Determinants of Tourist Length of Stay in Tanzania

Benedicto Kazuzuru
Lecturer
Department of Biometry and Mathematics
Sokoine University of Agriculture
P.O.Box, 3038
Morogoro

Abstract

This study examines factors influencing tourist length of stay in Tanzania as one of the key elements in tourism revenue generation. Globally such studies exist, but those which have applied survival analysis are very few especially in Tanzania. This study, using survival analysis on a cross-section of tourists surveyed in the years 2001, 2007 and 2008 by the Tanzania Tourism Sector Survey (TTSS) has established that trip-related characteristics of the tourists are more influential in determining tourist length of stay than their demographic characteristics and their destination attributes. More particularly tourist longer stay is associated with tourists who visit friends and relatives, those who are familiar with the destination and those familiar with English language. On the other hand tourist short stay is associated mostly with frequent visitors, visitors on business mission and visitors on leisure and recreation and those whose source of travel information is from the word of mouth. It is therefore advised that stakeholders such as hotel owners and tour operators should consider the mentioned factors in planning their activities. Further to this, efforts need to be done by the Tanzanian government and other stakeholders to diversify the country' tourism activities to encourage longer stay by some segments of tourists such as business visitors and visitors on leisure and recreation.

Keywords: trip-related characteristics, destination attributes, demographic attributes

1. Introduction

Among the key determinants of a country total revenue is the length of tourists' stay in the country (WTO, 2005). This comes from the fact that the longer the stay by tourists the higher the aggregate expenditure. Unfortunately to the best of the author's knowledge no study on determinants of tourist length of stay based in Tanzania has ever been done. knowing the determinants of tourist length of stay would be important to various stake holders such as hotel owners, tour operators and the Ministry of Natural Resources and Tourism .Therefore this study examines the determinants of tourist length of stay in Tanzania. The study is interested in testing the hypothesis that destination attributes as measured by season variation are more influential than demographic and the trip-related characteristics in determining tourist length of stay in Tanzania. There two main seasons of tourists' coming to Tanzania. The peak season (July-September) versus other seasons which could be regarded as low seasons. The different seasons would proxy for different destination characteristics much as in each season there are different tourist activities and different weather conditions. This is especially so for Tanzania, which has long tourist seasons and many activities (e.g. game viewing, bird watching) which offer different sightings with the changing seasons (MNRT, 2002).

The study has employed survival analysis to achieve its objectives. The use of survival analysis in tourism studies for the Tanzanian setting has not been attempted before. Globally, the first study to use survival analysis in studying tourist length of stay was by Govakali et al (2007) in Turkey. Other pioneers in the field have been Barros and Correia (2007), Garcia and Raya (2008) and Menezes et al. (2008).

2. Theoretical Literature Review

Length of stay is a commodity, of which its demand depends among other things on a tourist's income and how much the length of stay costs. The consumer theory puts great emphasis on the price determinants of a commodity. The non-price determinants (other things, being equal) can best be explained from the sociological point of view, using the theory of reasoned action (TRA) as given by Fishebein and Ajzen (1975).

TRA attempts to explain why a person pursues a certain action. TRA suggests that a person's behavioural intention depends on the person's attitude to the behavior and the subjective norms. Attitudes consist of beliefs about the consequences of performing the behaviour multiplied by his or her evaluation of these consequences. Subjective norms are seen as a combination of perceived expectations of relevant individuals or groups and intentions to comply with these expectations.

Barros and Correia (2007), in studying the determinants of tourist vacation length of stay, argue that the conceptual problem consists of understanding the relationship between vacation length of stay and expectations and attitudes in the behavioral intentions, as well as the relationship between intentions and subsequent behaviour. In order for a tourist to stay longer, he must have perceived the destination to be an attractive place.

These attitudes and beliefs concerning an action are not tangible, but are reflected in the characteristics of the individual and the characteristics of the action. In terms of consumer behaviour these attributes and subjective norms would be reflected in consumer characteristics, product characteristics and the environment where the purchase was made (Nzuki, 2006).

The application of TRA in tourism studies has led to the theory of the tourist decision process, which Barros and Correia (2007) describe as being influenced by four factors: tourist profiles, trip-related characteristics, trip awareness, and destination characteristics. Tourist profile is a reflection of a consumer's characteristics, which essentially refer to social-economic characteristics. Trip-related characteristics are comparable to product characteristics, while destination characteristics are comparable to environment of purchase.

Alegre and Pou (2006) as well as Garcia and Raya (2008) assert that the theoretical determinants of a tourist's length of stay can be looked at two perspectives. One is the determinants of consumer preferences, such as demographics, and the second is the price determinants, covering a consumer's income and the cost of travel time and the holiday time.

