Modelling multi-attribute reference dependence in a commuting choice experiment: Does one size fit all?

Amanda Stathopoulos, Department of Statistical and Economic Sciences, University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy, amandairini.blombergstathopoulos@phd.units.it

Stephane Hess, Institute for transport studies, University of Leeds, LS2 9JT, Leeds, UK, S.Hess@its.leeds.ac.uk

ABSTRACT

In contrast with the standard normative theory, empirical findings indicate that decision makers are sensitive to departures from reference points rather than states. The associated idea that value is generated from comparisons with reference points is one of the most solid and appealing findings to extend axiomatic views of choice behaviour. Several tests of predictions from the reference dependent preference framework have been carried out in experimental economics, and to a smaller extent in a choice modelling setting, to date. These empirical applications of referencing choice behaviour have left a series of unresolved questions. The most prominent shortcomings concern; the lack of a multi-attribute trading context, the use of homogenous modelling approaches and hypotheses for each type of attribute studied and not exploring the possibility of multiple reference points. This paper presents a practical application where multiple attributes of a different nature are modelled, allowing for gain-loss asymmetry, non-linearity, relative-absolute changes in utility and different reference points. The data used to carry out these tests is drawn from a choice experiment on intra-mode commuting choices of train and bus users in the UK. It is shown that allowing for reference effects and decreasing sensitivity improves modelling results, in particular when differentiated according to the nature of each considered attribute. The importance of modelling trading affected by reference effects should not be underestimated given the potential impacts on results later used for measuring welfare and assessing the effectiveness of policies. The consequences of the reference-dependent trade-offs modelled in this paper are reflected in the effect they have on wtp/wta measures.

Keywords: Choice modelling, discrete choice experiment, reference effects, non-linearity, gain/loss deviations, commuting

1 INTRODUCTION

The notion that value or utility is strongly influenced by reference points - and above all departures from reference points as defined in prospect theory - is currently accepted by researchers in a variety of disciplines. The basic idea of value generated in view of reference points has given rise to numerous additional theories, including asymmetrical utility drawn from gains and losses, non-linear probability evaluations, asymmetrical decreasing sensitivity and endowment effects to the status quo condition (Kahneman et al. 1991; Kahnemann and Tversky 1979). However, empirical tests of the theory of reference dependent behaviour have left a series of unresolved questions. In particular, the nature of the good, amount of familiarity and the presence of reference points other than those hypothesized by the researcher may alter the typical predictions of prospect theory. What is more, reference dependence has rarely been explored in situations with multiple and complex tradeoffs among a large number of attributes, a setting typical for real world choices and a prominent feature of most choice experiments. In this paper we compare evaluations of commuter trips in a choice experiment (CE), firstly as linear-in-attributes and parameters then progressively incorporating a separate evaluation of gains and losses and testing for presence of non-linear terms. It is shown that the specifications used to describe responses in a standard model may not be adequate for modelling departures from status quo conditions. What is more, responses may be dependent on the level of this departure point, for instance a status quo where a traveller is never delayed. The main novelty proposed concerns the possibility that attributes of a very different nature, including discrete percentage rates such as seating-availability, can be modelled allowing for reference effects. This paper will compare the use of four different modelling approached that draw on referencing behaviour and apply them in a multi-attribute setting. More specifically the current work tests for,

- the impacts of separating attribute reactions into gains and losses from the reference condition with possible asymmetry in sensitivity,
- the impact of absolute-relative and certain-uncertain attribute-level effects controlling for respondent specific experiences,
- non-linearity in responses referred to reference-free and relative responses.
- the possibility that referencing may not occur exclusively against current trip conditions but instead be guided by other cognitive anchors.

To account for this last possibility we test two additional plausible reference points against which gains and losses are modelled: the ideal and acceptable conditions for each trip attribute as defined respondents. This analysis has potentially important policy implications in that policy-makers will typically be interested in reactions to changes of current trip variables. The paper presents referencing effects for a wide range of trip-variables going beyond the most common time and cost attributes. What is more, the evidence brought forth to sustain the idea of multiple reference points suggests that there may be conditions under which the respondent's ideal trip, as opposed to current, does a better job of describing preferences and reactions to changes.

The paper is outlined as follows. The second section overviews literature on referencing behaviour and defines the specificities that are needed to understand commuter behaviour. Section three overviews the modelling approach. The data and survey instrument are overviewed in section 4. Results are reported in section 5 while section 6 closes with a conclusion.

2 LITERATURE

2.1 Modelling reference dependencies

A range of factors beyond the traditionally dominant idea of taste variations influence choices and explain heterogeneity in choice outcomes. McFadden (1999) classified these 'other' factors in four (overlapping) groups: context effects, reference point effects, availability effects and superstition effects. Notwithstanding the important motivation behind such efforts of extending traditionally modelled taste-and-budget-driven utility models no comprehensive modelling framework has been created for the use by applied discrete choice practitioners. The idea that reference dependence shapes individual utility is not new in social science disciplines such as economics and psychology. The underlying idea is that individual preferences are not generated or modified in a vacuum, but are dependent on comparisons against a frame of reference. The literature has identified several types of reference reliance effects and a number of these can be appropriately dealt with in a choice experiment setting. Zhang et al (2004) set out a general framework where utility is defined by the context in which the choice is made. These are referred to a) features of the choice set (alternative or attribute-specific), b) the background situation (circumstances surrounding the choice) and finally, c) individual features that influence decision-making, including the behaviour of peers and reference groups and past choice behaviour (individual-context). This approach inserts McFadden's classification into a framework of relative utility, where task, context and personal factors each influence decision making by providing a frame of reference. These three approaches, including a brief overview of experimental findings, are reviewed in the following.

2.1.1 References and alternative-specific context

In the first group of referencing behaviour two main modelling approaches can be identified. The first approach hypothesizes that the reference alternative captures unobservable influences that go beyond the inclination for the attributes present in the choice set. Along these lines, Adamowicz et al (1998) suggest that a significant coefficient for the alternative specific constant (ASC) can be interpreted as the utility drawn from the SQ alternative. Thus a significant ASC, for the SQ alternative or an opt-out alternative, suggests that a SQ effect is occurring (Scarpa et al. 2005). The main modelling challenge then lies in explaining the probability to select the SQ, which goes beyond what can be captured by varying attributes and levels across designed experimental alternatives (Ferrini and Scarpa 2007).

A second modelling approach attaches a distinct coefficient to positive and negative deviations from a reference situation. Examples from a transport setting include (De Borger and Fosgerau 2008; Hess 2008; Hess and Rose 2009; Hess et al. 2008; Masiero and Hensher 2010). When referencing occurs with regard to the SQ alternative, this may imply that respondents display a systematically different attitude toward alternatives coinciding with the current situation compared to other alternatives. Therefore a prominent modelling issue is to account for this tendency to view the status quo alternatives as substantially different in nature from the experimentally designed ones (Hess and Rose 2009). Scarpa et al (2005) model this disparity in SQ versus design responses by associating additional error components with the hypothetical alternatives. Typically the difference consists in smaller valuation or perception errors when dealing with known SQ alternatives. Concerning the explanations for tendencies to use referencing several propositions are given in the literature. SQ choosing may be caused by the unfamiliarity of the choice context, as noted by (Munro and Sugden 2003) or be a side-effect of task complexity (Moon et al. 2005). What is more, the proportion of SQ-choices can be strongly influenced by design properties in the CE (Ferrini and Scarpa 2007). In fact, concerning the task presentation, (Bateman et al. 2009) suggests that making the task easier by the use of virtual reality graphs may decreases the often-observed loss/gain asymmetry around the reference point.

In general, these studies illustrate that there are indeed important differences between evaluations of improvements and deteriorations of travel variables from respondent's current status. Mounting proof indicates that indifference curves for losses are steeper than for improvements, whereby a gap between willingness to pay and accept is generated. The issue of changes in sensitivity with the increase in attribute levels remains a less explored issue in the presence of reference points. Likewise the competing impact of absolute versus relative changes is poorly understood. Although there is a relatively large number of studies exploring attribute and level-context, the impact of a host of effects – non-linearity, asymmetry and reference bias – on multi-attribute choices are not yet clarified.

2.1.2 References and circumstantial context

A less explored line of research concerns the circumstances of the choice situation that are neither tied completely to person or inter-personal effects, nor to what is explicitly presented in the CE in terms of attributes, levels and ranges. Oppewal & Timmermans (1991) promote the idea that wider context effects, in their case interest rates, economic climate, housing market fluctuations, are likely to influence the choice of residential location. To test for such affects, some aspects of the background economic setting could be varied in the experiment along with the attributes, according to statistical criteria. In the same vein, albeit concerning effects that surround the immediate choice setting, some authors are exploring response latency effects. In a transport setting Bonsall et al (2009) explore the links between road pricing policy acceptance, effort and complexity. The authors find that pro-pricing attitudes led to quicker response times, but also to more evaluation errors. Haaijer (2000) and Rose and Black (2006) explore the effects of response times on response heterogeneity. A significant improvement in explanatory power is gained by accounting for this effect. On a different note, controlling for choice task lag and lead effects, Holmes and Boyle (2005) explore context and anchoring effects in a choice sequence. Concerning the relation between design effects and status quo bias, Moon et al (2004) illustrate how task complexity significantly increases the probability of choosing the SQ. Overall these findings need to be related to personal and choice-set factors to assess the expected impact of circumstantial effects on behaviour.

