MODELLING DETERMINANTS OF TOURISM DEMAND AS A 5-STAGE PROCESS. A DISCRETE CHOICE METHODOLOGICAL APPROACH

Juan L. Eugenio-Martin

Departamento de Análisis Económico Aplicado
Universidad de Las Palmas de Gran Canaria

Facultad de Ciencias Económicas y Empresariales. Módulo D.
C/ Saulo Torón. CP. 35017. Las Palmas de Gran Canaria. Spain.
Phone: +34 928 458216
Email: jluis@empresariales.ulpgc.es

Environment Department
University of York

Heslington, York. YO10 5DD. United Kingdom.
Phone: + 44(0) 1904 434067
Email: jlem100@york.ac.uk

Juan Luis Eugenio-Martin is currently undertaking PhD in Economic Analysis at Universidad de Las Palmas de Gran Canaria (Spain) and PhD in Environmental Economics and Environmental Management at University of York (UK).

MODELLING DETERMINANTS OF TOURISM DEMAND AS A 5-STAGE PROCESS. A DISCRETE CHOICE METHODOLOGICAL APPROACH

Abstract

In the tourists' destination choice there are multiple factors that affect their decision. Individuals or families with exactly the same socioeconomic and demographic characteristics may choose very different destinations. The paper deals with this heterogeneity problem recognising there are taste differences among tourists and that final destination choice is not an independent decision, but just last decision of a set of choices that are also determining it. In this sense, we argue that tourists face a 5-stage decision process. First of all, people have to decide whether or not travel within a period of time. Second, those who expect to travel need to estimate a budget for tourism expenses. Third, given the budget, they need to determine frequency and length of stay of their trips. Fourth, once a date and length of the stay is proposed, tourists need to choose which kind of tourist destination wishes to visit. And finally, among all the available destinations that satisfy tourist's conditions, final destination and mode of transportation are chosen. It is the purpose of this paper to propose a methodological framework for modelling each of these stages and their relationship.

Keywords: Tourism demand, Outbound tourism, Discrete choice models, Tourists' decisions, Tourism marketing

1. INTRODUCTION

For many regions tourism has become one of the most significant economic activities in terms of economic growth and employment. World tourism demand is still growing and new or current destinations may be developed or extended in order to satisfy such a growth. In this sense, tourism may be been seen as an opportunity of economic growth for developing regions. However, despite tourism demand is still growing, tourism supply is also growing in the same fashion and competition among destinations arises. Destinations may deal with competition from a two-level perspective. On the one hand, from a micro-level point of view, hospitality sector may change prices and quality and adapt the services offered to the preferences of their potential visitors. On the other hand, from a macro-level point of view, local, regional or national authorities may invest on the development of tourist resorts and promote them. In order to be efficiently applied, any of these policies requires a deep knowledge of the characteristics of potential visitors, their needs and the interrelations with other competitive destinations. Therefore, the question that needs to be answered is why people travel to different places. This paper aims to focus in this way. The main purposes are related with tourism marketing and tourism planning decisions from both micro and macro-level points of view. More precisely, we try to provide a methodological framework which may analyse the relative importance of different attributes for tourists' destination choice; estimate probability of visiting each kind of destination for different kind of tourists; and finally, generate a tool that lets us simulate changes in the demand under alternative scenarios.

In the tourists' destination choice there are multiple factors that affect their decision. For instance, age, labour conditions and income are usually conditioning tourists' choices. However, what makes modelling destination choice even more challenging is the fact that individuals or families with exactly the same socioeconomic and demographic characteristics

may choose very different destinations. The paper deals with this heterogeneity problem recognising there are taste differences among tourists and that final destination choice is not an independent decision, but just last decision of a set of choices that are also determining it. In this sense, we argue that tourists face a 5-stage decision process. First of all, people have to decide whether or not travel within a period of time[1]. Second, those who expect to travel need to estimate a budget for tourism expenses. Third, given the budget, they need to determine frequency and length of stay of their trips. Fourth, once a date and length of the stay is proposed, tourists need to choose which kind of tourist destination wishes to visit. And finally, among all the available destinations that satisfy tourist's conditions, final destination and mode of transportation are chosen. It is the purpose of this paper to propose a methodological framework for modelling each of these stages and their relationship.

2. TOURISTS' DESTINATION CHOICE PROCESS

When studying tourism demand we can consider two points of view. On the one hand, we can forecast number of tourists that are expected to arrive to a particular destination; i.e. we consider inbound tourism. But on the other hand, we can try to understand tourist destination choice of the inhabitants of a particular region; i.e. we analyse outbound tourism. The purpose of this paper is to provide a methodological framework which may estimate the main determinants of outbound tourism demand.

