Research Article

Tannaz Alizadeh Ashrafi*, Øystein Myrland

Determinants of trip duration for international tourists in Norway; a parametric survival analysis

https://doi.org/10.1515/ejthr-2017-0008
received 19 April, 2017; accepted 27 August, 2017

Abstract: How long a tourist stays in a host country acts as an indicator of tourism industry's contribution towards the national economy. The purpose of this study is to examine how socio-demographic characteristics of international tourists, their travelling purpose, tourism products and characteristics of the destination influence the length of stay in Norway, by estimating a parametric survival model. Total cost of trip, purpose of travel, type of accommodation and transportation, age of tourist and geographical area are key elements that explain the variation in the length of tourist stay in Norway. The Cox proportional hazard model with time-independent covariates indicates the survival probability of tourists with less budget constraints and younger ages is higher than that of low-spending tourists and elderly travelers. Moreover, tourists with the purpose of friend and family visitation are at lower risk of leaving Norway than are tourists with other purposes. In terms of tourism products, choosing camping sites as the type of accommodation and road transport as the mode of transportation are associated with the highest survival probability. Another key finding is that tourists stay longer in northern Norway than in southern Norway; hence, on average, tourists' overall expenditures are higher in northern Norway.

Keywords: Tourism, Length of stay, Survival analysis, Cox proportional hazard model, Weibull distribution

***Corresponding author: Tannaz Alizadeh Ashrafi,** Muninbakken 21, 9019, Tromsø, Norway, UiT– The Arctic University of Norway, Phone:004777620824, E-mail: tannaz.alizadeh@uit.no
Øystein Myrland, School of Business and Economics, UiT The Arctic University of Norway, Tromsø, Norway

1 Introduction

Tourism is an increasingly important economic activity. It functions as an invisible export to inject money into the economy. The integrative nature of tourism results in an intertwined relationship between the tourism sector and other relevant sectors, such as accommodation, transportation and food industries. Its multidisciplinary nature makes the tourism industry more effective in expanding business and income than any other sector (Candela & Figini, 2012; Hall et al., 2008; Holloway & Taylor, 2006). However, positive effects of tourism on the economy depend on the attractions, characteristics, capacity and potential of the host destination to generate tourist inflow (Candela & Figini, 2012; Holloway & Taylor, 2006; Lew, 1987). In this regard, a wide range of opportunities to enjoy the natural surroundings, such as a rich marine environment, extensive coastline and outstanding fjords together with economic and political stability, high social safety, well-developed infrastructure, highly educated workforce and the use of advanced technologies make Norway a fairly well-known travel destination.

Prior to the 20th century, researchers and experts had not discovered the importance of tourism as an industry (John Towner, 1988). Hence, the tourism industry was viewed neither as a scholarly field nor as a profitable industry (Lickorish & Jenkins, 1997; Sezgin & Yolal, 2012). After World War II, increased political and economic stability in Europe and changes in socioeconomic circumstances (e.g., increases in disposable income) gradually led to mass tourism, which has transformed the tourism industry into a very lucrative activity (Lickorish & Jenkins, 1997; Sezgin & Yolal, 2012). In response, tourism authorities across Europe began collecting tourism statistics to study the tourism market in hopes of establishing better management and planning. To do so, experts and scientists have been using different measurements of tourism inflow such as tourist arrivals, tourism expenditure and tourist length of stay in the host country (Lim, 1997; Song, et al., 2010; Witt & Witt, 1995). Among these, length of stay has received the least attention in the literature (Culiuc, 2014; Lim, 1997; Matias et al., 2009). What makes length

of stay appealing to the economy is its significant contribution to revenue generation and tourism expenditure, job creation, accommodation occupancy rate and retail growth in the destination (Alegre & Pou, 2006; Barros et al., 2010; Kazuzuru, 2014; Matias et al., 2009). Length of stay is more sensitive to real exchange rate movements in the tourist-receiving country, as it undoubtedly affects the choice of destination and duration of vacation in the host country (Barros et al., 2010; Barros et al., 2008; Culiuc, 2014; Davies & Mangan, 1992; Mok & Iverson, 2000; Nogawa, Yamaguchi, & Hagi, 1996).

In studies, which used length of stay as a tourism measurement, a large number of researchers employed traditional regression, considering length of stay as a dependent variable (Alegre & Pou, 2006; Fleischer & Pizam, 2002; Thrane, 2012). However, applying standard techniques such as linear regression to analyse the duration of events causes severe problems such as bias, which can lead to less reliable estimates (Aalen et al., 2008; Cleves et al., 2010). Hence, an alternative method that is capable of overcoming such shortcomings is required.

Survival analysis is a statistical technique that analyses longitudinal data on the occurrence of events (Aalen et al., 2008; Cleves et al., 2010). In tourism survival analysis, length of stay is a random variable with a stochastic behavior referring to the period during which the tourist stays in Norway as a tourist destination. Additionally, leaving Norway (returning home) is considered as the event of interest. The core of the survival analysis is time. Hence, survival analysis is the best suited statistical method for analysing the time-to-event type of data such as trip duration (Aalen et al., 2008; Cleves et al., 2010; Lancaster, 1992).

Given the aforementioned considerations, this study adopts a Cox proportional hazard model to develop a probabilistic model to study the effects of a given set of explanatory variables on variation in length of stay for the international tourist in Norway in 2012. In the selective review of the tourism research literature, which employed survival analysis Machado (2010), Barros et al. (2010), Barros et al. (2008), Gokovali et al. (2007), Hong and Jang (2005), Martínez-Garcia and Raya (2008) and De Menezes et al. (2008) study the relation between the length of stay of tourists in different destinations and the relevant factors such as economic variables (e.g. total cost of the trip, income), demographic characteristics of the tourists (e.g. age, gender, nationality, social class) and destination attributes and facilities (e.g. diversity of tourism product and their quality, availability of recreational activities and climatic features). The researchers found out that the aforementioned factors are the most significant variables in explaining variation in the trip duration.

