Journal of Air Transport Management 14 (2008) 275–279

Contents lists available at ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Treatment of reference alternatives in stated choice surveys for air travel choice behaviour

Stephane Hess

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

ARTICLE INFO

Keywords:
Discrete choice
Logit
Air travel behaviour
Stated choice
Stated preference

ABSTRACT

With the need for accurate forecasts of passenger demand, the airline sector is increasingly making use of behavioural models calibrated on data from stated choice surveys that allow for the analysis of hypothetical travel situations. To allow analysts to better frame the scenarios presented to respondents, the choice situations in such stated choice surveys often include a current trip as one of the travel options. Classically, these reference alternatives have been treated in the same way as the hypothetical alternatives. The applications presented in this paper show that this potentially leads to biased results, and that it is important to recognise the differences in the nature of the two types of alternatives.

1. Introduction

Accurate forecasts of air travel demand are a crucial requirement not only for airlines and airports, but also for transport authorities in many major metropolitan areas. Aside from long-term trends in demand, there is great interest in how this demand could be affected by more short-term changes in service levels or service characteristics. Airlines are for example likely to be interested in the potential impact of increases or reductions in air fares on passenger numbers, while airports may want to gauge the impact of reductions in minimum check-in or transfer times on the attractiveness (and hence usage) of an airport. Finally, urban transport planners require accurate forecasts of passenger levels and modal split for the access journeys to airports.

In the context there is growing interest in making use of state of the art modelling techniques to analyse air travel behaviour, with a particular reliance on discrete choice models (DCM) belonging to the family of random utility models (RUM).[1] RUM structures can be calibrated on two main types of data, revealed preference (RP) data describing actual real world choices, and stated choice (SC) data containing choices from hypothetical scenarios presented in travel surveys. With there being a strong interest in predicting behaviour across a range of hypothetical settings (e.g. fare reductions, new routes, new access modes) there is an increasing reliance on SC data.[2]

While SC data have an advantage over RP data in being able to look at choices in hypothetical settings, there has been considerable concern about response quality (Louviere et al., 2000), leading to attempts to increase the realism of SC choice situations. One possibility is to weaken the hypothetical nature of surveys by framing choice situations around a scenario known to the respondent. In a growing number of cases, this is achieved by including the current choice as one of the alternatives in the survey. Evidence suggests that such a framing approach makes preference revelation more meaningful at the level of the individual (Starmer, 2000) and has the advantage of allowing the analyst to determine what kind of incentives are required to get a respondent to move away from their current travel option. One example of exploiting this type of framing approach is the air travel survey data collected by Resource Systems Group Inc. (2003) in the US.

However, while this type of survey design advantages in terms of framing the choice situation, potentially crucial in the context of complex air travel decisions, it is not immediately clear whether standard modelling approaches are appropriate for use on such data. Indeed, the two types of alternatives included in the choice situations in these surveys are inherently different (hypothetical versus actually experienced) and it could be suggested that these differences need to be accommodated in the modelling framework.[3]

E-mail address: s.hess@its.leeds.ac.uk

[1] For a discussion of discrete choice models, see Train (2003).

[2] Recent examples including Proussaloglou and Koppelman (1999), Adler et al. (2005) and Hess et al. (2007).

[3] The issues dealt with here are different from similar discussions of studies combining separate RP and SC datasets in a joint analysis (Ben-Akiva and Morikawa, 1990). The difference lies that, with the present data, a real world alternative is included as one of the options in the SC survey, but only a single dataset is used.

0969-6997/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jairtraman.2008.06.001

2. Data

The analysis makes use of SC data collected via the Internet by Resource Systems Group. Specifically, we make use of the 2005 version of the survey, with a sample of 4256 observations collected from 532 randomly selected travellers who had recently undertaken a domestic air trip. Prior to the SC survey, information was collected on a traveller's most recent air trip, along with detailed socio-demographic information. The traveller is then faced with eight binomial choices, where in each case, a choice is offered between the current flight and an alternative. While the attributes of the reference alternative remain fixed across the eight choice sets, those of the second alternative are varied according to an experimental design. The airports and airlines used for this second alternative are selected on the basis of information gathered from respondents in terms of a ranking of the airports and airlines available to them.

Aside from the airport and airline names, the attributes used to describe the alternatives in the SC survey include flight time, number of connections, air fare, arrival time (used to calculate schedule delays), aircraft type, and on-time performance of the various services. Access cost is not included (in the absence of an actual specification of the mode choice dimension), and no choice is given between travel classes.

3. Methodology

Three main modelling approaches are used, with different degrees of recognising the specific nature of the two alternatives included in the choice sets. All models have a multinomial logit (MNL) structure at the core but normally distributed error components with a zero mean were included to account for each respondent facing with eight choices. These random terms are distributed identically and independently across respondents and alternatives, but not across observations for the same respondent. This allows for an individual-specific effect that can be interpreted as a random scale with the aim of avoiding bias in the standard errors, where such bias commonly exhibits itself in the form of underestimated standard errors when failing to recognise the repeated choice nature of SC data (Ortúzar et al., 2000).