2.1.3 Referencing and person-specific context

A last, and largely unexplored, area of research concerns the link between reference effects and personal and interpersonal behaviour. Schwanen and Ettema (2009) study the impact of referencing in a context of parents collecting children in presence of uncertain travel-times. The study underscore the importance of considering socially imposed reference points, and the deviations from these, rather than just focussing on current travel conditions. (Mahmassani et al. 1990) indicate that workers in a context where late arrivals are not tolerated do not change their departure time in view of deteriorated traffic conditions. Likewise the acceptance of road-pricing policy is highly influenced by perceived control and opinions of significant others (Schade and Baum 2007). Several papers find the role of internal or social norms to be influential for mode choices (VanVugt et al. 1996).

To conclude the overview of the defined reference effects it is worth noting that most studies still focus on a single context effect in specific choice situations defined by a narrow set of choice characteristics. An example of including each potential effect is the following. To explain disproportionate SQ selection in the context of a biodiversity CE Meyerhoff and Liebe (2009) compare three main explanations: protest or pro-environment attitudes, perceived task complexity and socio-economic respondent features. Several modelling specifications led to the conclusions that the probability of choosing the SQ (with no biodiversity improvement) increases with protest attitudes and decreases along with favourable attitude toward the environment. The influence of perceived choice task complexity is found to be less clear-cut. In the following chapter some of the context specific variables that need to be assessed to study commuter choices are overviewed.

2.2 Prospects and commuting behaviour

Prospect theory is built around the idea that utility is drawn from changes in endowments, not states (Kahnemann and Tversky 1979). This foundation has solved several systematic violations of expected utility theory. To start with, outcomes below the reference point are defined as losses, while improvements are framed as gains. Utility is concave for gains, implying risk aversion and with a steeper convex slope for losses in line with risk seeking and higher sensitivity observed in experimental tests. The originally developed theory viewed prospects as a simple one-attribute choice with probabilistic (risky) outcomes. Instead the extension to riskless choice incorporated a set of important innovations (Tversky and Kahneman 1991). Namely, alternatives are decomposed into a multiple attribute evaluation where each attribute has a distinct value function and reference point. Thus the three fundamental features of the value function are:

- Reference dependence: deviations determine value, not states
- Loss aversion: discrepancy between what agents are willing to accept to give up an attribute and what they are willing to pay to acquire it, where losses incur a steeper inclination in the value function.
- Diminishing sensitivity: marginal value of both gains and losses decrease or dampen with increase in the level of the attribute.

In a commuter setting this has a series of consequences that will be briefly overviewed in the following paragraphs.

2.2.1 Habits and mental budgets

One theoretical construct with a large impact for evaluation of time and money is the notion that there exists a mental Travel Time Budget (TTB). That is, a commuter will allocate an amount of daily time for transportation, which will tend to remain very stable over time. This goes against a common wisdom in transportation analysis that travel-time is a necessary (derived) evil (Mokhtarian and Chen 2004). Similar observations are made for a budget for travel fare expenditure (Gunn 1981). As an associated idea, commuters might form an ideal travel-time budget (Mokhtarian and Salomon 2001; Redmond and Mokhtarian 2001). In line with these observations we have devised three distinct reference points for each of the treated attributes. Beyond the standard current trip situation that incorporates a tendency for path dependence, two additional mental reference points are modelled, acceptable and ideal conditions. In line with the theories on mental time and cost budgets for commuting travelling, behaviour might be strongly influenced by changes from these set points. It cannot, a priori, be excluded that the frame of reference used is a negotiated acceptable level or an ideal condition, which possibly does not coincide with the current trip variables. Modelling these three reference points for every single attribute used in the experiment will offer a large set of tests of the reference-free utility attribute evaluation hypothesis.

2.2.2 Multiple reference points

If we accept the idea that behaviour depends on reference levels, then the predictions generated by models allowing for reference-dependence will depend crucially on what the reference level is assumed to be. Unfortunately, theorizations of which reference points should be employed is much more limited than the research concerning how actors react to changes from these reference points. What is more, the choice of reference point appears to be guided by data-availability and is not always given a theoretically solid justification. The point of reference that effectively guides behaviour is likely to change in view of the choice context. For instance, a finding from behavioural economics is that taxi-drives use their daily earning reference to determine the number of daily work hours (Camerer et al. 1997). In a transport setting there appears to be a reference dependence with regard to both un-tolled and tolled road, once habituation changes the status quo (Schade and Schlag 2003). Instead, personal aspirations, along with loss aversion, have an important role in explaining individual health behaviour van Osch et al. (2006). In a commuting setting the natural reference point, readily available in most studies would be the currently experienced trip duration or time of arrival. However a first point of complexity is that of variability in the phenomenon. That is, respondents are typically asked to respond to SP experiments, carrying a recent or typical trip in mind, with little empirical grounds for which of these is more representative and which of these is more likely to be the effective reference per the person in making decisions. De Borger & Fosgerau (2008), in the context of a car-commuter survey, argue that the current trip is the most plausible reference point to assess gains and losses of time and money. This idea will be tested by controlling for the impact on modelling performance when different reference points are employed.

2.2.3 Risky and non-risky travelling time

A final point to be explored empirically in this paper is whether a probabilistic prospect, such as the number of crowding events and delays out of ten trips are treated differently than more predictable and stable features of the trip such as average travel-time and cost. The definition of the crowding and delay attributes also allows for an explicit modelling of certain and uncertain segments of an attribute. That is, it is possible to control for the impact of never/always standing, as opposed to intermediary levels. As a further point of interest these ideas can be carried over to a reference-based framework where the relative changes from an individual reference point in these attributes is analysed.

3 DESIGN & DATA

The study is based around data from a choice experiment on intra-mode commuting choices of train and bus commuters in the UK. The selection of attributes to include drew on scientific literature and official statistical reports on transport user service evaluation. Beyond standard attributes such as travel time and fare, several innovative service quality features were introduced. The questionnaire was extensively pre-tested with simulation and using a pilot consisting of 50 respondents, including analysis of pilot respondents comments and estimation of various econometric models. The service attributes in the final selection included availability of seating, frequency of delays, extent of delays and the availability of an information service to update on traffic conditions. The definition of these qualitative attributes in a choice-experiment friendly presentation proved difficult. Few studies have dealt contemporarily with the problems of delay and crowding. (Cantwell et al. 2009) is an exception, however their SP choice situation is characterized by a given level of crowding / variability in travel time for a generic commuter trip. This means that no weight is given to how representative this trip is for the comprehensive commuter experience or how often it occurs. Instead in this paper both attributes are represented as the number of occurrences out of ten typical trips (a week worth of commuting). Due to the repeated nature of the trip undertaken this allows us to represent some of the uncertainty that is likely to influence the evaluation of such quality attributes. Regarding the **crowding** attribute, previous evidence is limited and seating or crowding considerations are often included as one among several 'comfort' features (Ben-Akiva and Morikawa 2002). The focus is mostly on stress induced by the commuting in crowded condition, where passengers do not necessarily travel on foot but experiences a saturated vehicle (Litman 2008). For the present study, initial attribute definitions differentiated between the amount the respondent had to stand during a trip along with the frequency of having to stand (Hensher et al. 2003). Such a definition proved too complex to represent along with five other variables. Finally the attribute was defined as the number out of ten typical trips the respondent would have to stand, without considering the amount of standing in that/those occasions. In this sense, the attribute describes the number of times the commuter is standing in a typical workweek. In one of the first studies on transit

reliability, the idea was expressed in terms of day-to-day consistency in operational performance (Polus 1978). There is a rich recent literature on how passengers respond to delays and uncertainty in travel-time in a SP setting (Asensio and Matas 2008; Bhat and Sardesai 2006; Brownstone and Small 2005; Li et al. 2010; Noland and Polak 2002). One of the most prominent findings is the difficulty to represent variability without losing track of the nuisance brought on by uncertainty. Respondents indeed appear to be more sensitive to nonrecurring travel time compared to predictable time (Small et al. 2005). Several authors have hypothesized what type of relationship incurs between travel time and delayed time. Litman 2009 describes a 'penalty' on unscheduled delays approximately equal to three to five times the value of in-vehicle time.

Table 1 - Attributes and levels used in the design

Attributes	n. design levels	description of levels (bold=SQ)	values that attribute can take
Time	5	20, +10, +0, -10, -20	>= 20
Fare	5	20 +10 +0 -10 -20	> 0
Crowding	5	2 +1 +0 -1 -2	standing in 0-10/10 trips
Rate of delay	5	2 +1 +0 -1 -2	delayed for 0-10/10 trips
Extent of delay	5	30 +15 +0 -15 -30	no restrictions
Information service availability	3	no service, free service, charged service	charged service: 15 p for bus users, 30 p for train users or as stated by respondent

The basic efficient design ensured utility balance through the use of conditionals in the code to avoid unrealistic and dominant tasks, while also ensuring a good degree of attribute level-balance. In total, 60 choice scenarios were created, which were blocked into 6 different sets of 10 tasks, minimizing correlation with the blocking variable. The survey presented respondents with three trip options where one always corresponded to the current conditions as declared by each respondent in the first stage of the survey. The remaining options are pivoted around the status quo condition but presented to the respondent in real values to facilitate comparisons. A series of socio-demographics along with indications of different reference points and other information to aid modelling of non-linear preference structures were gathered. The final data was collected through an internet panel survey gathering a total of 400 respondents (not counting the initial 50 responses used for the pilot). Of these respondents, 32 did not fulfil a set of reasonableness criteria and were not used in the analysis. The main respondent characteristics are given Table 2.

