Modelling Choice of Information Sources

AGHA FAISAL HABIB PATHAN*, KAMRAN ANSARI*, AND ASHFAQUE AHMED MEMON*

RECEIVED ON 20.12.2012 ACCEPTED ON 20.03.2013

ABSTRACT

This paper addresses the significance of traveller information sources including mono-modal and multimodal websites for travel decisions. The research follows a decision paradigm developed earlier, involving an information acquisition process for travel choices, and identifies the abstract characteristics of new information sources that deserve further investigation (e.g. by incorporating these in models and studying their significance in model estimation). A Stated Preference experiment is developed and the utility functions are formulated by expanding the travellers' choice set to include different combinations of sources of information. In order to study the underlying choice mechanisms, the resulting variables are examined in models based on different behavioural strategies, including utility maximisation and minimising the regret associated with the foregone alternatives. This research confirmed that RRM (Random Regret Minimisation) Theory can fruitfully be used and can provide important insights for behavioural studies. The study also analyses the properties of travel planning websites and establishes a link between travel choices and the content, provenance, design, presence of advertisements, and presentation of information. The results indicate that travellers give particular credence to government-owned sources and put more importance on their own previous experiences than on any other single source of information. Information from multimodal websites is more influential than that on train-only websites. This in turn is more influential than information from friends, while information from coach-only websites is the least influential. A website with less search time, specific information on users' own criteria, and real time information is regarded as most attractive by the travellers.

Key Words: **Choice modelling, Travel information, Mixed Logit, Random Regret Model.**

1. INTRODUCTION

Traveller information has long been used to improve the traffic and network conditions. Many types of information services and products are already in the market and research continues to introduce new generations of these systems by using portable personal communication devices. An abundant body of literature is available on the impacts of new technologies since the last decade or so (see among many others [1-6]). Most of the studies focused on the usage of these ATIS (Advanced Traveller Information Systems) technologies and their impact on the behaviours of the people. There are

* Assistant Professor, Department of Civil Engineering, Mehran University of Engineering & Technology, Jamshoro.

substantial empirical studies that study the travellers' process of information acquisition and its effect on their travel decisions (especially [7-11]). Very few studies (notably [12-14]) actually focused on the process of information acquisition together with its effects simultaneously in one behavioural framework. But the studies carried out so far largely deal with only few information sources and the effect of one rather powerful and common source, i.e. the internet, is rather ignored. There have been few attempts to study information gain in an ICT (Information and Communication Technology) rich environment. There are certain gaps in the study of all sources considered in the choice set that goes beyond partial studies on the determinants of some en-route or pre-trip sources. Furthermore most of the studies were inclined towards the linking of above mentioned determinants with the route choice decisions. It is surprising to see that literature is lacking on studies that comprehensively link the various sources of information with the mode choice decisions of the travellers. There are numerous sources of information which provide information about the different modes of travel alternatives and it is important to study the influence of these information sources in the mode choice decisions of the travellers under various circumstances. Researchers in the past have done extensive research about the influence of different message contents upon travel decisions. However, the new information sources have different other attributes including their special design, advertisements, and graphics which have their own influence on the traveller and his travel decisions. Effects of these important determinants must be studied in order to improve the understanding of the influence of these sources on the overall modal shift. The issue of credibility as mentioned above was dealt with by a number of researchers. But the researchers focused on the reliability of the information for only the information source under study mostly VMS (Variable Message Signs) signs. There is a gap in the literature with respect to the variation of the credibility from one source to another. Moreover, the issue about the treatment of the sources with low credibility by travellers and what are the determinants that affect the credibility of the sources was also not fully understood in the past.

2. STRUCTURE OF THE PAPER

Following the introduction, Section 3 begins by discussing the issues raised by the literature review and identifying hypotheses to be tested. It discusses in detail the conceptual representation of information search and travel choice process. Based on this general framework a modelling framework is identified that simplifies the general conceptual framework into an estimable model. Two models are developed namely the Source Model and the Mode Model. The section then discusses a number of analytical issues including the choice of appropriate model specification to be estimated.

Section 4 briefly discusses the design of the questionnaire and includes the details of the SP (Stated Preference) experimental design and explains the choice of SP scenario, attributes, and levels for the selected variables for both source and mode choice experiments. It also describes the selection of the CATI (Computer Assisted Telephonic Interview) technique and the role of simulation and pilot studies for the development of the final SP exercise.

Section 5 details the calibration and estimation of RP (Revealed Preference) and SP models for information source choice. Section 5 then estimates combined RP and

SP models for source and mode. Four alternative model specifications are calibrated from the collected data. The four model specifications include the multinomial logit, the nested logit, the mixed logit and the random regret minimisation models. All the models were estimated to explain source choice decisions. The Jacknife method was applied to correct the SP MNL (Multinomial Logit) models for repeated measurement bias and a specification with random taste variation was applied to the mixed logit models.

In Section 6, the disaggregate choice models developed in Section 6 are applied to generate forecasts of information source. The forecasts are made on both the estimation sample and the expanded sample taking account of the NTS (National Travel Survey) 2006 long distance travellers' demographics.

Finally Section 7 provides the summary of the achievements of this work and identifies areas that would benefit from further research.

3. CONCEPTUAL REPRESENTATION OF INFORMATION SEARCH AND TRAVEL CHOICE

In the light of the discussion presented above, it is evident that information is required by travellers to make informed decisions about their travel. Not only this, provision of information actually influences people's decisions by informing them about various less tangible characteristics of different alternatives. The previous section also listed the important variables which should be included in any analysis/model in order to comprehend the importance of information, travel decisions and their inter dependence. These variables could be grouped in four categories: variables associated to decision makers, attributes associated with the information sources, characteristics of the travel options, and other external circumstances.

In order to understand the importance of these different variables, it is necessary to understand the information search and travel decision process together with its logical flow. The process of information acquisition for a traveller starts with her intention to travel. If, as an ideal case, the traveller has prior knowledge and experience, she is aware of every alternative and their attributes she may decide not to seek any further information. Similarly, if she is satisfied with the available information, she will put no value on obtaining additional information. In either case, the search process will not start at all and she decides about the travel alternatives straightaway. However, the decision to acquire information also depends on external circumstances, for example if she thinks that the network is not congested at all she may not see any reason to explore available information, but if, on the other hand, the weather conditions are poor, she may think it wise to seek further information. Secondly, it depends upon the individual traveller's preferences and attitudes. For example, if she habitually uses a particular route, this habit will influence her to put in less efforts of obtaining information. Thirdly it also depends upon personal circumstances and individual preferences/attitudes about the traffic and driving tasks, all of which may vary with personal characteristics.

The above mentioned ideal situation is unrealistic; one deviation could be that the traveller is not aware of all available travel alternatives and hence if she decides to acquire information the aim of that information will be alternative generation, similarly if a traveller feels that she

has incomplete information about the characteristics of the available alternatives, she may desire additional information about those alternatives. Here it must be noted that the randomness is not only in the environment but also there is uncertainty in the perception of the traveller about the non-stationary environment.

Once the traveller decides to acquire information and recognizes its value, she tries to search available potential sources of the information. From these sources she will decide to acquire information from a particular source depending upon awareness, its credibility, accessibility, information source attributes, and her own degree of understanding of the available sources. Awareness corresponds to the possibility of knowing the presence of a particular source and degree of success in understanding the information offered by that source. The traveller considers the information credible if it is provided by a trusted and reliable source, and also based on her experience with that particular source, her assessment of the likelihood of it meeting her needs, her trust in the motives of the information provider and the degree of its authentication by other reliable sources. The accessibility of a source depends on its cost and the ease with which it can be accessed and used. After gathering new information from a selected source the traveller tries to understand the acquired knowledge and rethinks the credibility of the new information and processes it as per her needs. She also synthesizes the new information with her beliefs and experiences prior to acquisition. If the acquired information satisfies her aspiration she will use this information in her travel decisions. On the other hand, if she is happy with the source of information but still require more information; she will try to acquire it from that source if the cost of acquisition is less than the disutility of delay and uncertainty. Alternatively, if she is not happy with the information source, she will select another source and will acquire information from it.

