promoting access to White Rose research papers

Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

This is an author produced version of a paper published in **Transportation Research Part A: Policy and Practice.**

White Rose Research Online URL for this paper:

http://eprints.whiterose.ac.uk/76917/

Paper:
Stathopoulos, A and Hess, S (2012) *Revisiting reference point formation, gains-losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment.* Transportation Research Part A: Policy and Practice, 46 (10). 1673 - 1689.

http://dx.doi.org/10.1016/j.tra.2012.08.005

White Rose Research Online
eprints@whiterose.ac.uk

Revisiting reference point formation, gains-losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment

Amanda Stathopoulos*[†][1] and Stephane Hess[‡][2]

[1]Transport and mobility laboratory (TRANSP-OR), ENAC, École polytechnique fédérale de Lausanne (EPFL)
[2]Institute for Transport Studies, University of Leeds

May 29, 2012

Abstract

In contrast with expected utility theory, empirical findings indicate that decision-makers are sensitive to departures from reference points rather than states. Several tests of the reference-dependent preference framework have been carried out in experimental economics, and to a smaller extent in a choice modelling setting, to date. However, these empirical applications have generally focussed on a single behavioural phenomenon using uniform modelling approaches. This paper aims to broaden existing work by presenting a multi-attribute framework, allowing contemporarily for gain-loss asymmetry, non-linearity and testing for several possible reference points. The framework is applied in the context of commuter choices and reveals important gains in model fit and further insights into behaviour compared to standard modelling approaches. Of particular relevance for future research is the functional form of fare sensitivity that varies significantly with the reference point used.

Keywords: Choice modeling, discrete choice experiment, reference-dependence, non-linearity, gain/loss deviations, commuting
JEL: C25, C9, D03, R49

1 Introduction

The notion that value or utility is strongly influenced by reference points - above all departures from reference points as defined in prospect theory - is accepted by researchers in a variety of disciplines. This has given rise to numerous corollaries, including asymmetrical utility drawn from gains and losses, non-linear probability evaluations, asymmetrical decreasing sensitivity and endowment effects to the *status quo* condition (Kahneman and Tversky, 1979, Kahneman et al., 1991). Several recent papers have looked at incorporating reference-dependence in a choice modelling setting (De Borger and Fosgerau, 2008, Hess et al., 2008, Lanz et al., 2010, Senbil and Kitamura, 2004, Delle Site and Filippi, 2011). Results indicate improved model fit along with large impacts for welfare measures when referencing is accounted for. However, extant empirical tests of reference-dependent behaviour have left a series of unresolved questions. In particular, there is scarce evidence on how referencing influences different attributes and whether other reference points matter apart from currently experienced levels. What is more, in transportation,

---

*Corresponding author.
[†]GC B3 425, bâtiment GC, Station 18, 1015 Lausanne, tel +41 21 693 2435, e-mail amanda.stathopoulos@epfl.ch
[‡]LS2 9JT, Leeds, UK, e-mail s.hess@its.leeds.ac.uk

reference-dependence is typically tested only for travel time and fare and has rarely been explored in situations with complex trade-offs among multiple attributes, a typical feature of real world choices.

In this paper, we compare evaluations of commuter trips in the context of a stated choice (SC) survey on commuting choices. We start with a linear-in-attributes utility specification, progressively incorporating insights from a reference-dependent approach, namely:

- non-linearity and decreasing sensitivity in responses,
- asymmetries when separating attribute reactions into gains and losses from the reference,
- referencing occurring against other cognitive anchors (apart from current conditions).

To account for this last possibility, gains and losses are modelled against additional plausible reference points, namely *ideal* and *acceptable* travel conditions.

The paper controls for co-occurrence of these dimensions allowing for differences across attributes. Findings indicate sizeable improvements when these effects are accounted for, in terms of model fit as well as significant shift in willingness-to-pay (WTP) and willingness-to-accept (WTA) measures. What is more, our findings show that the valuation of service improvements differs significantly depending on which reference points is used. This analysis has potentially important policy implications in that analysts, such as policy-makers or public transport operators, are typically interested in reactions to changes of current trip variables, not states.

The paper is organised as follows. The second section presents a review of existing literature, and discusses reference-dependence in the context of commuter behaviour. The data and survey instrument are described in section 3. Section 4 presents the modelling approach. Results are reported in section 5, while section 6 presents the conclusions.

2 Literature review

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.

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.

Prospect theory (PT) is built around the idea that utility is drawn from changes in endowments, not states (Kahneman and Tversky, 1979). This foundation has solved several systematic empirical violations of expected utility theory. The three fundamental features of the PT value function are: i) reference-dependence where deviations determine value, not states; ii) loss aversion with discrepancy between what agents are willing to accept to give up a choice feature and what they are willing to pay to acquire it, where losses incur a steeper inclination in the value function; iii) diminishing sensitivity whereas marginal values of both gains and losses decrease, or dampen, with higher attribute levels.

The extension of prospect theory from simple one-attribute choices with probabilistic (risky) outcomes to risk-less choice (Tversky and Kahneman, 1991) is essential in the context of the current study. Indeed, alternatives are decomposed into multiple attribute evaluations where each attribute has a distinct value function and reference point.

The literature has identified several types of reference effects and a number of these can be appropriately dealt with in a choice experiment setting. Zhang et al. (2004) set out a framework where utility is defined by the decision context. This includes 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 past choice behaviour (social/individual reference). 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.

2.1 Existing work on non-linear sensitivities

Transportation researchers are increasingly questioning the wisdom of relying on linear-in-attributes utility functions (Tapley, 2008). Early examples in transportation analysis used non-linear transformations (Koppelman, 1981) and piece-wise functions (Ben-Akiva and Lerman, 1985) to relax this assumption. Enduring evidence indicates there may be effects of damping, particularly for cost, with increasing journey distances (Daly, 2010). Recent contributions in a choice experiment setting propose non-linear models, mainly in the context of freight. Drawing on Swait (2001), Danielis and Marcucci (2007) model a kink in the utility for several freight service attributes. Separating attribute sensitivity below and above the respondent-defined maximum acceptable values significantly improves models. Masiero and Hensher (2010) frame the non-linearity around respondents' current reference values and extend the analysis to control for piece-wise marginally decreasing sensitivity. Similarly, Rotaris et al. (2012) compare a wide set of non-linearities and marginally changing attribute sensitivity in freight service evaluation. Such findings have provided valuable insights regarding non-linearities in behaviour.

2.2 Existing work on asymmetrical preference formation

Choice modelling typically allows for reference-dependence in two main ways. A first approach focusses on a differential treatment of specific alternatives, in particular reference or *status quo* (SQ) alternatives, either through the use of constants (Adamowicz et al., 1998), or by explicitly recognising that attitudes towards current alternatives may be different (cf. Ferrini and Scarpa, 2007). This recognition requires a careful treatment of such alternatives in a modelling context, either using error components or alternative-specific coefficients (cf. Scarpa et al., 2005, Hess and Rose, 2009).

A second modelling approach focusses on attributes, and associates different coefficients with positive and negative deviations from the reference. Examples from a transport setting include De Borger and Fosgerau (2008), Hess et al. (2008), Hess (2008), Masiero and Hensher (2010). These studies illustrate that there are indeed important differences between evaluations of improvements and deteriorations from a respondent's current status. Mounting proof indicates that indifference curves for losses are steeper than for improvements, generating a gap between WTP and WTA. However, the issue of sensitivity to changes in absolute versus relative levels (i.e. considering a specific reference-point) for different types of attributes is still poorly understood.

A last, largely unexplored, area of research concerns the link between referencing and personal and interpersonal behaviour. The papers cited until now in Section 2.2 rely on current status as the personal reference. In a social reference setting Schwanen and Ettema (2009) underscore the importance of socially imposed reference points, and deviations from these, in the timing of collecting children. Mahmassani et al. (1990) look at departure time adjustments in view of tolerance by colleagues of late arrival at work. Similarly, attitudes to measures such as road-pricing are shown to be highly influenced by opinions of significant others (Schade and Baum, 2007).

