I T L S

WORKING PAPER

ITLS-WP-10-01

Search based internet surveys: Airline stated choice

By

Andrew T Collins, John M Rose & Stephane Hess*

*** ITLS Sydney and ITS Leeds**

January 2010

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-10-01

TITLE: **Search based internet surveys: Airline stated choice**

ABSTRACT: Stated preference (SP) experiments are becoming an increasingly popular survey methodology for investigating air travel choice behaviour. Nevertheless, some evidence suggests that SP experiments do not mirror decisions in real markets. In this paper we introduce a novel survey methodology that aims to make air travel surveys more consistent with real world settings, with the aim of obtaining more realistic results. The survey is modelled on the interface and functionality of an online travel agent (OTA). As with a real OTA, many ticket options are presented. Sort tools allow the options to be reordered, search tools allow options to be removed from consideration, and a further tool allows attributes to be hidden and shown. Extensive use of these tools is made by the 462 respondents, with the captured data revealing some attribute preferences at the individual level, and significant heterogeneity of preference across individuals. A traditional SP component was also completed by the respondents. Mixed multinomial logit models were estimated on data from both the traditional SP and OTA components, with the later exhibiting greater willingness to pay (WTP) heterogeneity.

KEY WORDS: ***Airline choice; stated choice experiment; information processing strategies; and survey realism***

AUTHORS: Andrew T Collins, John M Rose, Stephane Hess

CONTACT: Institute of Transport and Logistics Studies **(C37)**

The Australian Key Centre in Transport Management

The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071

Facsimile: +61 9351 0088

E-mail: itls@sydney.edu.au

Internet: http://www.itls.sydney.edu.au

DATE: January 2010

1. Introduction

Stated preference (SP) experiments have grown to become the predominant data paradigm in the elicitation of behavioural responses of individuals, households and organisations over diverse choice situations and contexts. One partial explanation for this is research evidence suggesting that SP experiments are capable of replicating decisions made in real markets (see e.g., Burke et al., 1992; Carson et al., 1994). Several studies have shown that SP experiments are able to reproduce the behavioural outputs, such as willingness to pay (WTP) measures, obtained from revealed preference (RP) choice experiments (e.g., Carlsson and Martinsson, 2001; Lusk and Schroeder, 2004). Nevertheless, contradictory evidence also exists that calls into question whether results obtained from SP experiments do in fact mirror those obtained from real markets. For example, Wardman (2001) and Brownstown and Small (2005) found significant differences between WTP values derived from RP and SP choice studies. In both these studies, values of travel time savings (VTTS) from SP experiments were found to be undervalued in comparison to the results from RP studies. Interestingly however, the opposite is typically observed in traditional contingent valuation studies where WTP values have been found to over value those observed in real markets (see e.g., Hensher (2009) for a detailed overview of differences obtained between WTP values from different survey methodologies).

Given such a divergence of evidence, of particular research interest is to determine firstly, to what extent SP experiments are able to replicate real market decisions, and secondly, assuming a difference between SP and RP results does exist, what factors can bridge the gap. Rose and Hensher (2006) argued that one such factor is the degree of realism imposed in SP surveys, where realism in SP experiments arises not simply from the fact that respondents are asked to undertake similar actions as they would in real markets, but also in how much the experiment is made to look and feel like choices made in real market transactions. Yet despite this argument, researchers are yet to examine the problem in detail.

The vast majority of SP research has typically applied a traditional survey response mechanism where respondents are presented with grids consisting of columns of alternatives and rows of attributes. Respondents are then asked to review the grid before deciding which alternative appeals to them the most. In most cases, only one response mechanism is used per study making a comparison of different survey task representations impossible. As such, little research has focused on making SP experiments more realistic. Indeed, research has tended to centre on identifying sources of cognitive burden placed upon respondents undertaking SP tasks, rather than improving the realism of the task itself (e.g., Arentze et al., 2003; DeShazo and Fermo, 2002) as well as reducing the cognitive load placed on those same respondents (e.g., Louviere and Timmermans, 1990; Wang et al., 2001).

Nevertheless, a small number of studies have examined differences in responses to SP experiments produced when attributes are represented in alternative forms. For example, Nelson and Towriss (1995) found differences between text and visual representations of attributes in SP study results, however Jones and Bradley (2006) suggested that the contrary is true, with little evidence to support that the use of pictures, etc., produced different results to the use of text descriptors. Nevertheless, Jones and Bradley (2006) called for urgent research into the area.

In this paper, we examine how web-based surveys can be used to promote greater realism in SP survey tasks. Centred around an application of airline choice, the paper presents results from a study in which respondents complete both a traditional SP task and a SP task made to look and feel exactly like the website of an online travel agent (OTA) (e.g., Expedia and Travelocity). The use of a web-based OTA look-alike survey instrument, as opposed to the traditional SP grid representation, presents respondents with a more realistic choice setting, given that many air travellers will have had experience with such systems in the past.

The remainder of the paper is structured as follows. In the following section a brief review of the literature on air travel choice behaviour is presented, along with a discussion of OTA choice environments, before section 3 describes the current survey. Section 4 describes analysis performed on the survey data, and presents results, before conclusions are drawn in Section 5.

2. Literature on air travel behaviour modelling and introduction to OTA choice environments

A wide range of studies have investigated air travel choice behaviour using both SP and RP methods. Kanafani and Sadoulet (1977) modelled the choice among fare types for long haul journeys. Proussaloglou and Koppleman (1995) examined the choice of airline for recent trips using mail-in RP data. In recent years, a majority of studies have used the SP methodology. Bradley (1998) used SP data to examine the choice of departure airport and route from Schiphol, Brussels and Eindhoven airports. Hensher et al. (2001) used SP data for airline choice between New Zealand and Australia. Hess et al. (2007) and Hess (2007) also made use of SP data collected via the internet and retrieve effects of a number of attributes which often cause problems in RP data (fares, frequent flier benefits).

The dimensions and nature of SP tasks are well suited to how people may choose air tickets with a travel agent. A travel agent is provided with some initial requirements, and then returns with several options. At this point, one ticket is chosen, a choice is made not to purchase any of the tickets, or the requirements are adjusted and another search asked of the travel agent. An SP task can be framed as a single choice in what is potentially an iterative process. Proussaloglou and Koppleman (1999) performed a novel air travel SP survey that incorporated one way that travellers may search for information when talking to a travel agent on the phone. Presented with a travel scenario, the respondents had to elicit from the interviewer the available flights as described by schedule and fare. Flights could be revealed in any order the respondent wished – according to schedule or fare, and a choice could be made at any stage. The interviewer had a record of what flights had been revealed when the choice was made. Key methodological aspects of this study include the ability of the respondent to drive the search process prior to making a decision, and the use of flights that were close to real world options. The study had a high level of realism as it extended the survey choice task to include a search for information.

