some covariates were statistically significant, namely, tourists' nationality, age, level of education, type of education, type of accommodation, the season, and the geographical area.

Gokovali et al. (2006) applied survival analysis to model the length of stay of tourists in Bodrum, Turkey, using two survival model specifications: the Cox model and the Weibull model. The covariates used were nationality, age, job type, socio-economic characteristics, package-vacation type (full-board, bed and breakfast, full-package), past visits, quality, level of hospitality, attractiveness, night life, accommodation, image, promotion and publicity, and recommendation. The main conclusion is that the covariates explain the length of stay. Hong

and Jang (2005) analyzed the duration of visits to a casino with survival models.

According to the best of the authors' knowledge, the present study is the first to adopt this technique in tourism, allowing for sample selection. Moreover, a parsimonious procedure is adopted, in order to extract meaningful and practicable policy implications. Therefore this paper contributes to tourism literature adopting a sample selection survival model to measure tourism vacations duration.

Theoretical Framework

The survival model estimated in this study is based on the theory of consumer behavior developed by Lancaster (1966) and the concept of hedonic prices (Rosen, 1974). The economic theory of consumer behavior assumes that a consumption decision faced by an individual aiming to maximize utility, subject to budgetary restrictions, is taken on the basis of prices and income (Varian, 1987:95). This traditional framework, however, does not allow for circumstantial conditions and social factors which are known to play an important role in shaping tourism demand, given the composite or differentiated nature of the services that comprise a tourist destination (Eugenio-Martin, 2003; Morley, 1992; Morley, 2009; Sakai, 1988). Therefore, the demand for tourism is derived from the demand for the various goods and services offered by a tourism destination (Ben-Akiva & Lerman, 1985), adopting the Lancaster approach. Moreover, the duration of the length of stay is estimated by traditional demand models (Crouch, 1994).

Method

In this study, the vacation length of stay is analyzed using a survival modeling approach (Allison, 1984; Cleeves, Gould, & Gutierrez, 2002; Cox & Oakes, 1984; Hosmer & Lemeshow, 1999; Kalbfleisch & Prentice, 2002; Yamaguchi, 1991). Survival analysis, also known as the duration model, measures the duration of an event, which is the time elapsed until a certain event occurs, or is completed. The length of a vacation is an example of a duration event. Time was traditionally measured in economics using the Poisson or negative binomial models (Palmer-Tous, Riera-Font, & Rosseló-Nadal, 2007). However, with the development of survival models, these were applied to the analysis of time (Martínez-García & Raya, 2008; Gokovali et al., 2006). The former model analyzes the discrete variable, i.e., how many times (how many days) an event occurs, whereas the latter measures the time elapsed between two events, which is a continuous variable.

The use of survival models in order to model duration is based on the fact that the error distribution in this context, by necessity, must be skewed to the right (Hosmer & Lemeshow, 1999). The survival model regresses the duration of an activity on covariates. Such models are generally adopted in economics on the basis of two facts: (i) time, as a dependent variable, is strictly positive, and therefore the use of the

traditional Gaussian distribution is not adequate to capture the characteristics of the time variable, and (ii) in clinical trials, censoring occurs when an individual participant in the initial phase of the study subsequently dies. Survival analysis can adequately accommodate the loss of observations when censoring occurs.

Traditional regression models are unable to resolve this issue and therefore, survival models, such as the Cox model and the Weibull model, have appeared (Hosmer & Lemeshow, 1999). The dependent variable of interest is the number of days that the respective tourist spends on vacation, which is regressed against covariates. Survival models in tourism include Kaniovski et al. (2008), Barros et al. (2008), Menezes et al. (2008), Martínéz-García and Raya (2008) and Gokovali et al. (2006).

Three issues must be addressed when analyzing survival models: 1) identification of the data set (i.e., cross-section vs. panel data); 2) censoring of the data; and 3) sample selection. With regard to the first issue, the present study adopted a cross-sectional approach. Therefore, time-invariant modeling, known as proportional hazard models, was adopted (Wooldridge, 2002). In terms of censoring, Gokovali et al. (2006) concluded that questionnaire data is uncensored, because the tourists were observed at the end of their vacations.