2.3 Which reference point?

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, research into which reference points should be employed is much more limited than the research concerning how actors react to shifts from reference-values. While Kőszegi and Rabin (2006) suggest that individual reference points may coincide with expectations of future consumption, the choice of reference point in current empirical work appears to be guided by data availability rather than theoretically solid justifications. Moreover, the point of reference that effectively guides behaviour is likely to change in view of the choice context (Loomes et al., 2009).

In a transport setting, Knetsch (2007) argues that the reference will coincide with the expected or normal state of travel for the majority of respondents. Thus, a first point of complexity is that of variability in the phenomenon. That is, respondents are typically asked to respond to SC experiments, carrying a recent or typical trip in mind, with little empirical grounds for which of these is more likely to be the actual reference for their decision making. In transportation analysis there has scarcely been any empirical exploration of variations in reference points across respondents, and the majority of published literature seems to rely on using current trip conditions as the frame of reference. Along these lines, De Borger and Fosgerau (2008) argue, in the context of a car-commuter survey, that the current trip is the most plausible reference point to assess gains and losses of time and money.

To some extent, the use of current conditions as a reference point is justified on the basis of the theory of mental Travel Time Budgets (TTB), which can also be extended to a stable mental budget for travel fare expenditure (Gunn, 1981). For instance, in the British context, surveys indicate little change in travel time and proportion of household income allocated to travel over the last 35 years (Metz, 2010). A possible explanation is that of habit-based travel decisions, where repeated commuting decisions become non-deliberate over time (Verplanken et al., 1997). On the other hand, Mokhtarian and Chen (2004), drawing on work by Mokhtarian and Salomon (2001) argue that commuters might form an *ideal* (albeit realistic, i.e. non-zero) travel time budget which may not coincide with the actual daily trip duration. In this vein, Páez and Whalen (2010) propose a study of commuter satisfaction where the dependent variable is defined as the ratio of $ideal$ to $actual$ commute time. A notable exception to the use of a sole reference point is Masiero and Hensher (2011) where a current and shifted reference point for cost, time, and punctuality is presented to freight operators. The shifted reference points are however not defined by respondents but formulated by the researchers and presented directly in the choice tasks.

2.4 Gaps in existing work

With only a handful of exceptions, applied work has focused on the use of a common reference point, namely the current travel conditions. Moreover, any asymmetry in gains and losses are assumed to follow the same specification, with identical marginal changes in sensitivity. Additionally, the same treatment in terms of reference-dependence and any non-linearity is typically used for all attributes. Indeed, to date, there has been little overlap between studies looking at reference formation and studies looking at non-linear sensitivities, despite the obvious risk of confounding between the two effects. These shortcomings form the motivation for the present work.

3 Survey work

The study draws on data from a UK stated choice survey on intra-mode commuting choices of train and bus users from 2009. Beyond standard attributes such as travel time and fare, a number of service quality features were introduced, namely availability of seating, frequency of delays, extent of delays and the availability of an information service alerting on delays. The attributes and levels are described in Table 1.

Given the large number of attributes, a highly detailed representation of crowding (Hensher et al., 2003) or reliability (see e.g. Bates et al., 2001, Batley et al., 2011) was not feasible. The final survey used a specification corresponding to a week worth of commuting: the number out of

Table 1: Overview of attributes

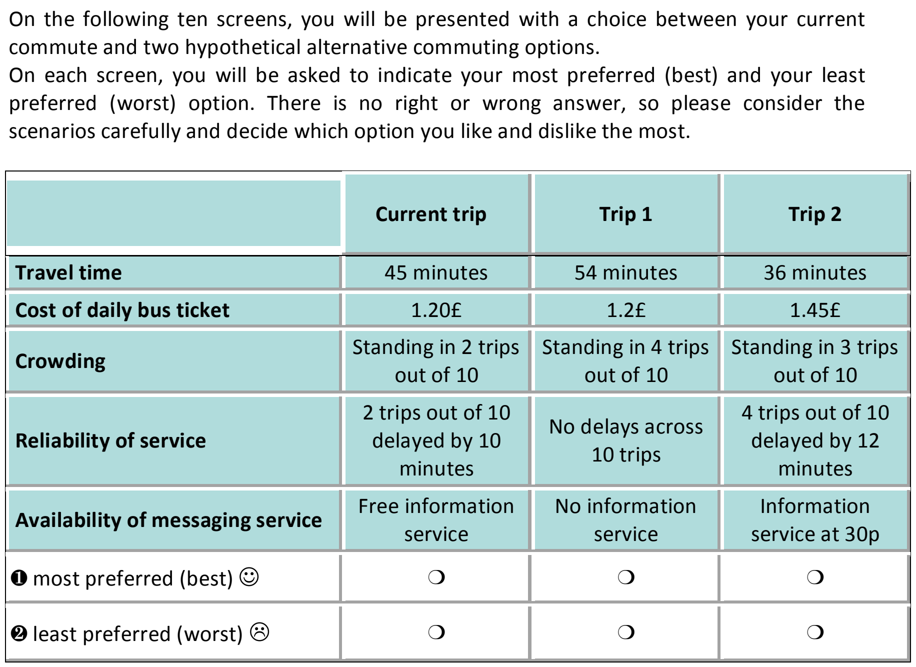
Attributes	Attribute index	N. design levels	Description of levels (bold=SQ)	Possible attribute values
Travel time (min)	TT	5	-20%, -10%, **+0%**, +10%, +20%	≥ 20
Fare (£)	FA	5	-20%, -10%, **+0%**, +10%, +20%	> 0
Crowding rate (frequency of having to stand out of 10 trips)	CR	5	-2, -1, **+0**, +1, +2	standing in 0/10-10/10 trips
Rate of delay (frequency of delays out of 10 trips)	RA	5	-2, -1, **+0**, +1, +2	delayed for 0/10-10/10 trips
Extent of delay (min)	RB	5	-30%, -15%, **+0%**, +15%, +30%	≥ 0
Information service availability (level,£)	I_NO, I_CH, I_FR	3	no service, charged service, free service	charged service: 15p for bus users, 30p for train users

ten typical trips for which the respondent would have to stand or the trip was delayed, along with the average delay duration across such trips.

A key distinction between the present work and past studies on reference-dependence is the inclusion of both certain attributes (e.g. fare) along with uncertain attributes (frequency of crowding and reliability). This allows us to study whether a probabilistic prospect is treated differently than more predictable and stable features such as average travel time and cost. Furthermore, even for the probabilistic attributes, we can look at the sensitivity to "certain" outcomes, namely situations with perfect occurrence (10 out of 10) and situations with no occurrence.

The survey used a D-efficient design created in Ngene software with appropriate conditions to avoid dominant alternatives (Rose and Bliemer, 2009). In total, 60 choice scenarios were blocked into 6 different sets of 10 tasks, minimising correlation with the blocking variable. In each task, the survey presented respondents with three trip options, with the first alternative corresponding to the current respondent-specific conditions. The remaining options were pivoted around the SQ alternative. Respondents were asked to indicate the best and worst alternative, where only the response in terms of the best trip was used in the current analysis. An example choice screen is shown in Figure 1. The data was collected through an internet panel yielding 400 respondents where 368 were used in the analysis. Socio-demographic information was gathered, with the main respondent characteristics summarised in the appendix (Table 6). The aim was not to obtain a representative sample, but instead to collect data from respondents who currently commute either by rail or bus to ensure that they could relate to the experiment.

Given the focus on analysing gains and losses from different cognitive anchor points, data on two additional reference points were collected, namely *acceptable* and *ideal* conditions for each trip attribute. To enhance realism respondents were explicitly instructed to consider technical constraints and the high usage rate of the public transport network. Results for these reference points for travel time and fare are presented in Table 2. Consistent with findings by Redmond and Mokhtarian (2001) regarding travel time and in line with expectations, the *ideal* values are lower than the current though rarely equal to zero. Furthermore, a large majority indicate *acceptable* levels as intermediate between current and ideal. Similar to the above study a small portion of respondents however declared acceptable value greater than the current (10% for time and 6%



On the following ten screens, you will be presented with a choice between your current commute and two hypothetical alternative commuting options.
On each screen, you will be asked to indicate your most preferred (best) and your least preferred (worst) option. There is no right or wrong answer, so please consider the scenarios carefully and decide which option you like and dislike the most.