Table 2 – Descriptive statistics for the sample

Attributes	Definition	Mean	St.dev	% rates
Age	Average of mean age within 7 age bands	34.612	10.950	
Income	Average of mean annual income within 9 income bands	25136.183	16143.170	
Sex	0=male, 1=female	0.617	0.487	
Education reached	1=mandatory school, 2=high school, 3=university,	1.810	0.750	40 % university
Information service	0= not available, 1=available at charge, 2=available for free	0.791	0.950	36% free info.service
Car availability	1=no car availability, 2=car availability	1.508	0.501	51% has car
Current tt	Average stated travel-time	45.793	26.719	
Current fare	Average stated daily fare	2.858	3.801	
Current delay (freq)	Average stated number of delays in 10 trips	3.405	2.525	
Current delay (min)	Average stated delay across delayed trips	10.073	9.248	
Current crowding	Average stated number of times having to stand in 10 trips	3.332	3.072	

Along with the controls proposed in this paper for reference-dependence beyond the current commute trip description, further questions were inserted to assess other plausible reference points. The respondents were solicited to define an ‘acceptable’ level for each trip attribute. Care was spent to ensure the evaluation was realistic and the respondent was explicitly instructed to consider technical constraints and the high usage rate of the transit network. The acceptable and ideal were presented in separate screens to avoid carry-over effects between the responses.

4 MODELLING & HYPOTHESES

The point of departure is a base model hypothesizing linear, reference-free attribute specifications, with identical alternatives. The model is defined as follows.

$$V_{j,n,t}=\beta_{TT}TT_{j,n,t}+\beta_{FA}FA_{j,n,t}+\beta_{CR}CR_{j,n,t}+\beta_{RA}RA_{j,n,t}+\beta_{RB}RB_{j,n,t}+\beta_{INF-L}IL_{j,n,t}+\beta_{INF-H}IH_{j,n,t} \qquad (1)$$

In line with random utility theory individual *n* chooses among alternatives *j* where *t* indicates the string of choice situations faced. Each attribute is defined to be linear while the information service attribute is effects coded to represent the availability of a free information service (INF-H) and a charged service (INF-L) compared to the omitted baseline situation where the service is not available.

4.2 Modelling non linearity

Non-linearity is modelled in two different ways depending on the nature of the attribute. For the finite and discrete attributes, namely crowding and reliability a segmented modelling approach was devised. Thus the non-linearity was modelled by fitting separate coefficients to the segments of the attribute levels (standing or being delayed in 0-10 trips), i.e. making use of a piecewise linear approach. To make these coefficients comparable to the simple linear scalar estimate, however, the scale of each segment is normalized. As a result the adequately transformed non-linear attribute specification can easily be compared to the linear version of the attribute to assess and interpret any differences in responses to different attribute-segments.

Instead the continuous attributes will be analyzed using a continuous non-linear transformation. This can be achieved through a Box-Cox transformation (Mandel et al. 1994). The below example illustrates this point using travel-time.

$$\mathrm{TT}^{\lambda}_{j,n,t} = \begin{cases} \dfrac{(\mathrm{TT}^{\lambda}_{j,n,t} - 1)}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(\mathrm{TT})_{j,n,t} & \text{if } \lambda = 0 \end{cases} \qquad TT_{j,n,t} > 0 \tag{2}$$

For the situations where the response can reach zero a Box-Tukey transformation needs to be used, adding a constant to the previous term, $\dfrac{((\mathrm{TT}^{\lambda}_{j,n,t} + constant) - 1)}{\lambda}$.

4.3 Modelling gains, losses and asymmetry

Referent dependent theory states that losses and gains are defined according to a reference, or aspiration point. Two methods are used to explore asymmetrical sensitivity towards gains and losses. Firstly, a piece wise linear function can assess the penalty associated with losses as opposed to reference-free non-directional attribute utility. The functional form for the example of travel-time can be written as (Ben-Akiva and Lerman 1985):

$$V(\mathrm{TT}_{j,n,t}) = \begin{cases} \beta_{TT}\mathrm{TT}_{j,n,t} \\ \beta_{TT(inc)}(\mathrm{TT}_{j,n,t} - \delta_{TT}) if(\mathrm{TT}_{j,n,t} \geq \delta_{TT}) \end{cases} \tag{3}$$

Where the first component applies to all possible values the variable can take, thus simply represents the slope of a coefficient with no way to measure any asymmetric impact of gains or losses. The second component of the function estimates a possible change in slope at the point of interest δ, which may be the current travel-time or other possible reference points. The third component is an indicator function to describe whether the condition (increase beyond a reference point) occurs. A significant $\beta_{TT(inc)}$ coefficient would then stand to indicate a significant change in gradient of the utility drawn from the attribute beyond the reference point. It is important to note that the multiplication by the observed x, or difference in x, ensures that the function is linear in the β's, piece wise linear if $\delta \neq 0$, but continuous in utility. This means that the model can be estimated using standard discrete choice methods. Moreover, willingness-to-pay according to different reference points and utility segments can be calculated correctly.

An alternative approach is presented in Hess et al where the attribute-modelling is split between increases and decreases from a reference value (2008). This extension is illustrated in equation 3 where the fare attribute is described as losses and gains compared to the current levels.

$$V_j \begin{cases} V_{REF} = ASC_{SQ}SQ + \beta_{TT}\mathrm{TT}_{n,t} + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} + \beta_{RB}\mathrm{RB}_{n,t} + \beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \\ V_{Alt2} = ASC_{Alt2}ALT2 + \beta_{TT}\mathrm{TT}_{n,t} + \beta_{FA(inc)n,t}if(\mathrm{FA}_{Alt2} > \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{Alt2} - \mathrm{FA}_{\mathrm{ref}}) + \\ \beta_{FA(dec)n,t}if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{ref} - \mathrm{FA}_{Alt2}) + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} + \beta_{RB}\mathrm{RB}_{n,t} \\ +\beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \\ V_{Alt3} = \beta_{TT}\mathrm{TT}_{n,t} + \beta_{FA(inc)n,t}if(\mathrm{FA}_{Alt3} > \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{Alt3} - \mathrm{FA}_{\mathrm{ref}}) \\ +\beta_{FA(dec)n,t}if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{ref} - \mathrm{FA}_{Alt2}) + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} \\ +\beta_{RB}\mathrm{RB}_{n,t} + \beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \end{cases} \qquad (4)$$

Where $\beta_{FA(inc)n,t}if(\mathrm{FA}_{Alt2} > \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{Alt2} - \mathrm{FA}_{\mathrm{ref}})$ = the extent to which the price of alternative 2 exceeds that of the reference alternative, thus representing a loss.
and $\beta_{FA(dec)n,t}if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{\mathrm{ref}})(\mathrm{FA}_{ref} - \mathrm{FA}_{Alt2})$ = the opposite case where the second alternative presents a lower level than the reference one. A similar procedure can be applied to each of the remaining attributes (CR, RA, RB) where increases are all interpreted as losses. A slightly different approach, using effects-coding of gains and losses is used to model the qualitative attribute of information availability. The advantage of this approach is that separate parameters for gains and losses can be uncovered, along with an indicator of their significance. Loss aversion occurs if the magnitude $\beta_{FA(inc)n,t} > \beta_{FA(dec)n,t}$.

It is straightforward to note that the estimated slope-penalty parameter in (3) is equal to the difference between $\beta_{FA(inc)n,t}$ and $\beta_{FA(dec)n,t}$ from equation (4). Thus a significant slope-penalty indicates the exact confidence with which gains and losses can be viewed as statistically different.
To control whether the gain and loss coefficients are significantly different from each other it is necessary to correct for the opposite sign before comparing the coefficients accounting for the robust standard errors (Hess et al 2008). A possible work-around the lack of comparability is to invert the sign of either gains or losses in the estimation phase or work with absolute numbers in calculation the statistic. A further change compared to equation 1 is that each alternative, except one for identification, is fitted with a separate constant to assess any bias for selecting the current alternative.

4.4 Multiple reference points

The modelling of reference effects, in the presence of a design reference alternative, as in the current case, entails inserting the $\beta_{(inc)}$ and $\beta_{(dec)}$ only for the non-reference alternatives. That is, deviations from the status quo alternative can only occur for alternatives two and three. However to account for reference effects with regard to other reference points (defined in section 3) the deviations are incorporated in the reference alternative as well (see eq. 5).

$$V_j \begin{cases} V_{REF} = ASC_{SQ}SQ + \beta_{TT}\mathrm{TT}_{n,t} + \beta_{FA(inc.acc)n,t}\,if(\mathrm{FA}_{Alt2} > \mathrm{FA}_{acc})(\mathrm{FA}_{Alt2} - \mathrm{FA}_{acc}) + \\ \beta_{FA(dec.acc)n,t}\,if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{acc})(\mathrm{FA}_{ref} - \mathrm{FA}_{acc}) + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} + \beta_{RB}\mathrm{RB}_{n,t} \\ +\beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \\ V_{Alt2} = ASC_{Alt2}ALT2 + \beta_{TT}\mathrm{TT}_{n,t} + \beta_{FA(inc.acc)n,t}\,if(\mathrm{FA}_{Alt2} > \mathrm{FA}_{acc})(\mathrm{FA}_{Alt2} - \mathrm{FA}_{acc}) + \\ \beta_{FA(dec.acc)n,t}\,if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{acc})(\mathrm{FA}_{acc} - \mathrm{FA}_{Alt2}) + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} + \beta_{RB}\mathrm{RB}_{n,t} \\ +\beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \\ V_{Alt3} = \beta_{TT}\mathrm{TT}_{n,t} + \beta_{FA(inc.acc)n,t}\,if(\mathrm{FA}_{Alt3} > \mathrm{FA}_{acc})(\mathrm{FA}_{Alt3} - \mathrm{FA}_{acc}) \\ +\beta_{FA(dec.acc)n,t}\,if(\mathrm{FA}_{Alt2} < \mathrm{FA}_{acc})(\mathrm{FA}_{acc} - \mathrm{FA}_{Alt2}) + \beta_{CR}\mathrm{CR}_{n,t} + \beta_{RA}\mathrm{RA}_{n,t} \\ +\beta_{RB}\mathrm{RB}_{n,t} + \beta_{INF-L}IL_{n,t} + \beta_{INF-H}IH_{n,t} \end{cases} \tag{5}$$

Where $\beta_{FA(inc.acc)n,t}$ represents increases beyond what is considered to be the maximum acceptable increase in the given variable, according to the respondent. An equivalent procedure is used to evaluate the ideal trip conditions.