Since the time of their study, online travel agents (OTAs) have emerged as viable competitors to traditional travel agents, and account for a significant percentage of market share. In 2007, more travel was purchased online (through both OTAs and airline websites) than offline in the U.S. (PhoCusWright, 2007). OTAs have the advantage of bringing together highly detailed information on a large number of options that a traveller may choose from. To help customers make sense of so much information a range of tools are provided. Searches can be refined on a range of criteria, and the alternatives can be sorted on many of the attributes. The level of control over the search process varies across OTAs, as does the attributes used in the description of the options, where for example information on seat pitch and entertainment options are only gradually being included.

Often, the number of ticket options by one airline presented on an OTA may appear disproportionately high. Airlines with several flights close in time on each leg of a multi-leg journey are able to present a ticket for every combination of the flights on each leg. This is further accentuated when an airline has code sharing partners operating on part of the route, as their code share partners' tickets can be included in the combinations of tickets provided. Here, the provision of search tools is important to allow the customer to reduce the number of options, and eliminate dominated options.

This paper will present a survey methodology that mimics the interface and functionality of an OTA. There are a wide range of motivations for implementing a survey of this type. As a means of determining air travel preferences across the population, the OTA survey interface boosts the level of realism for people who use or are comfortable with real OTAs. Rose and Hensher (2006) argued that this can help an SP experiment replicate market decisions. The survey methodology is an acknowledgement that the choice process includes information search as well as the final choice. Information search data provides an additional data source to the observed alternatives and choices. This additional data may not only aid in model specification but also provide crucial insights into choice behaviour. This study is a first attempt at an OTA based survey and forms the basis for further analysis.

3. Survey description

The SP scenarios in the current study ask respondents to choose a ticket for travel from Australia to a European destination. A long haul route was used as it is believed that travellers will be more discerning of attributes such as in flight entertainment and seat pitch on such routes. London and Paris were made available as destinations in the SP, with Sydney as departure. The choice was framed as a leisure trip as business trips are usually paid for by someone else resulting in very different willingness to pay measures. In the interests of survey simplicity, respondents were only presented with economy ticket options.

The survey presented to respondents contained two choice components; a traditional SP component with four choice tasks and a practice task, and a ticket search component modelled on OTAs, with four search tasks and a practice search. The order of the two components was randomised, as was the order of the tasks within each component.

For the traditional SP task, three unlabelled alternatives were included (although an attribute indicated the airline) alongside a "no choice" option. For the OTA task, the number of alternatives varied across task and respondent, ranging from 12 to 22. The same attributes were used for both presentation formats, and are listed in Table 1. All prices were displayed in Australian dollars. The average exchange rates for February 2008 (the time of the survey) were AUD1 = $US0.91 and AUD1 = €0.62. While an experimental design was used for the SP tasks, the search tasks made use of information from real world flights (where available). Two price components were shown: a carbon tax, and the ticket price excluding the carbon tax. Real airline names were displayed, always with their logo visible. Some of the comfort related attributes are not typically shown on ticket booking websites. Here, our survey presents respondents with more detailed information while still allowing them to eliminate these attributes to simplify the tasks. In real decision environments, a decision about both the departing and return flights must be made. In the interests of simplicity, for both presentation formats we only required a choice for the departing flight, and asked the respondent to assume that the return flight would have similar service levels.

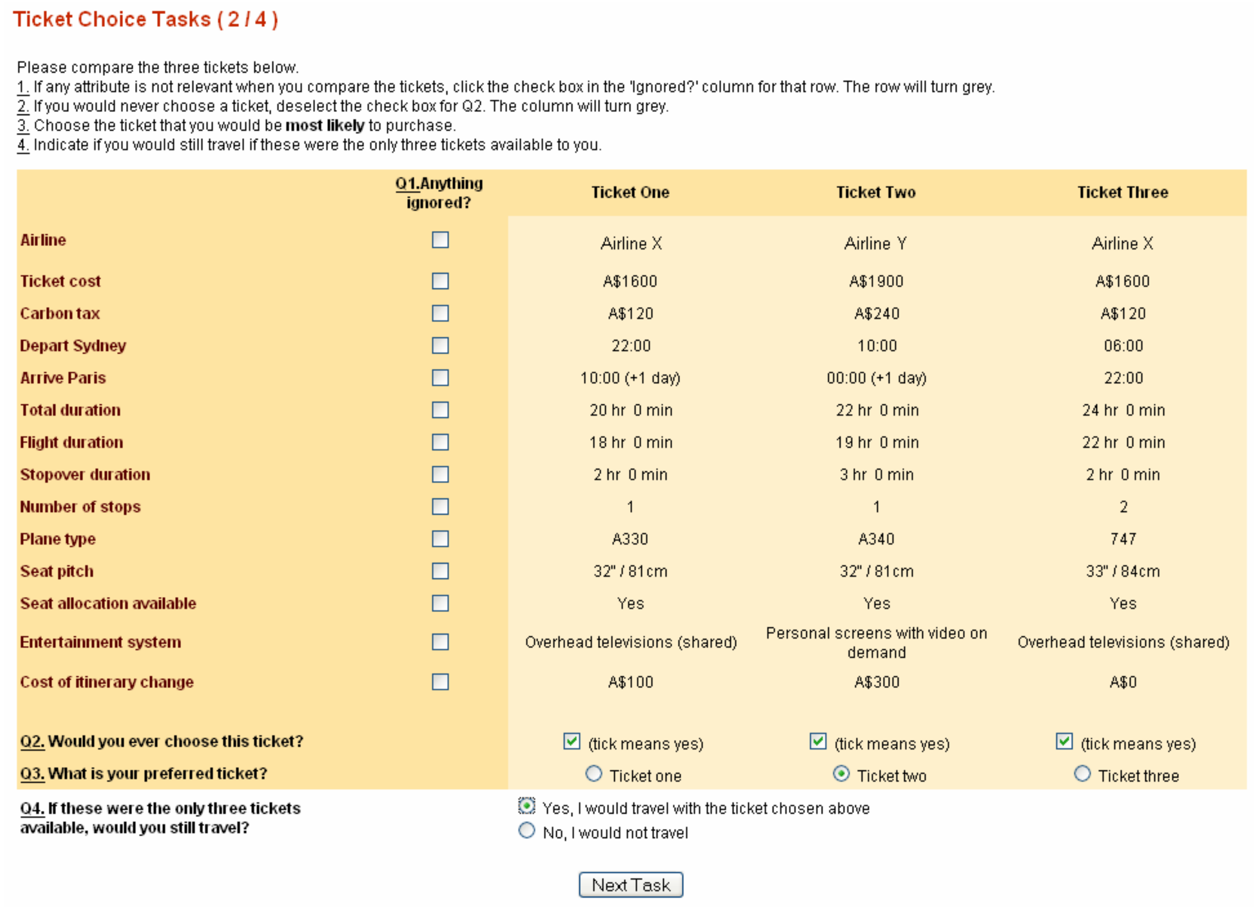
Table 1: Attributes in SP and search tasks

