Attribute Exclusion Strategies in Airline Choice: Accounting for Exogenous Information on Decision Maker Processing Strategies in Models of Discrete Choice

John M. Rose
David A. Hensher
Institute of Transport and Logistics Studies
Faculty of Economics and Business
The University of Sydney
NSW 2006 Australia
J.Rose@its.usyd.edu.au
D.Hensher@its.usyd.edu.au

William H. Greene
Department of Economics
Stern School of Business
New York University
New York USA
wgreene@stern.nyu.edu

Simon P Washington*
Director
Traffic Safety Center
Institute of Transportation Studies
University of California, Berkeley
Simon.washington@its.berkeley.edu

14 June 2009 (Revised 8 February 2010, 2 July 2010)
(in press Transportmetrica)

Abstract

We examine the impact of individual-specific information processing strategies (IPS) on the inclusion/exclusion of attributes on the parameter estimates and behavioural outputs of models of discrete choice. Current practice assumes that individuals employ a homogenous IPS with regards to how they process attributes of stated choice (SC) experiments. We show how information collected exogenous of the SC experiment on whether respondents either ignored or considered each attribute may be used in the estimation process, and how such information provides outputs that are IPS segment specific. We contend that accounting the inclusion/exclusion of attributes will result in behaviourally richer population parameter estimates.

* This research was undertaken when Simon Washington was a Visiting Professor at ITLS, The University of Sydney.

Key Words: Stated choice experiment, behavioural outputs, willingness to pay, information processing strategy

Acknowledgement The suggestions of three referees have materially improved the paper.

Introduction

Stated choice (SC) experiments have become a popular method to model choice behaviour in transportation contexts. The behavioural outputs of SC models, including elasticities, marginal effects and willingness to pay (WTP) estimates have been used extensively to derive demand forecasts for new and existing modes (Jovicic, et al., 2003; Hensher and Rose 2007), to understand and model route choice behaviour (e.g., Jou, 2001; Lam, et al., 2002), to model influences on travel behaviour (e.g., Peeta, 2000), to determine the viability of new infrastructure projects such as proposed toll roads (e.g., de Dios Ortúzar, et al., 2000; Hensher, 2001), and to test the implications on transport systems of proposed policies (e.g., Hensher and King, 2001). Given the risks often associated with such projects and the potential to suffer large financial losses if they fail, it has become increasingly important that the outputs of SC models be both reliable and unbiased estimates of the true population behavioural parameters which they purport to represent.

Aside from their ability to provide asymptotically efficient parameter estimates, one of the key reasons why SC methods have become so popular is their ability to mimic decisions made in real markets that otherwise could not be observed (Burke, et al., 1992; Carson, et al., 1994). Realism in SC experiments is captured through respondents being asked to undertake similar actions as they would in real markets (i.e., respondents are asked to make 'choices' between a finite but universal set of available alternatives, just as in real markets). However, for any individual respondent, realism may be lost if the alternatives, attributes and/or attribute levels used to describe the alternatives do not realistically portray that respondent's experiences or, in terms of 'new' or 'innovative' alternatives, are deemed not to be credible (e.g., Green and Srinivasan, 1978; Green and Srinivasan, 1990; Cattin and Wittink, 1982; Wittink and Cattin, 1989). With regards to the attributes and attribute levels used within a SC experiment, significant prior preparation on behalf of the analyst (including, amongst other things, extensive literature reviews and qualitative research in the form of focus groups and in-depth interviews) may reduce the possible inclusion of irrelevant or improbable product descriptors within the choice sets shown to respondents (Hensher, et al., 2005). Additionally, for quantitative variables, pivoting the attribute levels of the SC task from a respondent's current or recent experience is likely to produce attribute levels within the experiment that are consistent with those experiences, and hence, produce a more credible or realistic survey task for the respondent (see for example, Hensher and Greene, 2003). The selection of what alternatives to include within an SC experiment, whilst somewhat more difficult to manage, may also be handled within the design of SC experiments (Anderson and Wiley, 1992; Batsell and Louviere, 1991; Lazari and Anderson, 1994; Rose and Hensher, 2006).

Typically, SC studies have tended to rely on single pre-specified experimental designs with fixed numbers of alternatives, attributes and attribute levels. Significant research effort has therefore been expended on how to optimise the outputs derived from respondents completing choice tasks derived from these single design plans generated using statistical design theory (e.g., Bunch, et al., 1994; Huber and Zwerina, 1996, Kanninen, 2002; Kuhfeld, et al., 1994, Lazari and Anderson, 1994; Sandor and

Wedel, 2001), whilst minimising the amount of cognitive effort required of respondents (e.g., Louviere and Timmermans, 1990; Oppewal, et al., 1994; Wang, et al., 2001; and Richardson, 2002). Yet these research efforts appear to have developed without adequate recognition that respondents perhaps process SC tasks differently (i.e., there may exist heterogeneity in the information processing strategies (IPS) employed by respondents; this is borne out by the not uncommon observance of lexicographic choice behaviour in segments of respondents completing SC surveys) and therefore should be tailored to be as realistic as possible at the level of the individual respondent.