A survival time is described as censored when there is a follow-up time but the event has not yet occurred, or is not known to have occurred. For example, if the length of time being studied concerns a beach resort and the tourist will still be on vacation at the end of the data-gathering phase, then the start of vacation is observed, but the end time would be censored. If a tourist for some reason drops out of a study before the end of the study period, then that tourist's follow-up time would also be considered to be censored, since it is unobserved. Since our questionnaire surveyed tourists who had completed their vacations and were waiting to board their home-ward-bound flights, the length of stay is completely determined. With regard to the third issue, ignoring sample selection may bias the results (Cameron & Triverdi, 2005).

Sample Selection

Sample selection occurs when unobserved factors that influence the process of interest help determine whether that process is observed. When this happens, values of the dependent variable for uncensored observations are systematically unrepresentative of the population studied. If a tourist elects to answer a questionnaire selecting him/herself because he/she is satisfied with the tourist destination, while dissatisfied tourists decline to answer, we have a selection problem.

Selection bias can be avoided by estimating a model that accounts for the satisfaction while simultaneously estimating the factors that influence the duration. Relative to the length of stay, let us assume that some tourists have repeated their visit to Madeira and are therefore subjectively enthusiastic towards this destination. Such tourists will be

more likely to answer the questionnaire than tourists who are on their first visit and do not plan to return, therefore the answers are biased towards those who are repeating the visit. Under this conditions, ‘repeat’ will be endogenous. Controlling for repetitions will enable the researcher to correct the results.

The sample selection model used in this paper was proposed by [Boehmke et al. \(2006\)](#). The objective is to model a continuous duration variable y_i with an associate density $f(y_i)$. The selection depends on a set of covariates by letting $y_i = \exp(-X_i\beta)\varepsilon_i$. The selection process leads us to observe y_i only for observations for which the binary censoring variable c_i takes the value 1 (repeat) and d_i is an indicator of right censoring. The Weibull likelihood with right censoring is the following:

$$\begin{aligned} \ln L(\beta, \gamma, \alpha, p | X, W, y, c, d) = & \sum_{i=1}^n c_i(1 - d_i)[- \lambda_{1i} \\ & + \ln(1 + \alpha(2 \exp(-\lambda_{2i}y_i)^p - 1) \exp(\lambda_{1i}) - 1)) + \ln(p) \ln(\lambda_{2i}) \\ & + (p - 1) \ln(\lambda_{2i}y_i - (\lambda_{2i}y)^p) + c_i d_i[- \lambda_{1i} - (\lambda_{2i}y_i^0)^p \\ & + \ln(1 + \alpha(1 - \exp(-\lambda_{2i}y_i)^p)(1 - \exp(-\lambda_{1i}))) \\ & + (1 - c_i)[\ln(1 - \exp(-\lambda_{1i}))]] \end{aligned} \quad (1)$$

where β , γ , α are parameters to be estimated; p is the Weibull distribution shape parameter when the error term follows an exponential distribution; X is a vector of explanatory variables; W is the Weibull distribution; y is the length of stay; c is a censoring binary variable; and d is an indicator for right-censored variables. $\lambda_{1i} = \exp(-w_i\gamma)$ and $\lambda_{2i} = \exp(x_i b)$ ([Boehmke et al., 2006](#)).

Research Design: Hypotheses

As noted above, the duration of a vacation can be explained by several factors. The survey questionnaire therefore gathered data pertaining to: 1) socio-demographic characteristics; 2) nationalities; 3) expenditure; 4) destination attributes; 5) previous visits to the destination and 6) quality. Using the survey data on these characteristics, we tested the following hypotheses.

Hypothesis 1 (Socio-demographic characteristics): the length of stay is a positive function of individual socio-demographic characteristics such as age, gender and level of education, ([Goodall & Ashworth, 1988](#); [Weaver, McCleary, Lepisto, & Damonte, 1994](#); [Woodside & Lysonski, 1989](#); [Zimmer, Brayley, & Searle, 1995](#)).

Hypothesis 2 (Nationalities): the length of a vacation is a positive function of the nationality of the tourist, reflecting his/her income. This is a traditional hypothesis in tourism survival models ([Gokovali et al., 2006](#)).