5 RESULTS

5.1 Base results

A series of different models were used to test the aforementioned hypotheses. All models are estimated in the Biogeme package (Bierlaire 2003). The models progressively incorporate controls for status-quo bias, discrete and continuous non-linear impacts of attribute levels, gains and losses and decreasing sensitivity.

Model 1 (BASE)	- linear reference-free model
Model 2 (LN-FARE)	- natural log for fare attribute
Model 3 (REF)	- inclusion of reference-alternative coefficient
Model 4 (EXP-DEL)	- interacted terms
Model 5 (INF-GR)	- piece-wise reference groups for information attribute
Model 6 (SEGM-CR+RA)	- piece-wise absolute impact of crowding and reliability
Model 7 (BT-EXP-DEL)	- Box-Tukey transformation
Model 8 (G&L-REF)	- gain-loss asymmetry from current trip
Model 9 (G&L-ACC)	- gain-loss asymmetry from acceptable trip
Model 10 (G&L-IDE)	- gain-loss asymmetry from ideal trip

The first four models are summarized in Table 3. In model 1, although each coefficient is individually significant and of the right sign, the goodness-of-fit is poor. Drawing on the literature on cost damping (Daly 2010), or decreasing marginal utility with higher levels of the attribute, the classical candidates of time and fare damping were tested with log-linear formulations. Changing the specification for cost derives a significant improvement, albeit not for time. The log-transformation of the daily fare attribute generates an increase of 300 units of log-likelihood and a surge in ρ^2 from .08 to .16 (Model 2).

Model 3 sees the inclusion of a coefficient to denote the individual specific reference situation in line with Hess & Rose's suggestions to model pivoted reference alternatives (2009). Beyond the statistically significant improvement (LL ratio test statistic of 60 against a critical χ^2 of 5.99) an important finding is that including a distinctive coefficient for the reference alternative appears to stabilize the remaining part-worths. In fact, the coefficients for time and cost remain remarkably stable across more advanced specifications when this coefficient is present. The next notion to control concerns the impact of entering the socio-demographic variables in the models as interaction-terms with the different coefficients. All combinations of gender, income, education, car-availability and age were tested interacted in this manner with each of the attributes.

Table 3 - Base models

	Mod 1		Mod 2		Mod 3		Mod 4	
Variables	Beta	rob t-ratio	Beta	rob t-ratio	Beta	rob t-ratio	Beta	rob t-ratio
Asc-alt2	-	-	-	-	0.161	3.29	0.163	3.33
Asc-sq	-	-	-	-	0.384	7.81	0.390	7.92
CR	-0.175	-9.33	-0.229	-11.21	-0.220	-10.22	-0.223	-10.32
infH	0.179	6.12	0.267	8.78	0.252	8.21	0.251	8.15
infL	-0.101	-2.83	-0.272	-7.18	-0.168	-4.20	-0.171	-4.27
RA	-0.177	-9.69	-0.238	-12.08	-0.241	-11.53	-0.187	-7.08
RB	-0.033	-5.82	-0.040	-6.67	-0.042	-6.70	-0.029	-4.14
FA	-0.979	-9.23	-	-	-	-	-	-
TT	-0.036	-10.58	-0.044	-12.90	-0.047	-12.96	-0.047	-13.05
ln-FA	-	-	-5.600	-27.85	-5.970	-27.21	-6.000	-27.24
exp. delay	-	-	-	-	-	-	-0.062	-3.30
obs.	3680		3680		3680		3680	
n. variables	7		7		9		10	
Init. LL	-4042.89		-4042.89		-4042.89		-4042.89	
conv. LL	-3711.36		-3397.43		-3366.95		-3360.43	
ρ^2	0.082		0.160		0.167		0.169	
ρ^2 adj.	0.080		0.158		0.165		0.166	

Only marginal improvements were achieved in this manner thus these results are not reported as the main aim of the study is to explore referencing choice behaviour among multiple attributes. In view of the prior concerning the role of reliability, and the observation that the TT and RB had the exact same impact on utility lead to considering an interacted term where the frequency of delays is multiplied by the extent of those delays to shape a new variable defined as expected delay. The new attribute grasps the actual extent of delay-time that a traveller can expect to face in a week. However, the distribution of this delay-time is not identified, i.e. one delay of 40 minutes is modelled in the same way as four delays of 10 minutes. Model 4, incorporating this new coefficient, offers significant improvement over model 3.

5.2 Reference models

Until this point the proposed models have given no weight to the personal point of departure and the potential reference effects deriving from this. That is, the ideas set out in section two concerning whether the attributes are deteriorating or improving and whether there are specific reactions to be expected when the attribute levels approach its extreme points have not been explored. This second point is explored in models 5-7 (Table 4), whereas the notion of referencing with regard to gains and losses and multiple reference anchoring is described in table Table 8. Firstly the impact of individual point of departure for the information service attribute is analysed.

Table 4 - Referencing models

	mod 5		mod 6		mod 7	
Variables	Beta	rob t-ratio	Beta	rob t-ratio	Beta	rob t-ratio
Asc-alt2	0.160	3.25	0.163	3.31	0.164	3.33
Asc-sq	0.397	6.98	0.360	6.25	0.375	6.49
CR	-0.226	-10.41	-	-	-	-
RA	-0.187	-7.08	-	-	-0.0566	-1.80
RB	-0.029	-4.10	-0.017	-1.89	-	-
TT	-0.047	-13.09	-0.047	-13.03	-0.047	-13.11
Ln-FA	-6.010	-27.26	-6.020	-27.25	-6.060	-27.22
exp. delay	-0.062	-3.30	-0.081	-3.75	-0.327	-8.62
free-high	0.255	5.00	0.267	5.24	0.270	5.28
free-low	-0.293	-4.31	-0.308	-4.52	-0.318	-4.66
pay-high	0.226	4.93	0.229	4.92	0.227	4.91
pay-low	-0.116	-2.45	-0.112	-2.34	-0.114	-2.38
CR-higher	-	-	-0.692	-4.25	-0.707	-4.32
CR-lower	-	-	0.641	4.62	0.639	4.59
CR-null	-	-	1.250	8.99	1.25	8.99
CR-ten	-	-	-0.885	-4.45	-0.879	-4.42
RA-higher	-	-	-0.901	-3.64	-	-
RA-lower	-	-	0.553	5.41	-	-
RA-ten	-	-	-1.450	-4.57	-	-
λ exp. Delay	-	-	-	-	0.542	11.92
Obs	3680		3680		3680	
n. variables	12		17		15	
Init. LL	-4042.89		-4042.89		-4042.89	
Conv. LL	-3357.76		-3336.93		-3330.45	
ρ^2	0.169		0.175		0.176	
ρ^2 adj.	0.166		0.170		0.173	

5.2.4 Referencing information service

By comparing the preferences of the groups that currently have a free information service available to those that either had a charged service or no such service it is possible to assess the impact of current experience on utility for the different service options (free, charged, not available). This analysis is presented in model 5, where the first component of the variable name describes the condition of departure (free vs. other) and the second part the level presented to the respondent, where effects-coding requires the baseline coinciding with the charged alternative to be omitted. The most important observation to draw is that although the positive evaluation of having the service for free is very similar between the two groups, the disutility of having to pay for it is regarded as much stronger for individuals that currently get the service for free. This idea is in line with the aversion to pricing of a currently freely enjoyed consumption good, for instance pricing of ‘free’ urban roads.

Table 5 – Reference behaviour & information service availability

	Mod 4		Mod 5	
Group-attribute level	Wtp/wta	t-ratio wtp	Wtp/wta	t-ratio wtp
Free-high	-0.115	-4.978	-0.121	-5.199
Free-low	0.133	4.340	0.139	4.563
Pay-high	-0.102	-4.929	-0.103	-4.940
Pay-low	0.052	2.463	0.051	2.359

In terms of willingness to pay these considerations translate into a wta for the loss of the free service similar between the groups (12 and 10 pence respectively for free vs. other group or 27 and 23 pence in model 6). On the other hand, the wtp for having the service at a charge is 2.5 - 2.7 times higher for the group where it was originally free. This indicates a higher sensitivity of this group to the introduction of a charged service whereas the group is unaffected by the absence of an information service. In brief, it is shown how offering a free information service with the aim of progressively introducing a charge for it may create a large reluctance in the group where it was once free. The wtp coefficient in this *Free-low* group can be interpreted as the amount of money that would be required to accept the loss of the free service, which in this sense coincides with the (un)willingness to pay. The wtp/wta gap therefore appears to be between 2-3 times larger for a loss of service compared to what the group that is already charged for the service is effectively willing to pay. Indeed, consumers are shown to value characteristics such as predictability and stability of pricing schemes and be aversive to the opposite (Bonsall et al., 2007) The fact that the revealed ratios are very close to the typical sums paid for this type of services and the high significance of all wtp measures gives further strength to the analysis.