Attribute	SP levels	Search levels or range	Search: From real flight?	Typical online travel agent attribute?
Price	AUD1600, AUD1900, AUD2200, AUD2500	AUD1809 – AUD6036	Yes	Yes
Carbon tax	AUD0, AUD120, AUD240, AUD360	AUD0 – AUD460.76	No	No
Airline	9 possible	13 possible	Yes	Yes
Departure time	6am, 10am, 5pm, 10pm	Continuous	Yes	Yes
Arrival time	Based on departure time and flight duration	Continuous	Yes	Yes
Total duration	20hrs, 22hrs, 24hrs, 26hrs	22hrs 20mins – 38hrs 40mins	Yes	Yes
Flight duration	Based on total and stopover duration	21hrs 20mins – 26hrs	Yes	Yes
Stopover duration	1hr, 2hrs, 3hrs, 4hrs	40mins – 14hrs 50mins	Yes	Yes
Number of stops	1, 2	1, 2, 3	Yes	Yes
Plane type	747, 777, A330, A340		No	Yes
Seat pitch	31", 32", 34"		No	No
Seat allocation available?	Yes/No		No	Yes
Entertainment system	Overhead televisions (shared), Personal screens with limited movie selection, Personal screens with video on demand		No	No
Itinerary change cost	AUD0, AUD100, AUD200, AUD300		No	Often hidden

A few other points need to be addressed. Frequent flyer membership is widely recognised to have a significant influence on choice (Chin, 2002; Proussaloglou and Koppelman, 1999; Hensher et al., 2001; Proussaloglou and Koppelman, 1995). Increasingly, airlines are recognising this and are attempting to encourage passengers to choose more expensive fare classes in return for bigger benefits (i.e. fewer miles with discount tickets). No such complication was used in this survey, but we did ask respondents what, if any, frequent flier programs they are a member of. Furthermore, unlike in some previous studies, airport and access mode choice were ignored, where the effect of this is possibly mitigated by the long haul nature of the flights presented. Furthermore, Sydney is only served by a single international airport.

3.1 Traditional SP tasks

The SP component consisted of four choice tasks, each with three alternatives described by all of the attributes listed in Table 1. Respondents were asked to indicate their preferred flight, but were also given a no choice option. Furthermore, respondents were directed to indicate if any attributes were ignored, and were asked if some of the alternatives would never be chosen. An example of the choice screen is shown in Figure 1 (with airline names masked). A D-efficient design (see e.g., Rose and Bleimer, 2008) was used, with 18 blocks of four choice tasks each.



Ticket Choice Tasks (2 / 4)

Please compare the three tickets below.
1. If any attribute is not relevant when you compare the tickets, click the check box in the 'Ignored?' column for that row. The row will turn grey.
2. If you would never choose a ticket, deselect the check box for Q2. The column will turn grey.
3. Choose the ticket that you would be **most likely** to purchase.
4. Indicate if you would still travel if these were the only three tickets available to you.

	Q1.Anything ignored?	Ticket One	Ticket Two	Ticket Three
Airline	☐	Airline X	Airline Y	Airline X
Ticket cost	☐	A$1600	A$1900	A$1600
Carbon tax	☐	A$120	A$240	A$120
Depart Sydney	☐	22:00	10:00	06:00
Arrive Paris	☐	10:00 (+1 day)	00:00 (+1 day)	22:00
Total duration	☐	20 hr 0 min	22 hr 0 min	24 hr 0 min
Flight duration	☐	18 hr 0 min	19 hr 0 min	22 hr 0 min
Stopover duration	☐	2 hr 0 min	3 hr 0 min	2 hr 0 min
Number of stops	☐	1	1	2
Plane type	☐	A330	A340	747
Seat pitch	☐	32" / 81cm	32" / 81cm	33" / 84cm
Seat allocation available	☐	Yes	Yes	Yes
Entertainment system	☐	Overhead televisions (shared)	Personal screens with video on demand	Overhead televisions (shared)
Cost of itinerary change	☐	A$100	A$300	A$0
Q2. Would you ever choose this ticket?		☑ (tick means yes)	☑ (tick means yes)	☑ (tick means yes)
Q3. What is your preferred ticket?		○ Ticket one	◉ Ticket two	○ Ticket three

Q4. If these were the only three tickets available, would you still travel?
◉ Yes, I would travel with the ticket chosen above
○ No, I would not travel