Hypothesis 3 (Expenditure): the length of a vacation is a negative function of the individual's expenditure. This is a traditional hypothesis in tourism demand models, in which price, income and budget constraints define the frontier of consumption possibilities for travel (Aguiló Perez & Sampol, 2000; De la Vina & Ford, 2001; Hay & McConnel, 1979; Nicolau & Más, 2005).

Hypothesis 4 (Destination attributes): the length of stay is a positive function of a destination's attractiveness in attributes such as wine production, traditions, casinos and nature. Woodside and Lysonski (1989) argue that a destination's image and its choice are influenced by destination attributes such as those considered in this study (Wanhill & Lundtorp, 2001).

Hypothesis 5 (Previous visits): the length of stay is positively influenced by previous visits. Festinger (1954) noted that satisfaction in relation to the destination influences future choices. Beerli and Martín (2004) demonstrated that sun-and-sand destinations with a favorable image have a high level of repeat visitors. Kozak (2001) demonstrated that overall satisfaction and the number of previous visits considerably influences the intention to return, especially in mature destinations. Kozak (2003) also concluded that destination attributes influence future behavioral intentions and satisfaction.

Hypothesis 6 (Quality): the length of stay is a positive function of the hotel quality. Hotel quality is a main factor in tourism length of stay (Gokovali et al., 2006) and quality is evaluated by the perceptions of tourists' families (Aguiló Perez & Rosseló Nadal, 2005).

Survey and Study Context. The research study was undertaken in March and April 2008. Tourists departing from Funchal Airport, Madeira were randomly approached to answer the questionnaire. Although the questionnaire took the individual tourist as reference, family context was taken into account. The sample was defined by the confidence interval approach (Burns & Bush, 1995). The formula used to estimate a 95% confidence level was:

$$n = \frac{z^2(pq)}{e^2} = \frac{1.95^2(0.5 \times 0.5)}{0.5^2} = 385 \quad (2)$$

where z is the standard error associated with the chosen level of confidence (95%); p is the estimated variability in population (50%); $q=1-p$; and e is the acceptable sample error $\pm 5\%$ (desired accuracy 95%). The amount of variability in the population is estimated to be 50%, which is frequently the case in sociological research (Chi & Qu, 2008).

The interviewer approached the randomly-selected tourist while he/she was waiting in the departure lounge to board the flight home.

Table 1. Characterization of the Variables

Variable	Description	Min ^a	Max ^b	Mean	Std. Dev
Stay	length of stay in days	1	60	9.01	4.74
Repeat	Repeated visit (1-no, 2-Once, 3-twice, 4-three times and 5-regularly)	1	5	0.49	0.68
Wine	<i>Buy wine (no=0,yes=1)</i>	0	1	0.40	0.49
Age	<i>Age of the individual</i>	16	79	44	15
Gender	Gender of the individual (female=0, male=1)	0	1	0.59	0.49
Education	<i>Level of education</i> (1-high school, 2-university, 3-Masters, 4-Ph.D)	1	4	2.95	0.72
British	British tourist (British tourist=1, elsewhere=0)	0	1	0.63	0.48
German	German tourist (German tourist=1, elsewhere=0)	0	1	0.09	0.29
Dutch	Dutch tourist (Dutch tourist=1, elsewhere=0)	0	1	0.03	0.18
French	French tourist (French tourist=1, elsewhere=0)	0	1	0.03	0.19
Casino	<i>Visit casino</i> (1-no, 2-Once, 3-twice, 4-three times and 5-regularly)	1	5	3.14	1.15
Nature	<i>Visit the island's flora & natural resources (No=0, Yes=1)</i>	0	1	0.65	0.47
Expenditure	<i>How much did you spend during your vacation? 1-less than 500 euros; 2-between 500 and 1000; 3-between 1000 and 2000; 4-between 2000 and 3000; 5-between 3000 and 4000; 6-more than 5000.</i>	1	6	3.03	1.18
Quality	Hotel quality (1-low quality ... 5-high quality)	1	5	3.86	0.91

^a Min—Minimum; ^b Max—Maximum.