As it is mentioned in the introduction of this paper, we consider that before deciding where to go on holidays, most of the tourists need to make multiple decisions. While for some people these decisions are perfectly planned, for other people these are improvised or hardly planned. Moreover, some people can decide all of them simultaneously or in different stages. Anyway,

[1] For our purposes we assume to be a one year period.

either if it has been planned or not, we argue that most of the people consider consciously or unconsciously a process of decisions concerning their holidays trips. For modelling purposes, we assume that tourists choose final destination depending on another four decisions. In this sense, we consider five stages: participation decision; tourism budget decision; frequency and length of stay decisions; kind of destination decision; and final destination and mode of transportation choice.

Methodology proposed is a general framework applicable to any place in the world of any size. For instance, it may be employed for a country, a region or a small town or village. Obviously, the larger the region analysed is, the more heterogeneity we will face. The main objective of the methodology proposed is to deal with this heterogeneity problem.

In order to apply this methodology we need two different datasets. On one hand, we need micro data on socioeconomic and demographic characteristics of a representative sample of population. This dataset must also include data on tourism trips, as for instance, places visited, number of trips, length of stay or expenditure on tourism. On the other hand we need data on the attributes of choice set. This is an objective dataset and it is easier to obtain. It usually includes variables as accommodation cost index, price index, development level or temperature.

Concerning the sample we need to comment some points. Sampling can be covered from the whole population, from on-site or from a combination of both. The main advantage of on-site sampling is that deeper and wider variability may be obtained compared to population sampling. Although for a destination choice analysis, on-site sampling might be more

convenient, for our purposes it would be incomplete because we may loose information concerning the reasons why people decide to travel or not.

Another issue related with sampling is the period of time the study covers. This could range from a season, a year or a set of years. The period chosen depends on the purposes of the analysis. A period of a season might be chosen if the region analysed is remarkably affected by the season and this effect is relevant. A longer period than a year is useful if researcher wants to trace tourist's behaviour over years. It is very interesting because it lets to reveal aspects as repetition patterns, risk aversion and tourism budget decision making. However the inconvenience is the possible mistakes interviewees may commit especially concerning with data related with trips made more than three years ago. A solution is to generate a panel dataset that trace individuals over time repeating interview for each period of time. However, for a general purpose, a period of one year time seems to be sensible[2].

Next section deals with each of the stages, explaining the objectives, variables considered, methodology proposed for the analysis. Moreover alternative methodologies are briefly discussed.

3. STAGE 1: PARTICIPATION DECISION

Introduction

The first decision any individual has to make, concerns the choice between travelling or not within a period of time. Researcher, depending on the purposes of the analysis, must set a time interval. Usually, this may range from a period of a year to a period that contains the whole life of the interviewees.

[2] One year period is also recommended by Morley (1995).

Objective

Mainly, this is an interesting issue for tourism marketing analysts, because it shows which the determinants of the decision to travel are. Similar models can be constructed depending on socioeconomic or demographic characteristics of individuals. The objective is to estimate different models attending to a segmentation criterion, such that we can compare the results from different models and draw some conclusions. For instance, we can estimate determinants of participation for different places of residence. This analysis may reveal that residents of a particular region are less interested in travelling than residents of other regions due to several aspects, as income differences or differences in especial facilities for recreation, which help residents to enjoy holiday's time in their own place of residence. Moreover, we can make segmentation in the sample attending to any variable and compare the relative importance for each segment of any variable. A common case to be analysed is the effect of the age. We can estimate relative importance of particular problems of each segment. For instance, we can compare how important is the income for youth segment with respect to the others, or how relevant is to have a child or not for mid-age tourists' decision or how significant are health conditions for elderly people's decision. Marketing effort can focus on different segments attending to their main determinants for travelling.

Main variables

Intuition suggests that variables as age, education, income, labour conditions, characteristics of the place of residence and size and composition of the household or family, may be significant when deciding to travel or not.

Methodology

In order to model participation decision, we consider it is a binary choice, denoted by T_i, such that, $T_i = 1$ if household or individual decides to travel and $T_i = 0$ otherwise. We want to model probability that $T_i = 1$, i.e. $\Pr(T_i = 1)$. We assume $\Pr(T_i = 1)$ is linked to a set of exogenous variables, which may be those already shown above. More precisely, for some appropriate function $g(\cdot)$, $\Pr(T_i = 1) = g\left(\alpha + \sum_{j=1}^{k} \beta_j SE_{ji}\right)$, where $0 \le g(\cdot) \le 1$, α denotes a constant, SE_{ji} denotes jth socioeconomic variable of household or individual *i* and β_j denotes associated parameter to jth socioeconomic variable.