5.2.1 Crowding

Crowding is treated as a linear scalar variable throughout models 1-5. One way to extend the analysis is to estimate separate coefficients for each level this qualitative attribute can take, using cardinal effects coding. This is a way to test for non-linear effects in the impact on individual utility of experiencing different levels of the attribute. Effects codes are favoured over dummy coding given the presence of an ASC-status quo coefficient. Thus, by using effects coding the constant term fitted in the model will reflect only the utility associated with the status quo alternative and not the impact of the omitted base-line of the coded coefficients. However estimating 11 distinct coefficients for each possible value the variable can take was regarded as uninformative and with limited utility for policy analysis. Instead, it is more plausible that some changes between levels are regarded as indifferent while other level-to-level jumps may cause a large increase/decrease in utility. This can be tested by using the approach described in section 4.2. To assess the impact of this procedure the scalar crowding coefficient from the linear model is reported for comparison with the normalized piece-wise specification (Figure 1).

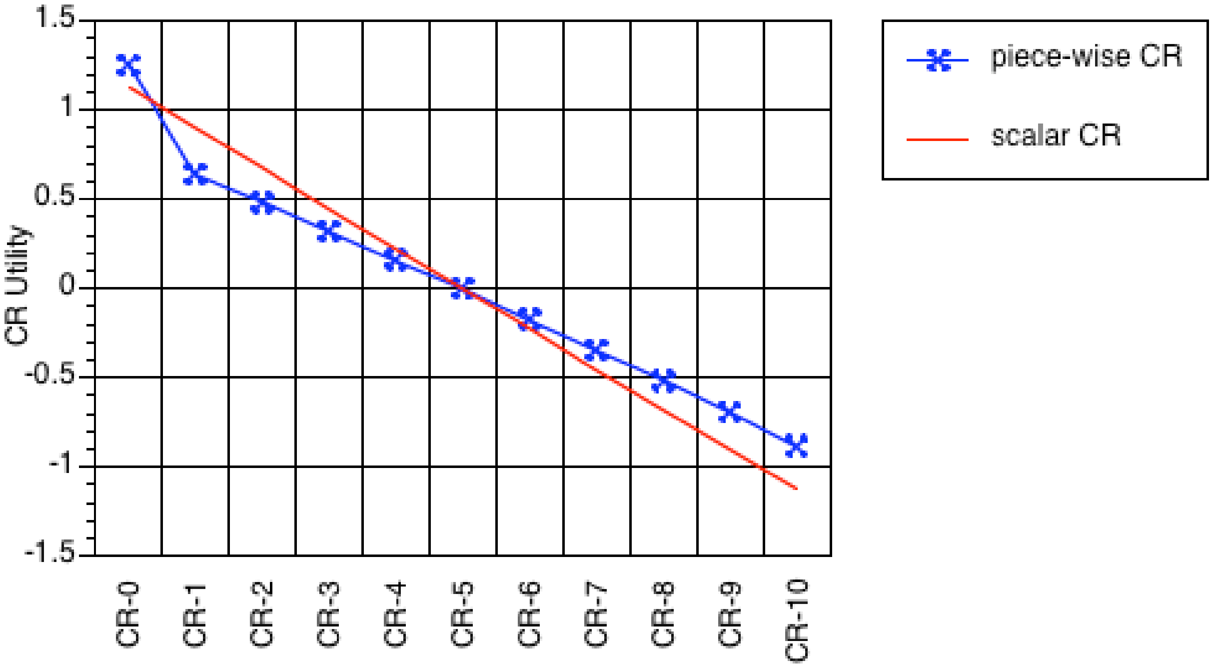

Figure 1 – Normalized scalar and piece-wise crowding coefficient

The model improvement is highly significant, providing an improvement in LL of 16 units. In view of the extreme similarity of the model with segmented crowding and the one where segmentation for reliability is added (model 6) only the latter is reported.

The most notable fact, from comparing the original albeit normalized crowding coefficient to the new segmented one is that not allowing for piece-wise effects tends to overestimate both the lower and higher segments. The lack of consideration for the significant positive impact of the condition of never having to stand (CR-0), followed by a steep drop in utility for 1 delay appears to be the most severe shortcoming of the linear specification.

Table 6 – Reference behaviour & crowding

	Mod 5	
	Wtp/wta	t-ratio wtp
CR-ten	0.400	4.461
CR-higher	0.313	4.238
CR-lower	-0.290	-4.655
CR-null	-0.565	-8.961

This observation is illustrated by the wtp estimates for the four crowding variables in Table 6. It may be noted that the wtp for crowding in the linear model is equal to 0,102 £ and is highly significant (t-ratio 10,53). This implies that an average consumer is willing to pay 10 cents in terms of daily ticket costs to improve crowding from the worst to the best condition. This approach implies that each step from 0-10 is valued equally, in this case worth 1 cent per step. Instead if the segmented approach is used it is possible to seize variations from this linear impact. A first thing to notice is that the values below the index or null point (5) are generating a stronger wta of the loss of lower crowding, in particular of the zero-crowding, that is not compensated by the wtp to improve ones position in the upper segment of trips. In terms of policy indications, this translates to a need to offer significant compensation to passengers that are losing a status-quo condition with no crowding, compared to any other loss in the crowding spectra. On the opposite extreme, the change from crowding in 9 to 10/10 trips does not have a comparably strong impact on well-being.

5.2.2 Rate of delays

Applying the same approach to the rate of delay attribute reveals a quite different picture. Model-fit improves when the delay frequency is transformed from linear to piece-wise form, albeit with a less pronounced improvement than for crowding As can be observed in Figure 2 reliability is discovered to have two important kinks where the indifference curves undergo significant changes in slope.

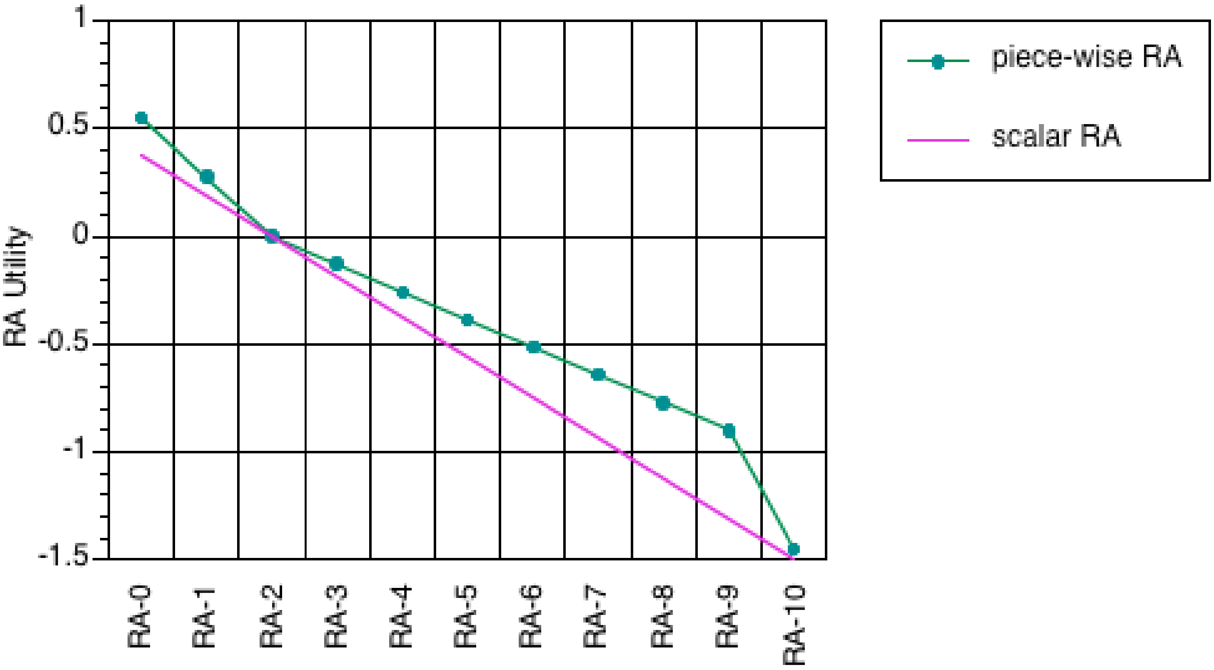

Figure 2 – Normalized scalar and piece-wise crowding coefficient

Compared to the scalar attribute definition, a lower sensitivity can be expected for the event where there are three or more delays. An important shortcoming of the scalar representation of utility is the failure to grasp the important differences between 3-9 delays and the large decrease in utility when 10/10 trips are delayed. For the lower range of the attribute scale the piece-wise version maps a greater tendency to pursue the most favourable levels, a tendency that the linear specification overlooks.

5.2.3 Expected delay & Travel-time

By multiplying the frequency of delayed trips with the number of minutes of average delay time that respondents have experienced across ten typical trips, we define an 'expected delay' (ED) variable. An important validation for the expected delay is to compare it to the sensitivity to travel-time. Numerous papers have explored the links between mean travel-time and variation of travel duration due to delays. Pinjari & Bhat (2006) found that in the first 15 minutes of travel-time commuters tend to be more sensitive to delays. Instead for longer travel-times there is a strong drop in the sensitivity to reliability. To test for non-linearity in these continuous variables they were tested against Box-Cox and Box-Tukey transformations. While for travel-time the minimum value was by definition 20 minutes since the survey placed a cut-off on shorter trips, a Box-Cox transformation was used. Instead, a constant needed to be added when testing for non-linearities in expected delay due to the many cases of a zero current delay.

The results indicate that there is no significant decreasing sensitivity for time, the coefficient not being significantly different from unity. Therefore it is concluded that travel-time is treated as a linear parameter. On the other hand, allowing for a non-linear specification of ED causes some conflict with the remaining delay attributes. In fact, once a lambda (of 0.5) is fitted with the expected delay coefficient the delay(min) variable can be dropped from the model without any penalty to model fit. Likewise, the segmented delay(freq) variables becomes insignificant and comparisons of model fit indicates that a linear scalar be used in its place. Thus, it appears that the log-linear form of ED can carry all the non-linearities inherent in the delay construct.