Budgetary restrictions limited the number of questionnaires to 550, assuming a response rate of 80% and the proportion of unusable questionnaires at 10% (385/0.7=550). The number of answered questionnaires obtained was 498, but only 346 were usable, due to errors and incomplete fields.

The general characteristics of the respondents were that they were male (52%) with an average age of 33, traveling with family-members (93.8%). On average, they were middle-class, with a family that includes one child. Other characteristics of the sample are summarized in Table 1, which also broadly indicates the items contained in the questionnaire.

The survey contained three types of variables: dichotomous variables, continuous variables and qualitative variables (using a 5-point Likert scale). The set of explanatory variables considered in this study sought to capture the key determinants of the tourist decision process, based on the theoretical framework and the literature review (Barros et al., 2008). The reference individuals was Portuguese tourists (20%).

In order to evaluate the degree of correlation among the explanatory variables, which can cause multicollinearity, the correlation matrix is estimated, concluding that it was low among the variables used in the analysis (these results are available on request from the authors).

Table 2. Survival Model Results

	Weibull model without sample selection	Weibull model with sample selection
		Selection
Repeat	-0.27 (8.10)	0.15 (1.83)*
Wine	-0.21 (7.12)	0.28 (1.37)*
Constant		-0.77 (1.18)*
		Duration
Age	0.01 (2.05)*	0.02 (1.54)*
Gender	0.04 (3.24)	0.53 (1.23)**
Education	0.09 (4.12)	0.34 (1.89)**
British	-0.55 (1.24)*	-0.06 (0.75)*
German	0.06 (3.92)	0.54 (1.56)*
Dutch	-0.73 (0.28)*	-0.97 (0.51)*
French	-0.22 (7.67)	-1.80 (1.32)*
Casino	0.03 (1.10)**	0.05 (1.10)**
Nature	0.21 (0.84)*	0.25 (0.79)*
Expenditure	-0.17 (0.63)*	-0.38 (0.87)*
Quality	0.24 (0.17)*	0.33 (0.31)*
Constant	-3.99 (1.05)*	-5.29 (0.94)*
Ln p	0.875	1.02
P (duration dependence)		2.77
Rho (error correlation)		0.24
Nobs	346	346
Log Likelihood	-204.46	-281.405
Wald chi2(3)	51.97	11.43

1—All models were estimated in Stata 10. Standard Errors in parentheses are below the parameters, those followed by * are statistical significant at 1% level. Those with ** are statistical significant at 5% level.

Results

Table 2 presents the results of the estimated duration models. The Weibull model is presented, both without and with sample selection, for comparative purposes. The dependent variable is the length of stay, measured in days, of tourists about to fly home from Madeira Airport after their vacations.

The results obtained from the two models are clearly not similar in their main effects. Given the model specification, positive values for the parameters implies that the length of stay increases with increasing values in the respective variable. A negative value for the parameters implies a negative relationship. The results across the two models demonstrate that the parameters have different signs for endogenous variables and the same signs but different values for exogenous variables, underlining the importance of controlling sample selection in survival models.

On the basis of the log likelihood statistic, the Weibull model with sample selection provides the superior fit to the data. The rationale for this result is that sample selection represents unobserved characteristics that influence the conditional probability of the length of stay in

Madeira which are not measured and therefore, not taken into account in the measurement errors of the variables; this can cause bias (Boehmke et al., 2006). The ancillary duration dependence parameter ρ is higher than one in sample selection model, implying that the length of stay increases with the events, but it is lower than one for the model without sample selection.

Taking as reference the Weibull model with sample selection, as hypothesized, the socio-demographic variables *age*, *gender* and *education* have a positive effect on the hazard, which means that older male and more educated tourists tend to stay longer. Moreover, different nationalities have different impacts on the hazard. British, Dutch and French tourists have a negative impact, while German tourists have a positive impact.

Furthermore, *expenditure* has a negative impact on hazards, which means that high-spending tourists stay for shorter lengths of time, which is rational for tourists on constrained budgets.

Moreover, some destination attributes such as old and the well-established reputation of Madeira *wine* have a positive impact on the hazard, signifying that these attributes increase the length of stay in Madeira.