Traditional linear probability model is not recommended to be used to estimate the probability function because it would present non normal errors, heteroskedasticity and logical inconsistency, since prediction of probabilities may lie out of range (0,1). It is well-known that the suggested model for binary choice estimations is latent variable model. This model considers the existence of a latent variable T_i^*. Since this latent variable is unobserved by the researcher we can consider it is composed by two parts: one observed by the researcher, which includes all the socioeconomic variables and another part that it is unobserved by the researcher and that corresponds to heterogeneity reasons among tourists. Thus the model can be represented as: $T_i^* = \alpha + \sum_{j=1}^{k} \beta_j SE_{ji} + \varepsilon_i$, where ε_i denotes unobserved part or error term. For our purposes, the latent variable will work as an index function, such that we will set $T_i^* = 1$ if $T_i^* > 0$ and $T_i^* = 0$ if $T_i^* \le 0$.

Let $S_i = \alpha + \sum_{j=1}^{k} \beta_j SE_{ji}$, such that $T_i^* = S_i + \varepsilon_i$.

Then, $\Pr(T_i = 1) = \Pr(S_i + \varepsilon_i > 0) = \Pr(\varepsilon_i > S_i) = 1 - \Pr(\varepsilon_i \le S_i) = 1 - F_\varepsilon(-S_i)$, where F_ε denotes cumulative density function of unobserved part. Due to a problem of identification of

location and scale of T_i^*, researcher needs to choose a distribution and a value for the variance of ε_i. The most common approaches assume ε_i is independently and identically distributed, either following a normal distribution with zero mean and variance of one, or following a logistic distribution with zero mean and variance of $\frac{\pi^2}{3}$. If we assume that ε_i follows the former distribution we are employing the well-known probit model, and if we assume the latter distribution we are employing the also well-known logit model. Any of these distributions can be employed for the participation decision and both present similar results. Finally, maximum likelihood estimation is applied to the model in order to estimate parameters of interest. Under correct specification, these estimates are consistent and asymptotically normal[3].

4. STAGE 2: TOURISM BUDGET CONSTRAINT

Introduction

Once a household or individual has decided to travel, they have to decide how much tourism expenditure may be. This decision depends mainly on income and preferences of individuals. If the analysis is performed with income in absolute terms, thus this variable is likely to dominate the estimated regression. In order to avoid a trivial result, we can estimate tourism budget as a percentage of income, i.e. as a ratio between tourism expenditure and income. This new formulation lets us to estimate how much people prefer to distribute their income for tourism purposes, or in other words, how much people like tourism.

[3] For a complete exposition of the methodology see Greene (2003).

Objective

The main purpose of this stage is to try to understand the main factors that push different individuals to spend part of their budget in tourism activities.

Main variables

We are interested in estimating the relative importance that variables as age, income, labour conditions, place of residence and size of the household or family possess in tourism budget decisions.

Methodology

Depending on the nature of the data we can apply different methodologies. Ideally, both income and tourism expenditure might be continuous variables. However, it is likely that the questionnaires consider discrete intervals for these two variables.

For the continuous variables case, we can estimate tourism expenditure as dependent on its own determinants as well as the demand of another goods and services. A traditional approach is to estimate this tourism expenditure as a demand function that is part of a system of demands which include all other goods and services. Deaton and Muellbauer (1980a) dealt with this issue developing a model known as Almost Ideal Demand System. Tourism literature has employed this approach frequently[4]. However, as far as we know, all the studies have used macroeconomic variables rather than micro data and the inconveniences already commented in previous sections applies. Micro data can be used to estimate this system of demands, but lack of data has usually been a common problem.

Nevertheless, we can assume that the whole set of commodities can be divided into different groups, such that preferences within groups can be described independently of the quantities in other groups. This assumption is known as weak separability[5] of the utility function. It is a plausible and common assumption in the tourism economics literature[6]. Under this assumption we can create a tree of commodities, where one of the main branches may be a group called entertainment, and tourism may be a further branch linked to entertainment. This structure allows us to analyse tourism expenditure allocation independently of expenditure levels in any other goods and services[7].

Consequently, under weak separability assumption, we can model tourism budget decision depending on socioeconomic variables already mentioned. However, since tourism budget is measured as a percentage of total income, such that its value must lie between zero and one, linear probability model cannot be applied for the same reasons we have already explained in the participation decision model. Several solutions can be adopted. A traditional way is to employ a translog model. This model usually works appropriately for independent functions. However, if model is used for a system of demand functions, it usually fails in testing additive and homothetic restrictions[8]. Another way to model tourism budget is through a double censored regression model. In this case, researcher can impose two censors to the potential freedom given by default to any endogenous variable in the traditional regression analysis. Obviously the lower censor might be set on zero and the upper censor on one, such that any estimated value is guaranteed to lie between these two values[9].

---

[4] See O´Hagan and Harrison (1984), Syriopoulos and Sinclair (1993), Papatheodorou (1999) or Divisekera (2003).
[5] For a further reference see Deaton and Muellbauer (1980b), pp:119-136.
[6] See for instance Rugg (1973).
[7] See Hultkrantz (1995), who considers a three-stage budgeting framework.
[8] See Bakkal (1991).