	Current trip	**Trip 1**	**Trip 2**
Travel time	45 minutes	54 minutes	36 minutes
Cost of daily bus ticket	1.20£	1.2£	1.45£
Crowding	Standing in 2 trips out of 10	Standing in 4 trips out of 10	Standing in 3 trips out of 10
Reliability of service	2 trips out of 10 delayed by 10 minutes	No delays across 10 trips	4 trips out of 10 delayed by 12 minutes
Availability of messaging service	Free information service	No information service	Information service at 30p
❶ most preferred (best) ☺	○	○	○
❷ least preferred (worst) ☹	○	○	○

Figure 1: Example choice task

for fare) where this proportion was marginal for ideal values (3% and <1%). These results point towards acceptable values being interpreted as a 'constrained' ideal solution. Importantly, the different natures of these reference levels have markedly different implications when controlling for gain-loss asymmetry in the modelling section. Indeed, we gain a richer representation of the degree and type of asymmetry that can be expected. For instance, we can seize on the reduced appeal of lowering fare from ideal conditions, compared to improving upon acceptable conditions.

4 Model specification

The data were analysed within the random utility framework (McFadden, 1974) which assumes that, in choice task t (with $t = 1, \ldots, T$), individual n chooses the alternative j that maximises their utility, where the utility for j is given by $U_{j,n,t}$, which is composed of a deterministic component $V_{j,n,t}$ and a stochastic component $\varepsilon_{j,n,t}$. The deterministic component is given by interactions between measured attributes and estimated sensitivities, where, in our case, the point of departure is a base specification hypothesising linear, reference-free attribute sensitivities, with no differential treatment across alternatives. We thus have that:

$$
\begin{aligned}
V_{\mathrm{j,n,t}} = {} & \beta_{tt}\mathrm{TT}_{j,n,t} \\
& + \beta_{fa}\mathrm{FA}_{j,n,t} \\
& + \beta_{cr}\mathrm{CR}_{j,n,t} \\
& + \beta_{ra}\mathrm{RA}_{j,n,t} \\
& + \beta_{rb}\mathrm{RB}_{j,n,t} \\
& + \beta_{i-ch}\mathrm{I-CH}_{j,n,t} \\
& + \beta_{i-fr}\mathrm{I-FR}_{j,n,t} \\
& + (-\beta_{i-ch} - \beta_{i-fr})\,\mathrm{I-NO}_{j,n,t}
\end{aligned}
\tag{1}
$$

Each attribute is linear while the information service attribute is effects-coded to represent the availability of a free (I-FR) and charged service (I-CH), compared to the omitted baseline situation where the service is not available (final line in Eq. 1).

Table 2: Respondents reported current, acceptable and ideal travel time and fare

Travel time (min)	Current	Acceptable	Ideal	$\Delta curr-acc$	$\Delta curr-ide$	$\Delta acc-ide$
mean	45.79	40.30	35.61	5.49	10.18	4.69
median	40	35	30	5	10	5
st.dev	26.72	23.39	21.94			
% current=acceptable	32%					
% current=ideal	21%					
% acceptable=ideal	31%					
Fare (£)	Current	Acceptable	Ideal	$\Delta curr-acc$	$\Delta curr-ide$	$\Delta acc-ide$
mean	2.86	2.25	2.03	0.60	0.83	0.23
median	1.75	1.48	1.25	0.27	0.50	0.23
st.dev	3.80	3.42	3.19			
% current=acceptable	17%					
% current=ideal	10%					
% acceptable=ideal	34%					

Note: The fare medians are fractions due to the transformation of the stated fare into daily values

We will now discuss the various departures from this base specification, looking in turn at non-linearity and asymmetric gains-losses sensitivity.

4.1 Modelling non-linearity

4.1.1 Continuous variables

Non-linearity is modelled in two different ways depending on the nature of the attribute. For the continuous travel time and cost attributes, a non-linear transformation was used. The point of departure was a Box-Cox transformation (Mandel et al., 1994), where e.g. for travel time, we have:

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

The transformations were used as a 'diagnostic tool' and drawing on the results attributes were included in the model linearly (e.g $\lambda = 1$) or as a log-transform in cases where λ was not significantly different from 0.

4.1.2 Discrete variables

For the crowding and reliability attributes non-linearity could be captured by estimating level specific coefficients. However, estimating 10 distinct coefficients (one being normalised) for each possible attribute level is uninformative and has limited utility for policy analysis. A different approach is proposed here, where non-linearity is modelled by fitting separate coefficients to segments of the attribute levels, i.e. making use of a piece-wise linear approach. To ensure comparability with the simple linear specification, the piece-wise specification was normalised by centering the estimate on a reference value. In particular, we make use of M different segments, characterised by $M+1$ different boundary points. Using crowding as the example, we estimate the value of the start and end points, i.e. β_{cr-0} and β_{cr-10}, meaning that $k_1 = 0$, and $k_{m+1} = 10$. This leaves $M-1$ additional coefficients, namely k_2 to k_m, where, for normalisation, we set $\beta_{cr-l} = 0$, for one value of l, with $2 \leq l \leq M$. The contribution of the crowding attribute to the utility of alternative j can then be written as:

$$\begin{aligned} V_{j,n,t,cr} = & \sum_{m=1}^{M+1} \beta_{cr-m} \mathrm{I}(\mathrm{CR_{j,n,t}} = \mathrm{m}) \\ & + \sum_{m=1}^{M} \mathrm{I}(\mathrm{k_m} < \mathrm{CR_{j,n,t}} < \mathrm{k_{m+1}}) \left(\beta_{cr-k_m} + \left(\beta_{cr-k_{m+1}} - \beta_{cr-k} \right) \frac{CR_{j,n,t} - k_m}{k_{m+1} - k_m} \right) \end{aligned} \tag{3}$$

As a result, for the specific break points identified by k_1 to k_{m+1}, the actual estimates for β_{cr-k_1} to $\beta_{cr-k_{m+1}}$ will be used, with interpolated values used in-between. It is important to note that the multiplication by the observed levels ensures that the function is piece-wise linear in the β parameters but continuous in utility, avoiding issues in estimation and willingness-to-pay computation.

4.2 Modelling gains and losses asymmetry jointly with decreasing sensitivity

For modelling asymmetry, we estimate separate coefficients for gains and losses (see e.g. Hess et al., 2008). We also propose a careful and flexible treatment of non-linearity. In particular, and in line with insights from reference-dependent preference formation, we incorporate a control for two different departures from linearity. The proposed formulation controls for the presence of changing marginal sensitivity as the shift away from the reference point increases, while also evaluating the impact of the specific point of departure of a given respondent on overall sensitivity. Defining $V_{j,n,t,fare}$ to be the contribution made by the fare attribute to the utility of alternative j, and using FA_{ref} as the reference point, we would have:

$$\begin{aligned} V_{j,n,t,fare} = & \beta_{fa(inc.ref)} I \left(FA_{j,n,t} > FA_{ref} \right) \left(FA_{j,n,t} - FA_{ref} \right)^{\gamma-inc.ref} \times \left(fa_{\mathrm{n}} / \overline{fa} \right)^{\lambda} \\ & + \beta_{fa(dec.ref)} I \left(FA_{j,n,t} < FA_{ref} \right) \left(FA_{ref} - FA_{j,n,t} \right)^{\gamma-dec.ref} \times \left(fa_{\mathrm{n}} / \overline{fa} \right)^{\lambda} \end{aligned} \tag{4}$$