Next Task

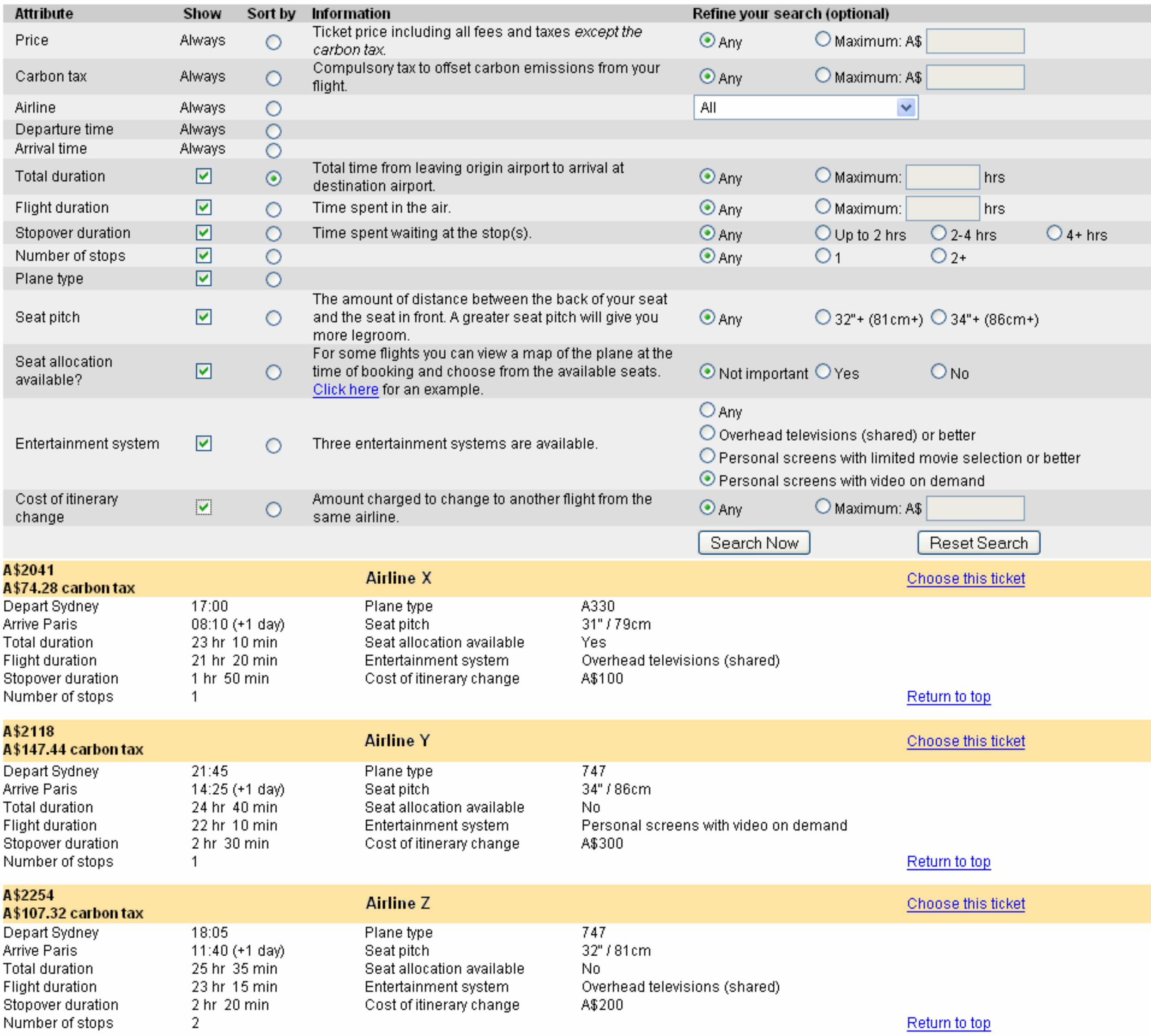
Figure 1: Stated preference task

3.2 Search tasks

The flights for the search tasks were based on real flights that were obtained from a popular OTA. To prevent extensive correlations within airlines, the plane type, seat pitch, seat allocation, entertainment system and cost of itinerary change attribute levels were not drawn from the real flights. Instead, for each attribute, each level was allocated an equal number of times. The levels were then swapped between flights such that the correlations between attributes were minimised.

Four search tasks were presented to the respondents, in addition to a practice search task which contained four flights only. Real flight prices vary over time for the same flight due to yield management systems. Also, travel at certain times of the year will always be more expensive due to high demand. Consequently, each of the four search tasks represent flights at different times in the future. This allows for a good coverage of flight prices in the sample. Flights were selected for departure in two weeks' time, in a month's time, in five months' time, and over Christmas. These timeframes were randomised in presentation order across respondents and explicitly mentioned to the respondents to help them understand why the average prices varied from task to task. Figure 2 shows how the tickets appeared in the search tasks, with all attributes shown in this example (and airline names masked in the screenshot).

The top of the search task screens contained a set of tools that included sort, search, a description of the attributes, and a means to hide some attributes. Figure 2 contains this set of tools. All attributes could be sorted on, with the best quality attribute shown first: lowest price, shortest duration, best entertainment system and so on. By default, the flights were sorted on price for the first choice set. Subsequent sort selections were preserved from one task to the next.



Attribute	Show	Sort by	Information	Refine your search (optional)
Price	Always		Ticket price including all fees and taxes *except the carbon tax.*	Any; Maximum: A$
Carbon tax	Always		Compulsory tax to offset carbon emissions from your flight.	Any; Maximum: A$
Airline	Always			All
Departure time	Always			
Arrival time	Always			
Total duration			Total time from leaving origin airport to arrival at destination airport.	Any; Maximum: hrs
Flight duration			Time spent in the air.	Any; Maximum: hrs
Stopover duration			Time spent waiting at the stop(s).	Any; Up to 2 hrs; 2-4 hrs; 4+ hrs
Number of stops				Any; 1; 2+
Plane type				
Seat pitch			The amount of distance between the back of your seat and the seat in front. A greater seat pitch will give you more legroom.	Any; 32"+ (81cm+); 34"+ (86cm+)
Seat allocation available?			For some flights you can view a map of the plane at the time of booking and choose from the available seats. Click here for an example.	Not important; Yes; No
Entertainment system			Three entertainment systems are available.	Any; Overhead televisions (shared) or better; Personal screens with limited movie selection or better; Personal screens with video on demand
Cost of itinerary change			Amount charged to change to another flight from the same airline.	Any; Maximum: A$

Search Now | Reset Search

A$2041
A$74.28 carbon tax — **Airline X** — Choose this ticket

Depart Sydney	17:00	Plane type	A330
Arrive Paris	08:10 (+1 day)	Seat pitch	31" / 79cm
Total duration	23 hr 10 min	Seat allocation available	Yes
Flight duration	21 hr 20 min	Entertainment system	Overhead televisions (shared)
Stopover duration	1 hr 50 min	Cost of itinerary change	A$100
Number of stops	1		

Return to top

A$2118
A$147.44 carbon tax — **Airline Y** — Choose this ticket

Depart Sydney	21:45	Plane type	747
Arrive Paris	14:25 (+1 day)	Seat pitch	34" / 86cm
Total duration	24 hr 40 min	Seat allocation available	No
Flight duration	22 hr 10 min	Entertainment system	Personal screens with video on demand
Stopover duration	2 hr 30 min	Cost of itinerary change	A$300
Number of stops	1		

Return to top

A$2254
A$107.32 carbon tax — **Airline Z** — Choose this ticket

Depart Sydney	18:05	Plane type	747
Arrive Paris	11:40 (+1 day)	Seat pitch	32" / 81cm
Total duration	25 hr 35 min	Seat allocation available	No
Flight duration	23 hr 15 min	Entertainment system	Overhead televisions (shared)
Stopover duration	2 hr 20 min	Cost of itinerary change	A$200
Number of stops	2		

Return to top

Figure 2: Search task

All attributes except for departure and arrival time could be searched on. All costs and most duration times could be searched on with a respondent specified maximum. Other attributes could be searched on by choosing a category. Searches on stopover duration were limited to distinct categories that did not overlap. This was done both for simplicity and to test whether some respondents wanted a minimal stopover time while others wanted some longer time period. Any number of searches could be performed. By default, no search criteria were applied, although the final search criteria in each task were preserved for the next task.

Price, carbon tax, airline name, departure time and arrival time were always shown. All other attributes could be hidden and shown as desired via the set of tools. We provided this option as a way to remove unnecessary clutter on the screen and help facilitate easier, faster decision making on attributes that matter to the respondent. We did not initially show the attributes that could be hidden, and forced the respondents to identify the attributes that mattered to them.

In order to find out how respondents use the sort, search and show/hide tools (which we will collectively refer to as the search tools), large amounts of data were captured by the survey instrument. In addition to the state of the search tools at the time of choice, every action performed using the search tools was captured in the database, as was the resulting choice set. This information allows the analyst to examine the numerous strategies that people employ to refine their search.

There are some notable differences between the OTA survey and real OTA choice environments. In the later, the preferred date is entered as part of the initial search criteria. The user can change the day of travel if the preferred day does not have a satisfactory ticket, or merely if they are curious to compare flights across days. A more complex extension of our survey could include searches across days and so capture more complex search processes.