3. Empirical Literature Review

TTSS (2001), used cross-tabulations for establishing the variations of tourists' length of stay in Tanzania. The study found that tourist length of stay varied with nationality and with travel arrangement with tourist on non package tour staying longer (11 days) than those on package tour (8 days). Similarly, Alegre and Pou (2007) using logistic regression also found nationality to be one among the determinants of tourist length of stay in the Balearic Islands. The study further found that age, type of jobs, type of accommodation, number of trips, visiting rate, size of party, daily cost of holiday and total party expenditure as among the key determinants of tourist length of stay. Most of these variables were found to be significant with type of job, nationality, single trip and total holiday expenditure having a positive influence on a tourist's length of stay.Gokovali et al. (2007), Barros and Correia (2007), Menezes, et al. (2007), Menezes et al. (2008) and Barros et al. (2010) analyzed the determinants of tourist length of stay and established factors similar to those established by Alegre and Pou (2007) .Other scholars such as Yang et al. (2011) in China have found travelling distance and accommodation type to be influential in length of stay while Chaiboonsri and Chaitip (2012) establishes that being less educated and having less income affects positively tourist length of stay. The finding on the negative impact of income on the length of stay is surprising and contrary to the finding by previous studies such as by Gokovali et al. (2007) as well as by Barros and Correia (2007).

From the reviewed literature, it is evident that among these studies none has been done from Tanzania. This study attempts to widen the literature on the determinants of tourist length of stay by examining the case in Tanzania. The remaining sections in the paper are organized as follows: Section 4 describes the methodology and data sources and section 5 outlines the findings and discussion. Section 6 concludes the study and section 7 provides the acknowledgements.

4 Methodology

4.1 Model Specification

The study adopts a discrete choice/continuous model by Durbin and Mcfadden 1984 as well as by Hanemann 1984, as described by Alegre and Pou (2006). This model takes account of the length of stay in a tourist's utility function.

According to the account given by Alegre and Pou (2006), the model assumes that a tourist utility function comprises three goods: *q*, a vector of consumer goods excluding tourism services, *z*, the vector of characteristics that define the holiday (the destination, type of accommodation, category of accommodation), *t* the length of holiday. A consumer chooses the values of *q* , *z* and *t* which maximize his utility, subject to income constraint *Y* and time constraint *T*. The time constraint *T* consists of time taken to travel to the destination and the time of staying at the destination. Accordingly budget *Y* constrains the expenditure on travel, expenditure at the destination and the expenditure on non-tourism goods *q*.

If one assumes weak separability in the utility function[1], the utility function can be maximized separately in the absence of non-tourism goods *q*. Under this weak separability assumption, the demand for length of stay can therefore be viewed as a function of the holiday characteristics, the price of traveling to a destination, the daily price of the holiday, the total expenditure available for the holiday, the maximum time available for the holiday, the characteristics of the consumer and the unobservable random effects. Following this discussion, Alegre and Pou (2006) specify the following demand function for length of stay:

$$D = f(p_{tour}, z, Y - pq - p_{travel}, T - t_{travel}, \tau, \varepsilon) \quad (1)$$

where

P_{tour} = price of the holiday, *z* are the trip characteristics, *p* is the price of the non-tourism goods, $p_{travel,}$ is the price of travelling to a destination, *T* is the total holiday time, t_{travel} is the travel time to the destination, τ are the consumer characteristics and ε is the random error term.

Garcia and Raya (2008) also give an account of this model. According to them, determinants of length of stay emanate from the tourist's preferences as given by his utility function as well as the utility constraint. The former encompasses the determinants of a tourist's preferences, such as demographics (age, education and gender), while the latter encompass the arguments involved in the constraint, which involve price of travel and price of the holiday time, total budget for the holiday and the total holiday time. A tourist therefore chooses time *t* which maximizes his utility given these constraints. As before, the maximization of the utility function given the length of stay must assume weak separability between the non-tourism good *q* and length of stay *t*, in the entire utility function.

In this study the same model was adopted. However, the nature of the data collected contained no information on the total holiday time of a tourist, the budget allocated for the holiday, and the prices of the travel and of the holiday. Nevertheless for total allocated budget, per capita GDP was used as a proxy whereas travel distance was used as a proxy for cost of travel. The missing variables (maximum holiday time and the daily costs of the holiday) were not the key ones in the study as the hypothesis was based on the consumer's demographic characteristics and trip-related characteristics and the destination attributes, for which data were available. Equally well the used proxies could not affect the test of the hypothesis.