The explanatory variables in this study are based on a survey conducted by Innovation Norway. Total cost of the visit (Etzel & Woodside, 1982; Machado, 2010; Silberman, 1985; Thrane, 2015), purpose of the trip (Thrane, 2015; Turner & Witt, 2001), preference of accommodation (Barros et al., 2010; Martínez-Garcia & Raya, 2008; Silberman, 1985; Turner & Witt, 2001) and transportation (Thrane, 2015; Turner & Witt, 2001), gender (Machado, 2010; Thrane, 2015) , age (Etzel & Woodside, 1982; Machado, 2010) and geographical area and its attributes (Barros et al., 2010; Machado, 2010; Martínez-Garcia & Raya, 2008; Lew, 1987) (i.e., southern or northern Norway) are chosen as explanatory variables. In order to undertake appropriate management of tourism industry in Norway, the present study separates geographical destination areas to southern and northern parts, as the attractions and characteristics in these areas are different. To the best of our knowledge, no study has attempted to study this factor.

Section 2 presents an overview of the historical development of the Norwegian tourism industry, beginning in the 19th century. Section 3 outlines the survival analysis and introduces our model. Data and variables of the study are presented in section 4. Section 5 includes the empirical results of the study. Section 6 includes implications of the findings.

2 History of the Norwegian tourism industry at a glance

Most historians and authors believe that the elite class of British travellers had a global impact on the emergence of the current form of the tourism industry (Ousby, 1990; John Towner, 1985, 1995; John Towner & Wall, 1991). In fact, wealthy and aristocratic British travellers were among the first to travel to Norway for pleasure and provided a fertile contribution to the Norwegian economy (Farr & Guegan, 2013; Fjågesund & Syme, 2003; Walchester, 2014). A successful exchange of language and traditions, cultural similarities, political alliance, close relation of royal families, great friendship and mutual respect, anglophile among Norwegians and existence of the regular steamships over the North Sea had provided the opportunity for British to enjoy the picturesque natural landscapes in Norway (Fjågesund & Syme, 2003; Ousby, 1990; Walchester, 2014). During the 1840s and 1860s, growing enthusiasm of the

British travellers for participating in outdoor recreational activities during their journeys made the mountainous zone of Norway a popular destination. This phenomenon, as a result, led to commercialization and further intensification of sports tourism in Norway (Holloway & Taylor, 2006; Lovelock, 2007; Swarbrooke & Horner, 2007). British travellers started renting or buying a second home in Norway, and sporting activities such as hunting reindeer, moose, red deer and ptarmigan, as well as catching salmon and trout in Norwegian rivers, became common pursuits among the British during their stay (Fjågesund & Syme, 2003; Lovelock, 2007; Walchester, 2014). This actually resulted in significant British tourist over-crowding and overharvesting of wildlife, so in 1888, the Norwegian government and authorities were forced to impose restrictions on British tourists against hunting and fishing to avoid natural ecosystem destruction (Lovelock, 2007).

Political events of the first half of the 20th century darkened the tourism industry throughout Europe, including Norway, especially due to the disruptive effects of World Wars I and II (Ousby, 1990; Sezgin & Yolal, 2012; Swarbrooke & Horner, 2007). After the ravages of the wars, the economy and politics in Europe started to rebuild, which resulted in the revival of a once-moribund tourism sector. Economic growth and growing disposable income among the middle class, tourism product development, innovation, development of roads in urban and rural areas, advances in means of transportation, elevation of social safety and public health in big cities, promotion in educational level and awareness of people all boosted the tourism industry throughout Europe (Boissevain, 1996; Ousby, 1990; Sezgin & Yolal, 2012). In tandem with the upward changes throughout Europe, an impetus for international travel developed in Norway. In particular, the demolition of the Berlin Wall and the fall of the Soviet regime opened the borders to travellers from Germany and post-communist states to visit Norway. Today, Germany is one of the most important tourism markets for Norway.

However, Norway's tourism boom ended soon after its oil and gas exploration efforts began in the 1970s. A strong oil and gas industry has shifted Norway to be more of an oil-driven economy (Enger et al., 2015; Müller & Grenier, 2011; Svalastog, 1992), leaving the Norwegian tourism sector behind. Oil and gas extraction transformed Norway from a low-cost to a high-cost country. Consequently, the purchasing power of international travellers in Norway, and proportionately the length of travelers stay, have both decreased. This resulted in travellers beginning to redirect their destinations to alternative, less expensive countries (Enger et al., 2015; Svalastog, 1992).

3 Survival analysis

Survival data, also known as 'time-to-event' data, are described by a 'time until failure process' measured in some discrete time units. Survival data have three main features: (1) The dependent variable is waiting time until a change in the current state of the unit takes place. This waiting time is a positive valued random variable, such as time to leave the destination or length of stay in the host country. (2) Units observed at some specific time are at the instantaneous risk of transitioning to a new state at any given point in time. In survival analysis, change or transition from one state to another is known as the 'event'. In a tourism context, the unit is a tourist in the host country, and an event is considered the period of time it takes a tourist to leave the destination country to return home. (3) The final characteristic of survival analysis concerns the effect of different explanatory variables on the time-to-event. Tourism researchers seek to assess the relationship between socio-economic covariates and survival time in the destination (Barros & Machado, 2010; Gokovali et al., 2007; Martínez-Garcia & Raya, 2008).

The core of survival analysis is time; hence, this method is the best suited for modelling the duration of events such as length of stay (Barros et al., 2008; Barros & Machado, 2010). Analysing survival data by applying traditional econometric models can cause severe problems such as bias and inadequacy in outcome information (Aalen et al., 2008; Cleves et al., 2010; David et al., 2008). In tourism studies with length of stay as a tourism measurement and dependent variable, the application of traditional regression such as ordinary least square (OLS) is abundant (Alegre & Pou, 2006; Fleischer & Pizam, 2002; Thrane, 2012). In this section, we briefly discuss the advantages of survival analysis in dealing with longitudinal data over traditional techniques of econometrics.