As we have mentioned, it is likely that micro data on income or tourism expenditure be collected in the interview process, as the choice among alternative predetermined intervals, rather than an exact number. Unfortunately, this kind of data collection is less efficient and therefore loose of information occurs. Anyway, we can estimate the model employing a discrete approach. For illustrative purposes, say we have income variable collected in ten intervals and tourism expenditure in five intervals. Therefore, we have fifty different combinations and potential groups. It is researcher's task to group all these intervals into new sensible sets, say for instance, five sets. As we have mentioned, we loose information in the process, however, these sets may represent how much people like tourism. In this example, these sets may group people, such that those individuals or households that belong to the first set with percentage of tourism expenditure over income, between zero and lowest threshold determined by researcher, means these individuals do not like tourism. As long as interval of the set corresponds to higher percentage, researcher can define other labels as do not like tourism much, like tourism, like tourism much and finally, like tourism very much. In order to deal with this ordered categorical variable, we can use what is known as an ordered probit model.

In the ordered probit model, we have different ordered multinomial outcomes, denoted by j, where $j = 1...m$. In our example the categories $j = \{$ do not like tourism, like tourism a bit, like tourism, like tourism much and like tourism very much $\}$. Traditional ordered probit model estimate thresholds itself, however, since researcher may predetermine them, as we have already commented, using percentage of tourism expenditure over income, then we can employ a superior regression known as interval or grouped data regression. This is no more than a variant in which the values of the thresholds are known. Because the thresholds are

[9] For a complete analysis of this technique, see Maddala (1983) or Greene (2003).

known, the estimates of the parameters are more efficient and it is possible to identify the variance of the error term.

Similar justification to the model presented in stage 1 applies in this case. Again, we consider a latent variable model $B_i^* = SE_i\beta + \varepsilon_i$, where B_i^* denotes the latent variable that reflects how much people like tourism assuming this is a function of tourism budget decisions, SE_i denotes socioeconomic variables corresponding to household or individual *i* and ε_i denotes unobserved part of the model or error term, which it is assumed to be normally distributed with zero mean and unitary variance. If we denote the thresholds which determine the limits within category *j* lies by μ_{j-1} and μ_j. The model assumes individual or household belongs to category *j* if $\mu_{j-1} < B_i^* \leq \mu_j$, $j = 1,...,m$. Since $B_i^* = SE_i\beta + \varepsilon_i$, then substituting into the inequality, we obtain $\mu_{j-1} < SE_i\beta + \varepsilon_i \leq \mu_j$, and furthermore, $\mu_{j-1} - SE_i\beta < \varepsilon_i \leq \mu_j - SE_i\beta$. Hence, we can estimate probability that any individual *i* belongs to any category *j* as the difference between two cumulative density functions as: $P_{ij} = \Phi\left(\mu_j - SE_i\beta\right) - \Phi\left(\mu_{j-1} - SE_i\beta\right)$.

Moreover, we can obtain how varies this probability under a marginal variation on any socioeconomic variable. As for instance, to be a year older or to have a baby. We need no more than to differentiate probability with respect to a marginal change on SE_i:

$$\frac{\partial P_{ij}}{\partial SE_i} = \phi\left(\mu_j - SE_i\beta\right)(-\beta) - \phi\left(\mu_{j-1} - SE_i\beta\right)(-\beta) = \beta\left[\phi\left(\mu_{j-1} - SE_i\beta\right) - \phi\left(\mu_j - SE_i\beta\right)\right].$$

These marginal effects will reveal us how robust enjoyability of different kinds of tourists is with respect to changes in any of the socioeconomic variables that define them.

5. STAGE 3: FREQUENCY AND LENGTH OF STAY

Introduction

Once we know that a household is participating in tourism activities and that a tourism budget constraint is assigned, we may extend the participation analysis to how often the household travels. Unfortunately, this analysis is not as straightforward as previous stages are. Complexity arises from two simultaneously dependent decisions. Given a budget constraint, individuals decide how often and how long stay in their trips. These two decisions are depending on each other because a longer stay may affect frequency of travelling and vice versa.