where $\beta_{fa(inc.ref)}$ is the coefficient associated with increases compared to the reference point FA_{ref}, while $\beta_{fa(dec.ref)}$ is the coefficient associated with decreases. Each time, the multiplication by the indicator function ensures that the correct coefficient is used, while, at the reference point, we have that $V_{j,n,t,fare} = 0$. Loss aversion occurs if $-\beta_{fa(inc.ref)} > \beta_{fa(dec.ref)}$.
The parameter γ amounts to an exponential transformation to measure decreasing sensitivity for shifts further away from the reference. Similarly to a Box-Cox transformation $\gamma = 1$ indicates a linear sensitivity, while $0 < \gamma < 1$ measures sensitivities going from strong damping (e.g the natural log-transform) to more linear sensitivities. Finally, $\gamma > 1$ implies the inverse situation of higher marginal sensitivity for values further from the $status\ quo$. In addition we allow the marginal rate of substitution to be different for gains and losses by estimating separate γ coefficients for increases and decreases. Although prospect-theory predicts that both directions of shifts are subject to uniform decreasing sensitivity, we hypothesise that losses have a much less pronounced damping than improvements.
Finally we look at specifications with two further reference points, namely the $current$ and $ideal$ values. Particularly, this implies substituting FA_{ref} for these additional reference-points. Here, it can be seen that when using the *current* value as the reference point, the contribution by the concerned attribute to the base alternative is zero. This is no longer necessarily the case with these

additional reference points, as the current value is typically different from declared $current$ and $ideal$ values. Next, fa_{n} delineating the respondent-specific current value for fare and $\overline{fa}$ giving the average across the whole sample. Thus the estimated λ indicates the impact of the currently experienced fare-level on the sensitivity to changes of the $status\ quo$. Here $\lambda = 0$ indicates a neutral effect where the current level has no impact on the sensitivities to shifts. Instead, estimates of $\lambda > 0$ means that as the base level increases, respondents become more sensitive to changes. Our prior is instead that $\lambda < 0$, indicating that at a higher base-level people will be less sensitive to a marginal shift in fare. Such findings may have large implications for the analysis of transport policy that gradually shift the reference value of respondents. The more negative the λ, the more pronounced is the reduction in sensitivity to variations.

5 Empirical results

A number of different models were estimated, progressively incorporating controls for status-quo bias, discrete and continuous non-linear impacts of attribute levels, and asymmetric utility drawn from gains and losses. Initial attempts to incorporate the impact of socio-demographic characteristics showed only marginal improvements in fit, and a generic (across respondents) specification was thus used throughout. A list of the models is given below.

Model 1: base specification with ln(fare)

Model 2: like 1, with non-linear specification for crowding and reliability and reference-dependence for information attribute

Model 3: like 2, with gain-loss asymmetry for fare from current trip

Model 4: like 2, with gain-loss asymmetry for fare from *acceptable* trip

Model 5: like 2, with gain-loss asymmetry for fare from *ideal* trip

All models were estimated using Biogeme (Bierlaire, 2008). The reported t-statistics are based on estimated robust asymptotic standard errors, where, to account for the repeated choice nature of the data, the panel specification of the sandwich estimator was used (Daly and Hess, 2011).

In line with the objective of accommodating multi-attribute dynamics, each trip characteristic was tested against the different modelling approaches. The specification search revealed the most appropriate specification to be; piece-wise non-linearity for crowding and reliability and continuous non-linearity for fare. Evidence of reference-dependence was found for fare and the information service. Decreasing sensitivity with asymmetry for gains and losses is relent for fare. Remaining modelling explorations drop back to a linear and symmetrical effect. Notably, this last case applies fully only for travel time.

5.1 Base specification

The search for a base specification implied the application of standard non-linear transformations for continuous attributes. The Box-Cox transform revealed a log transform for the fare attribute to be appropriate (β_{ln-fa}). This is in line with the literature on cost damping, i.e. decreasing marginal (dis)utility for higher levels of the attribute (see e.g. Daly, 2010). No evidence of significant decreasing marginal returns was found for the time attribute. The specification search used goodness-of-fit criteria. The model with logarithmic fare is not a generalisation of the model with linear fare, so that the likelihood-ratio (LR) test cannot be used for selection. However, the evidence from the adjusted ρ^2 statistics pointed towards a clear improvement in model fit. The results from the base specification, Model 1, are shown in Table 3. We see negative sensitivity towards increases in crowding, both reliability measures, fare, and travel time. We also note that a free delay information service is preferred to the base situation (i.e. no service), while a charged service is seen as less desirable than no service (omitted baseline). Two alternative specific constants are included, the first (δ_1) reveals a *status quo* effect, while the second (δ_2), associated with the middle alternative, captures left-to-right reading effects. Early specifications estimated separate parameters for the rate of delays (RA) and the average extent of delays across affected trips (RB). The final specification instead incorporates an interaction between these two variables, equating to the expected delay. The new coefficient $\beta_{exp.delay}$ has the expected negative sign, and its inclusion dampens the estimates for the two single effect coefficients. It should be noted that, given the nature of the data, one delay of 40 minutes is modelled in the same way as four delays of 10 minutes. Treating several smaller losses as equivalent to one larger is not necessarily consistent with real behaviour and prompts further work to distinguish between the situations.

Each of these features were included separately into the model and LR tests used as guidance in the process of specification (only final base model results are displayed for space reasons).

5.2 Models incorporating non-linearity and asymmetry

This section discusses the more advanced specifications that gradually incorporate additional non-linearities and asymmetries in the sensitivity to gains and losses. The results for non-linearity is displayed in Table 3 and the models with reference-dependence in Table 4.

5.2.1 Referencing information service

As a first step (model 2), we focus on the information service attribute, looking at differences in sensitivity depending on whether respondents currently have a free service available or not, where no significant differences were found between respondents with no service and a charged service. By comparing the preferences of the commuters that are currently experiencing a free information service (with the first subscript denoting experiment condition and the second the actual experience e.g. $\beta_{i-ch,free}$) to those that either had a charged service or no such service ($\beta_{i-ch,other}$), it is possible to assess the impact of current experience on utility for different service options (free, charged, unavailable).

The referencing for the information service obtains an improvement in log-likelihood by 2.67 units over the base specification, which, at the cost of 2 additional parameters, is significant at the 93% level (see appendix B for full breakdown of each new feature presented in Model 2). The most important observation is that although the positive evaluation of obtaining the service for free is very similar between the two groups, the disutility of having to pay is more pronounced for individuals who currently receive the service for free. This finding is in line with aversion to pricing of freely enjoyed consumption goods, for instance pricing of 'free' urban roads. On the other hand, for the other group, the implied benefit of a free service is slightly smaller, while no service is still just about preferred to a charged service ($-\beta_{i-fr,other} - \beta_{i-ch,other} = -0.117$).

5.2.2 Crowding and rate of delays

Our next step in model 2 is to explore non-linearities in the response to the rate of crowding and the rate of delays, making use of the specification described in section 4.1. The model gives us an improvement in log-likelihood by 20.83 units over a specification with linear crowding, at the cost of 5 additional parameters, which is highly significant, as is the improvement over models incorporating the non-linearity in either one of the two coefficients (see Appendix B, Table 7). The specification used for the non-linearity differs between the two coefficients, where the modelling was informed by detailed separate analysis. For crowding, we found that splitting the interval into four distinct segments was appropriate, with estimates for the extremes, breaks at the second highest and second lowest levels and a change in slope midway (5 trains out of 10, set to a base

Table 3: Estimation results for models 1 & 2

	Model 1		Model 2	
Parameters	est.	t-rat.	est.	t-rat.
δ_1	0.390	5.85	0.360	4.97
δ_2	0.163	3.30	0.163	3.30
β_{cr}	-0.223	-8.58	-	-
β_{ra}	-0.187	-5.96	-	-
β_{rb}	-0.029	-3.25	-0.017	-1.59
$\beta_{exp.delay}$	-0.062	-2.64	-0.081	-2.98
β_{ln-fa}	-6.000	-18.87	-6.020	-18.83
β_{tt}	-0.047	-9.50	-0.047	-9.47
β_{cr-0}	-	-	1.250	7.13
β_{cr-1}	-	-	0.641	3.73
β_{cr-5}	-	-	0	-
β_{cr-9}	-	-	-0.692	-3.77
β_{cr-10}	-	-	-0.885	-4.18
β_{ra-0}	-	-	0.553	4.13
β_{ra-2}	-	-	0	-
β_{ra-9}	-	-	-0.901	-3.16
β_{ra-10}	-	-	-1.450	-4.00
β_{i-fr}	0.251	6.01	-	-
β_{i-ch}	-0.171	-3.47	-	-
$\beta_{i-fr,free}$	-	-	0.267	3.97
$\beta_{i-ch,free}$	-	-	-0.308	-4.13
$\beta_{i-fr,other}$	-	-	0.229	3.92
$\beta_{i-ch,other}$	-	-	-0.112	-1.84
obs.	3,680		3,680	
par.	10		17	
LL(est.)	-3360.43		-3336.93	
ρ^2	0.169		0.175	
adj. ρ^2	0.166		0.170	

of 0). A different picture is revealed for the rate of delay attribute, where we find evidence of only three distinct segments. The base is set at a level of two out of ten trains, normalised to zero, with linear interpolation from the level at perfect reliability, i.e. β_{ra-0}. A further breakpoint is identified at the second highest level (i.e. 9 trains out 10).