First, the censored nature of the data in survival models is not handled properly when using a standard ordinary least square (OLS) procedure (Aalen et al., 2008; Cleves et al., 2010; Liu, 2012). Survival analysis is based on following a unit over time until experiencing the event during the observation period. If the subject does not experience the event, this will be considered right-censored data. Left censoring occurs when one does not observe the start of the event (Cleves et al., 2010; Liu, 2012). Left censoring occurs when the observer loses track of some tourists and gets incomplete information. In such cases, the observer knows the arrival date and time of such tourists, but does not know when they have left the country (Lancaster, 1992; Liu, 2012; Van den Berg, 2001).

Second, applying regression models to survival data only gives the mean duration, while one may also be interested in the effects of socio-economic variables on the probability of leaving the destination country (Aalen et al., 2008; Cleves et al., 2010; Liu, 2012). In survival analysis, the survival time or time-to-event is always greater than zero, while a linear regression can predict negative values (Aalen et al., 2008; Cleves et al., 2010; Liu, 2012). Table 1 provides an overview of the survival analysis terminology used in this study.

In survival analysis, our interest is to focus on and estimate the survival *S(t)* and hazard function *λ(t)*. Suppose *T* is the length of a tourist stay, which is a non-negative continuous random variable with a probability density function of *f(t)* and cumulative density function of *F(t)* (Lancaster, 1992; Van den Berg, 2001).

$$F(t) = \Pr(T \leq t) = \int_0^t f(u)\mathrm{d}u \tag{1}$$

In Equation (1), *F(t)* denotes the probability that a tourist exits from the current state (i.e., staying) and enters a new state (i.e., leaving) at time *t*. A simple transformation of the *F(t)* characterizes the survival function as following (Lancaster, 1992; Van den Berg, 2001):

$$S(t) = \Pr(T > t) = 1 - F(t) = \int_t^\infty f(u)\mathrm{d}u \tag{2}$$

In a tourism context, the survival function refers to the probability that a tourist will stay in Norway at least until time *t*. An alternative characterization of the distribution of any arbitrary *T* is given by the hazard function. In tourism studies, the hazard function *λ(t)* refers to the instantaneous probability that a tourist leaves Norway at time *t*, conditional upon the fact that he has been staying in that state until time *t*. Mathematically speaking, the hazard function is given by (Lancaster, 1992; Van den Berg, 2001)

$$\lambda(t) = f(t)/[1 - F(t)] = f(t)/S(t) \tag{3}$$

It can be shown that the hazard rate can be rewritten as

$$S(t) = \exp\left[-\int_0^t \lambda(u)\mathrm{d}u\right] \tag{4}$$

A vector of explanatory variables can affect the behaviour of the random variable *T*. The Cox proportional hazard model is one of the most popular classes of survival models that can be used to account for the effects of the covariates on the survival probability of a tourist (Cox, 1972). In proportional hazard models, the covariates are assumed to be time-independent (Cleves et al., 2010; Cox, 1972). In the presence of time-invariant covariates, hazard function at time *t* is conditional on the explanatory variables, and thus is given by (Ansell & Phillips, 1996; Lancaster, 1992; Van den Berg, 2001)

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t)\exp[\boldsymbol{\beta x}] = \lambda_0(t)\exp\left(\sum_{i=1}^n \beta_i x_i\right) \tag{5}$$

This is defined as the multiplication of a base hazard rate $\lambda_0(t)$ and a term describing the effects of explanatory variables *x*, which is often given using an exponential function, $\exp(\boldsymbol{\beta x})$. Here, $\lambda_0(t)$ has some functional form and

Table 1:Survival analysis terminologies in tourism context.

Terminology in survival analysis	Terminology in tourism demand concept	Description
Event	Event	The event of interest is leaving Norway.
Survival state	Survival state	The state referring to staying in Norway.
Failed state	Left state	The state referring to leaving Norway.
Time-to-event (survival time)	Time-to-leave (staying time)	The time during which a tourist stays in Norway. In this study, survival time is denoted by length of stay (LOS).
Survival probability	Staying probability	The probability that a tourist stays in Norway for a certain time under a given set of explanatory variables.
Failure probability	Leaving probability	The probability that a tourist leaves Norway before a certain time under a given set of explanatory variables.
Hazard rate (instantaneous risk)	Leaving rate (instantaneous risk of leaving)	The probability that a tourist leaves Norway slightly after the time he or she has spent in Norway. In other words, the probability that a tourist leaves Norway before time t_2 on the condition that he or she has stayed in Norway for the time t_1 in such a way that $\lim(t_2 - t_1) \to 0$.

the $\boldsymbol{\beta}$ is a vector expressing the coefficients of explanatory variable matrix x Using Equation (4), survival function for the Cox proportional hazard model can be defined as a function of time and covariates:

$$S(t,\boldsymbol{x}) = \exp\left[-\mathrm{e}^{\sum_{i=1}^{n}\beta i x i}\int_0^t \lambda_0(u)\mathrm{d}u\right] \quad (6)$$

Depending on the estimation technique of the baseline hazard, a proportional hazard model is divided into semi-parametric and parametric categories (Aalen et al., 2008; Ansell & Phillips, 1996; Liu, 2012). In connection with this, Oakes (1977) and Efron (1977) observed that parametric models provide more reliable estimation of the parameters than do semi-parametric models. Moreover, since parametric survival distributions put a particular structure on baseline hazard, parametric models embark richer information (Cleves et al., 2010; Murthy, Xie, & Jiang, 2004). Among the most popular parametric survival distributions include exponential, Weibull and Gompertz. In this study, we use the Weibull distribution as the baseline hazard, chosen on the basis of a minimum value of the Akaike information criterion (AIC) (Aalen et al., 2008; Ansell & Phillips, 1996; Cleves et al., 2010; Murthy et al., 2004). Moreover, the Weibull distribution is convenient because of its flexibility, which stems from shape and scale parameters. The hazard rate of a Weibull distribution and survival function with the shape and scale parameters of p and η is given by (Ansell & Phillips, 1996; Cleves et al., 2010; Liu, 2012; Murthy et al., 2004).