Following this discussion, equation 1 can be rewritten in a more detailed manner while considering the variables discussed in section 1:

$$\begin{aligned} L_{ict} &= \beta_o + \beta_1 Age_i + \beta_2 Females_i + \beta_3 GDP_i + \beta_4 Er + \beta_5 El + \beta_6 Childpresence \\ &+ \beta_7 childno + \beta_8 Tarra + \beta_9 VistFRD + \beta_{10} VistLSR + \beta_{11} VistBSN + \beta_{12} Adultno + \beta_{13} Tpartyno \\ &+ \beta_{14} Frvist + \beta_{15} Fadest + \beta_{16} Dist + \beta_{17} Nosites + \beta_{18} \mathrm{Pr}ice + \beta_{19} Peak + \beta_{20} Isource + Africa \\ &+ Asia + Europe + MEast + NAmerica + SAmerica + year1 + year2 + year3 + \varepsilon \end{aligned} \quad \cdots(2)$$

Where

L_{ict} =the length of stay by a tourist belonging to country c and observed in year t.

Age= Age of a tourist measured in years. Its coefficients β_1 is indeterminate.

[1] A utility function is weakly separable if the marginal rate of substitution between any two goods belonging to a group of goods, say tourism goods, is independent of any quantity of goods outside this group. This assumption is important for solving a tourist's maximization problem. When the vice-versa is true then one has a strong separable utility function.

Females=number of females in the travel party. Its coefficient β_1 is indeterminate.

Income =tourist's income level. In this study per capita GDP of a tourist's country of origin was used as a proxy for income, because tourists were not asked about their incomes during the surveys. Its coefficient β_3 is expected to be positive.

Ear=Exchange rate between Tanzania's currency and the tourist's country of origin. Currencies are expressed in terms of units of Tanzanian shillings per unit of a foreign currency.. Its coefficients β_4 is expected to be positive

EL=1 if a tourist is from an English-speaking country and 0 otherwise. Its coefficient β_5 is expected to be positive .An ability to communicate fluently may lead to the tourist becoming more familiar with the destination leading to longer stay other things being equal.

Child presence=1 if there is at least one child in the travel party and 0 otherwise. Its coefficient β_6 is expected to be negative.

Childno= Number of children in the travel party. This variable should follow the same sign as the former one in the sense that more children could imply more family commitments back home, hence leading to shorter length of stay. The only difference from the former one is that, it captures not only the direction but also the intensity of the influence of children on a tourist's length of stay. Therefore its coefficient β_7 is expected to be negative.

Tarra= 1 if a tourist is on package tour and 0 if a tourist is on a non- package tour. Its coefficient β_8 is expected to be negative. A tourist on package tour has limited time and would rarely extend his length of stay as everything is scheduled unlike the tourist on non-package tour. Govakali et al.(2007), Garcia and Raya(2008) find this coefficient to be negative.

VistFRD = 1 if a tourist is visiting friends and relatives, 0 otherwise. Its coefficient β_9 is indeterminate.

VistLSR= 1 if a tourist has come for leisure and recreation, 0 otherwise. Its coefficient β_{10} is indeterminate.

VistBSN=1 if a tourist is on business purpose. Its coefficient β_{11} is expected to be positive.

Adultno = Number of adults in the travel party. This variable has not been common in the literature examining tourist length of stay. However, most travelers would be adults rather than children leading to the possibility of it behaving more or less the same as the total number of visitors in the party. Therefore its coefficient β_{12} is expected to be negative.

Tpartyno=1 if there are at least 2 members in the travel party, 0 otherwise. Its coefficient β_{13} is expected to be negative. This is because most travellers in alarge party are on organized tours, which tend to be shorter than those of single travellers or smaller parties.

Frvist = 1 if a tourist has visited at least one African country before Tanzania, 0 otherwise. Its coefficient β_{14} is expected to be negative. A frequent visitor in this context must be either an explorer or a businessman, for these are the kind of people who travel frequently. These people in much the same way as frequent visitors, would rarely stay for a long time at a particular destination.

Fadests = 1 if a tourist has visited Tanzania at least once before the current visit and 0 otherwise. Its coefficient β_{15} is expected to be positive. The inference is that repeat visitor is very much attracted to the destination for one reason or another and therefore is likely to stay longer . Barros and Correia (2007) as well as Menes et., al(2008) proved that repeat visitors stay longer.