Finally, the decision is made on the basis of the acquired information and perceived modal and trip attributes. If user is satisfied with the decision, she forms beliefs and add/update her experiences about her travel and this particular information source. The process of information acquisition is summarised in Fig. 1.

From the understanding of the above conceptual framework and in the light of the literature review it can be deduced that the decision to acquire travel information depends on the external circumstances e.g. bad weather, congestion, incident etc. personal attitudes and preferences; and personal circumstances. Similarly, once a decision is made to acquire information, the choice of source depends upon its accessibility and credibility, the individuals' awareness of it and information source characteristics. The attributes are thus classified in four categories: information source attributes; respondent characteristics; scenario attributes; and travel mode characteristics.

The information source attributes are source owner, search time (time spent on search for required information), frequency of updating of information, type of information presented, presence of advertisements, type of information source, presentation of information (attractiveness of design of webpage), capabilities (how many probes used, CCTV camera installed, etc), price and subscription, spending on advertisements, coverage in search engines, year of start, coverage in newspapers and articles. The scenario attributes include trip purpose, bad weather, congestion, incident occurrence, and accompanying

travellers. The personal characteristics include personal segmentation, socio-economic characteristics, trip frequency, travel time, frequency of using source, attitudes to optimise the journey, habit and other individual characteristics.

In addition to the above mentioned variables associated with the information acquisition process, the literature also suggests that there are characteristics which are associated with the travel alternatives. These include travel time, travel cost for different alternatives, safety,

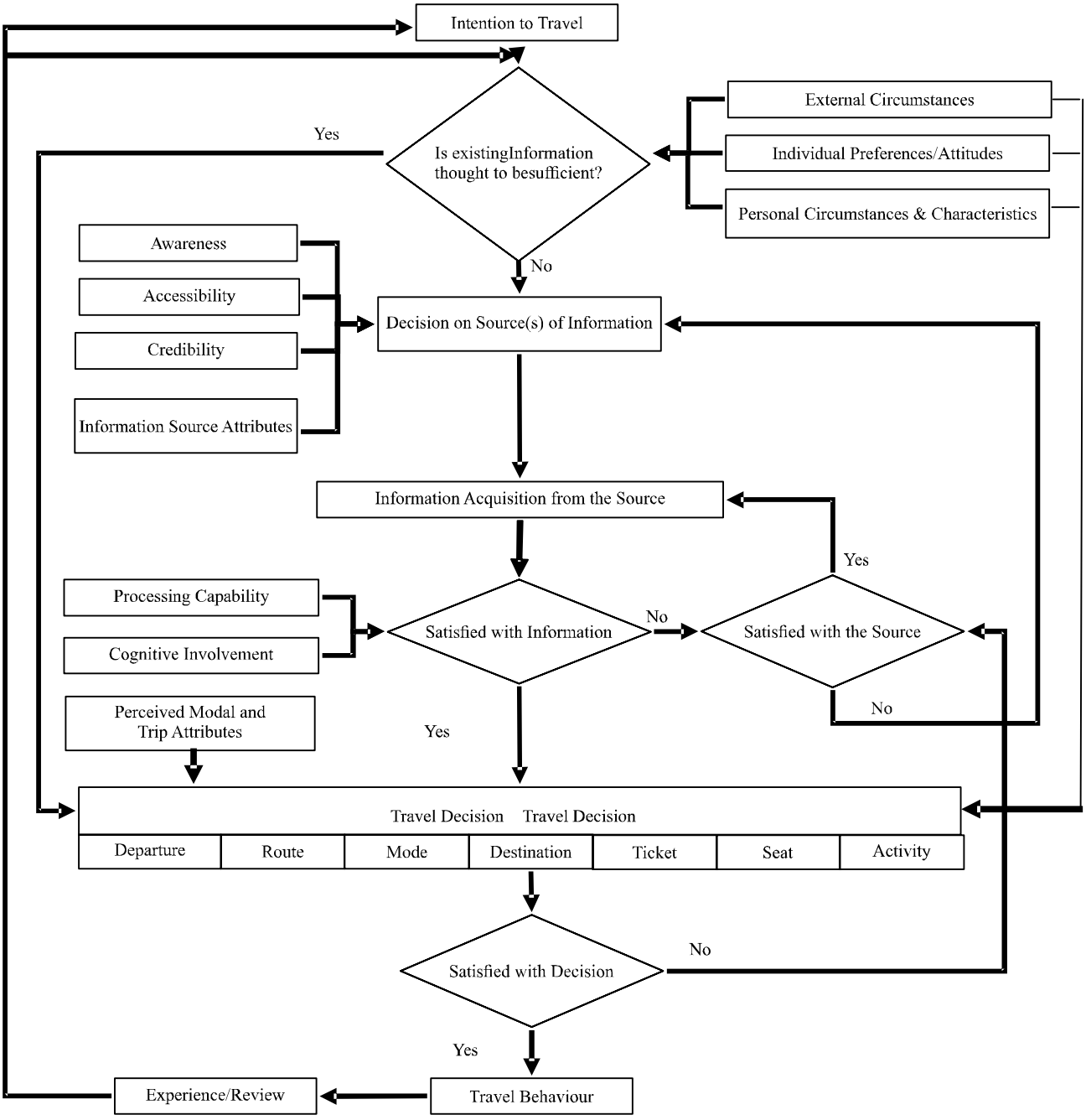

FIG. 1. CONCEPTUAL REPRESENTATION OF INFORMATION SEARCH AND TRAVEL CHOICE PROCESS

reliability, security and convenience. The details of these variables and the justification for their inclusion are also discussed in detail in the next section. The travel decisions as mentioned in Fig. 1 could be various i.e. choice of departure time, route, mode destinations, purchase of tickets, seat reservations and even cancellation/amendment of the activity itself. The study of all these travel decisions in one framework is quite difficult if not impossible and requires extensive data and resources. Hence, in this study, only mode choice decisions are taken into account.

In the light of the above discussion, the following research questions emerge to aid better understanding of the use of present day information sources.

- How does initial information effect subsequent search?
- Does credibility vary from source to source?
- How is information with low credibility treated?
- What affects credibility?
- Does presentation of information affect choice behaviour?
- What is the influence of information sources on the mode choice decisions?
- Does the design and content of an information source affect choice decisions?

4. DATA COLLECTION

The questionnaire developed in this study has four parts. The first part gathers RP data for the last long journey (over 50 miles) made by the respondents. The second part includes SP survey questions to investigate travellers' choice of information sources and their subsequent mode choice when making long journeys. The third part includes general questions about their attitudes towards different sources of information and on their normal search patterns. The final part contains questions about travellers' characteristics. The questionnaire could not be paper based because complex branching/routing is required for different segments of the respondents and the questionnaire needs to be interactive such that the appearance of a question sometimes depends on the previous response of the respondent. Hence the decision was made to recruit people at stations and arrange to interview them by telephone using CATI software to facilitate branching & routing in which the presence of an interviewer can reduce errors and the chance of the loss of a valuable respondent is reduced).