3.3 Collection of other information

In addition to the choice observations, information was collected on how many times the respondent had travelled domestically, internationally, and to Europe over the last three years, broken down by whether the ticket had been paid for by themselves or others. The number of unique airlines flown with over the previous three years was obtained, as was frequent flyer membership and the usual class of ticket purchased for international flights.

3.4 Survey deployment

Survey participants were recruited from an online sample of Sydney residents. To be eligible for the study, respondents were required to have travelled to Europe in the last three years. This restriction was an attempt to screen out respondents who would not make this choice in real life. Screening on the likelihood of travel in a future time period might be more suitable for future studies, especially as it is plausible (and testable) that travellers lacking recent experience will search more than experienced travellers. After screening for eligible respondents and some further data cleaning, a final sample of 462 respondents was obtained.

4. Analysis

4.1 Sort behaviour

In the OTA search tasks, the flights can be sorted on any attribute, with the initial default being a sort by price. Table 2 indicates both which attributes were sorted on at the time a choice was made, and how many times an actual sort was explicitly performed. Since sort information is preserved between tasks, for any given attribute there may be fewer sort actions than tasks that were sorted on that attribute at the time of choice. Furthermore, since many sorts can be performed before a choice is made, there may be more sort actions than tasks that were sorted on that attribute at the time of choice. Table 2 includes both the practice search task and the four main search tasks. Of the 1380 sort actions, 862 were performed in the practice task, which suggests that many of the sorts were performed experimentally or to establish a preferred sort preference.

Table 2: Sorting strategies

	Tasks with this sort at time of choice		Individuals with this sort at choice for all tasks		Sort actions performed	
Price	1019	44%	159	34%	539	39%
Price (by default)	793	34%	147	32%	-	
Carbon tax	63	3%	7	2%	134	10%
Airline	129	6%	17	4%	188	14%
Departure time	39	2%	5	1%	88	6%
Arrival time	43	2%	5	1%	60	4%
Total duration	45	2%	4	1%	88	6%
Flying duration	25	1%	1	0%	50	4%
Stopover duration	10	0%	0	0%	45	3%
Number of stops	8	0%	0	0%	27	2%
Plane type	7	0%	1	0%	25	2%
Seat pitch	37	2%	5	1%	33	2%
Seat reservation	24	1%	3	1%	37	3%
Entertainment system	48	2%	6	1%	39	3%
Ticket change charge	20	1%	2	0%	27	2%
Combination	-	-	100	22%	-	-
Total	2310	100%	462		1380	100%

Clearly price is the dominant sort attribute, with flights being sorted on price explicitly and by default for 78 percent of choice tasks. Cumulatively the remaining attributes account for 22 percent of sorts at choice, which is a significant minority. Sort preference for these remaining attributes is roughly equal, which indicates an overall heterogeneity of sort preference. There are more sorts on airline than any other individual non-price attribute, which hints that some respondents may have strong airline preferences. At the individual level, Table 2 shows that most individuals are consistent with their sort preference at time of choice. Only 22 percent varied their sort at choice over the five tasks (four 'real' and one practice).

4.2 Search behaviour

Table 3 shows, at the attribute level, the number of tasks for which a search criterion was applied at the time of choice. Whereas price was the dominant attribute for sorting, relatively few tasks included a search on price or carbon tax. Instead, searches were performed in greater numbers on the comfort attributes, including entertainment system (for 21 percent of all tasks), seat reservation (11 percent) and seat pitch (nine percent). Many searches were also performed on attributes concerned with stopovers: numbers of stops (eight percent) and stopover duration (seven percent).

The manner in which each attribute was searched is interesting. Some attributes have a clear preference sign, including price and entertainment system. Price limits were typically low but reasonable, and entertainment system searches were evenly split between restriction to video on demand and personal screens or better. Other attributes are likely to be considered in different ways across the population. The stopover duration levels were mutually exclusive, and searches on this attribute were split between a desire to minimise stop time (up to two hours) for 75 percent of cases, and a desire to have a more leisurely stop (2-4 hours) for 25 percent of cases. Either search strategy is plausible. The former would minimise total travel time, while the later would provide a lengthy break from a confined environment, or perhaps provide an opportunity for shopping. Taste heterogeneity can be captured in advanced models such as mixed multinomial logit (MMNL). However, the search tool can help reveal with some certainty the individual level preferences of those respondents who use it.

An area for future research is to consider how the search information may be incorporated in advanced model structures.

Table 3: Number of tasks with search criteria applied for each attribute at time of choice

	Number of tasks with search criteria applied at time of choice	(percent)
Price	96	(4%)
Carbon tax	36	(2%)
Airline	76	(3%)
Total duration	49	(2%)
Flying duration	27	(1%)
Stopover duration	167	(7%)
Number of stops	187	(8%)
Seat pitch	198	(9%)
Seat reservation	258	(11%)
Entertainment system	476	(21%)
Ticket change charge	40	(2%)
All tasks	2310	(100%)

Unlike sort selections, search criteria can be applied across multiple attributes concurrently. An analysis of the data showed that 18.3 percent of all tasks were completed with multiple search criteria applied. It is with these complex searches that the search tool is most useful. If only one search criterion is applied, it might be easier to just perform a sort. However, the sort tool is cumbersome and ineffective if more than one attribute is very important.

Whereas sort actions only reorder the flights on screen, search actions actually add or remove flights from view. This makes a search a stronger form of filter, as any flight that fails to meet the search criteria cannot be chosen. These reductions are quite large in absolute terms when some search tasks contain 22 potential flights. On average, the choice set size after applying search criteria was reduced to seventy-three percent of its original size, where for a quarter of respondents, it was reduced to under forty percent of its original size.

4.3 Showing and hiding of attributes

The price, carbon tax, airline, departure time and arrival time attributes were always visible and could not be hidden. All other attributes were not shown by default and had to be actively chosen for display. None of these attributes were shown for more than half of the tasks, with the least shown attribute being ticket change charges. At the individual level, 37 percent of respondents did not show any of the additional attributes for any of their tasks. This may have been due to satisfaction with the default attributes as the sole means of ticket differentiation, for example with highly price sensitive respondents. It also may have been due in part to a lack of engagement with or understanding of the survey.

4.4 Choice modelling methodology

The models presented in the remainder of this paper were estimated using mixed multinomial logit (MMNL) models. With the scope of the present paper being on survey methodology as well as results, we have consciously limited the detail on modelling methodology and refer the reader to excellent descriptions of MMNL methodology in Train (2003).

The MMNL model is an advanced type of discrete choice model in which the sensitivities to explanatory variables (e.g. fare, flight time) are allowed to vary randomly across respondents following a pre-specified distribution with estimated parameters. This gives the MMNL model a significant advantage over models assuming an absence of such random taste heterogeneity, such as the Multinomial Logit model. Additionally, and crucially for the present analysis, the MMNL model can also recognise the repeated choice (panel) nature of SP data directly in estimation, avoiding issues of biased standard errors common to other models in such situations (cf. Ortúzar and Willumsen, 2001) .