Adaptive-Choice-Based-Conjoint (e.g., see Toubia, et al., 2004) customizes the attribute levels of a SC experiment shown to a respondent using the previous choices made. This, however, is not the same as customising the actual alternatives or attributes in order to make the choice task more realistic or believable to the individual respondent. Rose and Hensher (2006) address the mapping of alternatives in terms of their presence or absence in reality to choice experiments at the individual respondent level, however, research addressing the presence or absence of attributes at the individual level is only just beginning to be considered within the literature (see Hensher 2010, Hensher and Rose 2009, Scarpa et al., 2009). This is somewhat surprising given that in real markets, there will likely exist heterogeneity in the information held with regards to the attributes and attribute levels of alternatives amongst decision makers, as well heterogeneity in terms of the salience of and preference for specific attributes. For example, one respondent may have perfect information with regards to the torque of alternative vehicles and posses a positive marginal utility for the attribute, whilst a second respondent may have no understanding of the attributes meaning (indeed, some respondents may not realise that such an attribute actually exists, or if so, whether more or less torque is desirable) and hence possess no marginal utility for the attribute at all. SC experiments assume that all respondents have perfect information (at least on the attributes included within the experiment) and that all respondents process these attributes in the same way.

Whilst advances in the econometric modelling of discrete choices, in the form of latent class and mixed logit models, may help in uncovering preference heterogeneity for attributes, experience suggests that, depending on the random parameter distribution, these models will likely assign non-zero parameter estimates to individual decision makers, even though their marginal utility for an attribute may strictly be zero[1]. Whilst this might apply to only a small number of decision makers, a bias in the population parameter estimates is still likely to exist. Therefore, the econometric models used to estimate SC outputs need to be conditioned to assign to those individuals who either ignore an attribute or do not have that attribute present, a zero parameter estimate.

Rather than rely solely on econometric models to uncover different IPS strategies and preference heterogeneity, an alternative strategy is to tailor the choice experiment to the individual so that each choice set includes only those alternatives and attributes that the respondent would have access to information on in real markets and which they would likely use in making their choices. Whilst we advocate this as the

[1] This will particularly be the case if the constrained triangular or log-normal distributions are used. Whilst these distributions force the parameter estimates to be of the same sign, they also ensure that few, if any, individual specific parameter estimates will be zero.

preferred strategy, the focus of this paper is to examine how we can use exogenous information on the IPS strategies employed by individual respondents undertaking SC tasks and how we can use such information to condition the parameter estimates derived from the econometric models fitted. Additional non-design information that may be captured in SC surveys to assist in revealing the IPS include the inclusion/exclusion plan for each attribute as well as an aggregation plan (e.g., the adding up of attributes such as components of travel time). In this paper, we concentrate only on the information processing attribute inclusion/exclusion strategy employed by individual respondents. In doing so, we recognise that as we start to appreciate the role of a set of IPS's in representing choice behaviour (in contrast to the default fully compensatory IPS), that we are adding an additional layer of complexity in model estimation and model application. In terms of applications, we would need supplementary models to identify the probability of a specific IPS being used by each individual in the application population. Given the growing evidence (as set out succinctly in Hensher and Layton 2010), that accommodating a heterogeneous set of IPS does impact non-marginally on WTP and elasticity outputs, this is a layer of complexity that cannot be assumed away.

The paper is organised as follows. The next section outlines the econometric model used in the paper. A brief overview of the empirical data used is then given followed by the set of model results comparing traditional SC models with those conditioned using information on the information processing attribute inclusion/exclusion strategies used. The substantive implications of the analysis are set out followed by some conclusions and directions for ongoing research.

Model Development

Consider a situation in which q=1,2,…,Q individuals evaluate a finite number of alternatives. Let subscripts j and t refer to alternative j=1, 2, …, J and choice situation t=1,2, ...,T. Random utility theory (RUT) posits that the utility for alternative j present in choice situation t may be expressed as:

$$U_{jtq} = \theta_q^{'} x_{jtq} + \varepsilon_{jtq} \tag{1}$$

where
U_{jtq} is the utility associated with alternative j in choice situation t held by individual q,
x_{jtq} is a vector of values representing attributes belonging to alternative j, characteristics associated with sampled decision makers q, and/or variables associated with context of the choice situation, t, and ε_{jtq} represents unobserved influences on utility. $\theta_q^{'}$ is a vector of parameters such that θ=θ_1,θ_2,…,θ_K where K is the number of parameters, corresponding to the vector x_{jtq}.

In the most popular choice model, multinomial logit, the probability that alternative i will be chosen is given as:

$$P_{itq} = \frac{\exp\left(V_{itq}\right)}{\sum_{i=1}^{J}\exp\left(V_{itq}\right)}, \tag{2}$$

where

$$V_{jtq} = \theta_q^{'} x_{jtq}. \quad (3)$$

Assuming a sample of choice situations, t = 1, 2, ..., T, has been observed with corresponding values x_{jtq}, and letting i designate the alternative chosen in situation t, the likelihood function for the sample is given as

$$L(\theta) = \prod_{t=1}^{T} P_{itq} \quad (4)$$

and the log likelihood function of the sample as

$$L^{*}(\theta) = \ln\left[L(\theta)\right] = \sum_{t=1}^{T} \ln\left(P_{itq}\right). \quad (5)$$

Equation (5) may be re-written to identify the chosen alternative i:

$$L^{*}(\theta) = \sum_{t=1}^{T}\left[V_{itq} - \ln\left(\sum_{j} e^{V_{jtq}}\right)\right]. \quad (6)$$

Given that θ is unknown, it must be estimated from the sample data. To do this, we use the maximum likelihood estimator of θ which is the value of $\hat{\theta}$ at which L(θ) is maximised. In maximising equation (6), it is usual to use the entire set of data for V_{jtq}. That is, it is assumed that across all t, all V_{jtq} and hence x_{jtq} are considered and as such, the levels assumed by each x in the x_{jtq} matrix are used in determining the value at which $\hat{\theta}$ maximizes the likelihood estimator of θ.