To represent the implications of the specification these results are illustrated in Figure 2 which compares the implied sensitivities to the estimates from the linear specification. To overcome potential scale differences between models, WTP and WTA measures are used for the presentation[1]. Thereby values below the baseline are framed as gains (WTP) and those above as losses (WTA). For crowding, the most notable change in slope is the sharp drop when moving from no crowding to a 10% risk of crowding, while, for reliability, the biggest change is the shift from 9/10 and a sure delay. Notably, the linear specification overstates the response to crowding for higher levels while strongly underestimating the lowest level (i.e. no crowding). Indeed, it is this lack of consideration for the significant positive impact of the condition of never having to stand (CR-0) that unduly affects the estimated slope in the linear specification. This finding replicates the certainty effect from PT where people display preferences for absolutes, and dislike for loss of certainty

---

[1] To facilitate comparison, the linear specification is shifted to coincide with 0 identical to the piece-wise approach, using the same baseline of $4/10$

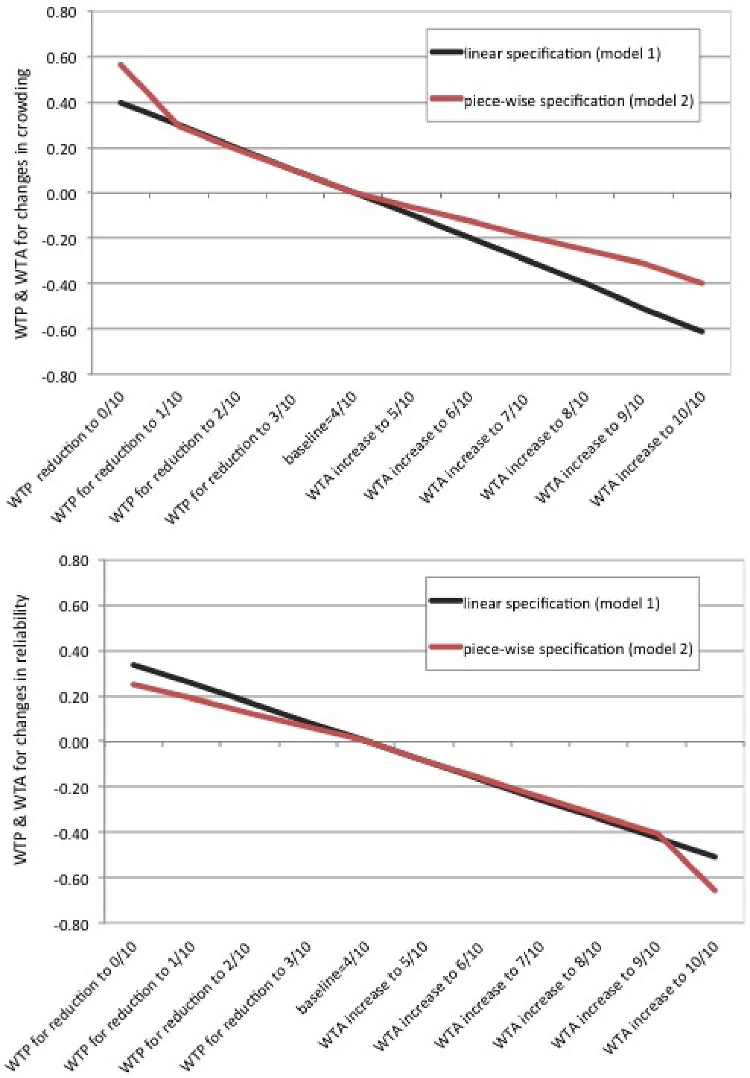


Figure 2: WTP & WTA for normalised scalar and piece-wise crowding and delay

(Kahneman and Tversky, 1979). For reliability the linear specification is a better approximation, but fails to seize on the disutility for universal delays. Interestingly, this epitomises the opposite implication of the certainty effect. In fact, for the loss domain, behaviour will tend to be risk seeking, which is exactly what we observe where people have a strong preference for the risky prospect of $9/10$ delays to avoid the disutility of a sure loss. The explanation for this opposite manifestation of the certainty effect may lie in the different nature of the two service features where crowding may allow for idealised levels of zero occurrence. Instead, the occurrence of delays is externally determined whereas it is more plausible to aspire to avoid bad outcomes.

It needs to be noted that the two frequency-based measures of crowding and rate of delays are most appropriate modelled using the interpolated segment approach and do not appear to display any consistent endence. It cannot, *a priori* be ruled out that the presentation format, using the occurrence out of 10 typical trips, influenced the observed behaviour. At the same time, it appears reasonable that commuters frame the events such as crowding and delays as frequency measures given around symbolic values (such as zero risk of standing) rather than their personal averaged experiences. Despite the significant role of absolutes in the evaluation, we still see an impact of reference-dependence for the average commuter with the manifestation of the certainty effect that implies pro-certainty for gains and pro-riskiness for the case of losses.

Table 4: Referencing models with asymmetric fare formulations

	Model 3		Model 4		Model 5	
Parameters	est.	t-rat.	est.	t-rat.	est.	t-rat.
δ_1	0.357	4.10	0.267	3.61	0.255	3.45
δ_2	0.176	3.48	0.169	3.39	0.170	3.42
β_{rb}	-0.016	-1.44	-0.014	-1.31	-0.014	-1.25
$\beta_{exp.delay}$	-0.079	-2.92	-0.080	-2.99	-0.080	-3.02
$\beta_{fa.dec}$	1.520	9.40	1.150	4.17	0.471	1.41
$\beta_{fa.inc}$	-1.340	-6.35	-2.420	-14.90	-2.100	-13.19
λ	-0.356	-3.46	-0.978	-11.12	-1.210	-11.78
γ_{dec}	0.375	-6.77†	0.841	-1.06†	0.664	-1.07†
γ_{inc}	0.403	-4.98†	1.000	0.00†	1.210	2.53†
β_{tt}	-0.050	-9.69	-0.049	-9.66	-0.049	-9.67
β_{cr-0}	1.490	8.08	1.250	7.00	1.270	7.09
β_{cr-1}	0.844	4.79	0.640	3.68	0.659	3.76
β_{cr-9}	-0.899	-4.86	-0.710	-3.86	-0.688	-3.78
β_{cr-10}	-1.120	-5.13	-0.900	-4.15	-0.887	-4.14
β_{ra-0}	0.636	4.71	0.567	4.22	0.570	4.25
β_{ra-9}	-1.230	-4.24	-0.891	-3.13	-0.882	-3.09
β_{ra-10}	-1.800	-4.95	-1.460	-3.98	-1.440	-3.91
$\beta_{i-fr.free}$	0.281	4.17	0.262	3.94	0.262	3.91
$\beta_{i-ch.free}$	-0.310	-4.09	-0.292	-3.92	-0.291	-3.85
$\beta_{i-fr.other}$	0.256	4.37	0.235	4.01	0.237	4.03
$\beta_{i-ch.other}$	-0.132	-2.16	-0.110	-1.83	-0.115	-1.91
obs.	3.680		3,680		3,680	
par.	21		21		21	
LL(est.)	-3,317.751		-3,317.219		-3,301.399	
ρ^2	0.179		0.179		0.183	
adj. ρ^2	0.174		0.174		0.178	
Asymmetry $\beta_{fa.dec}$ vs. $\beta_{fa.inc}$	0.88		2.10		4.46	
t-rat for $\beta_{fa.dec}$ vs. $\beta_{fa.inc}$	0.78		5.52		6.16	

† t-ratio refers to the test against rejecting the null of the coefficient being equal to unity (linearity)

5.3 Asymmetrical response to increases and reductions in continuous attributes

As a final step, we control for asymmetry and increasing/decreasing marginal returns. Asymmetrical response to gains and losses was only observed for the fare attribute (in addition to the earlier asymmetry for the delay information service).