In addition, *previous visits* has a positive effect on the hazard, indicating the number of repeat tourists.

Finally, quality increases the length of stay, having a positive effect on the hazard.

DISCUSSION AND CONCLUSION

The paper presents a sample selection duration model to measure tourists' length of stay in the vacation destination of Madeira (Boehmke et al., 2006). The sample selection accounts for unobserved characteristics that influence the conditional probability of the length of stay in Madeira. Sample selection in questionnaire data is a common characteristic which has not previously been applied in tourism research.

Conclusions drawn without taking sample selection into account may be erroneous (Heckman, 1979). In the present case, the sample selection variables change the sign and the value when sample selection is taken into account, signifying that traditional survival models may distort results when they do not consider the sample selection problem. *Previous visits*, a variable used in almost all the papers reviewed, is found to be sample-selected in the present research. This is an intuitive result, since repeat visitors are naturally enthusiastic about the destination and are keen to demonstrate this in the questionnaire.

The general conclusion is that older, male and educated German tourists are more likely to stay the longest in Madeira. This is a broadly intuitive result.

The results support the majority of the hypotheses. Socio-economic characteristics explain the length of stay, validating Hypothesis 1. Hence, age, gender and education increase the length of stay in Madeira.

Second, nationalities also explain the length of stay, validating Hypothesis 2. Specifically, Germans stay longer than the other nationalities (British, Dutch and French).

Third, expenditure explains the length of stay, validating Hypothesis 3. It is found that the higher the expenditure, the shorter the length of stay.

Fourth, destination attributes such as wine, casinos and nature explain the length of stay. However, wine is found to be sample-selected.

Fifth, previous visits increase the length of stay, validating Hypothesis 5. However, this variable is sample-selected.

Sixth, quality increases the length of stay, validating Hypothesis 6. Therefore, longer-stay tourists tend to consume Madeira wine, have visited Madeira at least once previously and tend to manage their expenditure in order to maximize their stay, given that the higher budgetary expenditure, the shorter the stay and the higher the hotel quality, the shorter the stay.

British tourists tend to stay shorter periods, a result that may be related to the low-cost flight opportunities available to them and to the economic crisis UK suffered in the period of analysis, whereas Germans tend to travel on charter flights and package tours.

The parsimonious results of the present research permit us to extract a clear, focused view of the length of stay in the tourism destination analyzed. Thus, parsimonious modeling is called for in this field of research.

With regard to the policy implications, while recognizing that it is not possible to control all the elements contributing to length of stay in Madeira, it is possible to concentrate on those identified, by means of advertising and promoting tourist attractions, such as nature, wine and the casino; in addition, marketing can be oriented towards more educated, elderly German men, and the service quality provided by the tourism infrastructure can be maximized (Chi & Qu, 2008). Therefore, the results suggest that tour operators seeking to attract tourists to Madeira for extended lengths of stay should target the older, male, educated German market. Moreover, as low-cost travel is related to shorter length of stay, promotion policies should be adopted that take this into account when campaigns seeking to increase the length of stay are designed.

The research found that a survival model that accounts for sample selection is superior to models that do not assume it (Boehmke et al., 2006). Therefore, sample selection should be taken into consideration in questionnaire data. With regard to the managerial implications, this research provides a deeper insight into the driving motives behind tourists' length-of-stay decisions and can thus aid the development of marketing strategies. The statistically positive significant variables should be used in promotional strategies to increase the length of stay.

The sample employed in this research was restricted to a single destination in a limited number of months, thus the conclusions cannot be generalized to other travel contexts. However, since this research was an exploratory study designed to investigate and demonstrate the suitability of survival modeling as a means of developing a deeper insight into the factors which drive tourists' length-of-stay decisions,

these limitations are not critical. Survival modeling has been shown to be a useful technique for this purpose.

With regard to comparisons of these findings with previous research, first, statistically significant variables were chosen to explain the length of stay, recognizing that the relevant attributes are specific to each tourism destination (Barros et al., 2008). Second, the results confirm that socio-economic variables such as age and income increase the length of stay in all cited studies (Menezes et al., 2008). The Gokovali et al. (2006) study found that nationality was the main explanatory variable.