All our models made use of Normal distributions for the random parameters, but the effects of these shape assumptions are mitigated by the calculation of individual specific willingness to pay indicators (cf. Greene et al., 2005). Additionally, in the computation of these willingness to pay indicators, we explicitly accounted for the fact that the survey made use of different cost components (fare, carbon and change) by using a weighted average of the associated cost coefficients.

4.5 Model results

Four models were estimated as part of this study and are reported in Tables 4 and 5. Table 4 reports two models estimated the traditional SP data (models M1 and M2) and Table 5 reports two additional models estimated on the OTA data set (M3 and M4). All models were estimated assuming that random parameters are normally distributed with 500 Halton draws used in model estimation. All four models produce decent model fits, however the models estimated on the OTA data produce far superior model fits to those estimated on the traditional SP data.

For the present study, we concentrate on two sets of models per data set. The first set of models estimated on each data set use the full set of collected data (M1 and M3), ignoring any additional information captured in the survey (such as attribute searches, sorts, or respondents choosing not to display certain attributes). The second set of models are estimated accounting for respondents either stating that they did not consider an attribute in the SP choice tasks (M2) or actually selecting not to display attributes in the OTA tasks (M4).

Table 4: M1 and M2 Model results

	M1				M2			
	Mean par.	(*t*-ratio)	Std Dev. Par.	(*t*-ratio)	Mean par.	(*t*-ratio)	Std Dev. Par.	(*t*-ratio)
Random Parameters								
Ticket Price	-0.003	(-17.58)	0.001	(5.15)	-0.002	(-15.50)	0.001	(6.09)
Travel Time	-0.002	(-6.37)	0.001	(5.54)	-0.001	(-3.45)	0.002	(9.25)
Number of Stops	-0.388	(-4.59)	0.606	(4.70)	-0.291	(-3.34)	0.688	(5.25)
Entertainment (on demand)	0.581	(5.42)	0.980	(5.92)	0.548	(4.79)	1.115	(6.22)
Seat Assignment	-	-	-	-	0.260	(3.09)	0.551	(3.39)
Departure (6am)	-0.152	(-1.50)	0.456	(2.00)	-	-	-	-
Departure (10am)	-0.058	(-0.48)	0.620	(2.67)	0.186	(1.22)	0.740	(3.34)
Non-Random Parameters								
Constant A	7.823	(7.22)	-	-	6.646	(11.66)	-	-
Constant B	7.886	(7.26)	-	-	6.707	(11.72)	-	-
Constant C	7.625	(7.04)	-	-	6.448	(11.33)	-	-
Airline 1	0.449	(3.60)	-	-	0.477	(3.37)	-	-
Airline 3	-0.574	(-4.18)	-	-	-0.294	(-1.96)	-	-
Airline 4	-	-	-	-	0.433	(2.97)	-	-
Airline 5	-0.309	(-2.38)	-	-	-0.306	(-2.14)	-	-
Airline 6	-0.345	(-2.66)	-	-	-0.272	(-1.95)	-	-
Airline 7	0.276	(2.13)	-	-	0.388	(2.67)	-	-
Carbon Tax	-0.001	(-2.73)	-	-	0.000	(-0.69)	-	-
Charge	-0.001	(-2.29)	-	-	0.000	(0.87)	-	-
Seat Pitch	0.077	(2.63)	-	-	0.027	(2.24)	-	-
Seat Assignment (1=yes)	0.230	(3.14)	-	-	-	-	-	-
Depart (5pm)	0.231	(2.03)	-	-	0.421	(3.37)	-	-
747 aircraft	-	-	-	-	0.368	(3.33)	-	-
A330 aircraft	-	-	-	-	0.256	(2.27)	-	-
Arrive (7-10am)	-	-	-	-	0.389	(3.09)	-	-
Arrive (11am-2pm)	-	-	-	-	0.329	(2.03)	-	-
Arrive(3-7pm)	-	-	-	-	0.388	(2.58)	-	-
Model Fits								
LL(0)	-2561.872				-2561.872			
LL(β)	-2081.124				-2141.275			
Number of Pars.	25				30			
ρ^2	0.188				0.164			
Adj. ρ^2	0.177				0.150			
Observations	462				462			
Respondents	1848				1848			

Table 5: M3 and M4 Model results

	M3				M4			
	Mean par.	(*t*-ratio)	Std Dev. Par.	(*t*-ratio)	Mean par.	(*t*-ratio)	Std Dev. Par.	(*t*-ratio)
Random Parameters								
Ticket Price	-0.007	(-19.63)	0.005	(23.82)	-0.006	(-20.17)	0.004	(19.33)
Travel Time	-0.004	(-10.13)	0.003	(7.09)	-0.004	(-6.74)	0.003	(3.26)
Seat Pitch	0.113	(4.15)	0.205	(3.81)	0.370	(5.92)	0.476	(4.85)
Carbon Tax	-0.004	(-8.59)	0.004	(7.03)	-0.003	(-7.30)	0.005	(8.46)
Charge	-0.001	(-4.07)	0.002	(2.54)	-0.002	(-3.58)	0.003	(3.64)
Non-Random Parameters								
1st alt. shown	1.831	(13.86)	-	-	1.656	(13.52)	-	-
2nd alt. shown	1.212	(9.51)	-	-	1.184	(9.98)	-	-
3rd alt. shown	0.985	(7.41)	-	-	0.890	(7.16)	-	-
4th alt. shown	1.017	(7.05)	-	-	0.787	(5.80)	-	-
5th alt. shown	0.688	(4.49)	-	-	0.361	(2.35)	-	-
6th alt. shown	0.683	(4.59)	-	-	0.423	(2.80)	-	-
7th alt. shown	0.697	(4.29)	-	-	0.406	(2.51)	-	-
8th alt. shown	0.646	(4.14)	-	-	0.466	(2.96)	-	-
Airline 2	-0.582	(-5.49)	-	-	-0.600	(-5.89)	-	-
Airline 4	-0.246	(-2.13)	-	-	0.080	(0.83)	-	-
Airline 5	-0.687	(-4.55)	-	-	-0.572	(-4.37)	-	-
Airline 6	-0.750	(-4.21)	-	-	-0.684	(-4.16)	-	-
Airline 7	0.440	(2.77)	-	-	0.436	(3.11)	-	-
Airline 8	0.460	(2.16)	-	-	0.348	(1.87)	-	-
Airline 9	1.753	(1.57)	-	-	1.402	(1.64)	-	-
Airline 10	-0.853	(-2.90)	-	-	-1.840	(-6.33)	-	-
Airline 11	-0.527	(-2.58)	-	-	-0.228	(-1.21)	-	-
Airline 12	2.622	(3.08)	-	-	-1.172	(-0.87)	-	-
Entertainment (on demand)	0.377	(4.40)	-	-	0.948	(7.11)	-	-
Entertainment (personal)	0.323	(3.65)	-	-	0.484	(4.01)	-	-
Seat Assignment (1 = yes)	0.234	(3.03)	-	-	0.628	(5.34)	-	-
Number of stops	-0.821	(-5.24)	-	-	0.112	(0.67)	-	-
A330 aircraft	-0.179	(-2.05)	-	-	-0.235	(-1.77)	-	-
Arrive (7-10am)	-0.620	(-4.09)	-	-	-0.249	(-2.42)	-	-
Arrive (11am-2pm)	-0.740	(-3.16)	-	-	-0.325	(-1.86)	-	-
Arrive(3-7pm)	-0.899	(-2.03)	-	-	-1.022	(-2.67)	-	-
Model Fits								
LL(0)	-5712.246				-5712.246			
LL(β)	-3313.299				-3310.062			
Number of Pars.	36				36			
ρ^2	0.420				0.421			
Adj. ρ^2	0.408				0.409			
Observations	462				462			
Respondents	1848				1848			