Assuming that over a sample of choice situations, t, not all k variables within the x_{jtq} vector are considered in the decision process, the value of $\hat{\theta}$ which is conditioned on the assumption that all x_{jtq} are considered, will likely be biased. For those choice situations in which an attribute, k, is excluded from consideration in the choice process, $\hat{\theta}_k$ should be equal to zero. Note that this is not the same as saying that the attribute itself should be treated as being equal to zero[2].

In cases where attribute k is indicated as being excluded from the decision process, rather than set the value for the k^{th} element in the x_{jtq} vector to zero and maximising

[2] To demonstrate, consider the situation where attribute x_{jtq} is the price for alternative j in choice situation t. For all but giffen goods, setting the price to equal zero will likely make that alternative much more attractive relative to other alternatives in which the price is not equal to zero. Further, the procedure for maximising $L^{*}(\theta)$ will be ignorant of the fact that setting $x_{jtq} = 0$ represents the exclusion of that attribute in the choice process and will estimate a value of $\hat{\theta}_k$ assuming that the value observed by the decision maker in choice situation t was zero for that attribute when indeed it was not. As such, setting $x_{kjt} = 0$ will not guarantee that the parameter for that attribute will be equal to zero for that choice situation. It is therefore $\hat{\theta}_k$ that should be set to zero in the estimation process, not x_{kjt}.

equation (6), the search algorithm in searching for the maximum of equation (5), excludes that x from the estimation procedure altogether and automatically assigns to it a parameter value of zero. The parameter estimate, $\hat{\theta}_k$, is then estimated solely on the sample population for which the variable was not excluded. In this sense, the process is analogous to selectivity models (which censors the distribution, as distinct from truncation). To demonstrate, consider a simple example in which there are only two variables, x_1 and x_2, associated with each of j alternatives. Denote N as the number of attribute processing strategies such that n = 1 represents those decision makers who consider only x_1 in choosing between the j alternatives, n = 2 represent those decision makers who consider only x_2, and n = 3 represent those decision makers who consider both x_1 and x_2. The likelihood is defined by the partitioning of observations based upon subset membership defined above. The likelihood function is therefore given as:

$$L^*(\theta) = \sum_{t=1}^{T}\sum_{n=1}^{N} \ln\left(P_{itq}\right). \tag{7}$$

The derivatives of the log likelihood for groups n_1 and n_2 have zeros in the position of zero coefficients and the Hessians have corresponding rows and columns of zeros. This partitioning of the log-likelihood function may be extended to any of the logit class of models, including the nested logit and mixed logit family of models. In the next section, we discuss the empirical application in which we estimate models of the form described above.

Empirical Application

The data reported herein was collected as part of a larger study examining differences between the temporal partitioning of the administration of stated choice data. The empirical setting for the study is a labelled SC experiment, the context of which was the choice of airline carrier for an interstate holiday. The experiment involved four alternatives, three labelled alternatives and a no choice alternative. Each labelled alternative was described by four attributes, each further described by four attribute levels. Within the labelled experiment, three existing airlines were named as part of the experiment. The first airline, which we report as Airline A, represents the dominant domestic airline carrier in Australia. The second airline, (Airline B) is an international carrier that is perceived within the Australian domestic airline market as the being the budget carrier. The third alternative airline (Airline C) within the experiment is a dominant international airline that competes with Airline A in terms of offering similar service levels within the marketing mix. Given that the experiment was a labelled choice experiment, the smallest possible experimental design (capable of estimating non-linear main effects in the marginal utilities of each attribute) consists of 16 treatment combinations (see Rose and Bliemer, 2007). Rather than generate a design with 16 treatment combinations, a $4^{(3\times4)}$ orthogonal fractional factorial experimental design with 40 treatment combinations was generated. This design allows for the estimation of non-linearities in the marginal utilities over the attribute levels for all main effects. The attributes and attribute levels are shown in Table 1.

Table 1: Attribute and attribute levels

Attribute	Attribute levels
Ticket Price	\$79, \$99, \$119, \$139
Flight Time (minutes)	40, 50, 60, 70
Departure Time	6.00am, 10.00am, 2.00pm, 6.00pm
Flight Time Variability	±5%, ±7.5%, ±10%, ±12.5%

In addition to the attribute columns, two additional orthogonal blocking columns were generated as part of the experimental design. The first blocking column of two levels, divided the design into two orthogonal halves. The second blocking column of four levels, divided the design into four orthogonal quarters. These two blocking columns were used to establish two of three experimental conditions. The first experimental condition, involving neither blocking column, consisted of respondents completing the entire design in a single session (i.e., respondents completed all 40 choice sets in one sitting). The second experimental condition, using the first blocking column, saw respondents complete the entire experiment over two sessions, completing each half fraction of the experiment as determined by the blocking column, spaced one week apart. The second blocking column was used in the third experimental condition, with respondents asked to complete each of the four quarters in separate sessions spanning a four week time frame. In each condition, the order of choice sets was randomized so as to avoid order effect biases. A second non-labelled choice experiment involving mobile phone choice was also conducted at the same time using the same principles described above (see Rose and Black, 2006).

Two hundred and thirty two first and second year marketing undergraduate students were recruited to complete the experiment. Recruited students were randomly assigned to one of the three experimental conditions. Of the 232 students, 61 were randomly assigned to the first experimental condition, 81 to the second experimental condition and 90 to the last experimental condition. Greater numbers of students were assigned to each successive experimental condition so as to compensate for expected attrition over sessions. Table 2 shows the number of respondents completing each experimental condition of the study and the number of observations thus obtained. Percentages shown represent the within condition completion/non-completion rates.