In terms of other variables common across studies, the general conclusion is that tourists seem to behave similarly in different contexts relative to some characteristics (socio-economic variables, nationalities, quality), but differences across studies are evident with respect to the role of a number of other variables (destination attributes). Therefore, different destination attributes define the tourism destinations, which are specific to each of them. In the present case, rather than sea and sun, the main destination attributes are found to be nature, wine, and the casino.

Finally, in the present research, a parsimonious procedure identified the small group of statistically significant variables among the many insignificant variables. There is, thus, a need for further research to apply survival modeling in analyzing the determinants of length of stay in a variety of other contexts. As the body of research into the determinants of length of stay grows, at some point in time it may be possible to investigate the differences in results across studies using meta-analytical techniques.

This paper has analyzed the length of stay of European tourists taking vacations on the island of Madeira. Two duration models were presented for comparative purposes, namely, the parametric Weibull model without sample selection and the Weibull model with sample selection. The Weibull model with sample selection displayed the superior performance on the basis of the log-likelihood statistics. It is concluded that the length of stay is positively related to *age*, *gender*, *education*, *German*, *buy wine* and *previous visit*. The length of stay is negatively related to other variables including *British*, *Dutch* and *French* and *expenditure*. The length of stay is positively related to *quality*. These results are intuitively consistent with the concept that economic affluence influences the length of stay. More research is needed to confirm the present findings. ■

Acknowledgements—UECE (Research Unit on Complexity and Economics) is financially supported by FCT (*Fundação para a Ciência e a Tecnologia*), Portugal. This article is part of the Multi-annual Funding Project (POCI/U0436/2006).

REFERENCES

- Aguiló, E., Alegre, J., & Sard, M. (2005). The persistence of the sun and sea tourism model. *Tourism Management*, 26, 219–231.
- Aguiló Perez, E., & Rosselló Nadal, J. (2005). Host community perceptions: A cluster analysis. *Annals of Tourism Research*, 32(4), 925–941.

- Aguiló Perez, E., & Sampol, C. J. (2000). Tourism expenditure for mass tourism markets. *Annals of Tourism Research*, 27(3), 624–637.
- Allison, P. D. (1984). *Event history analysis*. Beverly Hills, CA: Sage.
- Barros, C. P., Correia, A., & Crouch, G. (2008). Determinants of the length of stay in Latin American tourism destinations. *Tourism Analysis*, 13, 329–340.
- Beerli, A., & Martín, J. (2004). Tourists' characteristics and the perceived image of tourist destinations: A quantitative analysis—A case study of Lanzarote, Spain. *Tourism Management*, 25, 623–636.
- Ben-Akiva, Moshe, & Lerman, S. (Eds.). (1985). *Discrete choice analysis: Theory and applications to travel demand*. MIT Press: Cambridge.
- Boehmke, F., Morey, D. S., & Shannon, M. (2006). Selection bias and continuous-time duration models. *American Journal of Political Science*, 50, 192–207.
- Burns, A. C., & Bush, R. F. (1995). *Marketing research*. New Jersey: Prentice-Hall.
- Cameron, A. C., & Triverdi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge, N.Y.: Cambridge University Press.
- Chi, C. G. Q., & Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tourism Management*, 29, 624–636.
- Cleaves, M. A., Gould, W. W., & Gutierrez, R. G. (2002). *An introduction to survival analysis using Stata*. College Station TX: Stata Press.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London, UK: Chapman and Hall/CRC Press.
- Crouch, Geoffrey I. (1994). The study of international tourism demand: A review of findings. *Journal of Travel Research*, 33, 12–23.
- De la Vina, L., & Ford, J. (2001). Logistic regression analysis of cruise vacation market potential: Demographic and trip attribute perception factors. *Journal of Travel Research*, 39(4), 406–410.
- Eugenio-Martin, J. L. (2003). Modelling determinants of tourism demand as a five-stage process: A discrete choice methodological approach. *Tourism and Hospitality Research*, 3, 341–354.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Gokovali, U., Bahar, O., & Kozak, M. (2006). Determinants of length of stay: A practical use of survival analysis. *Tourism Management*, 28, 736–746.
- Goodall, B., & Ashworth, H. (Eds.). (1988). *Marketing in the tourism industry. The promotion of destination regions*. London: Routledge.
- Hay, M., & McConnell, K. (1979). An analysis of participation in nonconsumptive wildlife recreation. *Land Economics*, 55, 460–471.
- Heckman, James J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153–161.
- Hong, S. K., & Jang, H. (2005). Factors influencing purchasing time of a new casino product and its managerial implications: An exploratory study. *Journal of Travel Research*, 43, 395–403.
- Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis: Regression modeling of time to event data*. N.Y.: John Wiley.
- Kalbfleisch, J. D., & Prentice, L. (2002). *The statistical analysis of failure time data* (2nd ed.). New York: John Wiley and Sons.
- Kaniovski, S., Peneder, M., & Smeral, E. (2008). Determinants of firm survival in the Austrian accommodation sector. *Tourism Economics*, 14(3), 527–543.
- Kozak, M. (2001). Repeaters' behavior at two distinct destinations. *Annals of Tourism Research*, 28, 784–807.
- Kozak, M. (2003). *Measuring tourist satisfaction with multiple destination attributes tourism analysis*, 7, 229–240.
- Lancaster, Kelvin J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, 132–157.
- Martínez-García, E., & Raya, J. M. (2008). Length of stay for low cost tourism. *Tourism Management*, 29(6), 1064–1075.
- Menezes, A. G., Moniz, A., & Vieira, J. C. (2008). The determinants of length of stay of tourists in the Azores. *Tourism Economics*, 14, 205–222.
- Morley, C. L. (1992). A microeconomic theory of international tourism demand. *Annals of Tourism Research*, 19, 250–267.