$$\lambda(t,\boldsymbol{x}) = \frac{p}{\eta^p}\, t^{p-1}\mathrm{e}^{\sum_{i=1}^{n}\beta i x i} \quad (7)$$

$$S(t,\boldsymbol{x}) = \mathrm{e}^{-\left(\frac{t}{\eta}\right)^p \mathrm{e}^{\sum_{i=1}^{n}\beta_i x_i}} \quad (8)$$

Testing the proportionality assumption is a main concern when employing a proportional hazard model (Box-Steffensmeier & Jones, 2004). However, most of the previous studies on the application of the duration model in economic-related topics have not verified the proportionality assumption (Burger, Dohnal, Kathrada, & Law, 2001). The term 'proportional' refers to the underlying assumption that the ratio of the hazard rates for any two individuals of the tourist population will remain constant over time (Aalen et al., 2008; Cleves et al., 2010; David et al., 2008). In order to develop an accurate and highly reliable model, proportionality is tested for the all covariates visually (i.e., Kaplan–Meier curves). Later formal tests based on Schoenfeld residuals (Schoenfeld, 1982) and Martingale residuals (Therneau et al., 1990) have been provided. The finding shows that hazard ratios between any two travellers are strictly parallel over time; hence, there is no evidence of the violation of the proportional hazards assumption. Finally, in order to assess the overall fitness of the Weibull distribution, the present study uses Harrell's C concordance statistic (Harrell et al,. 1982). The value of Harrell's C concordance statistic for this study is 0.8, denoting that the model has high predictive power.

4 Data and variables

In this study, a tourist refers to an overnight visitor residing in Norway for at least 24 hours for pleasure, health, business or any other purpose. This study uses survey data to take the heterogeneity of tourists' preferences into account (Martínez-Garcia & Raya, 2008). Further, using individual-level data allows us to perform a sensitivity analysis. The survey data used in this study came from a cross-section tourist survey conducted by Innovation Norway Institute in 2012. Since the questionnaire targeted tourists who completed their vacation throughout the whole period of observation, censored data is not an issue in this study. In total, 2,848 tourists were interviewed. Responses with missing information and incomplete fields were discarded. The final sample included 1,321 respondents. The tourists were asked socio-demographic questions, their accommodation and transportation preferences, and their motivation for choosing Norway as their tourist destination.

The dependent variable of this study is the time spent in Norway before the tourist leaves the destination; i.e., length of stay, which has positive values for each tourist. In this study, we use the number of overnight stays as a proxy for a tourist length of stay.

According to the Innovation Norway Tourism Survey in 2012, the average number of overnights spent is 4 with a minimum of 1, maximum of 55, and standard deviation of 4.12 overnight stays. The estimated model contains both continuous and qualitative covariates. The descriptive statistics of the two continuous variables – age and natural logarithm of travelling cost to Norway – are presented in Table 2. Regarding the cost of the visit, after performing further tests (Martingale residuals and Harrell's C statistics), it is concluded that the natural logarithm of total cost is superior to the level of total cost. With an average of 4 nights spent in Norway, each tourist spends on average 5467 NOK. The average age of the study sampling

is approximately 51 years, with a maximum of 85 and minimum of 12 years old (see Table 2).

The categorical explanatory variables are purpose of the trip, type of accommodation, transportation preference, tourist gender and geographical area (i.e., South or North) (see Table 3). In addition, the interaction between the natural logarithm of total cost and visited area (North/South) is added as another explanatory variable. By including the interaction variable, we want to test the hypothesis that the relationship between the amount of expenditure in Norway and length of stay is different in the northern part than in the southern area.

It is necessary to select the reference (base) category for the categorical variables. In this regard, the reference categories are 'visiting family and friends', 'hotel', 'air transportation', 'male' and 'South'. Table 3 presents the descriptive statistics for the categorical variables used in this study.

The largest proportion of tourists (93.56%) are here for entertainment and pleasure reasons, while travelling with the purpose of visitation and transit are less common, only accounting for 3.71% and 2.73%, respectively. The foremost accommodation type among international tourists is hotel (83.42%), while 12% and 4.6% of the travellers prefer holiday centres and camping sites, respectively. The two most popular transportation means for tourists to come to Norway are air transport (53.3%) and road transport including cars, caravans, coaches, buses and motorcycles (totalling 41.2%). In contrast, sea (4.09%) and rail transportation (1.44%) are the least popular types of transportation for entering the country. In terms of gender, male tourists make up the higher proportion of tourists in Norway (about 54.88%), with an average age of 51.38 years old. The cross-tabulation analysis demonstrates that about 64.35% of the tourists choose southern Norway as a tourist destination, and 33.65% travel to northern Norway. However, the average number of nights that a tourist stays in northern Norway is 5.7, while the average is 3 overnight stays in southern Norway. On average, a tourist spends 7534.72 NOK in northern Norway and 4316 NOK in southern Norway.

Table 2: Characterization of the continuous explanatory variables selected for further analysis

Description	Variable	Min	Max	Mean	Std
Logarithm of total cost	Lntotcost	2.65	10.98	8.21	0.95
Tourist age	Age	12	85	51.38	13.8

Table 3: Characterisation of the categorical explanatory variables selected for further analysis

Description	Variable	Frequency	Percent	Cumulative Frequency
Purpose of travelling	Visit = 1	49	3.71	3.71
	Pleasure = 2	1236	93.56	97.27
	Transit = 3	36	2.73	100
Type of accommodation	Hotel = 1	1102	83.42	83.42
	Holiday centre = 2	61	4.62	88.04
	Camping site = 3	158	11.96	100
Type of transportation	Air = 1	704	53.29	53.29
	Road = 2	544	41.18	94.47
	Rail = 3	19	1.44	95.91
	Sea = 4	54	4.09	100
Tourist gender	Male = 1	725	54.88	54.88
	Female = 2	596	45.12	100
Destination area in Norway	South = 1	850	64.35	64.35
	North = 2	471	35.65	100
Interaction variable between the destination area (i.e., northern or southern Norway) and logarithm of total cost (i.e., Area#Lntotcost)	South#Lntotcost North#Lntotcost	-		

5 Analysis results of proportional-hazards regression

The purpose of Table 4 is to provide a straightforward overview of the empirical estimations of the Cox proportional hazard model.