Dist = the shortest distance by air between Tanzania and the tourist country of origin measured in miles. Based on the model formulation distance would be a proxy for travel cost to the destination and based on the model formulation, distance should reduce the total time to be spent at a destination. Therefore its coefficient β_{16} is expected to be negative.

Nosites=1 if a tourist had visited more than one site and 0 otherwise. Its coefficients β_{17} is expected to be positive. The logic here is much clearer than anywhere else. For a tourist to visit many sites he needs more time.

Price= the relative cot of living between Tanzania and the tourist's country of origin measured as price= $CPI_{TZ}^{t} / CPI_{O}^{t} * ER_{jt}$ where CPI_{TZ}^{t}, is the consumer price index in Tanzania and CPI_{O}^{t} is the consumer price index in a tourist's country of origin. Its coefficient β_{18} is expected to be negative. This variable can also be a proxy for the cost of a tourist's holiday time.

Peak=1 if a tourist travelled during the peak season (July -September) and 0 otherwise. Its coefficient is indeterminate.

Isource= 1 is source of information is from the word-of-mouth, 0 otherwise. Its coefficient β_{19} is indeterminate, depending on a number of factors. First, if the returning tourists depict a negative picture of the destination back home, a coming tourist would certainly plan to stay for a short while. But that will also depend on what he finds after reaching the destination, for he/she can always extend his/her length of stay. Therefore its coefficient β_{20} is indeterminate.

Africa, Asia, Europe, MEast, NAmerica, SAmerica, represents the dummies for Africa, Asia, Europe, the Middle East, North America and South America. year1, year 2 and year3 are dummies for the year 2001, year 2007 and year 2008.

Equation 2 could be analyzed by OLS, but there a number of defects in using OLS to analyze a time variable. Greene (2003) as well Cameron and Trivedi (2005) addresses the following problems of using OLS in analyzing a time variable. First is the lack of normality as most of the time observation is positively skewed. Second is the fact that in most surveys involving the time to an event, observations are censored. In other words, the individuals are observed before the study was completed or the study comes to an end before the event has occurred. The former is a case of left censoring while the latter is a case of right censoring. However, in this study, as it will be observed, tourists are interviewed during their departure and thus censoring is not there. Third there may be the issue that a covariate like age may change during the duration, and the assumption of .may be violated, resulting in inconsistent coefficients. If the duration is short change in ages may not be substantial. Fourth is the fact that there is no guarantee that OLS will predict positive values of time. This limitation could be serious in prediction.

Because of the above problems the study opted to use survival analysis instead of OLS.

4.2 An Overview of Survival Analysis

The review is based on the work by Cameron and Trivedi (2005). One may begin by considering the cumulative distribution of the variable time given as $F(t)$ and its density function given by $f(t)$. The relationship between the two is such that

$$f(t) = dF(t)/dt \qquad (3a)$$

or

$$F(t) = P(T \le t) = \int_0^t f(s)ds \qquad (3b)$$

An equally important concept in duration analysis is the *survival* function which is in fact the greater than or equal cumulative function, defined as

$$S(t) = P(T \ge t) = 1 - F(t) \qquad . \qquad (3c)$$

This is the probability that a particular duration equals or exceeds time t. Another key concept is the hazard function. This is an instantaneous probability of leaving a state conditional on survival to time t. It is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr[t \le T \le t + \Delta t / T \ge t}{\Delta t} = \frac{f(t)}{S(t)} \qquad (3d)$$

It follows from (18), that

$$\lambda(t) = -dlin(S(t))/dt \Rightarrow S(t) = \exp(-\int_0^t \lambda(u)du) \quad \square \qquad (3e)$$

A final related function is the cumulative hazard function or integrated hazard function define as

$$\Lambda(t) = \int_0^t \lambda(t)dt = -\ln S(t) \qquad (3f)$$

These functions can be estimated using both non-parametric and parametric approaches. Non-parametric estimation can be carried out as described below:

Let

d_j be number of durations (spells) ending at time j;

m_j be the number of spells censored in (t_j, t_{j+1})

r_j spells at risk at time tj.

Then accordingly the hazard rate is estimated as $\hat{\lambda}_j(t) = \frac{d_j}{r_j}$

and the survival function known as the Kaplan-Meier estimator as

$$\hat{S}(t) = \Pi_{t_j}^t (1 - \hat{\lambda}_j(t)) =$$

$$\hat{S}(t) = \Pi_{t_j}^t (\frac{r_j - \hat{\lambda}_j(t)}{r_j}) \qquad (3g)$$

The parametric estimation involves estimating the hazard function through regression analysis. As these functions are non-linear, the maximum likelihood method is used to estimate them. Among the popular hazard functions used in survival analysis are Exponential, Weibull, and Gompertz distributions, whose hazard functions are respectively, γ, $\gamma\alpha t^{\alpha-1}$ and $\gamma \exp(\alpha t)$. These are examples of proportional hazard models (PH), because their hazard functions can be written in the form $\lambda(t/x) = \lambda_o(t,\alpha)\phi(x,\beta)$, where $\lambda_o(t,\alpha)$ is the baseline hazard expressed as a function of time and $\phi(x,\beta)$ is the relative hazard expressed as function of the individuals' covariates.