- Morley, C. L. (2009). Dynamics in the specification of tourism demand models. *Tourism Economics*, 15, 23–39.
- Nicolau, J., & Más, F. (2005). Stochastic modelling: A three-stage tourist choice process. *Annals of Tourism Research*, 32, 49–69.
- Oliveira, P., & Pereira, P. T. (2008). Who values what in a tourism destination? The case of Madeira Island. *Tourism Economics*, 14(1), 155–168.
- Palmer-Tous, T., Riera-Font, A., & Rosselló-Nadal, J. (2007). Taxing tourism: The case of rental cars in Mallorca. *Tourism Management*, 28, 271–279.
- Rosen, S. (1974). Hedonic prices and implicit markets: Products differentiation in pure competition. *Journal of Political Economy*, 82, 34–35.
- Sakai, M. Y. (1988). A micro-analysis of business travel demand. *Applied Economics*, 20, 1481–1495.
- Varian, H. R. (1987). *Intermediate microeconomics: A modern approach*. New York: W.W. Norton.
- Wanhill, S., & Lundtorp, S. (2001). Resort life cycle theory: Generating processes and estimation. *Annals of Tourism Research*, 28(4), 947–964.
- Weaver, P., McCleary, K., Lepisto, L., & Damonte, L. (1994). The relationship of destination selection attributes to psychological, behavioral and demographic variables. *Journal of Hospitality and Leisure Marketing*, 2, 93–109.
- Woodside, A., & Lyonski, S. (1989). A general model of travel destination choice. *Journal of Travel Research*, 27(4), 8–14.
- Wooldridge, J. F. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
- Yamaguchi, K. (1991). *Event history analysis*. Newbury Park, CA: Sage.
- Yang, J. Y., & Liu, A. (2003). Frequent flyer program: A case study of China airline's marketing initiative—Dynasty flyer program. *Tourism Management*, 24, 587–595.
- Zimmer, Z., Brayley, R., & Searle, M. (1995). Whether to go and where to go: Identification of important influences on seniors' decisions to travel. *Journal of Travel Research*, 33(3), 3–10.

Submitted 16 September 2008. Resubmitted 26 March 2009. Resubmitted 17 August 2009. Final version 10 December 2009. Accepted 28 December 2009. Refereed anonymously. Coordinating Editor: Eugeni Aguilo

Available online at www.sciencedirect.com