4.1 Non SP Content

In the RP part of the questionnaire the data was gathered for the last long journey (over 50 miles) made by the respondents. An account of the journey was obtained by asking numerous questions including their frequency of travel to those destinations, purpose of visits and the chosen mode. In order to reconstruct the choice set, questions were asked regarding respondents' perception of attributes of different modes available to them when they were taking the travel decisions. The questionnaire also included questions regarding the external circumstances of the journey and about the use and effect of any information source used while taking travel decisions about that journey. Other non SP questions explored respondents' attitudes towards different sources of information, their normal search patterns and personal characteristics.

4.2 SP Content

There are two SP exercises in the second part of the questionnaire. The first deals with the choice of information source with different attributes. A number of alternative scenarios varying the trip destination, purpose and accompanying individuals were tested in the pilot surveys. The most influential scenario(s) were selected. A fractional factorial design was used with the fraction selected to avoid dominance (combinations in which the respondent is better off or worse off on every dimension). However, since dominated options can be used to identify non-logical responses, each SP exercise included one dominated option to check the consistency of responses. In the interests of simplicity, it was decided to present a binary choice regarding choice of information source.

4.2.1 Representation of the Scenarios

This research focuses on the role and use of information sources in long distance journeys. Hence in this research the scenario was a long journey between cities. Although it would have been easy to present respondents with a hypothetical journey without mentioning the destination, but if the use of information sources would have been unrealistic and respondents would be unable to conceptualise a need for specific information. Another option was to offer a journey from Leeds to London since this is a long journey and requires acquisition of information for a variety of conditions. However, the problem with this scenario would have been that this journey is a very common and so the choice of mode and information source would be highly influenced by the respondents' previous journeys to London and by their use (or otherwise) of information to assist their mode choice for that journey. A journey from Leeds to Cardiff was therefore selected thus avoiding all of the above mentioned problems. There were two scenarios selected for the SP exercise, one for a trip for personal business the other for a leisure trip. Hence, the first scenario for both the SP exercises stated that "imagine that you need to travel from Leeds to Cardiff in 3 days time on personal business with an appointment from 1130 till 1230 in Cardiff city hall". The scenario is not varied within one SP design, rather the scenario attributes are explored by adding questions in the RP part of the questionnaire. The other scenario for a separate set of respondents to explore the effect of information sources on leisure trips stated that "imagine that you need to travel from Leeds to Cardiff in 3 days time on a leisure trip with two of your friends and you need to arrive there at 1130".

4.2.2 Attributes and Levels of SP Design

SP Exercise for Selection of Information Source: The attributes are classified in three categories information source attributes; respondent characteristics; and scenario attributes. The following attributes were included in the pilot survey: ownership of the source (Government or Private); type of information; search time; presence of advertisements; updating of information; and subscription cost. The final attributes of the SP design were decided after the execution of the pilot surveys.

A source of information could be either owned by government or private company. This may influence the credibility of the source and might be an important consideration in the selection of the sources. For type of information about travel time and delay, three levels are considered: Descriptive Quantitative real time information (e.g. specific information relevant to the journey about routes/services/prices), Descriptive

Qualitative information (e.g. general information relevant to the journey about routes/services/prices), and Prescriptive information (e.g. specific information relevant to the journey based on users' criteria for routes/services/prices). For the search time three levels are considered, 5, 10 and 15 minutes. These levels were decided on the argument that nobody would like to spend more than 15 minutes for search of the information source and will lose his interest on it. On the other hand, 5 minutes was anticipated to be considered as a reasonable minimum search time (including any associated typing and browsing time). For the presence of advertisement there can only be two possibilities; either the advertisements are there or they are not. Hence two levels are considered; advertising present and no advertising. For frequency of information changes on the source three levels are considered; real time, daily and weekly. Although other frequencies (e.g. hourly) could have been specified, the argument for selecting only these three levels was that the design should be as simple as possible and that the difference between, say, hourly and real time, is not as great as that between the three frequencies identified. For subscription there are two levels; no subscription and £5 already paid. The value of £5 is selected here because any higher value would probably be considered too high for an infrequent trip of the type mentioned in our scenario.

This results in a design with 3 attributes with 2 levels and 3 attributes with 3 levels. A full factorial design would require 216 combinations of attribute levels. However, since the attributes are, arguably, not truly independent (e.g. advertising is less likely on a government site and on a subscription site than on a free commercial site), it is not necessary or appropriate to use the "full" design. The attributes of owner of source, subscription and advertisements are therefore combined together as a one attribute with four levels:

- A government source with no advertisements and no subscription.
- A private source with advertisements and no subscription.
- A private source with no advertisements but with subscription.
- A private source with no advertisements and no subscription.

The revised design thus has one attribute with 4 levels and three attributes with 3 levels. A full factorial design in this case would require 108 combinations of attribute levels. This number of combinations is still too high to be used and hence an orthogonal main effects fraction is used having 16 combinations. This means that every respondent gets 8 binary choices in the source choice SP. This allows the uncorrelated estimation of all the main effects under the assumption that all interactions are negligible.

4.3 Main Survey

Respondents were recruited at the main long distance transport interchanges in Leeds (i.e. the Coach Station and the Train Station). It was originally hoped that recruitment would also be possible at locations likely to be frequented by long distance drivers but this proved unpractical (see later for a discussion of the potential consequences of this outcome). The recruitment interview (conducted at the rail and coach station 150 to be recruited at each location) was brief and sought only their willingness for participation, whether they make journeys

over 50 miles for leisure or personal business, a telephone number and time at which it would be convenient to conduct the telephone interviews. Recruits were subsequently interviewed by CATI at a mutually convenient time. About 950 members of the public were recruited to achieve the target sample of 300 completed interviews. Table 1 shows the distribution of socio-economic characteristics of the samples and compares them with the Census data (HMSO-2001) for Leeds residents and NTS data of long distance travellers. The sample was not representative of the typical UK long distance travellers due to the fact that the respondents were recruited at public transport centres and thus the captive car travellers were under represented. However, car owners recruited were more than the typical for Leeds. This modal bias was compensated by expanding the sample for forecasts to make it representative of the long distance travellers. Moreover, it was also found that these variables (i.e. car ownership, income etc.) were not significant contributors of the long distance information seekers.

5. MODELLING CHOICE OF INFORMATION SOURCE

Previous sections discussed the important issues that should be included in any study focused on the travel information acquisition and travel decisions. In particular, Section 3 discussed the general conceptual framework of the information search process and subsequent travel decisions. The inclusion of all those determinants and choices in a single study is not practical and would be too ambitious. Hence in this study the choices are simplified and restricted to only two i.e. the choice of information source and the choice of mode. The selections of the determinants for these choices are also simplified and only the most relevant attributes are selected. Fig. 2 presents the simplified modelling framework conducted in this thesis. The choices are divided in two different sections, namely choice of information source and choice of mode. The two choices are calibrated in separate models using different data sets and are combined at a later stage.

The first choice is the selection of the source of information. The determinants for this choice are classified in three categories: Personal Characteristics and Attitudes including experience of sources; Journey Characteristics; and Source attributes. Fig. 2 presents these independent variables in detail. In this study MNL and NL (Nested Logit) models were calibrated from the data whereas, out of these models, only selected best performing models are estimated in the MML (Multinomial Mixed Logti) and RRM (Random Regret Minimisation) framework to check their robustness.