Table 2: Attribute and attribute levels

Condition (choice sets per condition)	Number of choice sets completed	Number of respondents	Number of choice observations
1 (40)	40	61 (100%)	2440
2 (20)	40	55 (67.9%)	2200
2 (20)	20	26 (32.1%)	520
3 (10)	40	34 (37.78%)	1360
3 (10)	30	29 (32.22%)	870
3 (10)	20	12 (13.33%)	240
3 (10)	10	15 (16.67%)	150
		Total	7780

Table 3 shows the demographic breakdown of the sampled respondents by experimental condition. The vast majority of those completing the survey were female and currently owned or had owned a mobile phone in the past.

Table 3: Demographic breakdown of sample

Condition (choice sets per condition)	Number of choice sets completed	Age (average)	Gender (female)	Percentage having experience with a mobile phone
1 (40)	40	20.54	47.46%	77.97%
2 (20)	40	20.30	64.00%	78.00%
2 (20)	20	20.97	65.63%	71.88%
3 (10)	40	20.38	61.54%	76.92%
3 (10)	30	19.9	65.52%	82.76%
3 (10)	20	19.9	72.73%	72.73%
3 (10)	10	20.3	58.33%	91.67%

Upon completing the choice tasks for a session, sampled respondents were asked which attributes they had ignored in making the choices that they had made whilst undertaking the choice experiment. The response metric for this question was a simple binary yes/no for each attribute. Although we use a simple binary indicator to define the inclusion or exclusion of an attribute in an individual's information processing strategy, we do not attribute the reason for the response herein, which could be due to cognitive burden or simply relevance (see Hensher, 2004). Table 4 summarises the number of times each attribute was stated as being ignored over experimental conditions. Ticket price was ignored in the choice process the least number of times and flight time variability the most number of times. Over the sample, flight time and departure time were ignored approximately the same number of times. Significantly, a check of the data showed no respondent ignored all attributes in making their choices.

Table 4: Number of respondents who did not consider an attribute by experimental condition
(Percentage of respondents in which an attribute were excluded from the choice process shown in brackets)

Condition (choice sets per condition)	Ticket Price	Flight Time (minutes)	Departure Time	Flight Time Variability
1 (40)	7 (11.48%)	12 (19.67%)	16 (26.23%)	37 (60.66%)
2 (20)	1 (1.82%)	8 (14.55%)	6 (10.91%)	40 (72.73%)
2 (20)	1 (3.85%)	6 (23.08%)	3 (11.54%)	21 (80.77%)
3 (10)	9 (29.03%)	6 (19.35%)	10 (32.26%)	21 (67.74%)
3 (10)	6 (20.69%)	5 (17.24%)	8 (27.59%)	15 (51.72%)
3 (10)	0 (0.00%)	0 (0.00%)	0 (0.00%)	10 (76.92%)
3 (10)	1 (6.67%)	8 (53.33%)	2 (13.33%)	9 (60.00%)
Total ignored	25 (10.78%)	45 (19.40%)	45 (19.40%)	153 (65.95%)

Empirical Results

Table 5 presents the model results for the experiment. The first two models were estimated using all data irrespective of whether a sampled individual indicated whether they had ignored an attribute throughout the experiment or not. This represents current practice whereby it is assumed that all attributes are relevant (to varying degrees) to all sampled respondents. The final two models are estimated using the procedure described earlier. Models 1 and 3 are MNL models, models 2 and 4 are mixed logit (ML) models. All four models were estimated using the pooled choice data from all three experimental conditions, irrespective of whether all 40 choice sets were completed or not. While we recognise that pooling data with different numbers of choice sets may have some influence on the parameter estimates, this is not the issue of interest in this current paper. Research by Hensher (2006), amongst others, has not found any systematic biases in parameter estimates due to the number of choice scenarios evaluated. Bias tends to be influenced by attribute range and levels which were the same throughout the entire data set. We have allowed for correlation between choice scenarios for each individual; however given the almost instantaneous nature of the 'panels' we suggest that the correlated structure has already been adequately handled, given it is still a very short 'panel'.

Table 5: Summary of Empirical Results for Models 1 through 4

(Random Parameters mean = spread parameter)

	Full Data				Partial Data			
	MNL		MMNL		MNL		MMNL	
	Coeff.	(*t*-ratio)	Coeff.	(*t*-ratio)	Coeff.	(*t*-ratio)	Coeff.	(*t*-ratio)
Ticket Price	-0.054	(-52.87)	-0.107	(-32.26)	-0.036	(-41.41)	-0.035	(-32.69)
Flight Time	-0.027	(-18.78)	-0.041	(-20.13)	-0.016	(-14.67)	-0.022	(-16.04)
Flight Time Variability	0.483	(0.78)	0.378	(0.46)	-6.488	(-11.26)	-9.905	(-12.56)
Departure Time (6am)	-0.533	(-17.36)	-0.686	(-17.29)	-0.424	(-8.26)	-0.452	(-8.29)
Departure Time (10am)	0.437	(14.16)	0.617	(14.49)	0.488	(9.86)	0.448	(8.78)
Departure Time (12pm)	0.089	(2.96)	0.121	(3.31)	0.159	(3.21)	0.123	(2.39)
Non-Random Parameters								
Constant A	7.171	(50.33)	14.814	(33.88)	4.650	(41.08)	5.220	((32.69)
Constant B	7.245	(50.97)	14.865	(34.17)	4.731	(41.66)	5.304	(33.08)
Constant C	6.952	(49.48)	14.451	(33.87)	4.490	(39.75)	5.061	(31.82)
Model Fits								
LL(0)	-10785.370		-10785.370		-10785.370		-10785.370	
LL(β)	-8538.611		-8158.476		-9502.17		-9441.21	
χ^2	4493.519		5253.788		2566.401		2688.324	
ρ^2	0.199		0.243		0.118		0.124	
Observations	7780		7780		7780		7780	