The first column of Table 4 represents the length of time tourists stay in Norway, sorted by number of nights. The second column lists the number of tourists corresponding to each overnight stay (i.e. the number of tourists at risk of leaving Norway at the beginning of each night). The failure column in Table 4 presents the number of tourists who have left Norway after spending a certain number of nights. To illustrate, at the beginning, there are 1,321 tourists, of which 262 leave Norway after staying one night. The rest stay for another night. In other words, at the beginning of the second day, the number of tourists at risk is 1,059. As time goes on, tourists leave Norway at random times. The last tourist has left Norway after staying 55 nights. The survival probability in the rightmost column of Table 4 gives the probability that a tourist will stay for a certain number of nights. For example, the probability that a tourist stays in Norway for 4 nights is 28.01%. The corresponding Kaplan–Meier curve based on the fraction surviving at each time is shown in Figure 1.

Table 4: Survival function list

Overnight stays	Sample in each night	Failure	Survivor probability
1	1321	262	0.8017
2	1059	345	0.5405
3	714	212	0.38
4	502	132	0.2801
5	370	100	0.2044
6	270	52	0.165
7	218	63	0.1173
8	155	38	0.0886
9	117	16	0.0765
10	101	28	0.0553
11	73	14	0.0447
12	59	17	0.0318
13	42	5	0.028
14	37	15	0.0167
15	22	6	0.0121
18	16	3	0.0098
19	13	1	0.0091
20	12	3	0.0068
25	9	1	0.006
30	8	3	0.0038
35	5	1	0.003
40	4	1	0.0023
44	3	1	0.0015
45	2	1	0.0008
55	1	1	0.0000

When $t = 0$, all the tourists are in a staying state; hence the survival function has a value of 1. As time passes, the number of tourists remaining 'at risk' of leaving Norway decreases. That means, survival function is a non-increasing monotone function of t.

Next, in order to capture the effect of different covariates on survival probability of the tourists in the destination of Norway, we need to fit the data to a Cox proportional hazard model, whose baseline failure rate has a Weibull form. The detailed results of the Weibull model estimation are presented in Table 5. Based on equations (7) and (8), if β_i is positive, an increase in x_i raises the hazard rate and thus reduces the survival probability. Similarly, for a negative β_i, an increase in x_i reduces the hazard rate and thus increases the survival probability.

First of all, the shape parameter of the Weibull distribution, p, is 1.75, indicating that the length of stay increases with increased experience in the event. The negative sign of the log cost variable implies that high-spending tourists tend to take longer travels than tourists with less flexibility in their budget. In our model, the amount of money that a tourist spends while staying in Norway is included in the hazard rate and survival probability through two different variables: the natural logarithm of total cost and the interaction variable of natural logarithm of total cost and the destination area in Norway. To determine to what extent an increase in the total cost can change the hazard rate, one can write:

$$\frac{\lambda(t,x)|_{Ltotcost_2}}{\lambda(t,x)|_{Ltotcost_1}} = \frac{e^{\beta_{Lntotcost}\times Ltotcost_2+\beta_{North\#Ltotcost}\times(Ltotcost_2\times 1)}}{e^{\beta_{Lntotcost}\times Ltotcost_1+\beta_{North\#Ltotcost}\times(Ltotcost_1\times 1)}} = e^{\beta_{Lntotcost}\times(Ltotcost_2-Ltotcost_1)+\beta_{North\#Ltotcost}\times(Ltotcost_2-Ltotcost_1)}$$

If $Ltotcost_2 - Ltotcost_1 = 1$,

$$\frac{\lambda(t,x)|_{Ltotcost_2}}{\lambda(t,x)|_{Ltotcost_1}} = e^{(\beta_{Lntotcost}+\beta_{North\#Ltotcost})\times 1} = \exp(-0.5567) = 0.5731$$

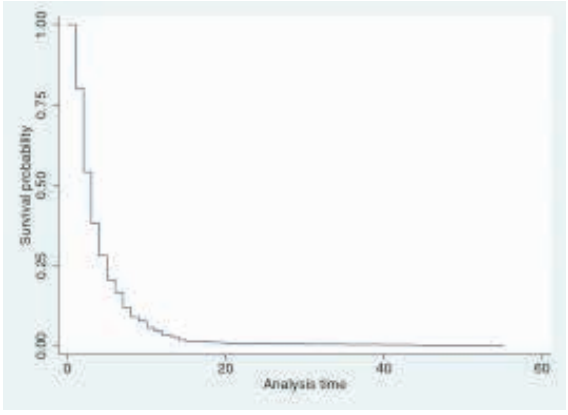

Figure 1: Kaplan-Meier survival probability function

This means that if the natural logarithm of total cost, $Ltotcost$, is increased by 1, the hazard rate reduces by a factor of 0.5731. In terms of changes in the total cost, an increase in $Ltotcost$ by unity is equivalent to an increase in *Cost* by a factor of exp1 = 2.718, as given by

$$\frac{Cost_2}{Cost_1} = \frac{\ln Cost_2}{\ln Cost_1} = \frac{1+\ln Cost_1}{\ln Cost_1} = \frac{\ln[Cost_1 \times \exp(1)]}{\ln Cost_1} = \exp(1) = 2.718$$

In other words, if the total cost, or tourist expenditure, is increased by a factor of 2.718, the hazard rate reduces by a factor of 0.5731. Moreover, the significant coefficient of the interaction variable demonstrates that the effect of cost on length of stay is different for different geographical areas (i.e., North or South).