The results of this process are summarised in Table 4, where we apply the formulation set out in eq. 4, additionally controlling for the use of three different respondent-reported reference points (current, acceptable and ideal). Before proceeding with a discussion of the results, it should be acknowledged that the use of respondent reported reference points could potentially lead to endogeneity bias, an issue that deserves further attention beyond this exploratory research. This could be resolved econometrically in a hybrid modelling framework treating the real reference points as latent and employs the stated reference points as indicators for these latent variables. This was however beyond the scope of the present work.

Starting with model 3, which uses the current fare as the reference point, we observe a LR statistic of 38.36, which, at the cost of 4 additional parameters over model 2, is significant above the 99% level of confidence. The difference in sensitivity between gains and losses $\beta_{fa.inc}$ and $\beta_{fa.dec}$ is not statistically significant (t-ratio=0.78). We note that γ_{inc} and γ_{dec} are significantly different from unity, indicating decreasing sensitivity, although there is no statistically significant difference between gains and losses in the degree of non-linearity. Finally, λ is moderately negative suggesting that for higher base fares the impact of changes decreases. The marginal utility for the specification from the point of view of a respondent with three different base fare levels (2£, 6£, 10£) is illustrated in Figure 3. In the top left figure we can observe that when using current fare as the reference the behaviour in the gains and losses domains is largely symmetrical, with decreasing sensitivity as shifts become larger, and also for higher base fares.

When using the respondent-reported *acceptable* value as the reference point (model 4), we observe an equally large improvement over model 2 as with the *current* value. Here, however, the degree of asymmetry is highly significant ($\left|\frac{\beta_{fa,inc}}{\beta_{fa,dec}}\right| = 2.10$) with a t-ratio of 5.52) showing that respondents view losses as more painful than equivalent gains. In addition, there is significantly less damping in either direction, with $\gamma_{inc} = 1$ implying linear sensitivity for losses and damping for gains $\gamma_{dec} = 0.84$ not significantly different from unity. As can also be observed from the top right graph in Figure 3, this gives a totally different description of behaviour where large losses, for instance an increase from a base of $\pounds 6$ to $\pounds 8$ giving twice the discomfort in the $acceptable$ compared to the $current$ model. The cost damping as a function of increases in the base (λ) is more marked in this model. This finding is consistent with the nature of the indications of acceptable fare levels, which in this setting in to be interpreted mainly as a constrained ideal value (which the commuter places near the ideal in our sample). Indeed, either improvements (towards the ideal) or deteriorations (towards the current level) incur a constant change in marginal utility, but retain a marked asymmetry. This is consistent with the notion that the indicated value is short of the ideal aspiration, thereby retaining the appeal of a lowered level, which is however matched by the well-known property of loss aversion.

Finally, using the respondent-reported *ideal* value as the reference point (model 5) leads to the best fit of the three models, with an improvement in log-likelihood over model 2 by 71.06 units, retrieving the largest ($\left|\frac{\beta_{fa,inc}}{\beta_{fa,dec}}\right| = 4.46$) and most significant ($t$-ratio of 6.16) degree of asymmetry. Notably, the difference in slope is matched by strong dissimilarities in the non-linearity. Indeed while gains undergo significant damping for larger shifts, the situation for losses is the opposite. As can be seen in the bottom graph of figure 3, for more distant increases in fare, sensitivity actually increases. This significant effect suggests that there is no habituation with losses. The cost damping as a function of the base (λ) is the most pronounced in this model. These findings have an elevated face validity, as we would expect that once a person has reached ideal values, further improvement become less appealing. Similarly, at the margin we observe similar behaviour. Indeed, each unitary decrease in fare is viewed less favourably with a more pronounced effect for people with a higher base fare.

The remaining parameter estimates remain largely unaffected across the three specifications. Using the *acceptable* and especially the *ideal* fares as the reference point not only leads to better model performance than with the commonly used *current* fare, but also indicates a higher degree of reported asymmetry. It is also worth noting that as the degree of asymmetry increases, the significance of $\beta_{fa,dec}$ reduces while that of $\beta_{fa,inc}$ increases. This is in part a result of the average *acceptable* fare being lower than the average *current* fare, while the average *ideal* fare is lower still. This means that with a change in the reference point, fewer gains (i.e. reductions in fare) will occur, with the opposite applying for losses (i.e. increases in fare).

Earlier findings concerning the role of fare evaluation in a reference-dependent preference framework offer some insight into this issue. In their work on preferences for flooding events, Lanz et al. (2010) found strongly asymmetrical response for cost (annual billing) along with asymmetry in the degree of marginal decreasing sensitivity for gains and losses, similar to the one in this paper. In fact, the coefficient for billing gains was not statistically significant with pronounced marginally

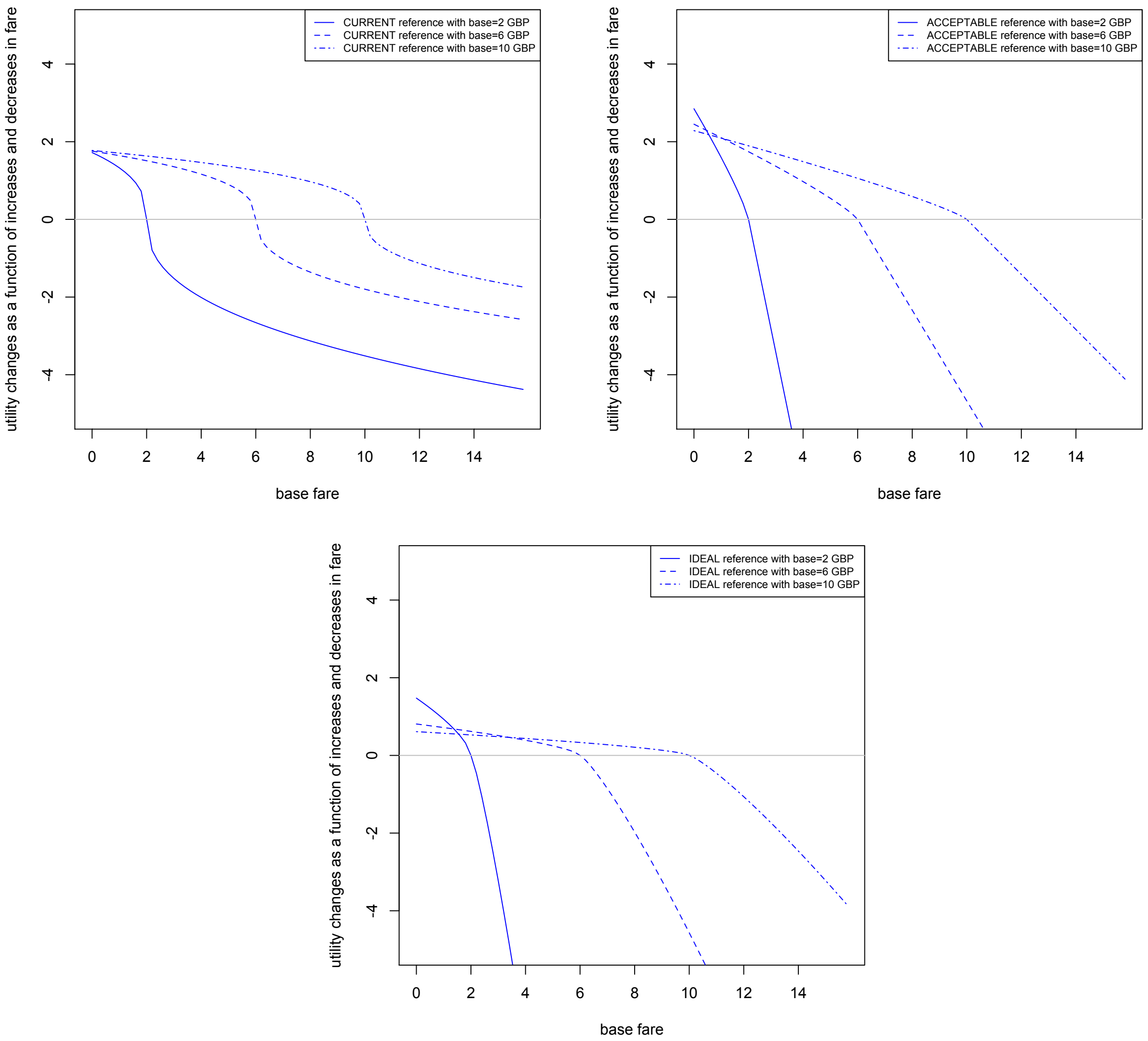


Figure 3: Utility for gains and losses of fare (with different reference-points and base values)