Nevertheless, for simplicity, we may assume that for every travel, each tourist possesses an optimal length of stay. Let me explain the basis for this assumption. Usually travelling provides satisfaction to the travellers. Extra days in the destinations increase satisfaction, however, these increases in satisfaction are less and less relevant as number of days in the destination increases, i.e. it does not provide the same satisfaction the first day of travelling than any of the last days of the travel. Generally speaking, this satisfaction is the main benefit of the traveller. Unfortunately for the traveller, he or she incurs in several costs. Costs can be decomposed between fixed and variable costs. Variable costs are the costs incurred by an extra day of travelling, while fixed costs are all other necessary costs to travel that are independent of the length of stay. For instance, among the fixed costs we can find transportation costs, travel time costs or travel planning costs, and among the variable costs we can find accommodation and other local services as food, local transportation and leisure activities. It seems obvious that optimal length of stay depends on how large fixed costs are. For instance, if fixed costs are large, stays of a week or two weeks are expected, while if fixed

costs are not large, stays of a weekend may be long enough. We argue that if individual or household has flexibility to set length of stay, according to the balance between satisfaction and costs incurred they will determine an optimal length of stay. Hence, given a destination choice, length of stay will be optimally determined depending on preferences of the individuals. Consequently, in the hierarchy of tourist's decision process, it seems plausible that first, tourists decide how many trips to make in a period of time, say a year, and then determine the destination to go, where each destination is conditioned to an optimal length of stay determined by each individual. This assumption allows us to concentrate on frequency decisions separately.

Objective

The purpose of this analysis is to determine the main factors that contribute to travel frequency within a period of time. As in stage 1, we consider a period of one year. This kind of study is of interest to tourism marketing analysts since it may reveal some new information about different segments of the market. It is of special interest the significance of variables as age, labour conditions and income.

Main variables

The following variables are expected to be relevant for the analysis: age, labour conditions, income, place of residence, size and composition of the household or family, education, health conditions and unobservable variables as risk aversion and propensity to travel.

Methodology

If we have a look to the frequency of travelling of people, we can see this follows a distribution that is skewed to the left and contains a large proportion of zeros and ones.

Besides this distribution, dependent variable is a non-negative integer-valued count and therefore we can employ count data methodology to estimate frequency of travelling.

A classical model for count data is Poisson process: $P(n_i) = \frac{\lambda^{n_i} e^{-\lambda_i}}{n_i!}$, where n_i denotes number of travels that household or individual *i* makes during a fixed interval, say for instance, a year; $P(n_i)$ denotes probability that household or individual *i* travels n_i times and $\lambda_i = E(n_i | SE_i) = \exp(SE_i \beta)$. An important feature of the Poisson model is that it imposes $E(n_i | SE_i) = Var(n_i | SE_i) = \lambda$, which is known as the equidispersion property. In some cases this property may be true but in other cases it is violated. If variance is greater than expectation, we have a case of overdispersion and Poisson model will tend to under-predict the actual frequency of zeros. Mullahy (1997) associates overdispersion with the existence of unobservable heterogeneity. He suggests to deal with overdispersion allowing for a specification which includes an error component which represents omitted variables or unobserved variables. For this purpose, negative binomial distribution is usually applied. It can be seen as a more general expression that also contains Poisson distribution as a particular case. The negative binomial distribution imposes $E(n_i) = \lambda$ and $Var(n_i) = \lambda + a\lambda^{2-k}$, such that when $k = 1 \Rightarrow Var(n_i) = \lambda + a\lambda$, known as negative binomial 1 model, when $k = 0 \Rightarrow Var(n_i) = \lambda + a\lambda^2$, known as negative binomial 2 model and when $a = 0 \Rightarrow Var(n_i) = \lambda$ we obtain Poisson model. An alternative model is known as zero inflated model, which employs a mixing specification which adds extra weight to the probability of observing a zero. The main inconvenient of these models is that variance is imposed exogenously by the researcher. If this variance works properly for the model of interest it will be sufficient but sometimes excess of zeros may not be associated with

increased dispersion but with an underlying tourist's behaviour. We refer to unobservable patterns in the behaviour of tourists, as fear of flying or high propensity to travel. If researcher believes this behaviour is relevant enough, then it requires a more complex model for the analysis.

In order to deal with unobservable heterogeneity we will consider briefly two different approaches.

On one hand, we can split population according to participation decision of stage 1 and assume frequency depends on two separate processes, one followed by those people who decided not to travel and other process that consider all those people who do travel. In the econometric literature this is known as hurdle model.

On the other hand, we can split population according to how much they like tourism, already obtained in stage 2. In this case, each segment follows an independent process[10].

The independent processes of both of these mechanisms are added up in the log-likelihood final function in order to be estimated by maximum likelihood. This is the reason why these models are known as mixture models.

6. STAGE 4: KIND OF DESTINATION

Introduction

In previous stages, we have studied participation decision, tourism budget decision and frequency of travelling. Once an individual or household have chosen a frequency for travelling, they must decide where to go in their different trips. Nowadays, for most of the developed countries, the choice set for travelling is quite large. There are multiple available

destinations. Each of them posses special features, i.e. some particular characteristics that make them unique. There are some destinations that satisfy almost every kind of general need, but maybe do not satisfy the needs of a minority of tourists. Once again, researcher faces a problem of heterogeneity. Tourists differ from what they consider it is an ideal destination and which are the needs they want to be satisfied.