5.1 RP Model for Source Choice

Multinomial models were estimated for the choice of mode in RP data. As mentioned earlier the data consists of respondents' past behaviour for the last long journey (over 50 miles). Variables used in the model were selected as discussed in the previous chapter and were added in the model incrementally. A MNL (Multinomial Logit Model) is constructed with the dependent variable being the choice among different important information sources as defined from the data i.e. onlyweb, friend+web, map+web, map+friend+web and onlymap. The choice set C_n of each individual thus consists of five alternatives. The utility functions are given by:

$$U_{(onlyweb)} = LFreq_1 * DLFreq_1 + GWeather_1 * DGWeather_1 + PeakP_1 * DPeakP_1 \quad (1)$$

$$U_{(friweb)} = LFreq_2 * DLFreq_2 + GWeather_2 * DGWeather_2 + PeakP_2 * DPeakP_2 + ASC_2 \quad (2)$$

$$U_{(mapweb)} = LFreq_3 * DLFreq_3 + GWeather_3 * DGWeather_3 + PeakP_3 * DPeakP_3 + ASC_3 \quad (3)$$

$$U_{(mpfrweb)} = LFreq_4 * DLFreq_4 + GWeather_4 * DGWeather_4 + PeakP_4 * DPeakP_4 + ASC_4 \quad (4)$$

$$U_{(onlymap)} = ASC_5 \quad (5)$$

The MNL estimates are shown in Table 2 whereas details of variables and associated coefficients used in all models are defined Annexure-I.

TABLE 1. DISTRIBUTION OF SOCIO-ECONOMICS

Characteristics	Survey		Census 2001 for Leeds	NTS-2006 (For Long Distance Travellers)
	Frequency	Percent		
Gender				
Female	108	39	52	39
Male	170	61	48	61
Total	278	100	100	100
Age				
Less than 20	6	2	25	
20-29	78	29	15	-
30-39	116	43	15	-
40-49	52	19	14	-
50-64	8	3	16	-
65 and Over	4	1	15	-
Total	270	100	100	-
Occupation				
Employed Part Time	12	4	-	12
Retired	2	1	12	13
Student	76	28	11	3
Looking After Home/Housewife	6	2	5	4
Without Work	2	1	13	1
Full Time Employed	170	63	59	66
Total	268	100	100	100
Car Ownership				
0	63	24	34	8
1	100	38	42	34
2 or More	102	38	24	58
Total	268	100	100	100
Household Size				
1	12	5	32	-
2	88	33	33	-
More than 2	164	62	35	-
Total	264	100	100	-
Income				
£20,000 or Less	94	34	-	18
£20,001-40,000	152	54	-	30
Over £40,000	12	5	-	52
No Response	20	7	-	-
Total	278	100	-	100

The alternative specific constants in the models reflect the average effects of variables which are omitted in the model. The negative sign for alternative specific constant in the model for friend + web implies that if all else being equal, travellers that are website users prefer more in the information provided by the website alone as compared to the information provided by friends. This suggests that travellers consider websites as reliable source of travel information. On the other hand travellers that use websites are more satisfied with the information provided to them as compared to the information provided by the maps only. This suggests that users exhibit an inclination to gather information from the websites. The analysis suggests that the frequent travellers require, in addition to website, the information from a person who has travelled before. The sign of PeakP is negative which suggests that Peak period travellers do like to stay with the maps only. This implies that in peak periods, travellers are more likely to take diversions from the usual route and thus require maps while they travel.

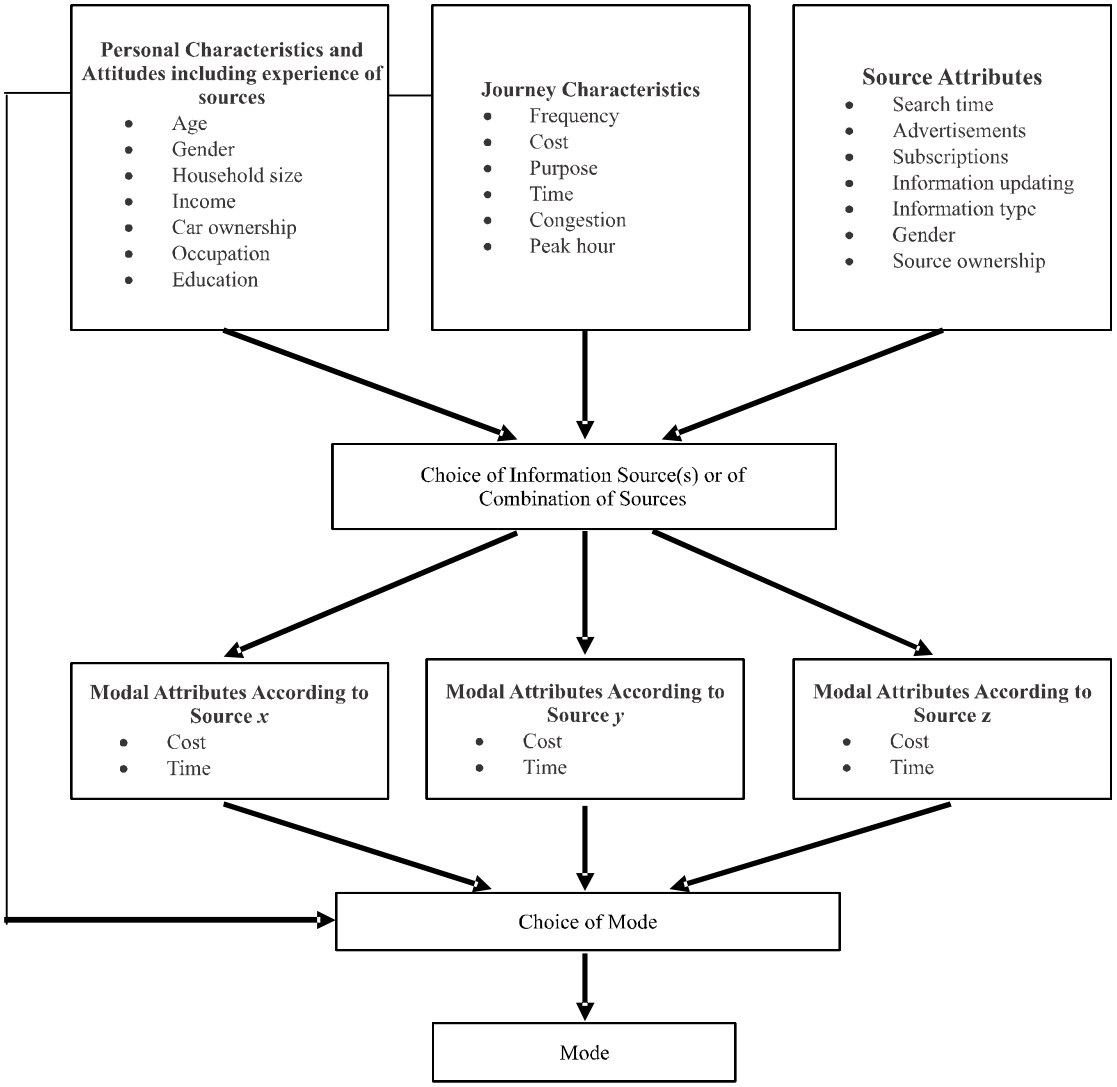

FIG. 2. FRAMEWORK FOR MODELLING CONDUCTED IN THIS RESEARCH

5.2 SP Model for Source Choice

Respondents were given two websites with different attributes as mentioned in Section 4. Multinomial Logit models were calibrated for the choice of website as source to analyse different important attributes and their influence on travellers. Variables used in the model were selected as discussed in Section 5, and were added in the model incrementally. A MNL was first constructed with the dependent variable being the choice among website one and website two. The base case for this model is website 1. The choice set C_n of each individual thus consists of two alternatives. The utility functions are given by:

$$U_1 = Seatime_1 * Seatime_1 + ComAdsub_1 * DComAdsub_1 + Com1 * DCom_1 + InfoSpec_1 * DInfoSpec_1 + InfoPersc_1 * DInfoPersc_1 + Updaily_1 * DUpdaily_1 \quad (6)$$