The purpose of the trip indeed has a significant effect on the trip duration. Different purposes have different impacts on the duration of the trip. The coefficient of the transit and pleasure purposes is positive. Thus, a tourist with a purpose of transit or pleasure has a higher hazard rate compared to the one whose purpose of travelling is visiting friends and family. Specifically, tourists who travel to Norway for the purpose of visiting friends and relatives tend to stay longer. For instance, the hazard rate of a tourist with a purpose of pleasure is higher that the hazard rate of a tourist with a purpose of visiting family and friends by a factor of exp(0.5536) = 1.74. Similarly, the probability that tourists will stay in Norway if they come for pleasure is higher than that of the tourists whose purpose is transit.

The type of accommodation is another categorical explanatory variable that can affect the duration of stay. According to estimation results, choosing camping sites as an accommodation establishment represents the highest probability of staying in Norway for a greater length of time, while choosing hotel accommodations equates to the shortest length of stay. This conclusion is also justified

Table 5: Estimation of the Weibull parameters

Variable, x_i	**Coefficient, β_i**	***p-value***	**exp(β_i)**
Constant	2.82	0.000	16.77
Lntotcost	−0.6776	0.000	50.78
Purpose of traveling:			
Visit = 1	–	–	–
Pleasure = 2	0.5536	0.000	1.7395
Transit = 3	0.5284	0.019	1.6962
Type of accommodation:			
Hotel = 1	–	–	–
Holiday center = 2	−1.2773	0.000	0.2787
Camping site = 3	−2.013	0.000	0.1335
Type of transportation:			
Air = 1	–	–	–
Road = 2	0.2234	0.001	1.2503
Rail = 3	−0.1485	0.525	0.862
Sea = 4	−0.005	0.973	0.995
Age	0.006	0.004	1.006
Gender group:			
Male = 1	-	-	-
Female = 2	-0.025	0.646	0.9753
Destination area:			
South = 1	–	–	–
North = 2	−1.6737	0.000	0.000
Area#Lntotcost interaction:			
South#Lntotcost			
North#Lntotcost	–	–	–
	0.1209	0.010	0.010
p (ancillary parameter)	1.7502		

based on the cheaper price of camping sites and holiday centers in comparison to hotels.

With regard to the type of transportation chosen, road transportation is a relevant parameter that positively affects the hazard rate. Considering the positive coefficient of road transportation, one can conclude that road tourists tend to stay in Norway for a longer time compared to those who take a flight (i.e. air transportation category). Rail and sea transports do not have explanatory power to illustrate a notable variation in length of stay.

The high *p*-value of the gender variable indicates that gender does not statistically affect the duration of stay in Norway.

With regard to age, the positive coefficient indicates that with an increase in tourist age, the hazard rate increases, and thus the probability of staying in Norway decreases. In other words, a positive coefficient indicates a certain trend towards a decreased probability of staying in Norway among older tourists. The corresponding coefficient refers to the increase in the logarithm of hazard for each one-year increase in age. As a result, the risk of leaving Norway increases by a factor of exp(0.006) = 1.006 for each year the tourist ages.

The negative sign of northern Norway as a tourist destination indicates that the hazard rate in the North is lower than in the South. We can also illustrate this fact by comparing survival experiences of different tourist groups in the northern and southern regions upon the whole curve and not upon specific points.

Figure 2 shows that the survival probability of the tourist population visiting northern area is always higher than that of the southern part. A log-rank test, with a p-value of 0.001, is used to verify that the survival times for the two regions are significantly different from one another.

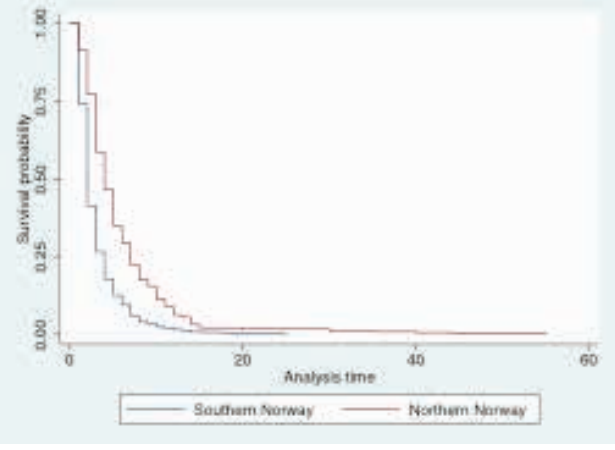

Figure 2: Kaplan–Meier survival estimate for tourists staying in northern and southern Norway

6 Discussion and conclusion

The length of a tourist's stay in a host country has managerial implications, as the time spent in the destination is in close relation with money being generated, jobs created, occupancy rates in tourist accommodation establishments and retail growth. Hence identifying the determinants of trip duration is important for governments, stakeholders, managers, executives and tour operators for planning, evaluative and promotional purposes. Based on this study, it is not surprising that total expenditure is negatively associated with trip duration. Generally speaking, high-spending tourists with easy affordability tend to stay longer in the destination than do tourists with higher budgetary restriction. Based on economic theory, an increase in disposable income leads to an increase in consumption, provided an elastic income elasticity. In tourism context, this means that the income elasticity of travel demand is elastic (Fouquet, 2012; Gallet & Doucouliagos, 2014). A study by Dadgostar and Isotalo (1992) shows that the easy affordability is not necessarily associated with longer vacations. He found out that residents of small towns take shorter vacations in near-home destinations a when their income level increases.

The purpose of the trip is another influencing factor on trip duration. Tourists with the purpose of visiting family and friends have the highest survival probability among tourists. However, not surprisingly, tourists with transit purposes tend to opt for shorter stays than those travelling with entertainment and visitation purposes.

Moreover, according to our results, focusing on the elements of tourism products such as accommodation and transportation is relevant to determining length of stay. In case of accommodation, the negative coefficient suggests that those who stay in camping sites have a higher survival probability than tourists staying at holiday centers and hotels. A study by Thrane (2012) has lent support to this finding. Due to the lower price of camping sites in comparison to holiday centres and hotels, it is expected that providing high-quality camping sites with comfortable facilities can prolong the length of stay and promote the contributions of the tourism industry to the benefit of the economy.