decreasing sensitivity. Recent work by Mabit and Fosgerau (2011) on vehicle choice with several car attributes found price to be the only feature to display significant asymmetry for gains and losses. On the other hand, in a car commuter setting, De Borger and Fosgerau (2008) found the asymmetry for cost to be smaller than for travel time. An important extension is the consideration of additional reference points to enhance the understanding of the cost attribute which is essential to study welfare effects.

The findings open a debate on the potential asymmetry in evaluations of travel costs. Redmond and Mokhtarian (2001) note, for the case of travel time, that similarity between *actual* and *ideal* travel time implies satisfaction with the commute experience whereas deviations in either direction represent dissatisfaction. However, the authors do not offer a detailed analysis of the asymmetry between the experience of such deviations. Instead, our analysis offers evidence that discrepancies between ideal, acceptable and current fare levels, does generate asymmetric effects on utility. As a general finding, falling short of ideal values is much more painful than it is favourable to obtain performances in excess of the ideal state. Importantly, the specification here offers a flexible view of the different functional form that gains and losses may display, depending on the reference-point used and the individual point of departure.

5.4 Implications for monetary valuations

The results in terms of implied willingness-to-pay (WTP) and willingness-to-accept (WTA) measures are reported in Table 5. Owing to the different specification of the fare coefficient across models we use two different methods to obtain monetary valuations. In models $1-2$, a log-transform on the fare attribute is used, making WTP a function of the fare level. Here, presented values are at the sample mean fare of $\pounds 2.72$. In models $3-5$, the WTP and WTA formulae become more complex still, given the nature of the partial derivative against the cost attribute of the full function described in Equation 4. Consistent with the presence of both marginal decreasing sensitivity and differences in the base as illustrated in figure 3 the actual WTP/WTA can be computed for every base and shift of each respondent. Consequently, to obtain the WTP, for each sample observation we include all the cases where a fare above the reference value is chosen, and take the average of the resulting WTP measures across these. Similarly, standard errors need to be calculated separately for each observation. An equivalent procedure is used to obtain WTA measures, for cases where respondents choose a fare below the reference.

Starting with the valuation of travel time, we have symmetrical WTP and WTA measures for models 1 and 2. This implies that the amount of money respondents are willing to pay to save one hour of travel time is the same as the amount of money they would require to accept an increase in travel time by one hour. In models 4 and 5, the WTA measure is higher than the WTP measure as a result of the asymmetry in the fare coefficient, with a greater sensitivity to increases than decreases. As previously discussed, the level of asymmetry is higher with the *acceptable* and especially *ideal* reference points. An interesting observation for the valuation of travel time is that WTP decreases but becomes more precise with significantly smaller standard errors when going from linear to the log-transform on fare. The estimated WTP/WTA measures may appear low in comparison with the official UK values of $\pounds 5.04/hr$ (cf. DfT, 2009), but need to be put in the context of the low average reported fares in the present data.

Turning next to crowding, the results are presented from the point of view of a respondent who currently experiences crowding on 4 out of 10 journeys. In the first model, a linear specification is used, leading to symmetrical response to increases and decreases from the starting point of 4 out of 10 journeys. The robust t-ratios are clearly also the same for each of the measures. The situation changes in model 2, where the higher sensitivity to the lower levels leads to higher WTP than WTA measures, especially for the lowest level of crowding, in line with the observations in Figure 2. It should be noted that these observations relate solely to non-linearity and are not the results of any gains-losses asymmetry as no such asymmetry was observed in the data, albeit that some may be captured by the non-linearity specification. In models $3-4$, the gap between WTP and WTA gradually increases as a result of the gains-losses asymmetry in the fare coefficient (with $\beta_{fa,inc}$ used for WTP and $\beta_{fa,dec}$ used for WTA), and in model 4, the extent of asymmetry for the fare coefficient leads to WTA being higher than WTP. The lower t-ratios in the WTA domain in model 4 are a direct result of the lower significance for $\beta_{fa,dec}$ in that model. In all cases the standard error associated with losses are more elevated than for gains. The opposite situation in model 5, where WTA measures have higher t-ratios, is due to the extreme asymmetry in the fare function where the elevated WTA make up for the higher standard errors.

The results for the rate of delays use a similar approach, once again based on a starting point of 4 out of 10 trains being affected by delays. The symmetrical specification in model 1 can be contrasted with the non-linearity in model 2 with the main effect being the big jump in WTP for avoiding a situation where all trains are affected by delays. In models $3-5$, the asymmetry between WTA and WTP becomes more pronounced as a result of the gains-losses asymmetry in the fare coefficient.

When looking at the WTP/WTA for average delays, notice that the use of a non-linear specification for the rate of delays in model 2 further reduces the role of β_{rb} and hence the resulting WTP/WTA measures. On the other hand, when looking at the WTP/WTA for expected delays, we see an increase as a result of moving to a non-linear specification for the rate of delays in model 2.

Table 5: Willingness-to-pay and willingness-to-accept measures

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
Travel time	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTP ($£/hr$)	1.28	9.60	1.28	9.53	3.19	8.43	1.26	2.40	1.34	1.38
WTA ($£/hr$)					3.10	7.52	2.47	1.85	9.94	4.14
Crowding (assume current level 4/10)	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTP for reduction to 0/10 ($£$)	0.40	8.40	0.56	6.95	1.60	4.21	0.54	2.36	0.58	1.40
WTP for reduction to 1/10 ($£$)	0.30		0.29	3.70	0.91	3.20	0.28	1.85	0.30	1.36
WTP for reduction to 2/10 ($£$)	0.20		0.19	3.70	0.60	3.20	0.18	1.85	0.20	1.36
WTP for reduction to 3/10 ($£$)	0.10		0.10	3.70	0.30	3.20	0.09	1.85	0.10	1.36
WTA increase to 5/10 ($£$)	0.10	8.40	0.06	3.67	0.19	5.01	0.12	1.79	0.47	4.12
WTA increase to 6/10 ($£$)	0.20		0.13	3.67	0.37	5.01	0.24	1.79	0.93	4.12
WTA increase to 7/10 ($£$)	0.30		0.19	3.67	0.56	5.01	0.36	1.79	1.40	4.12
WTA increase to 8/10 ($£$)	0.40		0.25	3.67	0.75	5.01	0.48	1.79	1.87	4.12
WTA increase to 9/10 ($£$)	0.51		0.31	3.67	0.94	5.01	0.60	1.79	2.34	4.12
WTA increase to 10/10 ($£$)	0.61		0.40	4.21	1.17	5.56	0.76	1.81	3.01	4.13
Rate of delays (assume current level 4/10)	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTP for reduction to 0/10 ($£$)	0.34	5.68	0.25	4.01	0.68	2.94	0.61	4.71	0.61	3.07
WTP for reduction to 1/10 ($£$)	0.25		0.19	4.01	0.51	2.94	0.46	4.71	0.46	3.07
WTP for reduction to 2/10 ($£$)	0.17		0.12	4.01	0.34	2.94	0.30	4.71	0.31	3.07
WTP for reduction to 3/10 ($£$)	0.08		0.06	4.01	0.17	2.94	0.15	4.71	0.15	3.07
WTA increase to 5/10 ($£$)	0.08	5.68	0.08	3.10	0.26	4.62	0.15	1.78	0.60	4.12
WTA increase to 6/10 ($£$)	0.17		0.16	3.10	0.51	4.62	0.30	1.78	1.20	4.12
WTA increase to 7/10 ($£$)	0.25		0.24	3.10	0.77	4.62	0.45	1.78	1.80	4.12
WTA increase to 8/10 ($£$)	0.34		0.33	3.10	1.03	4.62	0.60	1.78	2.40	4.12
WTA increase to 9/10 ($£$)	0.42		0.41	3.10	1.28	4.62	0.75	1.78	2.99	4.12
WTA increase to 10/10 ($£$)	0.51		0.66	3.93	1.88	5.19	1.23	1.81	4.89	4.13
Average delay	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTP ($£/hr$)	0.8	3.22	0.46	1.59	1.00	1.95	0.36	1.14	0.37	0.35
WTA ($£/hr$)					0.97	1.61	0.71	1.62	2.75	4.01
Expected delay	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTP ($£/hr$)	1.68	2.65	2.18	2.98	5.07	3.56	2.06	1.59	2.21	1.30
WTA ($£/hr$)					4.92	3.16	4.05	1.81	16.38	4.06
Delay information service	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
WTA for free service to charged service ($£$)	0.19	4.98	0.26	4.52	0.62	11.19	0.47	0.75	1.87	1.69
WTA for free service to no service ($£$)	0.15	4.69	0.10	1.95	0.26	3.50	0.20	0.75	0.79	1.69
WTP for no service to free service ($£$)	0.15	4.69	0.16	3.70	0.41	5.42	0.39	6.87	0.39	5.32
WTA for no service to charged service ($£$)	0.04	1.15	-		0.01	0.16	-		-	
WTP for no service to charged service ($£$)	-		0.00	0.04	-		0.01	0.55	0.01	0.28
WTP for charged service to free service ($£$)	0.19	4.98	0.15	3.13	0.42	5.13	0.37	6.56	0.38	5.20
WTA for charged service to no service ($£$)	-		0.00	0.04	-		0.01	0.27	0.02	0.69
WTP for charged service to no service ($£$)	0.04	1.15	-		0.01	0.09	-		-	