TABLE 2. RP SOURCE MODEL ESTIMATES (T-RATIOS IN BRACKETS)

Coefficients	RP Source Model (Bold Values Show 10% Significance)
$LFreq_1$	0.561(0.7)
$GWeather_1$	-0.465(-0.4)
$PeakP_1$	-1.27(-1.4)
$LFreq_2$	2.89(2.7)
$GWeather_2$	0.316(0.3)
$PeakP_2$	-0.732(-0.8)
$LFreq_3$	0.157(0.2)
$GWeather_3$	-0.390(-0.3)
$PeakP_3$	-1.02(-1.0)
$LFreq_4$	1.87(1.4)
$GWeather_4$	10.7(0.1)
$PeakP_4$	-2.44(-1.8)
ASC_2	-3.42(-3.4)
ASC_3	-1.35(-1.6)
ASC_4	-13.7(-0.1)
ASC_5	-2.96(-2.1)
Observations	121
Final log (L)	-145.9
DOF	16
$Rho^2(0)$	0.269
$Rho^2(c)$	0.088

$$U_2 = Seatime_2 * Seatime_2 + ComAdsub_2 * DComAdsub_2 + Com_2 * DCom_2 + InfoSpec_2 * DInfoSpec_2 + InfoPersc_2 * DInfoPersc_2 + Updaily_2 * DUpdaily_2 \quad (7)$$

Model estimates are shown in Table 3. As expected, search time for the website was found significant and negative which shows that, other things being equal, travellers will prefer an alternative which offers the lowest expected search time. Information type i.e. specific information available and information w.r.t own criteria were also found significant and positive which means more information increases the utility of a website. Updating of information was also as per priori expectation and negative which tells that if in comparison to real time information updating, daily and weekly have a negative effect on the utility. Commercial Ads No Sub; Commercial Ads Subs and Commercial No Ads No Sub were also found negative which means the base case i.e. government with no advertisements and no subscription has higher attraction and credibility within the respondents. Gender, full time employment and higher income and younger travellers were found insignificant - as were variables interacting with these dummies. This shows that the importance of information does not depend on the socioeconomic characteristics of the respondents.

5.3 Correction of Repeated Measurements

Although the model seems to be quite robust, it still has to be corrected for repeated measurement problems. In this study the Jack-knife method is selected to correct the repeated measurement problems inherent in the SP data. To undertake a Jack-knife, the sample is divided into n groups of individuals making n Jack-knife sub-samples, where each sub-sample is formed by deleting one of the groups from the sample. The number of sub-

samples is important in Jack-knife implementation because it improves the power of significance tests and make variance stable. The ideal number of sub-samples is the number of sample respondents (i.e. r=n). The program Alogit only allows the number of subsamples between 2 and 99. In this study, six models were estimated each with a different number of randomly selected sub-samples (5, 10, 20, 40, 60 and 90). The Jack-knife estimates show that, regardless of the number of sub-samples most coefficients on the Jack-knife method are very close to those of the uncorrected model estimates. These results show that the coefficients of the uncorrected model estimate were quite accurate despite of the repeated measurement problem. However the t ratios are slightly reduced in case of the Jack-knife method which indicates that the uncorrected model slightly overestimated the significance of the parameters, Jacknifed model estimates are shown in Table 3.

5.4 Mixed Logit Model

The SP MNL model was taken from previous section for the calibration of MMNL for SP mode choice model. This model has similar specification to the MNL but following notable differences.

Firstly, normally distributed coefficients were specified for search time to explain additional unobserved variation in the model (MMNL in Table 3). Secondly, in order to accommodate the repeated measurement problem, the search time was also identified to vary across individuals and not across observation per individual (Panel MMNL in Table 3) This was done by using Biogeme's panel data specifications (for details refer Biogeme manual).

The utility function is as follows:

$$U_1 = Seatime_1 * Seatime_1[Sigma1] + ComAd_1 * DComAd_1 + Com_1 * DCom_1 + InfoSpec_1 * DInfoSpec_1 + InfoPersc_1 * DInfoPersc_1 + Updaily_1 * DUpdaily_1 \quad (8)$$

TABLE 3. COMPARISON OF ESTIMATES BETWEEN MNL, JACK-KNIFED METHOD AND MIXED LOGIT SOURCE CHOICE MODEL (T-RATIOS IN BRACKETS AND BOLD VALUES SHOW 10% SIGNIFICANCE)

Coefficients	SP Source MNL	Jacknifed	MMNL	Panel MMNL
Seatime1	-0.194(-7.5)	-0.194(-7.9)		
Mean of Seatime1			-0.192(-8.0)	-0.202(-10.7)
SD of Seatime1			-0.0061(-0.3)	0.0116(0.7)
$infoSpec_1$	1.63(4.9)	1.64(4.2)	1.60(4.4)	1.70(5.4)
$infoPers_1$	3.91(5.8)	3.60(4.9)	3.85(5.9)	1.92(4.7)
$ComAd_1$	-1.27(-5.4)	-1.26(-5.0)	-1.26(-4.9)	-1.31(-6.5)
Com_1	3.55(8.1)	3.52(7.4)	3.53(8.2)	3.58(8.5)
$Updaily_1$	-1.79(-4.7)	-1.79(-4.7)	-1.75(-4.4)	-1.88(-6.3)
$infoSpec_2$	-0.619(-1.4)	-0.623(-1.6)	-0.565(-1.4)	-0.758(-2.6)
$infoPers_2$	0.288(0.5)	-0.0304(-0.1)	0.322(0.5)	-1.92(-6.4)
$ComAdSub_2$	0.227(0.5)	0.255(0.5)	0.170(0.4)	0.326(0.8)
Com_2	2.30(4.2)	2.32(4.9)	2.24(4.9)	2.47(7.1)
$Updaily_2$	-1.05(-3.7)	-1.06(-4.2)	-1.02(-3.9)	-1.13(5.1)
Observations	1056	1056	1056	1056
Final log (L)	-503.8	-503.8	-503.7	-503.8
DOF	11	11	12	12
$Rho^2(0)$	0.312	0.312	0.312	0.312
$Rho^2(c)$	0.312	0.312	0.312	0.312

$$U_2 = Seatime_2 * Seatime_2[Sigma1] + ComAdsub_2 * DComAdsub_2 + Com_2 * DCom_2 + InfoSpec_2 * DInfoSpec_2 + InfoPersc_2 * DInfoPersc_2 + Updaily_2 * DUpdaily_2 \quad (9)$$

The Model estimates are shown in Table 3 whereas details of variables and associated coefficients used in all models are defined Appendix-I. The results above show that the mixed logit calibration of the model increased the efficiency of the model with increased t stats. The inclusion of the error component terms improves the fit to the data, but only marginally. The results confirm the findings of the MNL model. The standard deviations of the normally distributed components are not significant in both MMNL and Panel MMNL this suggests that the MNL model specification represents the choice of the respondents adequately.

5.5 Random Regret Minimisation Estimations

The RRM model postulates that people aim to minimise their regret with respect to the foregone alternatives [15]. The regret based models are based on the theory that individuals minimise anticipated regret when given a choice instead of maximising anticipated utility. The RRM estimation procedures adopted in this section uses a specification suggested by Chorus, et. al. [15] which assumes that the regret is experienced with respect to all foregone alternatives that perform better than a chosen/ intended alternative in terms of one or more alternatives unlike previous assumptions that regret is only experienced with respect to the best of foregone alternatives. The specification produces intuitive estimation outcomes and satisfactory fit with available data [15].