Road transportation is another component of the tourism product that influences the duration of stay in Norway. A positive sign of this explanatory variable shows that tourists who prefer road transportation to air transportation are at lower risk of leaving Norway. This trend can be justified, to some extent, as road transportation

offers a higher degree of accessibility to a larger geographical territory and remote tourist sites. In terms of tourist attractions, Norway offers unspoiled natural areas and rich wildlife, which provide adventurous experiences for visitors. Hence, road transportation may prolong travel and duration of stay in the destination for nature-based tourist pursuits. However, the study by Thrane (2015) shows that trip duration for those travelling by airplane is longer than for those who prefer road transportation.

Furthermore, younger people tend to have a longer stay in Norway, which is contrary to common predictions. As we might expect, younger travellers are "experience seekers" and presumably attracted by Norway's offerings in adventurous outdoor activities, such as hiking, cycling, climbing, water sports and winter skiing, many of which suit younger tourists better. This particular result alludes to an imperative to target young tourists' preferences for tourism products and major activities during their stay in Norway. Due to the integrative nature of the tourism industry, detailed knowledge about young tourism behavior can help bring about added value in relevant industries such as gastronomy, transportation and lodging sectors. Hence, the youth tourism is a potentially vital resource for money injection and career development opportunities in the host country. Additionally, the development of youth tourism has another unique benefit for the destination. Today's young travellers tend to be respectful towards distinct cultures, well informed, educated and responsible for environmental protection. These characteristics speak to the potential for a promising sustainable tourism industry. However, most published studies have found out that the time spent in the destination is a positive function of age and, in general, older tourists are more likely to stay longer in the specific destinations (Dadgostar & Isotalo, 1992; Goodall & Ashworth, 1988; Machado, 2010; Thrane, 2015; Weaver et al., 1994). This conclusion may relate to higher purchasing power of the elderly tourist population than young tourists.

According to the empirical results, gender was not found to be significant in our study. However, in studies by Goodall and Ashworth (1988), Machado (2010), Thrane (2015) and Weaver et al. (1994), the general conclusion is that male tourists are more likely to have longer vacations than female travellers.

The choice of destination in Norway also plays a key role in the number of nights tourists spend in Norway. That is, the tourists who travel to the northern part of Norway tend to stay for a longer period compared to the ones staying in the southern part of Norway. Similarly, based on descriptive statistics, tourists visiting northern Norway spend more money than do travelers visiting the southern part. As would be expected, a longer stay in the destination is associated with higher tourist consumption. This particular outcome can be considered valuable information for authorities and officials to prioritize tourism development in northern Norway over the southern region. In a high-cost country like Norway, this is particularly valuable information to avoid investment decisions in a situation of trial and error. Additionally, identifying and ranking major tourist regions in northern Norway and providing detailed data for the tourist consumption component in this region can improve the nation's ability to maximize the tourism contribution to the Norwegian economy.

References

[1] Aalen, O., Borgan, O., & Gjessing, H. (2008). Survival and event history analysis: a process point of view. New Yrok: Springer Science & Business Media.
[2] Alegre, J., & Pou, L. (2006). The length of stay in the demand for tourism. Tourism management, 27(6), 1343-1355.
[3] Ansell, J., & Phillips, M. (1996). Practical Methods for Reliability Data Analysis (Vol. 159). Oxford Oxford Science.
[4] Barros, C. P., Butler, R., & Correia, A. (2010). The length of stay of golf tourism: A survival analysis. Tourism management, 31(1), 13-21.
[5] 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.
[6] Barros, C. P., & Machado, L. P. (2010). The length of stay in tourism. Annals of Tourism Research, 37(3), 692-706.
[7] Boissevain, J. (1996). Coping with tourists: European reactions to mass tourism (Vol. 1). Oxford: Berghahn Books.
[8] Box-Steffensmeier, J. M., & Jones, B. S. (2004). Event history modeling: A guide for social scientists. New York: Cambridge University Press.
[9] Burger, C. J. S. C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting — a case study of Durban, South Africa. Tourism management, 22(4), 403-409.
[10] Candela, G., & Figini, P. (2012). The economics of tourism destinations. Heidelberg: Springer.
[11] Cleves, M., Gould, W., Gutierrez, R. G., & Marchenko, Y. V. (2010). An Introduction to Survival Analysis Using Stata. Texas: Stata Press.
[12] Culiuc, A. (2014). Determinants of International Tourism - IMF Working Paper No. 14/82. Retrieved from Washington:
[13] Dadgostar, B., & Isotalo, R. M. (1992). Factors affecting time spent by near-home tourists in city destinations. Journal of Travel Research, 31(2), 34-39.
[14] David, W., Hosmer, S., Stanley, L., & Susanne, M. (2008). Applied survival analysis: Regression Modeling of Time to Event Data (Second ed.). Hoboken: Wiley Interscience.
[15] Davies, B., & Mangan, J. (1992). Family expenditure on hotels and holidays. Annals of Tourism Research, 19(4), 691-699.