Within this stage we consider tourists make two simultaneous decisions. One of them concerns the nature of the travel. In this sense, tourist must decide which kind of travel wishes. For instance, they may choose between a familiar, adventure or relaxing trip. The other decision concerns the kind of destination in terms of its physical attributes. In other words, for instance, tourists may decide if the kind of destination they prefer is a mountain, a city, a countryside resort or a seaside resort.

Objective

This stage is a previous stage to final destination choice and it helps in defining more appropriately the choice set for different kind of needs that different tourists are demanding. It is important to define properly the choice set because it will affect efficiency of estimates.

Main variables

Besides the socio-economic and demographic variables, we must create a tree structure that classify different kind of destinations depending on their physical attributes and kind of tourist environment that may possess.

---

[10] See for instance Deb and Trivedi (1997)

Methodology

An illustration of different criteria to classify the different kind of destinations that tourists may wish to go, is shown below:

a) Physical attributes of destination

   a. Seaside
   b. Countryside
   c. Mountain
   d. City

In this case, tree structure posses four branches and researcher needs to determine in which of these branches every destination must be classified.

b) Tourist environment

   a. Familiar
   b. Cultural
   c. Relax
   d. Party
   e. Adventure

Similarly to case a, researcher needs to classify each potential destination according to these five branches.

c) Both combined hierarchically

In this case, tree structure has two levels. In a first stage, tourists might choose a kind of destination, according to criterion a (or b) and in a second stage they might choose a kind of

tourist environment, according to criterion b (or a). In any case, we end up with 20 different kind of tourist destinations.

d) Both combined simultaneously

Similarly to previous classification, this proposition do not consider stages but tourist must decide simultaneously among the 20 different kind of tourists destinations.

The choice of any of these alternative ways to classify kind of tourist destination is relevant for the methodology of final destination choice. It is possible to model kind of tourist destination decision with a multinomial logit model if we attend criteria a, b, or d. For criterion c, we can model it with a nested multinomial logit model.

Any household, labelled *h*, may decide to travel not. Once the household has decided to travel, it has to choose the kind of tourist destination they wish, labelled *d*, among its choice set. In order to model the kind of destination choice, we follow a behavioural model where household chooses the alternative that provides higher level of utility, denoted by *U*. In this sense, household *h* would choose kind of destination *d* if and only if: $U_{hd} > U_{hr} \forall r \neq d$, where *r* denotes any other kind of resort. Nevertheless, these utility levels are unobservable for the researcher. The only aspects that we know are some socioeconomic variables of the household, denoted by SE_h, and some attributes of the set of destinations, denoted by A_d. From the information available, we can construct a function $V_{hd} = V\left(SE_h, A_d\right) \forall d$, which represents, the utility that site s provides individual. Obviously, this representative utility V_{hd} is an approximation to the current utility U_{hd} . Thus, we can state that utility can be decomposed as: $U_{hd} = V_{hd} + \varepsilon_{hd}$, where ε_{hd} denotes unobserved part of utility for household *h*

when visits destination *d*. Thus, the probability that a household *h* chooses to travel to kind of destination *d* is:

$$P_{hd} = \Pr\left(U_{hd} > U_{hr} \forall r \neq d\right) = \Pr\left(V_{hd} + \varepsilon_{hd} > V_{hr} + \varepsilon_{hr} \forall r \neq d\right) = \Pr\left(\varepsilon_{hr} - \varepsilon_{hd} < V_{hd} - V_{hr} \forall r \neq d\right)$$

For convenience, we assume that $\varepsilon_{hd} \sim iid$ extreme value. The advantage of this assumption is that the normalisation required in any discrete choice model, i.e. $\varepsilon_{hdr} = \varepsilon_{hr} - \varepsilon_{hd}$, imposes that the error differences ε_{hdr} is distributed logistically, which implies that, after algebraic manipulation[11], the probability formula can be obtained exactly as: $P_{hd} = \frac{e^{V_{hd}}}{\sum_r e^{V_{hr}}}$.

Nevertheless, we can link kind of tourist destination choice with final destination choice through a nested multinomial logit model. This is the purpose of last stage.

7. STAGE 5: DESTINATION AND TRANSPORTATION MODE CHOICE

Introduction

One of the most interesting decisions of all this choice process is final destination choice. As we have remarked along this paper, researcher needs to deal with the heterogeneity of tourists in order to obtain accurate estimates from the model proposed. Previous stages provides additional information to be included in the analysis of this final stage that help to obtain more efficient resuts. From stage 2 we can include information as how much each individual likes tourism, from stage 3, if significant, we can provide information as propensity to travel and stage 4 is very important for the methodological structure and minimisation of heterogeneity among tourists.