The SP Model for Source Choice was thus formulised in the RRM paradigm. Although when the choice sets are binary, RRM reduces to RUM's (Random Utility Maximisation's) linear additive binary logit model, the model was still calibrated to validate the applicability and performance of the RRM formulation. The results showed that the RRM model is almost equal to the MNL model which confirms the appropriateness of the RRM model.

The utility functions are as follows:

$$R_1 = ln(1+e^{Seatime1*(Seatime2-Seatime1)}) + ln(1+e^{ComAd1*(DComAd2-DComAd1)}) + ln(1+e^{Com1*(DCom2-DCom1)}) + ln(1+e^{InfoSpec1*(DInfoSpec2-DInfoSpec1)}) + ln(1+e^{InfoPersc1*(DInfoPersc2-DInfoPersc1)}) + ln(1+e^{Updaily1*(DUpdaily2-DUpdaily1)}) \quad (10)$$

$$R_2 = ln(1+e^{Seatime1*(Seatime1-Seatime2)}) + ln(1+e^{ComAdSub2*(DComAdSub1-DComAdSub2)}) + ln(1+e^{Com2*(DCom1-DCom2)}) + ln(1+e^{InfoSpec2*(DInfoSpec1-DinfoSpec2)}) + ln(1+e^{InfoPersc2*(DInfoPersc1-DInfoPersc2)}) + ln(1+e^{Updaily2*(DUpdaily1-DUpdaily2)}) \quad (11)$$

The results from the Table 4 show that the RRM model is almost equal to the MNL model which confirms the appropriateness of the RRM model. Although all the variables were first tried inside the Regret function, the resulting model lost its significance even on the important explanatory variables; hence only search time was tried in the regret function (presented in Table 4). The above table also shows that the signs of the dummy variables (outside the regret function) are reverse as compared to the MNL model, this confirms that the model shows the regret of an individual not a utility as in the other models.

6. FORECASTING

The Jacknifed SP model for source choice developed in section 5 of the previous section was used to predict the choice probabilities for the source forecasts based on the estimation sample. A model based on SP data alone cannot be used to forecast market shares. This requires additional support from observed (RP) data. The effects, calculated using the SP only model, are based on un-calibrated models

and as such the numerical magnitudes are only valid in comparisons across models. These models cannot be used to forecast market shares without calibration using revealed preference shares. There was no common variable between the SP source models and other RP and SP models and hence it was not possible to combine both data sets. However, the forecasting has been carried out on the SP model on the understanding that model was calibrated on the same sample and search time had the same value of time as that of other RP models.

The tests carried out on this model included the analysis of search time, type of the information provided, and whether the sources are commercial requiring subscriptions and with advertisements. These policy variables were selected because they were found significant in the model estimations and needed further analysis. Fig. 3 shows the forecasts for different values of the search time compared to the original models for the estimation sample. The methodology employed in this exercise involves forecasting of use of information source (website) with different values of the search time and resulting market share, everything else remaining same. For each respondent in the sample, a forecast is made for the probability that it will choose website A or website B as presented to them in the SP exercise. The choice probabilities for each respondent are then aggregated to generate forecasts. The first curve represents various search times for LHS website and corresponding market share keeping the RHS website with the levels as presented in the SP exercise whereas second curve shows search times for LHS website and corresponding market share by restricting the levels of search time for RHS website as 30 minutes. Both curves show slightly nonlinear trends. The slope of the first curve increases as the search time decreases, the market shares increase from 5 % at 30 minutes of search time to 79 % at search time of 1 minute. This shows the increase

TABLE 4. COMPARISON OF ESTIMATES BETWEEN UNCORRECTED METHOD, JACK-KNIFED AND REGRET SOURCE CHOICE MODELS (T-RATIOS IN BRACKETS AND BOLD VALUES SHOW 10% SIGNIFICANCE)

Coefficients	ModelSPS6	Jacknifed	RRM
$Seatime_1$	-0.194(-7.5)	-0.194(-7.9)	- 0.194(-8.3)
$infoSpec_1$	1.63(4.9)	1.64(4.2)	-1.63(-4.5)
$infoPersc_1$	3.91(5.8)	3.60(4.9)	-3.91(-5.9)
$ComAdSub_1$	-1.27(-5.4)	-1.26(-5.0)	1.27(4.9)
Com_1	3.55(8.1)	3.52(7.4)	- 3.55(-8.2)
$Updily_1$	-1.79(-4.7)	-1.79(-4.7)	1.79(4.5)
$infoSpec_2$	-0.619(-1.4)	-0.623(-1.6)	0.618(1.5)
$infoPersc_2$	0.288(0.5)	-0.0304(-0.1)	- 0.291(-0.5)
$ComAdSub_2$	0.227(0.5)	0.255(0.5)	-0.227(-0.5)
Com_2	2.30(4.2)	2.32(4.9)	-2.30(-4.9)
$Updaily_2$	-1.05(-3.7)	-1.06(-4.2)	1.05(3.9)
Observations	1056	1056	1056
Final log (L)	-503.8	-503.8	-503.8
DOF	11	11	11
$Rho^2(0)$	0.312	0.312	0.312
$Rho^2(c)$	0.312	0.312	0.312

in the importance of search time with every minute saved in the search. The trend is somehow reverse for the curve B where search time for website B is restricted to be 30 minutes for every case. Here the curve is concave in contrast to the first curve which suggests that as search time of website A approaches that of website B, the effect of unit change increases. Similarly Fig. 4 shows the forecasts for different values of the search time compared to the original models on expanded sample which reflects the demographics of the long distance travellers as revealed by the NTS. As expected, the figure reveals that the search time is very important characteristic of a website. Reduced search time attracts more people to use the website as an information source. Fig. 4 reveals that the adjusted sample reflects even more people to use websites as compared to the previous forecast of the estimation sample and the market shares increase from 5% at 30 minutes of search time to 97% at search time of 1 minute. This shows additional 18% users as compared to the previous forecast. This confirms that the adjustment of the sample with the higher income people increases the use of website as higher educated people are at ease with the websites as an information source. This also gives an interesting insight to the website owners and transport planning authorities to attract more users to implement their transport policy by putting more efforts in reducing the search time of the information.

Another important variable is the presentation of the information about the journey. Table 5 explains the importance of the information and its effect on market share of the website. It reveals that the choice of type of information is also related to the search time of the websites. The result in normal font relate to the estimation sample while those in italics relate to an expanded sample which reflects the demographics of the long distance travellers as revealed by the NTS. The Table 5 reveals that, on estimation sample, If the website provides information as per the traveller's criteria, the market share of the website raises from 41% when no information is provided to 63%. But when the