Other approaches include the log-logistic, log-normal and gamma distributions, which fall under the Accelerated Failure Time model (AFT). They are called the accelerated time failure rate because, unlike the proportional hazards, the covariates lead to changes in the baseline hazards. The hazards are formed when modelling the natural log of time rather than time itself. In other words, when modeling $\ln(t) = x\beta + \mu$, the hazard will result in either log-logistic, log-normal or gamma, depending on the specification of the distribution of µ.The hazards for the log-logistic, Gamma and log-normal distributions are respectively:

$$\frac{\alpha\gamma^{\alpha}t^{\alpha-1}}{[1+(\gamma t)^{\alpha}]}, \frac{\gamma(\gamma t)^{\alpha-1}\exp[-(\gamma t)]}{\Gamma(\alpha)[1-I(\alpha,\gamma t)]}, \frac{\exp(-(\ln t-\mu)^2/2\sigma^2}{\sigma\sqrt{2\pi}} \Big/ [1-\varphi((\ln t-\mu)/\sigma)]$$

Two of the proportional hazards mentioned before also follow under AFT. These are the Exponential and Weibull hazards.

The survival analysis of equation (2) can now be formulated as follows:

$\lambda(t/x) = \lambda_o(t,\alpha)\phi(x,\beta)$ where

$\lambda_o(t,\alpha)$ = the baseline hazard function

β = a vector of covariates so that

$$x'\beta = \beta_o + \beta_1 Age_i + \beta_2 Females_i + \beta_3 GDP_i + \beta_4 Er + \beta_5 El + \beta_6 Pcapita + \beta_7 Childpresence + \beta_8 childno + \beta_9 Tarra + \beta_{10} VistFRD + \beta_{11} VistLSR + \beta_{12} VistBSN + \beta_{13} Adultno + \beta_{14} Tpartyno + \beta_{15} Frvist + \beta_{16} Fadest + \beta_{17} Dist + \beta_{18} Nosites + \beta_{19} \Pr ice + \beta_{20} Peak + \beta_{21} Isource + Africa + Asia + Europe + MEast + NAmerica + SAmerica + year1 + year2 + year3 + \varepsilon \quad \cdots(4)$$

4.3 Variables and their Sources

The study used survey data from TTSS. Although there were six years of survey by the TTSS (2001, 2004, 2005, 2006, 2007, 2008), the study mainly used data for the years 2001, 2007 and 2008, which had similar and relatively more explanatory variables than the rest of the years. These three years made a total of 30,782 observations. But only 25,880 observations out of 30,782 were used, the rest being excluded due to missing values. Some Other variables were obtained from other sources. These included, Tourist country's GDP from IMF (2009), Tourist country's exchange rate from the Economist website, English language proficiency (El) by a tourist from (www.yahoo.com), tourist country distance to Tanzania from the internet using the online distance calculator, and relative cost of living between a tourist country of origin and Tanzania (Price) from the IMF (2009).

5 Results and Discussion

5.1 Summary Statistics

Table 1 provides summary statistics of the variables used in the analysis

Table 1: Summary Statistics of the Variables Used in the Analysis of Tourist Length of Stay

Variable	Mean	Std. Dev.	Min	Max	CV
Lstay(Days)	12.69	13.60	1	360	107.17
Age(years)	36.44	10.94	18	55	30
Females	0.89	0.90	0	21	101.12
GDP(Mil.USD)	29123.30	15379.39	54.62	113044	52.81
Er	965.49	603.92	0.030067	3229.16	62.55
El	0.54	0.50	0	1	94.30
Pcapita(USD)	191.69	206.92	9.5	1017.98	107.95
Childno	0.135	0.58	0	13	429.63
Tarra	0.53	0.50	0	1	94.34
VistFRD	0.08	0.27	0	1	337.5
VistLSR	0.77	0.42	0	1	54.54
VistBSN	0.10	0.30	0	1	300
Adultno	1.75	1.28	1	31	73.14
Tpartyno	0.56	0.50	0	1	89.18
Frvists	0.53	0.50	0	1	94.34
Fadest	0.37	0.48	0	1	129.73
Dist(Miles)	5287.09	2389.07	419	9527	45.19
Nosites	0.65	0.48	0	1	73.85
Price	0.05	0.58	3.24E-05	26.007	1160
Peak	0.74	0.44	0	1	59.46
Isource	0.39	0.49	0	1	125.64
Africa	0.09	0.29	0	1	322.22
Asia	0.08	0.27	0	1	337.5
MEast	0.009	0.09	0	1	1000
SAmerica	0.008	0.09	0	1	1125
year1	0.62	0.49	0	1	79.03
year2	0.10	0.30	0	1	300
year3	0.28	0.45	0	1	160.71