Marginal Effects				
	Airline A	*Airline B*	*Airline C*	*None*
Ticket A	-2.988	-3.376	-2.046	-1.737
Ticket B	-2.923	-3.325	-1.999	-1.698
Ticket C	-3.098	-3.666	-2.095	-1.698
Flight Time A	-0.828	-0.853	-0.449	-0.568
Flight Time B	-0.809	-0.840	-0.441	-0.557
Flight Time C	-0.851	-0.906	-0.462	-0.589
Flight Time Variability A	0.023	0.012	-0.122	-0.173
Flight Time Variability B	0.023	0.013	-0.120	-0.171
Flight Time Variability C	0.025	0.014	-0.128	-0.181
Departure Time (6am) A	-0.103	0.010	-0.068	-0.011
Departure Time (6am) B	-0.109	0.016	-0.071	-0.012
Departure Time (6am) C	-0.104	0.006	-0.066	-0.007
Departure Time (10am) A	0.084	-0.012	0.079	0.020
Departure Time (10am) B	0.089	-0.009	0.081	0.022
Departure Time (10am) C	0.085	0.019	0.076	0.027
Departure Time (12pm) A	0.017	-0.001	0.025	0.004
Departure Time (12pm) B	0.018	-0.002	0.026	0.004
Departure Time (12pm) C	0.017	0.001	0.024	0.004

For all four models, all parameters associated with the design attributes are specified as generic random parameter estimates. With the exceptions of the flight time variability parameters of models 1 and 2, all parameters associated with the design attributes are statistically significant and of the expected sign. In specifying the ML models, the parameters associated with the design attributes were drawn from a constrained triangular distribution. Hensher and Greene (2003) have shown that for the triangular distribution, when the mean parameter is constrained to equal its spread (i.e., $\beta_{jk} = \beta_k + |\beta_k|\ T_j$, where T_j is a triangular distribution ranging between -1 and +1), the density of the distribution rises linearly to the mean from zero before declining to zero again at twice the mean. Therefore, the distribution must lie between zero and some estimated value (i.e., the β_{jk}). As such, all individual specific parameter estimates are constrained to be of the same sign. Empirically the distribution will be symmetrical about the mean which not only allows for ease of interpretation, but also avoids the problem of long tails often associated with drawing from a log-normal distribution. The random parameter estimates of the ML models were drawn using 500 Halton draws.

Comparison of models 1 and 2, and 3 and 4 reveal significant differences in the parameter estimates of the four models. The parameter estimates for the ticket price and flight time attributes for models 1 and 2 suggest that when the information processing attribute inclusion/exclusion strategy is not accounted for, the sample population is much more sensitive to both increases in price and flight times than when the information processing attribute inclusion/exclusion strategy of sampled respondents is considered during the modelling process. The flight time variability parameter estimates which were not significant and of the incorrect sign when information processing attribute inclusion/exclusion strategy is ignored become highly significant and of the correct sign when estimated only for those who considered the attribute. This clearly illustrates that including or excluding attributes is an important segmentation criterion. The departure time attribute, which was effects coded (see Hensher et al., 2005), produces roughly similar population moments whether all data is used in the estimation process or only data for those who considered the attribute during the choice experiment.

In interpreting the parameter estimates for models 3 and 4, it is important to note that the parameter estimates are specific only to sample population segments who consider an attribute whilst undertaking the choice experiment. For those who do not consider an attribute, the parameter estimate for that individual is zero. As such the parameter estimates of models 3 and 4 are not inclusive of the entire sample population. That is, the parameter estimates are specific to each information processing attribute inclusion/exclusion strategy. In terms of segmentation and benefits studies, this is an important development. Assuming that respondents only consider attributes which they perceive a benefit when making choice decisions, the parameter estimates shown in models 3 and 4 may be interpreted as those for the specific needs benefits segments. In traditional models, these IP or benefits segments may be lost if the segment is small relative to the total population size. This is demonstrated with the flight time variability attribute in which only a small segment of the sampled population considered this attribute in the choice process. When the parameters are estimated ignorant of the information processing attribute inclusion/exclusion strategy employed, the flight time variability parameter is not significant (indeed it is of the wrong sign) which would result in the analyst wrongly assuming that the parameter is not important in the choice process for the entire population when in fact, for a small proportion of the sampled population, the attribute is a highly significant determinant of airline choice.

The impact of the information processing attribute inclusion/exclusion strategy carries through to the behavioural outputs derived from models of discrete choice. As well as the parameter estimates, Table 5 shows the direct marginal effects for the four estimated models. Supporting our earlier observations, ignoring the information processing attribute inclusion/exclusion strategy employed by sampled respondents tends to increase the sensitivities for the sampled population to increases in airline ticket prices and flight times. Indeed, the marginal effects for model 4 are approximately half those for model 2. Non-marginal changes are observed for the marginal effects for the flight time variability attribute when the information processing attribute inclusion/exclusion strategy is accounted for in the model estimation process compared to when the information processing attribute inclusion/exclusion strategy is ignored. Only marginal changes are observed within the magnitudes of the departure time effects coded attribute however several sign reversals are noted.