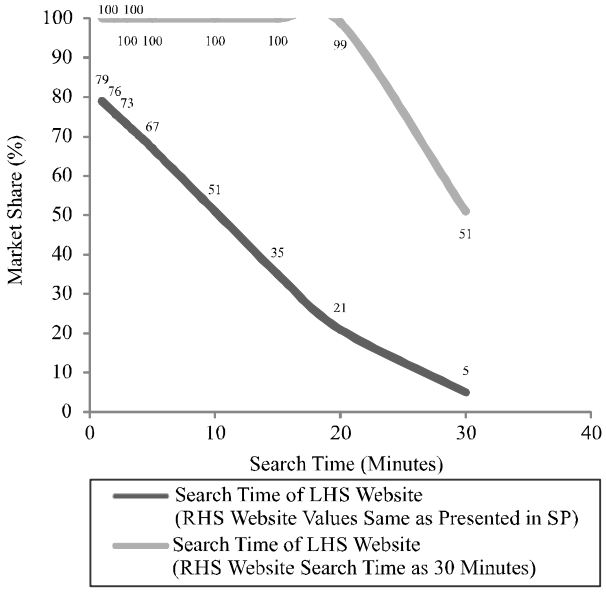

FIG. 3. SEARCH TIME AND CORRESPONDING MARKET SHARE ON ESTIMATION SAMPLE

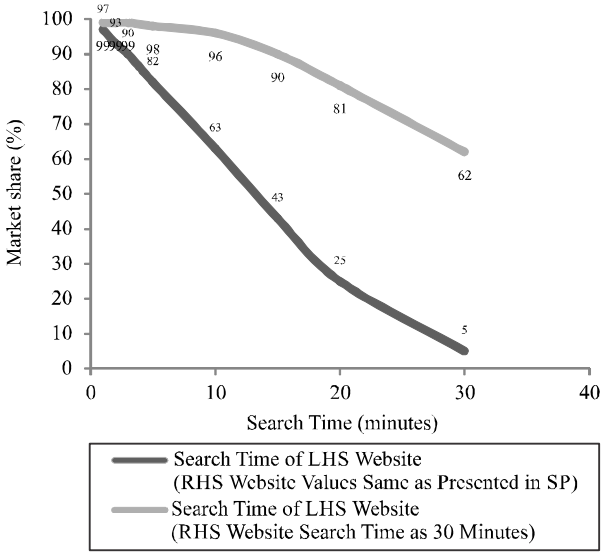

FIG. 4. SEARCH TIME AND CORRESPONDING MARKET SHARE ON EXPANDED SAMPLE

search time of the competing website is increased the importance of specific information reduces. Similarly prescriptive information also has the same tendency. This could be due to the complete reliance of the decision maker on the website as there is no competition between the websites and the selected website is much better and credible. On the other hand, on expanded sample, it reveals that if the website provides information as per criteria of the traveller, the market share of the website raises from 51% when no information provided to 77% and even more people are attracted if the search time of the competitive website is increased and the market share reaches 94%. On the other hand, prescriptive information also has a large effect on the market share of a website and it even goes to 100% of the market . This shows that with the increase of the higher income share in the population the credibility of websites increases and more travellers believe in the information prescribed by the website. This again is a very interesting result for the policy makers as they can prescribe different modes to distribute the travellers to obtain maximum performance of the available resources.

Table 6 reveals that, when a commercial website has advertisements on it, its market share decreases from 60-41%. On the other hand, if a commercial website has no advertisements and no subscription its share increases to 92%. Updating of the information also plays an important part in this phenomenon. And the real time information updating increase the share of a website from 38-63% when the information was updated only once daily. The italics figures in Table 6 reveals that for the expanded sample, when any commercial websites has advertisements on it, its market share decreases from original 74-48%. Whereas, if a commercial website has no advertisements and no subscription its share increases to 100%. This shows that the irrelevant information or subscription costs are, by large, not popular in the website users.

7. CONCLUSION

The above mentioned results and conclusions from the selected models give detailed insights for the choice of sources to make travel decisions. The results help to understand UK long distance travellers' behaviour in selecting the information sources and modes. The study found out that the previous experience or initial information significantly affects the utility of the selected mode. The selected Model ModelSPM16 suggests that UK travellers with a previous experience with a mode give more importance to it as compared to other sources. The model suggests that the traveller

TABLE 5. EFFECT OF TYPE OF INFORMATION ON MARKET SHARE (ESTIMATION SAMPLE AND EXPANDED SAMPLE)

Attribute Level	Market Share of Website A (%)			
	Search Time as Presented in SP		Website B Search Time 15 Minutes	
Information as per Criteria of the Traveller (InfoSpec) No Information Provided	41	*51*	54	*66*
As Presented in SP	51	*62*	67	*82*
Information Provided	63	*77*	76	*94*
Prescriptive Information (InfoPerc)				
No Information Provided	32	*39*	47	*58*
As Presented in SP 51	62	*67*	82	
Information Provided	86	*100*	93	*100*

anticipate more credence with previously experienced sources then with any other source. The model also suggests that Credibility does vary from source to source. It was found that UK travellers give varying degrees of importance to different information sources. It was found that travellers give positive credence to government owned sources. Similarly travellers give most importance to their own previous experiences followed by multimodal websites, train websites, friends and coach websites respectively. Moreover, the credibility of an information source increases significantly when it was correlated by another. UK Travellers try to gather more information from another source if they find that the information provided by a particular source is less credible. It was also found that a number of factors affect credibility of a source, they include past experience of the UK traveller with that source/mode, ownership of the source, and presence of advertising or irrelevant information. It was also found that multimodal websites are generally considered more credible as compared to mono modal website. A website offering specific information on users' own criteria increases the utility of that website. Prescriptive information also increases the utility of choice of an information source. Similarly, the presentation of real time and updated information has increased influence on the UK travellers' choice. The market share of the modes increases when information sources show decreased travel time and cost values and the maximum results are achieved when different information sources give the same information to the UK travellers. A UK website with comprehensive information about the intended travel and offering specific information on users' own criteria increases the utility of that website. On the other had presence of irrelevant information and advertisement significantly reduce the choice of that website and hence decrease its market share. Another important factor is that the government websites with no advertisements and no subscription have higher attraction and credibility within the respondents.

TABLE 6. EFFECT OF ADVERTS AND SUBSCRIPTION ON MARKET SHARE (ESTIMATION SAMPLE AND EXPANDED SAMPLE)

Attribute Level	Market Share of Website A (%)			
	Search Time as Presented in SP		Website B Search Time 15 Minutes	
Commercial Website with Adverts and Subscription				
Commercial Website no Adverts but with Subscription	60	*74*	77	*95*
As Presented in SP	51	*62*	67	*82*
Government with Adverts with Subscription	41	*48*	56	*69*
Commercial Website no Adverts no Subscription				
Government with Adverts and Subscription	46	*57*	64	*79*
As Presented in SP	51	*62*	67	*82*
Commercial no Adverts no Subscription	92	*100*	97	*100*
Daily Updating of Information				
Only Daily Updates	38	*46*	52	*64*
As Presented in SP	51	*62*	67	*82*
Real-Time Updating	63	*78*	76	*93*