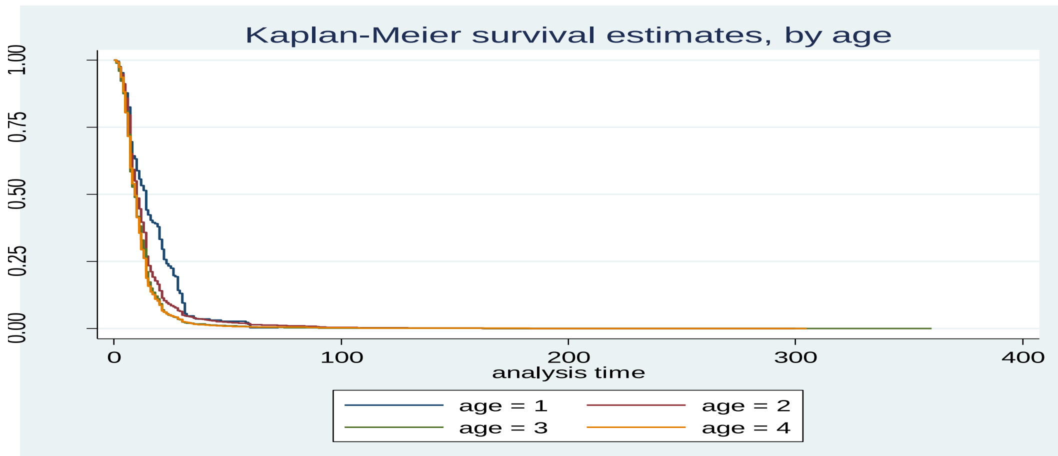
Table 1 indicates that most of the variables have a reasonable variation except for age which has a coefficient of variation less than 50%. It is interesting to note that length of stay is positively skewed (skewness=7.9).This observation is consistent with what is described in the literature regarding the distribution of a time variable (Cameron and Trivedi 2005, Greene, 2003).This skewness justifies the use of logarithmic transformation when modeling length of stay and/or the consideration of survival models which accommodates among other things the skewed nature of time distribution.

5.2 Results from the Survival Analysis

(i) . Non-parametric estimation results

Figure 1 gives the Kaplan Meir survival functions of tourists by age group. Note that the numbers 1, 2, 3 and 4 stand for age groups <18, 18-35, 36-55, and 55+ respectively (refer to section 2.1).

Figure 1: Kaplan Meier Survival Function by Age



Source: Author’s drawing based on statistics from TTSS (2001-2008)

Figure 1, suggests that the assumption of proportional hazard is not correct as the indicated curves cross each other. The Schoenfeld test on the residuals of the Cox PH model confirms this suggestion as it rejects the null hypothesis of proportional hazard (chi2 (24) =1057, p=0.000). This statistical test suggests that our consideration should be entirely focused on the AFT hazards models, which covers among others Exponential, Weibull, Log-normal, Log-logistic and Gamma distributions. In order to choose the best models from among these, Akaieke[2] Information Criteria (AIC) is employed, because some of the models do not nest in each other. According to the AIC, a model estimate with the least value of AIC is the best. Table 2 presents the AIC values for the five mentioned models as well as the log-likelihood values.

Table 2: AIC and Log-Likelihood Values from the Survival Models

Model	Log-like hood	AIC
Exponential distribution hazard	-31491.08	63032
Weibull distribution hazard	-28675.64	57403
log-normal distribution hazard	-24841.62	49735
Lo-Logistic distribution hazard	-23995.62	48043
Gamma distribution hazard	-23759.97	49573

Based on the AIC values, the log-logistic model outweighs the rest as it has the least AIC value (48043). Therefore Table 3 provides parametric estimation results from the log-logistic along side those of the Cox regression model for comparison purposes.