Figure 1 shows the willingness to pay (WTP) distributions for the flight time attribute estimated from the two ML models reported in Table 5. These WTP distributions were derived from the conditional individual specific parameter estimates obtained using methods outlined in Train (2003) and Greene et al. (2005). At the individual specific level, the estimation procedure assigns a zero parameter estimate to those who did not consider an attribute but assigns a parameter estimate from the assigned distribution for those who did, using the procedures described in Train (2003) and Hensher, et al. (2004). For derivation of WTP distributions, this poses problems if one or both of the parameters in the WTP ratio are equal to zero. If the cost parameter is equal to zero, the denominator of the ratio is equal to zero and the WTP measure becomes infinite. This is similarly the case if both parameters are equal to zero. If on the other hand, the parameter located in the numerator of the WTP calculation is zero, the WTP estimate becomes zero. These issues do not arise if the information processing attribute inclusion/exclusion strategy is not accounted for in the estimation process. In deriving the WTP distributions shown in Figure 1, we have removed those WTP measures which are infinite or which are equal to zero due to one or more of the individual-specific conditional parameter estimates are equal to zero. We discuss this in a later section. The WTP based on individual parameters are summarised in Table 6 for the mixed logit models. All WTP have a distribution in the positive range.

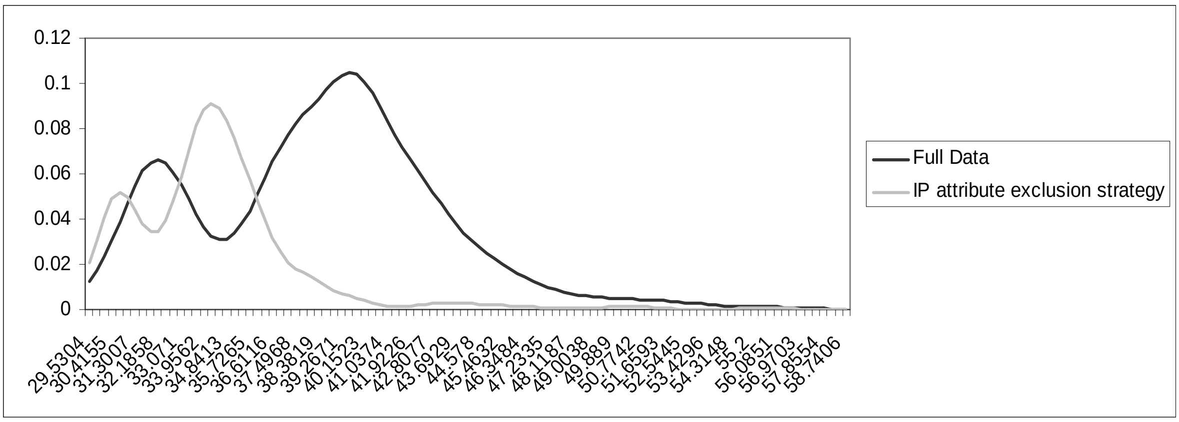

Figure 1: Willingness to Pay Kernel Density Functions for Flight Times

Table 6: Summary of Empirical WTP values from models 2 and 4

	WTP	Mean	Stand. dev.	Range
Full data	Flight Time (minutes)	\$25.14	\$7.10	\$16.20-\$76.16
IP att. exclusion strategy	Flight Time (minutes)	\$38.4	\$11.75	\$3.26-58.34

The above analysis demonstrates the importance of accounting for the processing strategies of respondents undertaking SP work, where the current paper has made use of exogenous information related to the ignoring patterns of respondents during the choice survey. Questions remain however as to how such results can be translated into the real world beyond the modelled results. For example, the above models, which for example have identified that a small minority of respondents in a choice survey are sensitive to flight time variability, will not assist airline managers to understand precisely who these people are, and hence will not allow for any direct marketing campaigns to target such individuals. To overcome this shortcoming, it is possible however to estimate a series of binary logit models based on the surveyed respondents stated IPSs when additional covariate information, also captured in the data, is used as independent variables. Using binary logit models where the dependent variable is one if an attribute was ignored and 0 otherwise, it is possible to determine whether particular covariate classes are more or less likely to ignore a specific attribute in making their choices. In the current dataset we do this using the respondent's age and gender as well as an interaction term between the two. Results for the four binary logit models are presented in Table 7.

Table 7: Binary Logit models of ignoring patterns

	Ignore Ticket Price		Ignore Flight Time		Ignore FT Var.		Ignore Dep. Time	
	Par.	(*t*-ratio)	Par.	(*t*-ratio)	Par.	(*t*-ratio)	Par.	(*t*-ratio)
Constant	-4.290	(-3.24)	-2.604	(-1.61)	-1.705	(-1.02)	-2.896	(-1.86)
Age	0.112	(1.77)	0.067	(0.86)	0.019	(0.24)	0.170	(2.22)
Gender (female = 1)	-0.905	(-2.61)	-7.262	(-2.90)	-2.104	(-1.99)	5.020	(2.56)
Age×Gender	-0.448	(-1.28)	0.335	(2.79)	0.097	(0.94)	-0.233	(-2.42)
Model Fits								
LL(0)	-137.507		-239.746		-241.120		-303.321	
LL(β)	-131.837		-224.684		-238.106		-298.993	
ρ^2	0.041		0.063		0.013		0.014	
Adj. ρ^2	0.033		0.055		0.004		0.006	

From Table 7, it can be seen that, ceteris paribus, females are less likely than males to ignore ticket prices, flight time and flight time variability, but are much more likely to ignore departure times. Examining the Age×Gender interaction however suggests that older females are more likely to ignore flight time but are less likely to ignore departure time than younger females. Irrespective of gender, older respondents are more likely to ignore departure time than younger respondents but no differences exist between ages in terms of the remaining attributes. Given such results, airline managers may specifically target their marketing campaigns and pricing strategies at certain specific groups.