ANNEXURE-I. VARIABLES AND ASSOCIATED COEFFICIENTS USED IN THE MODELS

Variables	Coefficients	Definition
$Time_n$	Time	(Generic, in Minutes)
$Cost_n$	Cost	(Generic, in Pence)
$DBPur_n$	$BPur_n$	(Dummy, Business Purpose = 1, Otherwise = 0)
$DLFreq_n$	$LFreq_n$	(Dummy, Trip Frequency less than 13/Year =1, Otherwise = 0)
$DGWeather_n$	$GWeather_n$	Dummy, Good Weather =1, Otherwise = 0)
$DPeakP_n$	$PeakP_n$	(Dummy, Travelled in the Peak Period = 1, Otherwise = 0)
$DReasMode_n$	$ReasMode_n$	(Dummy, Reason for Selecting Mode is Time/Cost = 1, Otherwise = 0)
$DTravAlone_n$	$TravAlone_n$	(Dummy, Travelling Alone = 1, Otherwise = 0)
$DImpSfty_n$	$ImpSfty_n$	(Dummy, Safety Important = 1, Otherwise = 0)
$DImpCmfrt_n$	$ImpCmfrt_n$	(Dummy, Comfort Important = 1, Otherwise = 0)
$DImpSeats_n$	$ImpSeats_n$	(Dummy, Seat Availability Important = 1, Otherwise = 0)
$DMale_n$	$Male_n$	(Dummy, If Male =1, Otherwise = 0)
$DEduc_n$	$Educ_n$	(Dummy, Left Full Time Education at or After 20 = 1, Otherwise = 0)
$DEduc25_n$	$Educ25_n$	(Dummy, Left Full Time Education at or After 25 = 1, Otherwise = 0)
$DFEmpl_n$	$FEmpl_n$	(Dummy, Full Time Employed =1, Otherwise = 0)
$DInc_n$	Inc_n	(Dummy, If income over £30,000 =1, otherwise = 0)
$DAge_n$	Age_n	(Dummy, If Age Less than 50 =1, Otherwise = 0)
$DSubs_n$	$Subs_n$	(Dummy, Subscribed to a Website = 1, Otherwise = 0)
$DImpAd_n$	$ImpAd_n$	(Dummy, Free from Advertisements Important = 1, Otherwise = 0)
$DImpRealt_n$	$ImpRealt_n$	(Dummy, Real time information important = 1, otherwise = 0)
$DImpGInfo_n$	$ImpGInfo_n$	(Dummy, General Information Important = 1, Otherwise = 0)
$Carfrit_n$	$Carfrit_n$	(Time by car Information by Friend, in Minutes)
$Carmult_n$	$Carmult_n$	(Time by Car Information by Multimodal Website, in Pence)
$Carfric_n$	$Carfric_n$	(Cost by Car Information by Friend, in Minutes)
$Carmultc_n$	$Carmultc_n$	(Cost by Car Information by Multimodal Website, in Pence)
$Trainwt_n$	$Trainwt_n$	(Time by Train Information by Train Website, in Minutes)
$Trainext_n$	$Trainext_n$	(Time by Train Information by Past Experience, in Minutes)
$Trainwc_n$	$Trainwc_n$	(Cost by Train Information by Train Website, in Pence)
$Trainexc_n$	$Trainexc_n$	(Cost by Train Information by Past Experience, in Pences)
$Coawt_n$	$Coawt_n$	(Time by Coach Information by Coach Website, in Minutes)
$Coawc_n$	$Coawc_n$	(Cost by Coach Information by Coach Website, in Pences)
$Coamult_n$	$Coamult_n$	(Time by Coach Information by Multimodal Website, in Minutes)
$Coamulc_n$	$Coamulc_n$	(Cost by Coach Information by Multimodal Website, in Pence)
$DComAd_n$:	$ComAd_n$:	(Dummy that Equals 1 if Commercial Ads No Sub; Otherwise 0)
$Timeweb_n$	Timeweb	(Time by Modal Web/Multimodal Website, Generic, in Minutes)
$Costweb_n$	Costweb	(Cost by Modal Web/Multimodal Website, Generic, in Pence)
$DComAdSub_n$:	$ComAdSub_n$:	(Dummy that Equals 1 if Commercial Ads Subs; Otherwise 0)
$DCom_n$:	Com_n:	(Dummy that Equals 1 if Commercial No Ads No Sub; Otherwise 0)
$Seatime_n$	$Seatime_n$	(Search Time in 5, 10, 15 Minutes)
$DUpdaily_n$:	$Updaily_n$:	(Dummy that Equals 1 if Website Updates Daily; Otherwise 0)
$DUpweekly_n$:	$Upweekly_n$:	(Dummy that Equals 1 if Website Updates Weekly; Otherwise 0)
$DInfoSpec_n$:	$InfoSpec_n$:	(Dummy that Equals 1 if Specific Info Available; Otherwise 0)
$DInfoPerc_n$:	$InfoPerc_n$:	(Dummy that Equals 1 if Info wrt Own Criteria; Otherwise 0)
$DYoung_n$	$Young_n$	(Dummy, If Age Less than 40 =1, Otherwise = 0)
	ASC_n	(Alternative Specific Constant)

ACKNOWLEDGEMENTS

Firstly, authors are pleased to acknowledge the contributions of Ph.D. supervisors Prof. Peter Bonsall, and Prof. Gerard de Jong, Institute for Transport Studies, University of Leeds, UK, and much of the text of this study owes to them and author's Ph.D. Thesis.

REFERENCES

[1] Ben-Akiva, M., Palma, A.D., and Kaysi, I., "Dynamic Network Models and Driver Information Systems", Transportation Research Part-A, Volume 25A, No. 5, pp. 251-266, 1991.

[2] Mahmassani, H., and Jayakrishnan, R., "System Performance and User Response Under Real Time Information in a Congested Traffic Corridor", Transportation Research Part-A, Volume 25, No. 5, pp. 293-307, 1991.

[3] Bonsall, P.W., "Analyzing and Modelling the Influence of Roadside Variable Message Displays on Drivers' Route Choice", 7th World Conference on Transportation Research, pp. 16-21, Sydney, 1995.

[4] Al-Deek, H., Khattak, A.J., and Thaanjevan, P., "A Combined Traveller behaviour and System Performance Model with Advanced Traveller Information Systems", Transportation Research Part-A, Volume 32, No. 7, pp. 479-493, 1998.

[5] Mahmassani, H., and Liu, Y.S., "Dynamics of Commuting Decision behaviour Under Advanced Traveler Information Systems", Transportation Research Part-C, Volume 7, pp. 91-107, 1999.

[6] Stough, R.R., "Intelligent Transport Systems: Cases and Policies", Edward Elgar, Chetenham, 2001.

[7] Polak, J., and Jones, P., "The Acquisition of Pre-Trip Information: A Stated Preference Approach", Transportation, Volume 20, pp. 179-198, 1993.

[8] Emmerink, R., Nijkamp, P., Rietveld, P., and Van Ommeren, J.N., "Variable Message Signs and Radio Traffic Information: an Integrated Empirical Analysis of Drivers' Route Choice Behaviour", Transportation Research, Volume 30A, pp. 135-153, 1996.

[9] Khattak, A.J., Yim, Y., and Prokopy, L.S., "Willingness to Pay for Travel Information", Transportation Research, Volume 11C, pp. 137-159, 2003.

[10] Kenyon, S., and Lyons, G., "The Value of Integrated Multimodal Traveller Information and its Potential Contribution to Modal Change", Transportation Research, Volume 6F, pp. 1-21, 2003.

[11] Abdel-Aty, M., and Abdulla, M.F., "Modeling Drivers Diversion from Normal Routes under ATIS Using Generalized Estimating Equations and Binomial Probit Link Function", Transportation, Volume 31, No. 3, pp. 327-348, 2004.

[12] Chorus, C.G., Molin, E.J.E., Arentze, T.A., Timmermans, H.J.P., and Van Wee, G.P., "Travellers' Need for Information: An Empirical Study into the Role of Knowledge", 85th Annual Meeting of the Transportation Research Board, Washington, DC, 2006.

[13] Chorus, C.G., Walker, J.L., and Ben-Akiva, M.E., "Travel Information Use and Effects", 11th International Conference on Travel Behaviour Research, Kyoto, August, 2006.

[14] Chorus, C.G., Molin, E.J.E., and van Wee, G.P., "Use and Effects of Advanced Traveller Information Services (ATIS): A Review of the Literature", Transport Reviews, Volume 26, No. 2, pp. 127-149, 2006.

[15] Chorus, C.G., Arentze, T.A., and Timmermans, H.J.P. "Spatial Choice: A Matter of Utility or Regret?", Environment and Planning Part-B, Volume 36, No. 3, pp. 538-551. 2009.