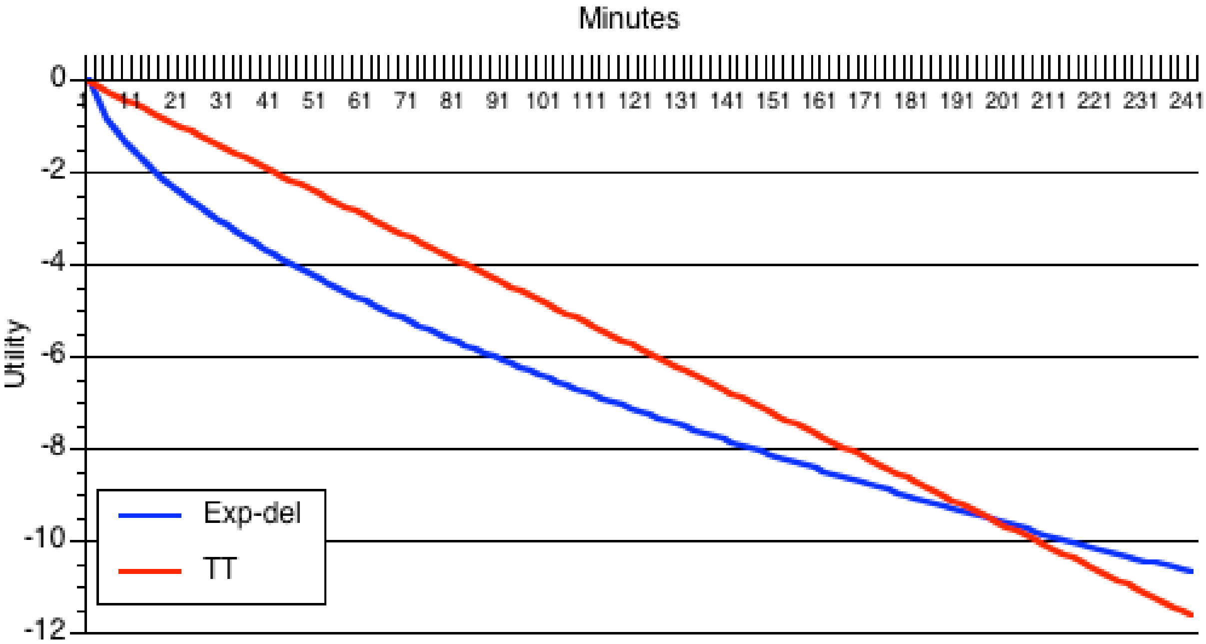

Figure 3 – Comparisons of travel-time and expected-delay

As can be noted from the graph, for the segment of trips considered ED 'prevails' until travel-time reaches a mean duration of 200 minutes. Thus the mean travel-time becomes more important than expected delay for very long (repeated) trips. On the other hand the risky prospect of lateness prevails for shorter mean commutes.

In terms of wtp, shown in Table 8, the evaluation of the expected delay interaction increases progressively with more advanced modelling structures, specifically when the rate of delays is treated non-linearly.

Table 7 – Reference behaviour, travel-time and reliability (£ per hour)

Attribute	Mod 4		Mod 6		Mod 7	
	Wtp/wta	t-ratio wtp	Wtp/wta	t-ratio wtp	Wtp/wta	Wtp/wta
Exp-delay	1.684	3.294	2.182	3.366	8.806	8.592
TT	1.281	13.617	1.277	13.582	1.279	13.683

The wtp measures estimated may appear low but are in line with DfT's appraisal value of 5.04£ for an hour of commuting time (DfT 2004). Caution should be applied before interpreting the high value of expected delay in model 7. It needs to be underlined that this measure, although apparently high is dampened by a factor of λ = 0.542 which indicates a moderate amount of decreasing sensitivity.

5.3 On gains and losses

The following models illustrate the effect of fitting separate coefficients for gains and losses to the benchmark model (Model 6). Significant differences between the impacts of increases and decreases indicate asymmetry in evaluation. If the asymmetry leans towards the loss-evaluation this conveys the idea of loss-aversion.

Table 8 - Referencing models with asymmetric fare

	mod 8		mod 9		mod 10	
Variables	Beta	rob t-ratio	Beta	rob t-ratio	Beta	rob t-ratio
Asc-alt2	0.166	3.36	0.168	3.42	0.169	3.40
Asc-sq	0.271	3.78	0.314	5.36	0.321	5.51
Relb	-0.015	-1.70	-0.015	-1.76	-0.015	-1.71
exp. delay	-0.079	-3.66	-0.081	-4.18	-0.081	-3.76
free-high	0.265	5.21	0.266	5.16	0.262	5.13
free-low	-0.299	-4.38	-0.301	-4.42	-0.302	-4.43
pay-high	0.229	4.93	0.237	5.19	0.229	4.93
pay-low	-0.107	-2.24	-0.114	-2.40	-0.117	-2.44
CR-higher	-0.687	-4.21	-0.714	-4.49	-0.687	-4.21
CR-lower	0.629	4.51	0.643	4.80	0.674	4.82
CR-null	1.240	8.84	1.250	9.13	1.280	9.10
CR-ten	-0.876	-4.39	-0.904	-4.68	-0.888	-4.42
RA-higher	-0.874	-3.53	-0.902	-3.88	-0.905	-3.66
RA-lower	0.555	5.44	0.564	5.68	0.564	5.51
RA-ten	-1.420	-4.48	-1.470	-4.82	-1.480	-4.61
Ln-fa(dec)	5.380	14.15	3.710	8.65	2.010	3.26
Ln-fa(inc)	-7.070	-11.93	-6.780	-27.18	-6.590	-27.17
Tt	-0.048	-13.00	-0.048	-13.39	-0.048	-13.15
Vars	18		18		18	
init. LL	-4042.890		-4042.890		-4042.890	
Conv. LL	-3335.007		-3318.264		-3312.137	
ρ^2	0.175		0.179		0.181	
ρ^2 adj.	0.171		0.175		0.176	
sig. g/l asymm fare	1.99		5.95		6.9	

Several models were tried including each of the attributes in a gain/loss formulation. The models incorporating gains and loss separation for the crowding and the frequency of delay attributes did not significantly improve fit. Instead, the piece-wise modelling approach used above provides a superior model fit. This indicates that the impact of the absolute levels used in the piece-wise models offer a better description of commuter behaviour for these attributes. Instead time and fare could be modelled profitably, albeit model fit improved only marginally. Comparing model 8 to model 6 with log-linear fare, the LL-ratio test gives a statistic of 3.852 against a critical χ^2 of 3.841, implying a narrow improvement. It can also be seen from the last row in Table 8 that the significance of the observed asymmetry is 1.99 implying a significant refusal of the null of no change in slope between gains and loss evaluation. Instead for travel-time there is no significant difference between improving or deteriorating the current travel-time (see appendix). To further explore the impact of reference points other than the current trip conditions two further mental anchor points are tested. Respondents are asked to express their acceptable (under the constraints imposed by the transportation system) and their ideal levels for each attribute. As described in section 4.4 in terms of modelling this means including the two deviations from acceptable and ideal conditions for the reference alternative as well as the current conditions may not coincide with these new points of reference.

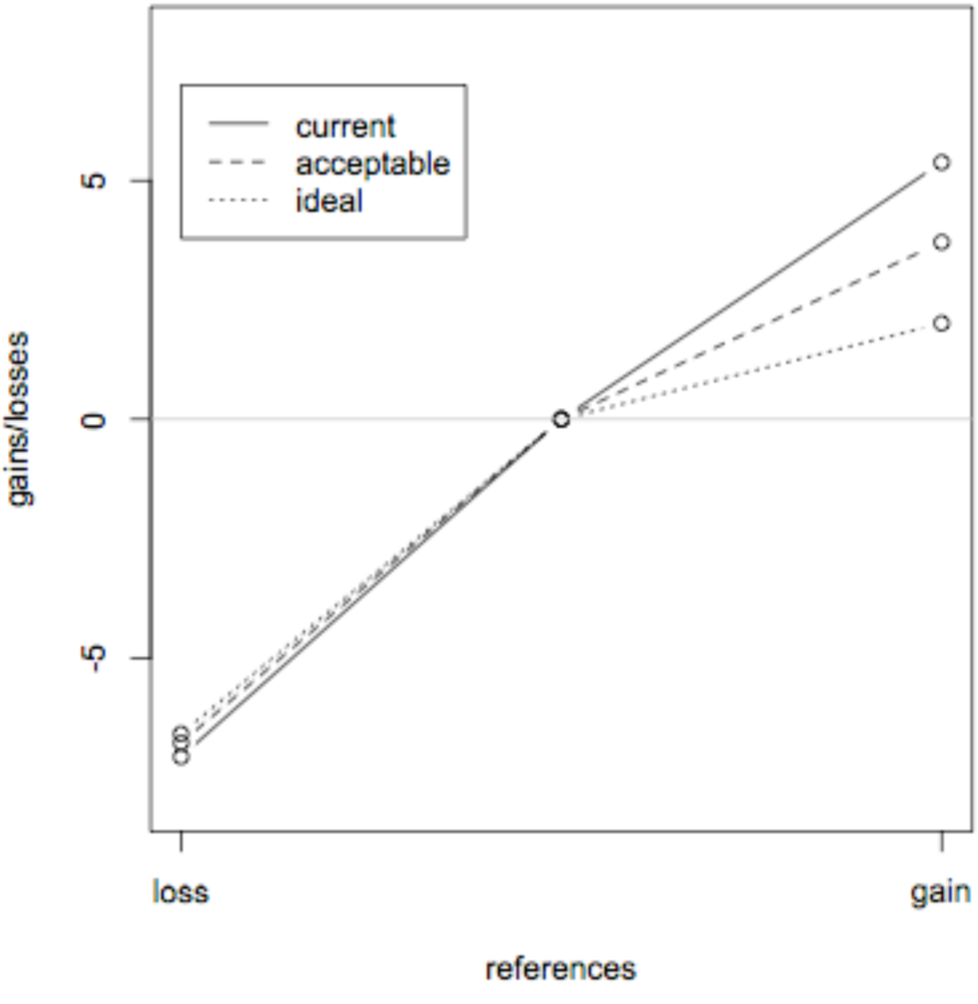

Figure 4 – Comparisons of gain/loss models for ln-fare (three separate reference points)