Discussion and Conclusion

This paper has examined the issue of information processing attribute inclusion/exclusion strategies and their effect on the parameter estimates and other behavioural outputs of models of discrete choice. We have shown that accounting for individual specific information on the processing of attribute inclusion/exclusion strategies results in significant differences in the parameter estimates and other behavioural outputs of models of discrete choice. These differences arise from a form of respondent segmentation, the basis of which is respondent IPS. Through partitioning the log-likelihood function of discrete choice models based on the IPS of individual respondents, the outputs of the models we estimate, represent those of these IPS segments, rather than those of the entire sample population. In this way, we are able to detect the preferences for different segments within the sample population based on the IPS strategies existing within the sampled population. In traditional choice models, such segments will likely go undetected.

The particular question asked as to whether an attribute should be excluded from model estimation for a specific respondent is critical to the method and the results. We recognise that there may be other ways of defining the behavioural rule for including or excluding an attribute. We also recognise that it is important to understand whether the attribute was excluded simply because of cognitive burden in the survey task in contrast to a genuine behavioural exclusion in respect of the relevance of the attribute in making such choices in real markets. It could be the case that cognitive burden associated with the survey instrument may indeed be real but so it can be in real markets with information acquisition and processing and so care is required in separating out and accounting for all these reasoning processes. Clearly they are all legitimate members of an individuals IPS.

Ultimately, our preferred strategy would be to tailor the SC experiment to the individual based on the IPS of the respondent. How best to do this is a matter of research. One question is whether the IPS strategy should be determined a priori and the SC experiment fixed for each respondent over the course of the experiment or whether the IPS strategy is determined for each distinct choice set. The former approach is appealing for reasons of simplicity, the latter for completeness given that the IPS strategy may be linked not only to the attributes, but the attribute levels of the experiment. The strategy we outline here, whereby we employ an SC experiment derived from a single design plan, represents the more traditional approach to conducting SC experiments; however, we are able to account for the IPS strategy exogenously, without having to tailor the SC experiment to each individual. Research, however, is required as to whether it is best to ask each respondent which attributes were ignored at the end of the experiment, as we did here, or upon completion of each choice task. As with the tailoring of the SC task, the former approach is appealing for reasons of simplicity as well as the probable limiting of cognitive burden experienced by respondents, whilst the later may represent a more complete approach, given that the attributes that are ignored or considered may be a function of the attribute levels of the alternatives as well as a function of experience or fatigue as the number of choice tasks completed increases. Whatever is the subsequent empirical evidence, we strongly promote the need to recognise that the choice experiment per se can influence the selection of an IPS, and hence the choice response, and adjustments mad where evidence from alternative choice experiments are being compared. Hensher (2006a) developed a formula to make such adjustments.

A further point of future research interest is on the WTP estimates derived from the processes outlined within this paper. In this paper, we have excluded the WTP outputs for those who stated they ignored one or both of the two attributes used in the WTP calculation. Whilst it may be impossible to do anything else, this does not suggest that such individuals WTP measures are zero (or infinity). What is known is that the WTP measures of such individuals are zero (or infinity) given the attribute levels used in the SC experiment. Had other levels been used, these individuals may have a non-zero (or non-infinite) WTP measure. This may point to a miss-specification of the attribute levels used in the experiment, which if shown to have occurred, suggests that WTP measures derived from models which fail to account of individual specific IPS strategies, are likely to be biased. This is an important issue, still needing to be satisfactorily resolved. One promising avenue for future exploration in this area may be the estimation of models in WTP-space rather than preference space (as recently undertaken by Train and Weeks 2005, Scarpa et al., 2008, and Hensher and Greene in press).

We conclude by noting that the proposed modelling approach discussed here applies equally to models estimated using revealed preference (RP) data. Researchers collecting RP data must pre-specify the data collected and assume, as with SC data, that the attributes of RP data are processed homogenously over the sampled population. As with SC data, this need not be the case.

References

Anderson, D.A. and Wiley, J.B. (1992) Efficient Choice Set Designs for Estimating Cross-Effects Models, *Journal of Marketing Research*, 3 (4), 357-370.

Batsell, R.R. and Polking, J.C. (1985) A New Class of Market Share Models, *Marketing Science*, 4, 177-198.

Bunch, D.S., Louviere, J.J., and Anderson D. (1996) A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models, Working Paper, Graduate School of Management, University of California, Davis.

Cattin, P. and Wittink, D.R. (1982) Commercial Use of Conjoint Analysis: A Survey, *Journal of Marketing*, 46 (3), 44-53.

de Dios Ortúzar J., Iacobelli, A. and Valeze, C. (2000) Estimating demand for a cycle-way network, *Transportation Research A*, 34 (5) 353-373.

Green, P.E., and Srinivasan, V. (1978) Conjoint Analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research*, 5 (2), 103-23.

Green, P.E., and Srinivasan, V. (1990), Conjoint Analysis in Marketing Research: New Developments and Directions," *Journal of Marketing*, 54, (4), 3-19.

Greene, W., Hensher, D.A. and Rose, J.M. (2005) Using Classical Simulation Based Estimators to Estimate Individual Willingness to Pay Values: A Mixed Logit Case Study of Commuters, in *Applications of Simulation Methods in Environmental and Resource Economics*, Riccardo Scarpa and Anna Alberini editors, Springer Publisher, Dordrecht, The Netherlands, 17-34.

Hensher, D.A. (2001) Measurement of the valuation of travel time savings, *Journal of Transport Economics and Policy* (Special Issue in Honour of Michael Beesley), 35 (1), 71-98.