A number of qualitative attributes required dummy coding in estimation. Firstly, respondents where given the choice of three different entertainment system levels (shared, personal or on-demand). In estimating the model, we treated the shared entertainment system as the base. As is expected, for all models, the shared entertainment system was less preferred than the personal or on demand entertainment systems. For models M1 and M2 estimated on the traditional SP data, the personal entertainment system was not statistically significant and hence removed from final estimation. The personal entertainment system was statistically significant in models M3 and M4 and hence retained. In this case, the personal entertainment system was still preferred to a shared entertainment system but less preferred to an on-demand system.

Different aircraft types were also dummy coded. In all models, the base aircraft type was A340, however the 777 aircraft proved not to be statistically significant in any model and hence was dropped from the analysis. In this way, the 777 also acts as a quasi base for the dummy variable. Departure and arrival times were similarly dummy coded with arrivals after 7pm and departures after 5pm representing the base levels. Where a departure or arrival time dummy was not statistically significant in all models estimated on the SP and/or OTA data, they were removed from the final analysis.

One significant difference between the OTA and traditional SP data lies in the ability of the respondent to sort the alternatives in the OTA data. To account for this, additional dummy variables representing the order that an alternative appears in the final screen used when the respondent made their choice were created. In both models M3 and M4, an option appearing as one of the first eight alternatives shown has a higher likelihood of being chosen than those shown after eight, ceteris paribus, with diminishing impacts within the first eight as one moves from the first shown to the eighth shown.

Table 6 shows the WTP values derived from the four estimated models. Here, special care needs to be taken. In the case of undesirable attributes, such as flight time, the trade-offs are given by a ratio of two negative coefficients, and should be interpreted as a willingness-to-pay for reductions in the undesirable attribute. In the case of a trade-off involving a desirable attribute, such as greater seat pitch, the ratio between the coefficients would be negative. Here, a sign change is required to give the willingness to pay for improvements in this attribute, where for desirable attributes, improvements equate to increases.

For some trade-offs, the situation is however not as clear cut. For example, looking at specific airlines, the associated dummy terms may be positive or negative depending on the airline. Working on the basis of a willingness to pay to fly on a specific airline (as opposed to avoiding flying on a specific airline), we would change the sign of the ratio of the dummy and the cost coefficient, meaning that for a positive dummy (i.e. desirable airline), the trade-off would be positive, while, with a negative dummy (i.e. undesirable airline), the trade-off would be negative, indicating a negative willingness to pay to fly on that airline which equates to a requirement of a reduction in fare to accept flying on that airline.

The explicitly ignored patterns by respondents (which defines M2) by and large tends to increase the WTP estimates when compared to those obtained in model M1. This result partly supports the findings of Wardman (2001) and Brownstone and Small (2005) in that SP data not accounting for such strategies tends to undervalue the WTP values found in real markets. This suggests that accounting for respondent reported information processing strategies, at least in this instance, may be an issue for the traditional SP data set. Nevertheless, for both the traditional SP and OTA data sets, accounting for these processes comes largely at the cost of increasing heterogeneity within the WTP model outputs, rather than reducing it for the models estimated on the SP data. The opposite effect would be expected suggesting issues may exist in using respondent stated information as to the information processing strategies for SP experiments. Similar issues have been reported in Hess and Hensher (2008). A comparison of models estimated on OTA data (M3 and M4) suggests reductions in WTP heterogeneity for some but not all attributes when accounting for what attributes respondents actually saw when making their choices (M4) as opposed to simply assuming they saw all attributes (M3). Given that the information gathered on this data was not obtained by asking the respondent, but rather from observing which attributes they elected to view, this finding is somewhat surprising, as it would be expected that WTP heterogeneity should be reduced for all attributes in model M4 in comparison to those obtained in model M3. A further interesting finding is that by and large, the WTP values for models M1 and M4 tend to be more similar than those from models M2 and M3, although some significant differences to this pattern are observed.

Table 6: WTP results

	M1		**M2**		**M3**		**M4**	
	Average	**Std Dev.**	**Average**	**Std Dev.**	**Average**	**Std Dev.**	**Average**	**Std Dev.**
WTP to avoid undesirable attribute								
Travel Time (hour)	$51.80	$17.17	$57.37	$23.07	$60.23	$20.78	$42.98	$18.06
Number of Stops	$204.79	$152.64	$225.76	$169.18	$171.73	$36.96	-$28.04*	$6.42
WTP to obtain a certain characteristic								
Airline 1	$236.26	$56.78	$274.34	$64.06	-	-	-	-
Airline 2	-	-	-	-	-$121.35	$26.02	-$150.20	$34.40
Airline 3	-$302.00	$72.58	-$168.84	$39.43	-	-	-	-
Airline 4	-	-	$249.29	$58.21	-$51.32	$11.00	$19.92*	$4.56
Airline 5	-$162.39	$39.03	-$176.08	$41.12	-$143.37	$30.75	-$143.19	$32.80
Airline 6	-$181.49	$43.62	-$156.38*	$36.52	-$156.58	$33.58	-$171.39	$39.25
Airline 7	$145.08	$34.87	$223.27	$52.14	$91.80	$19.69	$109.26	$25.02
Airline 8	-	-	-	-	$96.08	$20.60	$87.09*	$19.94
Airline 9	-	-	-	-	$365.74*	$78.43	$351.15*	$80.42
Airline 10	-	-	-	-	-$178.04	$38.18	-$460.86	$105.55
Airline 11	-	-	-	-	-$109.98	$23.58	-$57.08	$13.07
Airline 12	-	-	-	-	$547.16	$117.33	-$293.66*	$67.26
Entertainment (on demand)	$309.93	$285.82	$341.88	$300.45	$78.75	$16.89	$237.46	$54.38
Entertainment (personal)	-	-	-	-	$67.46	$14.47	$121.31	$27.78
Seat Assignment	$121.05	$29.09	-	-	$48.89	$10.48	$157.30	$36.02
Seat Pitch (inch)	$40.41	$9.71	$15.69	$3.75	$17.11	$9.73	$76.83	$56.76
747 aircraft	-	-	$215.41	$50.07	-	-	-	-
A330 aircraft	-	-	$150.02	$34.87	-$37.31	$8.00	-$58.74*	$13.41
Departure (6am)	$86.76*	$76.34	-	-	-	-	-	-
Departure (10am)	-$34.41	$117.18	-$40.98*	$125.52	-	-	-	-
Depart (5pm)	$132.08	$18.73	$258.68	$59.98	-	-	-	-
Arrive (7-10am)	-	-	$225.29	$52.77	-$129.39	$27.75	-$62.31	$14.27
Arrive (11am-2pm)	-	-	$190.19	$44.54	-$154.48	$33.11	-$81.52*	$18.70
Arrive(3-7pm)	-	-	$224.49	$52.58	-$187.63	$40.24	-$256.06	$58.64