The new reference points offer significantly better model fit and reveal the evaluation of fare to be significantly asymmetrical, with losses in the ideal model being three times larger than the coefficient for gains. Improvement in fit amounts to 17 and 22 units for the acceptable and ideal respectively, over the current fare pivoting. The relation between the reference points is shown in Figure 4. The most notable fact is that, while the impact of losses remains constant, the slope for the gains flattens when moving towards the ideal model. The latter models give results more in line with the predictions of prospect theory concerning the larger disutility caused by loss compared to utility drawn from gains. A possible explanation for the observed asymmetry is that for monetary expenses only deviations that bring the cost beyond the planned budget are perceived. Instead the commuter is indifferent to improvements, at least in the range proposed in the current experiment.

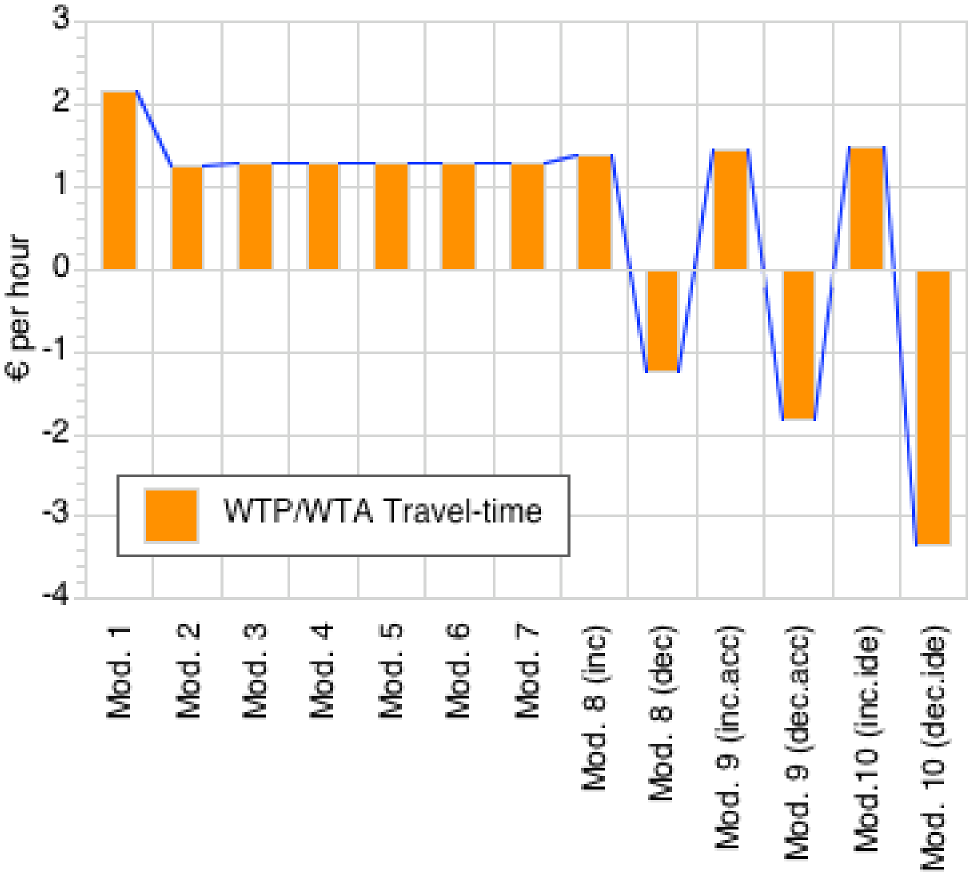

Figure 5 – Comparisons of gain/loss models for fare (three separate reference points)

The impact of these asymmetries in the cost evaluation has some interesting consequences for the VOT. As can be observed in Figure 5, the VOT evaluation is very stable across modelling specifications 2-7. However the large disparities observed for improvement in the fare levels determines an exponential increase in the willingness-to-accept deteriorations in travel-time compensated by a lower fee level. That implies that the low sensitivity in the acceptable and ideal models for a fare decrease causes a surge in the amount of compensation the commuter wants to accept an increase in travel-time. This peculiar situation, caused by the stable and linear evaluation of time and the asymmetrical role of fare requires further studies to assess its solidity. As for the studied sample, the findings appear to indicate that commuters are constant in their trade-offs between time and money for the situation of symmetrical increase, as in the standard assumption of wtp compensation. However, policy-makers should be weary of thinking that the same simple trading patterns apply for increase in travel-time compensated by lower fares. The average decision-maker instead would require three times the fare decrease to accept a unitary increase in travel-time, compared to an equivalent trade-off concerning travel time savings.

6 CONCLUSIONS

This paper sets out a series of discrete choice modelling formulations to account for different ways that referencing behaviour influences choices in a commuting setting. Special attention is paid to extending the empirical tests of reference dependent decision making to a multi-attribute context. This means not simply applying a uniform modelling treatment to all attributes but explicitly controlling for non-linear absolute sensitivity, asymmetry in reference point deviations and decreasing sensitivity. What is more we allow for multiple reference points, in line with the idea that constrained acceptable or ideal trip conditions may be the actual point of reference to determine utility of the evaluated options. This flexible treatment of the commute attributes reveals a series of interesting points on how changes in these attributes are perceived.

The somewhat surprising finding that utility for expected-delay is marginally decreasing for longer travel durations, while sensitivity to travel-time is linear, may be explained by the repeated nature of the commute. Indeed, the prospect of expected delays might be considered not as separate risky events but as a proportion of delays included in a mental commuter budget. Thus the mean travel-time appears to be perceived as consistently more important while the risk of lateness prevails for shorter mean trip durations.

Likewise the lack of asymmetry in gains and losses of travel-time indicates that once a specific amount of time is stably allocated for commuting purposes, deviations, at least in the short run, do not generate significant impacts on utility. In this sense, travel-time is bi-directionally neutral and linear implying that there is no change in sensitivity for different levels or for gains and losses. The contrasting asymmetry and decreasing sensitivity for the daily fare however does give a more complex picture when ratios of time and cost are concerned. Indeed respondents display a pronounced un-willingness to accept increases in travel-time in exchange for fare compensation. Evaluations of the frequency of delays and crowding reveal non-linearities in the sensitivity of going from the extreme of no crowding/delays to a situation of constant crowding/delays. A linear specification consistently overestimates the sensitivity to crowding while it fails to quantify the positive impact of never having to stand. For the frequency of delays the linear attribute specification significantly over-estimates the sensitivity to a high frequency of delays while it fails to assess the large penalty for reaching a situation of a sure delay (10 out of 10 trips). For these probabilistically described attributes there is no important improvement derived from modelling gains and losses separately. This confirms the notion that in evaluating risk of crowding and delays as a frequency measure, the current experience plays little role in defining utility for alternatives. Instead, it appears that reaching absolute levels of crowding/delay looms larger.

It has been shown in this paper how different types of attributes are impacted by reference effects in a non-uniform manner. There are large potential impacts for policies derived from the findings in this paper. For one, the evaluation of the commuter experience is affected by a variety of non-linear sensitivities as for the cases of crowding and the frequency of delays. What is more certain attributes are evaluated in terms of deviations from a reference point rather than absolute stand-alone service features, as for the case of fares. Appropriately accounting for these effects can improve the appraisal of the welfare drawn from (changes in) service features. Future research in this field needs to extend these analyses to encompass a wider variety of situations characterized by habitual and novel choices to understand the impact of reference-dependent choices. A further extension that would improve the applicability of these findings is relating the modelling findings to personal features, attitudes, task-perception and other context and personality effects. Furthermore, it would be valuable to associate the heterogeneity in reference-effects to models exploring heterogeneity in taste and scale to assess the relative impact of each of these explanatory models.

7. ACKNOWLEDGEMENTS

The second author acknowledges the support of the *Leverhulme Trust* in the form of a *Leverhulme Early Career Fellowship*.

8 REFERENCES

Adamowicz, W., Boxall, P., Williams, M., and Louviere, J. (1998). "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation." *American Journal of Agricultural Economics*, 80(1), 64-75.

Asensio, J., and Matas, A. (2008). "Commuters' valuation of travel time variability." *Transportation Research Part E*, 44, 1074–1085.

Bateman, I. J., A P. Jones, Jude, S., and Day, B. H. (2009). "Reducing gain–loss asymmetry: A virtual reality choice experiment valuing land use change ", 58(1), 106-118

Ben-Akiva, M., and Morikawa, T. (2002). "Comparing ridership attraction of rail and bus." *Transport Policy*, 9(2), 107-116.

Ben-Akiva, M. E., and Lerman, S. R. (1985). *Discrete choice analysis : theory and application to travel demand*, MIT Press, Cambridge, Mass.