3.1. Base specification

A standard specification is used for the base model, with all parameters entering the utility function in linear fashion. A common coefficient is used for all levels of memberships in frequent flier programmes, and no distinction is made between flights with a single connection and flights with two connections.[4] The observed utility (V) for the reference alternative (R) is given by

$$\begin{aligned} V_{\mathrm{R}} = {} & \beta_{\mathrm{current}} + \beta_{\mathrm{access\ time}} \times \mathrm{access\ time}_{\mathrm{R}} + \beta_{\mathrm{air\ fare}} \times \mathrm{air\ fare}_{\mathrm{R}} \\ & + \beta_{\mathrm{flight\ time}} \times \mathrm{flight\ time}_{\mathrm{R}} + \beta_{\mathrm{OTP}} \times \mathrm{OTP}_{\mathrm{R}} \\ & + \beta_{\mathrm{connecting}} \delta_{\mathrm{connecting,R}} + \beta_{\mathrm{FF}} \delta_{\mathrm{FF,R}} + \beta_{\mathrm{closest\ airport}} \delta_{\mathrm{closest\ airport,R}}, \end{aligned} \tag{1}$$

where all β parameters are to be estimated.

The parameter, β_{current}, is an alternative specific constant (ASC) for the reference alternative that, amongst other things, captures inertia. Parameters, $\beta_{\mathrm{access\,time}}$, $\beta_{\mathrm{air\,fare}}$ and $\beta_{\mathrm{flight\,time}}$ are marginal utility coefficients that capture the disutility associated with an increase by one unit (1 min or \$1) in access time, air fare and flight time. β_{OTP}, relates to the on-time performance (in percentage points) of an alternative. For the reference alternative, two levels are used, depending on whether the flight was on time (100%) or not (zero), while, for the second alternative, five levels between 50% and 90% are used. The variable $\delta_{\mathrm{connecting,R}}$ is set to unity for flights with at least one connection, while $\delta_{\mathrm{FF,R}}$ is set to one if the respondent holds some form of frequent flier (FF) membership with the airline. Finally, $\delta_{\mathrm{closest\,airport,R}}$ is set to unity if the airport used for the trip is that closest to the respondent's home.

The utility function for the second alternative is specified in a similar fashion, with the absence of ASC (β_{current}), and with the hypothetical, as opposed to reference, values for the various attributes and dummy variables.

3.2. Differential response to attribute values of reference alternative

To test the validity of the assumption that respondents treat the attributes of the reference alternative in the same way as those of the hypothetical alternatives, we use a specification in which all coefficients are alternative-specific:

$$\begin{aligned} V_{\mathrm{R}} = {} & \beta_{\mathrm{current}} + \beta_{\mathrm{access\ time,R}} \times \mathrm{access\ time}_{\mathrm{R}} + \beta_{\mathrm{air\ fare,R}} \times \mathrm{air\ fare}_{\mathrm{R}} \\ & + \beta_{\mathrm{flight\ time,R}} \times \mathrm{flight\ time}_{\mathrm{R}} + \beta_{\mathrm{OTP,R}} \times \mathrm{OTP}_{\mathrm{R}} \\ & + \beta_{\mathrm{connecting,R}} \delta_{\mathrm{connecting,R}} + \beta_{\mathrm{FF,R}} \delta_{\mathrm{FF,R}} + \beta_{\mathrm{closest\ airport,R}} \delta_{\mathrm{closest\ airport,R}}. \end{aligned} \tag{2}$$

The corresponding specification for the second alternative again lacks a constant, with the remaining seven coefficients being specific to the alternative. This specification not only allows for differences in how respondents react to the attribute values of the two alternatives, but also accounts for differences in the on-time performance attributes for the alternatives (a simple distinction between on-time and delayed flights for the reference alternatives, with percentage rates of on-time arrival for the second alternative).

3.3. Asymmetrical preference formation

Finally, we develop a model based on concepts taken from prospect theory, where the attribute levels of an alternative are evaluated relative to those of the base alternative (Hess et al., 2008), while allowing for a differential response to increases and decreases (gains and losses) compared to these base levels:

$$V_{\mathrm{R}} = \beta_{\mathrm{current}} \tag{3}$$

and

$$\begin{aligned} V_{\mathrm{S}} = {} & \beta^{+}_{\mathrm{access\ time}} \times \delta_{\mathrm{access\ time\ inc}} \times (\mathrm{access\ time}_{\mathrm{S}} - \mathrm{access\ time}_{\mathrm{R}}) \\ & + \beta^{-}_{\mathrm{acce\ ss\ time}} \times \delta_{\mathrm{access\ time\ dec}} \times (\mathrm{access\ time}_{\mathrm{R}} - \mathrm{access\ time}_{\mathrm{S}}) \\ & + \cdots, \end{aligned} \tag{4}$$

where we only show the coefficients associated with access time.

With this specification, the coefficients in the utility function for the second alternative interact with the difference between the attribute values for the two alternatives. Separate coefficients are used for increases and decreases relative to the attribute value for the base alternative, with $\beta^{+}_{\mathrm{access\ time}}$ and $\beta^{-}_{\mathrm{access\ time}}$, for example, giving the coefficients for increases and decreases in the access time attribute. The additional term $\delta_{\mathrm{access\,time\,inc}}$ is set to one only when the access time is longer for the second alternative than for the base alternative, with the same applying for $\delta_{\mathrm{access\,time\,dec}}$ in the case of decreases relative to the base alternative. The assumption of a symmetrical response can be tested by looking at the difference between coefficients for increases and decreases, say the difference between $\beta^{+}_{\mathrm{access\ time}}$ and $\beta^{-}_{\mathrm{access\ time}}$ in the case of access time.

[4] Very few alternatives with two connections were included.