* Numerator statistically insignificant

5. Conclusions

In this study we have presented an innovative survey environment for investigating air travel choice behaviour. By mimicking the interface of an OTA, we are able to boost realism and capture additional information on how people handle choice environments that contain large amounts of information. Extensive use was made of the sort, search and hide/show tools, with data from these tools indicating heterogeneity in how respondents processed the realistically large amount of ticket information. Modelling with OTA data that excluded non-viewed attributes resulted in much greater WTP heterogeneity than models that used the SP data. Accounting for ignoring patterns in the SP data led to an increase in most WTP estimates. The survey methodology shows promise as a viable alternative to traditional SP surveys for capturing preference in a variety of travel choice scenarios.

Acknowledgements

The authors would like to thank Jun Zhang for all his work in coding the internet survey.

References

Arentze, T., Borgers, A., Timmermans, H., and DelMistro, R. (2003) Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part E*, 39, pp. 229–244.

Bradley, M. A. (1998) Behavioural models of airport choice and air route choice, In: J. de D. Ort´uzar, D. Hensher & S. R. Jara-Diaz, eds, Travel behaviour research: updating the state of play (IATBR 94), Elsevier, Oxford, pp. 141–159.

Brownstone, D. and Small. K. (2005) Valuing time and reliability: assessing the evidence from road pricing demonstrations, *Transportation Research Part A*, 39, pp. 279-293.

Burke, R.R., Harlam, B.A., Kahn, B.E. and Lodish, L.M. (1992) Comparing Dynamic Consumer Choice in Real and Computer-Simulated Environments, *Journal of Consumer Research*, 19 (June), pp. 71–82.

Carlsson, F and Martinsson, P. (2001) Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environmental Economics and Management*, 41, pp. 179-192.

Carson, R., Louviere, J.J., Anderson, D., Arabie, P., Bunch, D., Hensher, D.A, Johnson, R., Kuhfeld, W., Steinberg, D., Swait, J., Timmermans, H., and Wiley, J. (1994) Experimental Analysis of Choice, *Marketing Letters*, 5 (October), pp. 351-367

Chin, A. T. H. (2002) Impact of frequent flyer programs on the demand for air travel, *Journal of Air Transportation*, 7(2), pp. 53–86.

DeShazo, J.R. and Fermo G. (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency, *J Environmental Economics and Management*, 44(1), pp. 123-143.

Greene, W.H., Hensher, D.A. and Rose, J. M. (2005) Using Classical Simulation Based Estimators to Estimate Individual Willingness to Pay Values, Invited paper for Applications of Simulation Methods in Environmental and Resource Economics, Anna Alberini and Riccardo Scarpa (eds.), Kluwer Academic Publisher, New York.

Hensher DA (2009) Hypothetical bias, stated choice studies and willingness to pay. Spec Issue Transp Res B (in press)

Hensher, D., Stopher, P. R. & Louviere, J. (2001) An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application, *Journal of Air Transport Management*, 7(6), pp. 373–379.

Hess, S. (2007) Posterior analysis of random taste coefficients in air travel choice behaviour modelling, *Journal of Air Transport Management*, paper accepted for publication (February 2007).

Hess, S., T. Adler and J. W. Polak (2007) Modelling airport and airline choice behaviour with stated-preference survey data, *Transportation Research Part E: Logistics and Transportation Review*, 43, pp. 221–233.

Hess, S. and Hensher, D.A. (2008) Using conditioning on observed choices to retrieve individual-specific attribute processing strategies, paper submitted to *Transportation Research Part B*, April 2008.

Jones, P.J. and Bradley, M. (2006). Stated Preference Surveys: An Assessment, Stopher, P.R. and Stecher, C. (eds.) Survey Methods, Elsevier Science, Oxford.

Kanafani, A. & Sadoulet, E. (1977) The partitioning of long haul air traffic – a study in multinomial choice, *Transportation Research*, 11(1), pp. 1–8.

Louviere, J.J., Timmermans, H.J.P. (1990) Hierarchical information integration applied to residential choice behaviour, *Geographical Analysis*, 22, pp. 127–145.

Lusk, J. and Schroeder, T. (2004) Are choice experiments incentive compatible? A test with quality differentiated beef steaks, *American Journal of Agricultural Economics*, 86(2), pp. 467-482.

Nelson, P., and Towriss, J. (1995) Stated Preference Techniques and the Representation of Attributes Presented to the PTRC Summer Conference.

Ortúzar, J.deD. and Willumsen, L.G. (2001) Modelling Transport, 3rd ed., John Wiley and Sons, Chichester.

PhoCusWright (2007) PhoCusWright's U.S. Online Travel Overview Seventh Edition, November 2007.

Proussaloglou, K. & Koppelman, F. S. (1995) Air carrier demand: an analysis of market share determinants, *Transportation*, 22(4), pp. 371–388.

Proussaloglou, K. & Koppelman, F. S. (1999) The choice of air carrier, flight, and fare class, *Journal of Air Transport Management*, 5(4), pp. 193–201.

Rose J.M. and Bleimer, M.C.J. (2008) Stated Preference Experimental Design Strategies, In Hensher, D.A. and Button, K.J. (eds) Handbook of Transport Modelling, Pergamon Press, Oxford.

Rose, J.M. and Hensher, D.A. (2006) Accounting for individual specific non-availability of alternatives in respondent's choice sets in the construction of stated choice experiments, Stopher, P.R. and Stecher, C. (eds.) Survey Methods, Elsevier Science, Oxford.

Train, K. (2003) Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, MA.

Wang, D., Jiuqun, L., Timmermans, H.J.P. (2001) Reducing respondent burden, information processing and incomprehensibility in stated preference surveys: principles and properties of paired conjoint analysis, *Transportation Research Record*, 1768, pp. 71–78.

Wardman, M. (2001) A review of British evidence on time and service quality Valuations, *Transportation Research E*, 37, pp. 91-106.