[11] Train (2003) includes the full algebraic manipulation.

We assume that when tourists determine a destination to travel, this decision is linked to the transportation mode choice. In international tourism, plane is the expected transportation mode and decision about transportation mode is already taken. However, there are many other destinations where tourist may choose how to go there. To capture this effect, we propose to include in tourist's choice set, every sensible combination between destination and transportation mode.

Objective

This final stage focus in the estimation of the main determinants of tourists' destination choice. This is the more complex stage in terms of variables that may influence tourists. We consider characteristics of the household or individual and attributes of the destinations. The purpose is to determine the relative importance of any of these variables with respect to the final decision. Moreover we want to obtain a framework which may allow us to simulate how current distribution of tourists would be affected by any change in any variable of interest.

Main variables

The main variables that we expect to be relevant are, for illustrative purposes shown in Table 1:

(Table 1 should be here. It is located at the end of this document.)

Methodology

As we have mentioned in the previous section, the methodology proposed links stage 4 and 5 employing a nested multinomial logit model[12]. This is a superior modelling with respect to the multinomial logit model because it deals more appropriately with unobservable heterogeneity. Moreover, multinomial logit model imposes, by construction, a restriction known as independent of irrelevant alternatives (IIA) property. For our purposes, this property may cause an inconvenient. More precisely, IIA implies that, given a change in any attribute of the alternatives, cross elasticities of the probabilities of choosing any alternative are exactly the same for every alternative. If the variables that explain the behaviour of tourists are stable over time this implication is not problematic as it is the case of the stage 4. However, in stage 5 it is likely that attributes of the destinations vary quite often. Moreover, if we wish to simulate changes in these attributes, same cross elasticities assumption would create a bias in the results. Nested multinomial logit deals with IIA, such that it allows for different cross elasticities between different nests but not within the same nest. Generally speaking, this specification would be flexible enough in order to obtain sensible simulations.

Nevertheless, more flexible alternative models can be applied. If IIA is still a relevant problem in the model, researcher may use heteroskedastic extreme value models (HEV). Furthermore, if researcher wishes to apply full flexibility to the estimates, such that, the model can provide particular estimates for each kind of representative individual, random parameters model or mixed logit model can be employed. The inconvenience of these models is that probabilities can not be obtained with integrals which posses a closed-form but they require simulation and consequently estimation is more complicated.

An specific problem of this stage that requires special care is the definition of the choice set. Three aspects must be taken into account: criteria to define different alternatives, flexibility and size of the alternative destinations.

Concerning the definition of different alternatives, we may follow different criteria. We can consider destinations as a political division; a natural division, in terms of kind of territory or weather; and a division according to the kind of activities that can be practised in the destination.

Since tourism literature is still limited in this area, we can benefit from outdoor recreation studies. Parsons and Hauber (1998) showed that for recreational fishing trips, 94% of individuals choose sites within one hour and a half travel time distance. Consequently, for this kind of recreation spatial boundaries for choice set must be established in order to optimise the efficiency of the estimation. Moreover, it is interesting to see Thill (1992), because he considers different approaches to capture the true choice set and Haab and Hicks (2000), who have provided a survey of the way that choice sets have been considered in recreation demand models.

Researcher might also think about the flexibility of the choice set. This can be predetermined and fixed by the researcher or it can be endogenously determined by the model, i.e. defining a particular choice set for each individual. Endogeneity may be explained, among other variables, through level of information, number of trips or age. General awareness of tourist destinations it is likely to widen the choice set. We can consider four levels of awareness: International, national, regional and local level. The more an individual travels at any of these levels, it is expected that, the wider his or her choice set will be.

---

[12] Eymann and Ronning (1997) applies nested multinomial logit model for the study of German outbound

Finally, we also need to adjust the size of the destinations, such that, most of them be homogeneous in size. Otherwise, we may have a bias with respect to the effect of the attributes. For instance, despite France and Switzerland may have in the Alps very similar attributes, France have many more visitors because it is a bigger country. In this sense, attributes of Switzerland may be undervalued by the estimation with respect to France due to the differences in the size of both countries. We need either to split up the country into homogeneous destinations or adjust the number of arrivals with respect to the size of the country (arrivals per km square).

8. CONCLUSIONS

The methodology proposed assume that tourists' destination choice is conditioned to another four decisions. We argue that decision process follows a hierarchy with five stages: participation decision; tourism budget decision; frequency and length of stay decisions; kind of tourist destination decision; and final destination and mode of transportation choice. This structure responds to the necessity to deal with the heterogeneity among tourists.

For the first stage, the participation decision, we suggest the employment of a probit or logit model. From this model, we may estimate the probability of an individual or household to travel and the effects on this probability of any change in any socioeconomic variable.