Bhat, C. R., and Sardesai, R. (2006). "The impact of stop-making and travel time reliability on commute mode choice." *Transportation Research Part B: Methodological*, 40(9), 709-730.

Bierlaire, M. (2003). "BIOGEME: a free package for the estimation of discrete choice models." Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bonsall, P., Schade, J., Roessger, L., and Lythgoe, B. (2009). "Factors Affecting People's Engagement with the Assessment of Road Charges in an Experimental Setting." Paper presented at the 12th International Conference on Travel Behaviour Research, Jaipur, Rajasthan, India, December 13-18, 2009

Brownstone, D., and Small, K. A. (2005). "Valuing time and reliability: assessing the evidence from road pricing demonstrations." *Transportation Research Part A: Policy and Practice*, 39(4), 279-293.

Camerer, C., Babcock, L., Lowenstein, G., and Thaler, R. (1997). "Labor supply of New York taxi drivers: one day at a time." *Quarterly Journal of Economics*, 112, 356-370.

Cantwell, M., Brian Caulfield, and O'Mahony, M. (2009). "Examining the Factors that Impact Public Transport Commuting Satisfaction." *Journal of Public Transportation*, 12(2).

Daly, A. (2010). "Cost Damping in Travel Demand Models: Report of a study for the Department for Transport." RAND TR-717-DFT, 52.

De Borger, B., and Fosgerau, M. (2008). "The trade-off between money and travel time: A test of the theory of reference-dependent preferences." *Journal of Urban Economics* 64(1), 101-115.

DfT. (2004). "Values of Time and Operating Costs." TAG Unit 3.5.6, Department for Transport.

Ferrini, S., and Scarpa, R. (2007). "Designs with a-priori information for nonmarket valuation with choice-experiments: a Monte Carlo study." *Journal of Environmental Economics and Management* 53, 342-363.

Gunn, H. F. (1981). "Travel budgets—a review of evidence and modeling implications." *Transportation Research A*, 15, 7–23.

Haaijer, R., Kamakura, W., and Wedel, M. (2000). "Response latencies in the analysis of conjoint choice experiments." *Journal of Marketing Research* 37(3), 376-382.

Hensher, D. A., Stopher, P., and Bullock, P. (2003). "Service quality - developing a service quality index in the provision of commercial bus contracts." *Transportation Research Part A*, 37(6), 499-517.

Hess, S. (2008). "Treatment of reference alternatives in stated choice surveys for air travel choice behaviour." *Journal of Air Transport Management*, 14(5), 275-279.

Hess, S., and Rose, J. M. (2009). "Should reference alternatives in pivot design SC surveys be treated differently?" *Environmental and Resource Economics*, 42(3), 297-317.

Hess, S., Rose, J. M., and Hensher, D. A. (2008). "Asymmetric preference formation in willingness to pay estimates in discrete choice models." *Transportation Research Part E: Logistics and Transportation Review*, 44(5), 847-863.

Holmes, T. P., and Boyle, K. J. (2005). "Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions." *Land Economics*, 81(1), 114-126.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). "The endowment effect, loss aversion, and status quo bias." *Journal of Economic Perspectives*, 5(1), 193-206.

Kahnemann, D., and Tversky, A. (1979). "Prospect theory: an analysis of decisions under risk." *Econometrica*, 47(2), 263-91.

Li, Z., Hensher, D. A., and Rose, J. M. (2010). "Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence." *Transportation Research Part E*, 46, 384–403.

Mahmassani, H., Caplice, C., and Walton, M. (1990). "Characteristics of Urban Commuter behaviour switching propensity and use of information." *TRB*, 1285, 57-69.

Mandel, B., Gaudry, M., and Rothengatter, W. (1994). "Linear or Nonlinear Utility Function in Logit Models? The Impact on German High-Speed Rail Demand Forecasts." *Transportation Research Part B: Methodological*, 28(2), 91–101.

Masiero, L., and Hensher, D. A. (2010). "Analyzing loss aversion and diminishing sensitivity in a freight transport stated choice experiment " *Transportation Research Part A: Policy and Practice*, Volume 44(5), 349-358.

McFadden, D. (1999). "Rationality for Economists?" *Journal of Risk and Uncertainty*, 19(1-3), 73-105.

Meyerhoff, J., and Liebe, U. (2009). "Status Quo Effect in Choice Experiments: Empirical Evidence on Attitudes and Choice Task Complexity." *Land Economics*, 85(3), 515-528.

Mokhtarian, P. L., and Chen, C. (2004). "TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets." *Transportation Research Part A*, 38, 643–675.

Mokhtarian, P. L., and Salomon, I. (2001). "How derived is the demand for travel? Some conceptual and measurement considerations." *Transportation Research A*, 35(8), 695–719.

Moon, A., Boxall, P. C., and Adamowicz, W. (2005). "Complexity in choice experiments: choice of the status quo alternatives." Working paper. Department of Rural Economy, University of Alberta.

Munro, A., and Sugden, R. (2003). "On the theory of reference-dependent preferences." *Journal of Economic Behavior and Organization*, 50, 407–428.

Noland, R. B., and Polak, J. W. (2002). "Travel Time Variability: A Review of Theoretical and Empirical Issues." *Transport Reviews*, 22(1), 39-54.

Oppewal, H., and Timmermans, H. (1991). "Context effects and decompositional choice modeling." *Papers in regional science*, 70(2), 113-131.

Pinjari, A., and Bhat, C. (2006). "Nonlinearity of response to level of service variables in travel mode choice models." *Transportation Research Record*, 1977, 67–74.

Polus, A. (1978). "Modeling and measurements of bus service reliability " *Transportation Research A*, 12(4), 253-256

Redmond, L., and Mokhtarian, P. (2001). "The positive utility of the commute: Modeling ideal commute time and relative desired commute amount." *Transportation*, 28(2), 179–205.

Rose, J. M., and Black., I. R. (2006). "Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments." *Marketing Letters* 17(4), 295-310.

Scarpa, R., Ferrini, S., and Willis, K. G. (2005). "Performance of error component models for status-quo e
ects in choice experiments." Applications of simulation methods in environmental and resource economics, Springer, ed., 247.

Schade, J., and Baum , M. (2007). "Reactance or acceptance? Reactions towards the introduction of road pricing." *Transportation Research Part A: Policy and Practice*, 41(1), 41-48

Schade, J., and Schlag, B. (2003). "Acceptability of urban transport pricing strategies." *Transportation Research Part F: Traffic Psychology and Behaviour.*, 6(1), 45-61.

Schwanen, T., and Ettema, D. (2009). "Coping with unreliable transportation when collecting children: Examining parents' behavior with cumulative prospect theory." *Transportation Research Part A*, 43, 511–525.

Small, K. A., Winston, C., and Yan, J. (2005). "Uncovering the distribution of motorists' preferences for travel time and reliability." *Econometrica*, 73, 1367–1382.

Tversky, A., and Kahneman, D. (1991). "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics*, 106(4), 1039-1061.

van Osch, S., van den Hout, W. B., and Stiggelbout, A. M. (2006). "Exploring the reference point in prospect theory: Gambles for length of life." *Medical Decision Making*, 26, 338–346.

VanVugt, M., Van Lange, P. A. M., and Meertens, R. M. (1996). "Commuting by car or public transportation? A social dilemma analysis of travel mode judgments. ." *European Journal of Social Psychology*, 26, 373-395.

Zhang, J., Timmermans, H. J. P., Borgers, A. W. J., and Wang, D. (2004). "Modeling Traveler Choice Behavior Using the Concepts of Relative Utility and Relative Interest." *Transportation Research B*, 38, 215–234.

Appendix - Referencing models TC & TT

	mod 8		mod 9		mod 10	
Variables	Beta	rob t-ratio	Beta	rob t-ratio	Beta	rob t-ratio
Asc-alt2	0.164	3.31	0.167	3.37	0.169	3.39
Asc-sq	0.252	3.11	0.310	5.20	0.312	5.31
relb	-0.015	-1.69	-0.015	-1.71	-0.015	-1.70
exp. delay	-0.079	-3.66	-0.081	-3.76	-0.081	-3.77
free-high	0.264	5.20	0.266	5.21	0.262	5.13
free-low	-0.298	-4.36	-0.300	-4.38	-0.301	-4.41
pay-high	0.229	4.92	0.236	5.07	0.228	4.89
pay-low	-0.106	-2.22	-0.114	-2.39	-0.116	-2.41
CR-higher	-0.686	-4.21	-0.713	-4.34	-0.688	-4.21
CR-lower	0.629	4.51	0.643	4.60	0.672	4.81
CR-null	1.240	8.85	1.250	8.90	1.280	9.08
CR-ten	-0.874	-4.38	-0.902	-4.45	-0.884	-4.40
RA-higher	-0.876	-3.54	-0.902	-3.64	-0.906	-3.66
RA-lower	0.556	5.45	0.564	5.52	0.565	5.52
RA-ten	-1.420	-4.48	-1.470	-4.59	-1.480	-4.62
Ln-tc(dec)	-5.380	-14.15	-3.710	-8.57	-2.040	-3.31
Ln-tc(inc)	-7.070	-11.92	-6.780	-25.16	-6.590	-27.15
TT(dec)	0.045	-6.47	0.046	-5.40	0.037	-3.40
TT(inc)	-0.052	-6.35	-0.049	-10.65	-0.050	-12.34
init. LL	-4042.89		-4042.89		-4042.89	
conv. LL	-3334.87		-3318.21		-3311.44	
Rho sq	0.175		0.179		0.181	
Rho sq adj.	0.170		0.175		0.176	
sig. tc asymm	1.98		5.94		6.85	
sig. tt asymm	0.52		0.30		1.13	