[2] $AIC = \left(-2\ln \hat{L}(M_{\beta}) + 2P\right)/N$ Where P is the number of parameters in the model, N is the number of observations and $\hat{L}(M_{\beta})$ is the likelihood of the fitted model. The smaller the value of AIC the better is the fit of the model (Scott long, 1997).

(ii) Parametric Estimation

Table 3: Log-Logistic and Cox Regression of the Length of Tourist Stay

Variable	Cox-Regression Hazard Ratio	Std.Err	Log-logistic Coefficients.	Std.Err
Age	1.17***	0.011	-0.09***	0.005
Females	0.97***	0.011	0.02***	0.006
GDP	1.00***	0.000	2.66e-06***	0.000
Er	1.00***	0.000	-3.1e-05***	0.000
El	0.94***	0.019	0.05***	0.012
Childno	1.04***	0.013	-0.02***	0.007
VistFRD	0.96	0.035	0.07***	0.023
VistLSR	1.63***	0.049	-0.25***	0.019
VistBSN	1.62***	0.057	-0.52***	0.023
Adultno	1.03***	0.008	-0.02***	0.005
Tpartyno	1.10***	0.018	-0.02**	0.009
Frvists	1.27***	0.017	-0.19***	0.008
Fadest	0.87***	0.012	0.06***	0.008
Dist	1.00***	0.000	-2.3e-05***	2.9e-06
Nosites	0.64***	0.009	0.34***	0.008
Peak	0.91***	0.014	0.08***	0.009
Price	1.01	0.009	-0.01	0.007
Isource	1.21***	0.016	-0.12***	0.008
Africa	1.65***	0.082	-0.33***	0.031
Asia	1.18***	0.034	-0.18***	0.017
MEast	1.29***	0.088	-0.15***	0.040
SAmerica	1.04	0.082	-0.05	0.046
year1	0.99	0.030	-0.03*	0.017
year3	1.14***	0.028	-0.15**	0.014
Constant	NA		2.78	0.040
likelihood	-237029		-22995.62	
Chi	3454.68		5470.21	
P>Chi	0.0000		0.0000	
n	25880		25,880	

NB: *significant at 10%, **significant at 5%, ***significant 1%

Table 3 indicates that the two models are highly significant. The results of the Cox-regression are given in terms of the hazard ratio, whereas the log-logistic results are given in terms of the coefficients of the explanatory variables. It is important to note that a hazard ratio smaller than 1 implies a negative impact of the covariate on the hazard ratio, which is equivalently to a positive impact of the covariate on length of stay and the vice versa.

It is evident that the null hypothesis of destination attributes being more influential than the other attributes is rejected in both models. Rather the trip-related characteristics, as measured by a tourists' frequency of traveling (*frvist*), business visits (*VistBSN*) and leisure visits (*VistLSR*) are more influential than the other attributes. The joint test of hypotheses proved this results beyond doubt as it showed that in totality trip-related characteristics had relatively higher influence (-0.89) compared to peak season(0.08) and demographics(-0.019). Nevertheless, destination attributes as measured by season are quite significant (p=0.000, logit coefficient=0.08, hazard ratio=0.91) and having a positive influence implying that during the peak season (July –September) tourists stay much longer than otherwise. In other words, these tourists must have perceived the peak season as the most ideal period for recreation and other tourist activities. As regards the trip-related characteristics, tourist purpose of visit as measured by business visits and those for leisure purposes are the most influential variables in the three categories of variables. The coefficients of both visits for business and leisure visits appear with negative and significant signs, implying that these types of tourists do not stay long unlike tourists visiting for other purpose who stay much longer. Menezes et al. (2008) established a similar result for business visits when studying tourist length of stay in the Azores.

The next most influential trip-related characteristic is frequency of traveling, which has a negative and significant influence (-0.19), indicating that a frequent traveller is associated with shorter stays than an ordinary traveller. In fact this variable increases the hazard ratio for that kind of a tourist (1.3). As previously postulated, a frequent visitor would allocate holiday time to various destinations compelling him to spend as little time as possible at a particular destination. This result however, is contrary to the study by Govakali et al. (2007), who found that an experienced tourist spends much more time than others.

The demographic variables included in the model and which had significant influence include age (-0.09), per capita GDP (.2.66e-06) and exchange rate (-3.1e-05), of which all were less influential than the trip related characteristics. Other variables were number of females (0.02) number of children (-0.02), travel party number (-0.02), number of adults (-0.02), distance (-0.00002) and number of sites visted (0.34)

6. Conclusion

The paper has highlighted the determinants of tourist length of stay in Tanzania. Among the key determinant are the tourists' trip-related characteristics which appear to be the most influential thus requiring the most attention by the stakeholders in the industry. To promote tourism revenue based on the tourists' length of stay the stakeholders need to diversify tourism activities to encourage business visitors and those on leisure and recreation to stay longer. Equally well, more tourists' sites need to be promoted especially those found in the western and Southern parts of the country which apparently receives relatively fewer tourists than the sites found in the country's northern circuit.