In the second stage, the objective is to determine the main factors that push different individuals to spend part of their budget in tourism activities. Rather than estimate tourism

---

tourism demand.

expenditure in absolute terms we prefer to model percentage of tourism expenditure over income. This transformation offers more relevant results and lets us estimate how much people like tourism.

Third stage analyses the main factors that contribute to define travel frequency within a period of time. In this stage we discuss the role of length of stay in the frequency decisions. We conclude that, provided individuals or household posses some flexibility to choose the length of stay, then this will optimally be chosen depending on the destination chosen and the amount of the fixed costs incurred. For frequency decisions, the model suggested depends on the dispersion of the frequency. If frequency is equidispersed, Poisson process may be applied. However, if frequency is overdispersed, this may need to be modelled because it would respond to a case of unobservable heterogeneity. For this purpose, negative binomial model, zero-inflated model, hurdle model and mixture model are discussed.

In the fourth stage, we propose a tree structure which may classify different segments of tourists depending on their needs. Two criteria which define this classification are the physical attributes of the destination and the tourist environment wished. Methodology proposed in this stage is linked with last stage, such that both decisions are modelled within a nested multinomial logit. Last stage, corresponds to destination and transportation mode choice. We propose to include in tourist's choice set, every sensible combination between destination and transportation mode. Special care needs to be considered with the choice set definition for tourism analysis. In this sense, we remark three main aspects to be taken into account: criteria to define different alternatives, flexibility and size of the alternative destinations.

Finally, we end up with a complete methodological framework that disaggregates tourists' decisions and obtain the determinants for each of these decisions and allows us to simulate how current situation may change under alternative scenarios.

References

Bakkal, I. (1991). Characteristics of West German demand for international tourism in the northern Mediterranean region. Applied Economics, 23, 295-304.

Deaton, A. and Muellbauer, D. (1980a). An almost ideal demand system. American Economic Review, 70 (3), 312-326.

Deaton, A. and Muellbauer, D. (1980b). Economics and consumer behavior. Cambridge: Cambridge University Press.

Deb, P. and Trivedi, P.K. (1997). Demand for medical care by the elderly: A finite mixture approach. Journal of Applied Econometrics, 12, 313-336.

Divisekera, S. (2003). A model of demand for international tourism. Annals of Tourism Research, 30 (1), 31-49.

Eymann, A. and Ronning, G. (1997). Microeconometric models of tourists' destination choice. Regional Science and Urban Economics, 27, 735-761.

Greene, W. (2003). Econometric Analysis. Fifth edition. New Jersey: Prentice Hall.

Haab, T.C. and Hicks, R.L. (2000). Choice set considerations in models of recreation demand: History and current state of the art. Marine Resource Economics, 14, 271-281.

Hultkrantz, L. (1995). On determinants of Swedish recreational domestic and outbound travel, 1989-93. Tourism Economics, 1 (2), 119-145.

Maddala, G. (1983). Limited dependent and qualitative variables in econometrics. New York: Cambridge University Press.

Morley, C.L. (1995). Tourism demand: characteristics, segmentation and aggregation. Tourism Economics, 1 (4), 315-328.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. Journal of Applied Econometrics, 12, 337-350.

O´Hagan, J.W. and Harrison, M.J. (1984). Market shares of US tourist expenditure in Europe: an econometric analysis. Applied Economics, 16, 919-931.

Parsons, G.R. and Hauber, A.B. (1998). Spatial boundaries and choice set definition in a random utility model of recreation demand. Land Economics, 74 (1), 32-48.

Papatheodorou, A. (1999). The demand for international tourism in the Mediterranean region. Applied Economics, 31, 619-630.

Syriopoulos, T.C. and Sinclair, M.T. (1993). An econometric study of tourism demand: the AIDS model of US and European tourism in Mediterranean countries. Applied Economics, 25, 1541-1552.

Rugg, D. (1973). The choice of journey destination: A theoretical and empirical analysis. Review of Economics and Statistics, 55, 64-72.

Thill, J. (1992). Choice set formation for destination choice modelling. Progress in Human Geography, 16 (3), 361-382.

Train, K. (2003). Discrete choice methods with simulation. Cambridge: Cambridge University Press.

Table 1. Main variables of stage 5: Destination and transportation mode choice

Characteristics of the household	**Attributes of the destinations**	**Mixed variable**
• Disposable income • Budget for tourism expenditure • Disposable time for tourism • Labour conditions • Frequency of travelling • Size of the household • Age of the oldest and youngest members of the household • Education • Place of residence • Size of community: rural or city • Risk aversion • Party size of travellers	• Relative prices (PPP): Prices and exchange rate • Accommodation cost index • Weather • Safety • Crowding • Development and facilities • Size of the country	• Transportation cost • Travel time cost (mode of transportation) • Available information • Language • Suitability of destination (loss function) • Marketing in the country of origin