Hensher, D.A. (2004) Accounting for Stated Choice Design Dimensionality in Willingness to Pay for Travel Time Savings, *Journal of Transport Economics and Policy*, 38 (2), 425-446.
Hensher, D.A. (2006) How do Respondents Process Stated Choice Experiments? – Attribute consideration under varying information load, *Journal of Applied Econometrics*, 21 , 861-878
Hensher, D.A. (2006a) Towards a method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions, Transportation Research A. 40 (10), 829-840.
Hensher, D.A. (2010) Attribute Processing, Heuristics and Preference Construction in Choice Analysis. *Invitational Keynote Paper for Choice Modelling Conference*, Leeds UK. March 30-April 1 2009 (Linked ARC DP 0770618; 07-09) Synthesis paper. To appear in Hess, S. and Daly, A. (eds.) *Choice Modelling*, Emerald Press, UK.
Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: state of practice, *Transportation*, 30 (2), 133-176.
Hensher, D.A. and Greene, W.H. (in press) Valuation of Travel Time Savings in WTP and Preference Space in the Presence of Taste and Scale Heterogeneity, *Journal of Transport Economics and Policy*
Hensher, D.A. and King, J. (2001) Parking demand and responsiveness to supply, pricing and location in the Sydney central business district, *Transportation Research A*, 35 (3) 177-196.
Hensher, D.A. and Layton, D. (2010) Parameter Transfer of Common-Metric Attributes in Choice Analysis and Cognitive Rationalisation: Implications for Willingness to Pay *Transportation*
Hensher, D.A. and Rose, J.M. (2007) Development of Commuter and non-Commuter Mode Choice Models for the Assessment of New Public Transport Infrastructure Projects: A Case Study, *Transportation Research A*, 41 (5), 428-433.
Hensher, D.A. and Rose, J.M. (2009) Simplifying Choice through Attribute Preservation or Non-Attendance: Implications for Willingness to Pay, *Transportation Research E*, 45, 583-590
Hensher, D.A., Rose, J.M. and Greene, W.H. (2005) *Applied Choice Analysis*: *A Primer* Cambridge University Press, Cambridge.
Huber, J. and Zwerina K. (1996) The Importance of utility Balance and Efficient Choice Designs, *Journal of Marketing Research*, 33 (3), 307-317.
Kanninen, B.J. (2002) Optimal Design for Multinomial Choice Experiments, *Journal of Marketing Research*, 39 (2), 214-217.
Jou, R. (2001) Modelling the impact of pre-trip information on commuter departure time and route choice, *Transportation Research B*, 35 (10) 887-902.
Jovicic, G. and Hansen, C.O. (2003) A passenger travel demand model for Copenhagen, *Transportation Research A*, 37 (4) 333-349.
Kuhfeld, W.F., Tobias, R.D., and Garratt, M. (1994) Efficient Experimental Design with Marketing Research Applications, *Journal of Marketing Research*, 21 (4), 545-557.
Lam, S.H. and Xie, F. (2002) Transit Path Models that use RP and SP data, *Transportation Research Record 1779*, paper no. 02-3052.
Lazari, A.G. and Anderson, D.A. (1994) Designs of Discrete Choice Experiments for Estimating Both Attribute and Availability Cross Effects, *Journal of Marketing Research*, 31 (3), 375-383.

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

Oppewal, H., Louviere, J.J., Timmermans, H.J.P., (1994) Modeling hierarchical information integration processes with integrated conjoint choice experiments, *Journal of Marketing Research,* 31, 92–105.

Peeta, S. Ramos, J.L. and Pasupathy, R. (2000) Content of Variable Message Signs and on-line Driver Behavior, *Transportation Research Record 1725*, pap no. 00-970, 102-103.

Richardson, A.J., (2002) Simulation study of estimation of individual specific values of time using an adaptive stated preference Survey, *Paper presented at the TRB Annual Meeting*, Washington D.C., USA.

Rose, J.M. and Bliemer, M.C.J. (2004) The Design of Stated Choice Experiments: The State of Practice and Future Challenges, Working Paper, University of Sydney, April.

Rose, J.M. and M.C.J. Bliemer (2007): 'Stated preference experimental design strategies' in Hensher, D.A. and K. Button, (eds.) *Transport Modelling*, Second Edition, Handbooks in Transport, Vol. 1, Elsevier Science, Oxford, Chapter Rose, J.M. and Black, I., (2006) Means Matter, but Variance Matter Too: Decomposing Response Latency Influences on Variance Heterogeneity in Stated Preference Experiments, *Marketing Letters*, 17(4), 295-310.

Rose, J.M. and Hensher, D.A. (2006) Handling 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, pp347-371.

Sandor, Z. and Wedel, M. (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, 38 (4), 430-444.

Scarpa, R., Thiene, M. and Train, K. (2008) Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps, *American Journal of Agricultural Economics*, 90(4):994-1010. (also seeAppendix: Utility in *WTP* space: a tool to address confounding random scale effects in destination choice to the Alps. Available at: http://agecon.lib.umn.edu/).

Scarpa, R., Gilbride, T., Campbell, D. and Hensher, D.A. (2009) Modeling attribute non-attendance in choice experiments for rural landscape valuation: does it matter? *European Review of Agricultural Economics*, 36(2).

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

Train, K. and Weeks, M. (2005) Discrete choice models in preference space and willing to-pay space, *in* R. Scarpa and A. Alberini, eds., *Applications of Simulation Methods in Environmental and Resource Economics,* Springer Publisher, Dordrecht, chapter 1, 1–16.

Toubia, R., Hauser, J.R., and Simester, D.I. (2004) Polyhedral Methods for Adaptive Choice Based Conjoint Analysis, *Journal of Marketing Research*, 41 (1), 116-131.

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, 71–78.

Wittink, D.R. and Cattin, P. (1989) Commercial Use of Conjoint Analysis: An Update, *Journal of Marketing*, 53 (3), 91-96.