[16] De Menezes, A. G., Moniz, A., & Vieira, J. C. (2008). The determinants of length of stay of tourists in the Azores. Tourism Economics, 14(1), 205-222.
[17] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association, 72(359), 557-565.
[18] Enger, A., Sandvik, K., & Kildal Iversen, E. (2015). Developing scenarios for the Norwegian travel industry 2025. Journal of Tourism Futures, 1(1), 6-18.
[19] Etzel, M. J., & Woodside, A. G. (1982). Segmenting vacation markets: The case of the distant and near-home travelers. Journal of Travel Research, 20(4), 10-14.
[20] Farr, M., & Guegan, X. (2013). The British Abroad Since the Eighteenth Century : travellers and tourists (Vol. 1& 2). Newcastle: Palgrave Macmillan.
[21] Fjågesund, P., & Syme, R. A. (2003). The Northern Utopia: British Perceptions of Norway in the Nineteenth Century (Vol. 10). Amesterdam: Rodopi.
[22] Fleischer, A., & Pizam, A. (2002). Tourism constraints among Israeli seniors. Annals of Tourism Research, 29(1), 106-123.
[23] Fouquet, R. (2012). Trends in income and price elasticities of transport demand (1850–2010). Energy Policy, 50, 62-71. doi:https://doi.org/10.1016/j.enpol.2012.03.001
[24] Gallet, C. A., & Doucouliagos, H. (2014). The income elasticity of air travel: A meta-analysis. Annals of Tourism Research, 49, 141-155.
[25] Gokovali, U., Bahar, O., & Kozak, M. (2007). Determinants of length of stay: A practical use of survival analysis. Tourism management, 28(3), 736-746.
[26] Goodall, B., & Ashworth, G. J. (1988). Marketing in the Tourism Industry: The promotion of destination regions. London: International Thomson Publishing Services.
[27] Hall, C. M., Muller, D. K., & Saarinen, J. (2008). Nordic tourism: Issues and cases. Bristol: Channel View Publications.
[28] Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. Journal of the American Medical Association, 247(18), 2543-2546.
[29] Holloway, J. C., & Taylor, N. (2006). The business of tourism. Harlow: Prentice Hall.
[30] 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(4), 395-403.
[31] Kazuzuru, B. (2014). Determinants of Tourist Length of Stay in Tanzania. International Journal of Business and Social Science, 5(9(1)), 204-214.
[32] Lancaster, T. (1992). The econometric analysis of transition data. Cambridge: Cambridge university press.
[33] Lew, A. A. (1987). A framework of tourist attraction research. Annals of Tourism Research, 14(4), 553-575.
[34] Lickorish, L. J., & Jenkins, C. L. (1997). An Introduction to Tourism. Oxford: Heinemann
[35] Lim, C. (1997). Review of international tourism demand models. Annals of Tourism Research, 24(4), 835-849.
[36] Liu, X. (2012). Survival Analysis: Models and Applications. West Sussex: Higher Education Press.
[37] Lovelock, B. (2007). Tourism and the consumption of wildlife: Hunting, shooting and sport fishing. New York: Routledge.
[38] Machado, L. P. (2010). Does destination image influence the length of stay in a tourism destination? Tourism Economics, 16(2), 443-456.
[39] Martínez-Garcia, E., & Raya, J. M. (2008). Length of stay for low-cost tourism. Tourism management, 29(6), 1064-1075.
[40] Matias Alvaro, Nijkamp Peter, & sarmento Manuela. (2009). Advances in Tourism Economics-New Developments. London: A Springer Company.
[41] Mok, C., & Iverson, T. J. (2000). Expenditure-based segmentation: Taiwanese tourists to Guam. Tourism management, 21(3), 299-305.
[42] Murthy, D. N. P., Xie, M., & Jiang, R. (2004). Weibull Models. New Jersey: John Wiley & Sons.
[43] Müller, D. K., & Grenier, A. A. (2011). Polar Tourism: A Tool for Regional development Quebec Presses de l'Université du Québec.
[44] Nogawa, H., Yamaguchi, Y., & Hagi, Y. (1996). An empirical research study on Japanese sport tourism in sport-for-all events: Case studies of a single-night event and a multiple-night event. Journal of Travel Research, 35(2), 46-54.
[45] Oakes, D. (1977). The asymptotic information in censored survival data. Biometrika, 64(3), 441-448.
[46] Ousby, I. (1990). The Englishman's England: taste, travel and the rise of tourism. Cambridge: Cambridge University Press.
[47] Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. Biometrika, 69(1), 239-241.
[48] Sezgin, E., & Yolal, M. (2012). Golden Age of Mass Tourism: Its History and Development. Ankara: INTECH
[49] Silberman, J. (1985). A demand function for length of stay: the evidence from Virginia Beach. Journal of Travel Research, 23(4), 16-23.
[50] Song, H., Li, g., Witt, S. F., & Fei, B. (2010). Tourism demand modelling and forecasting: how should demand be measured. Tourism Economics, 16(1), 63-81.
[51] Svalastog, S. (1992). Tourism in a changing Norway. Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 46(2), 109-116.
[52] Swarbrooke, J., & Horner, S. (2007). Consumer behaviour in tourism. New York: Routledge.
[53] Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale-based residuals for survival models. Biometrika, 77(1), 147-160.
[54] 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.
[55] Thrane, C. (2015). Research note: The determinants of tourists' length of stay: some further modelling issues. Tourism Economics, 21(5), 1087-1093.
[56] Towner, J. (1985). The Grand Tour, A Key Phase in the History of Tourism. Annals of Tourism Research. , 12(3), 297-333.
[57] Towner, J. (1988). Approaches to tourism history. Annals of Tourism Research, 15(1), 47-62.
[58] Towner, J. (1995). What is tourism' s history? Tourism Managemen, 16(5), 339-343.
[59] Towner, J., & Wall, G. (1991). HISTORY AND TOURISM. Annals of Tourism Research, 18(1), 71-84.
[60] Turner, L. W., & Witt, S. F. (2001). Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited. Tourism Economics, 7(1), 21-38.

[61] Van den Berg, G. J. (2001). Chapter 55 - Duration models: specification, indetification and multiple durations. In J. J. Heckman & E. Leamer (Eds.), Handbook of econometrics (Vol. 5): Elsevier.
[62] Walchester, K. (2014). Gamle Norge and Nineteenth-century British Women Travellers in Norway. London: Anthem Press.
[63] Weaver, P. A., McCleary, K. W., Lapisto, L., & Damonte, L. T. (1994). The relationship of destination selection attributes to psychological, behavioral and demographic variables. Journal of Hospitality & Leisure Marketing, 2(2), 93-109.
[64] Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. Interational journal of forecasting, 11(3), 447-475.
[65] Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society-Series B (methodological), 34(2), 187-220.