7. Acknowledgements

I am highly indebted to Professor Asmerom Kidane and Professor Nehemiah Osoro of the University of Dar es Salaam for their invaluable contribution which has enabled me to produce this work. The other appreciation should go to Tanzania Tourism Sector Survey (TTSS), which supplied me with the cross section survey data on tourists.

References

Alegre, J and Pou, L (2006) “The length of stay in the demand for tourism”, Tourism Management , 27(6):1343-1355

Barros, C and Correia, A (2007), “Survival Analysis in Tourism Demand: The length of stay in Latin American destinations”, working paper series No.2007/03 from the Department of Economics of the School of Economics and Management (ISEG),technical University of Lisbon.

Barros, C., Butler, R and Correia, A (2010), “The length of stay of golf tourism: A Survival analysis’, Tourism management, 31(1) :13-21

Belleau, B (2007), “Theory of Reasoned Action: purchase Intention of Young Consumers”, Clothing and textiles Research Journal, 25(3): 244-257

Cameron, A.C and Trivedi, P.K. (2005), Micro econometrics methods and Applications**, New** York, Cambridge University press, 40 West 20 th Street, NY 10011-4211, USA

Chaiboonsri, C and Chaitip, P (2012), “The Modeling International Tourism Demand for Length of Stay in India: Social Development, Economics Development”, Journal of Knowledge Management, Economics and Information Technology

Fishebein, M and Ajzen, I (1975), Belief, Attitude, Intention and Behavior: An introduction to theory and Research . Reading, MA: Addison-Wesley

Garcia, E and Raya, J (2008), “Length of stay for low-cost tourism”, Tourism Management volume, 29(6): 1064 - 1075

Govakali,U., Bahari, O and Kozak, M (2007), “:Determinants of length of stay: A practical use of Survival analysis”, Tourism Management,28(3):736-746

Greene, W.H (2003) **“**Econometric Analysis: fifth edition**”** Pearson Education, Inc., Upper Saddle River, New Jersey, 07458

Gujarat, D.N (2003), “Basic Econometrics: fourth edition” McGraw-Hill Companies.inc.

International monetary Fund (IMF) (2009), “Word Economic Outlook”, IMF

Mathieson, A. & Wall, G. (1984): Tourism economic, physical and social impacts**.** Third edition, Longman Group limited, N.Y, USA

Menezes, A.G., Moniz, A and Viera, J.C (2007), “The Determinants of length of stay of Tourists in the Azores: A count model Approach”, presented at Ponta Delgada, Portugal Observatorio Regional do Tourismo.

Menezes ,A. G., Moniz, A and Viera, J.C (2008), “ The determinants of length of stay of tourists in the Azores”, Tourism Economics, 14(1): 205-222.

MNRT (2002), “Tourism in Tanzania: Investment for Growth and Diversification”, Washington Multilateral Investment Guarantee Agency/World Bank Group

Nzuki, A (2006), “A consumer situational Analysis: “The influence of young children on the purchase of Travel to overseas destinations”, Unpublished PhD Dissertation, Graduate School of Clemson University.

Scott long (1997), Regression models for Categorical and limited dependent Variables: Advanced Quantitative Techniques in the Social Sciences, California, SAGE Publications, Inc. 2455 Teller Road, Thousand Oaks, California 91320

TTSS (2001), “The 2001 International Visitors’ Exit Survey Report”, The 2001 Tanzania Tourism Sector Survey

TTSS (2004), “The 2004 International Visitors’ Exit Survey Report”, The 2004 Tanzania Tourism Sector Survey

TTSS (2005), “The 2005 International Visitors’ Exit Survey Report”, The 2005 Tanzania Tourism Sector Survey

TTSS (2006), “The 2009 International Visitors’ Exit Survey Report”, The 2006 Tanzania Tourism Sector Survey Exit Survey Report”,

World Tourism Organization (WTO) (1995), “Technical manual no.2: The collection of Tourism Expenditure Statistics” WTO, Madrid

Yang, Y; Wong, K.F; Zhang, J (2011), “Determinants of Length of Stay for Domestic Tourists: Case Study of Yixing”, Asia Pacific Journal of Tourism Research, 16(6): 2011