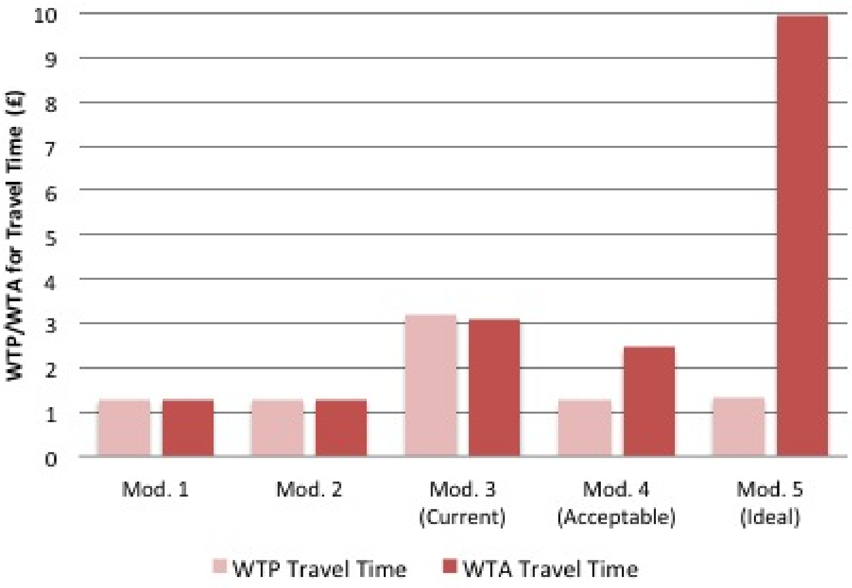


Figure 4: VOT of all models

The observations in relation to the gains-losses asymmetry as a result of the reference-dependent fare coefficient in models $3-5$ are in line with results for the other trade-offs.

For the delay information service, a number of different values can be computed. In the first model, generic coefficients are estimated independently of whether respondents currently have a delay information service or not. Here, the free service is valued higher than not having a service, which, in turn, is preferred to a charged service. As a result, we can compute a WTP for moving from a charged service to either no service or a free service, and a WTP for moving from no service to a free service. The three WTA measures are equal to their WTP counterparts, given not just the symmetrical fare coefficient, but specifically also the generic treatment independently of the current availability or not of a delay information service. This changes in model 2 (with two different points of departure) and already creates asymmetries as e.g. the move from free to charged is valued more negatively than the move from charged to free. In models 3, 4, and 5, these asymmetries are influenced further by the loss aversion in the fare coefficient. In all but three of the models, the charged service is valued more negatively than not having a service, leading to a WTP for moving from charged to no service, or a WTA for moving from no service to a charged service. In models 2, 4 and 5, this situation is reversed for those respondents who currently do not have a service or have a charged service. Overall, we see a strong aversion for respondents with a free service to move to a charged service, where in the reference-dependent models, the associated WTA measure is substantially higher than the corresponding WTP for moving from a charged service to a free service. This shows that offering a free information service with the aim of progressively introducing a charge for it may lead to undesired effects.

The impact of these asymmetries in the cost evaluation has some interesting consequences for the value of time (VOT) measures. As can be observed in Figure 4, the VOT evaluation is stable across models 1 and 2. However, the large disparities observed for improvement in the fare levels lead to a significant increase in the WTA for deteriorations in travel time in models 4 and 5. Albeit limited to one dataset, these results should serve as a warning to practitioners. Apparent stability in VOT measures despite changes in specification and associated improvements in fit could be deceptive and could be the result of not allowing for appropriate asymmetries in sensitivities. It remains to be seen whether the stability of the WTP measures (as opposed to the WTA measures) is specific to the data at hand.

6 Conclusions

This paper sets out a discrete choice modelling framework to account for different ways that referencing 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. In practice this means not simply applying a uniform modelling treatment to all attributes but instead choosing the most appropriate specification for each attribute. The proposed framework moreover offers proof concerning the important shifts when allowing for evaluations against several potential reference points. Reference-dependence with regard to points other than current trip conditions lead to important improvements in fit and further insights into the asymmetry of WTP/WTA measures.

Overall, the flexible treatment of the commute attributes reveals a series of interesting points on how changes in these attributes are perceived. In fact, the findings from this paper clearly show the importance of an attribute-by-attribute treatment of specification issues such as non-linearity and reference-dependence. At the same time, there are potentially important impacts for public transportation policies derived from the findings in this paper. Given the focus on a dedicated reference-dependence modelling approach for each attribute it is suitable to discuss the findings and relevant policy indications at this level.

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 sensitivity to higher frequencies of crowding while it fails to quantify the positive impact of never having to stand. For the frequency of delays the linear attribute specification instead fails to assess the large penalty for reaching a situation of a sure delay (10 out of 10 trips). For these attributes there is no important improvement derived from modelling gains and losses from current states. This confirms the notion that in evaluating risk of crowding and delays, defined as probabilistic frequency measures, the current experience plays little role in defining utility for alternatives. Instead, it appears that reaching absolute levels of crowding/delay is more important, particularly when it comes to the extremes. From the point of view of policy formulations this suggests that the aim of service quality improvement schemes should be to focus their message on symbolic ideal values, such as eliminating the risk of standing rather than providing general measures that improve travellers positions across-the-border. At the same time, caution should be applied to avoid falling short of such extreme promises given the non-linear weighing of different levels of performance for crowding and frequency of delays.

Commuter preferences for a delay information service, a qualitative categorical measure, was modelled using segmentation to compare sensitivities for groups with different experiences. Results revealed that depending on their current experience with the information service commuters radically change their evaluation. The most prominent policy indication to emerge is the path-dependence in preferences where respondents in a situation with charged or no service are comparatively insensitive to the service charge. The reluctance of the commuters who currently enjoy a free service cautions against the irreversibility issue, where the introduction of a free service will yield similar utility for all groups but the discontinuation generates highly asymmetrical response.

The linearity alongside symmetry in gains and losses of travel time indicates that once a specific amount of time is stably allocated for commuting purposes, deviations are perceived the same way for improvements and deteriorations. In a policy context this would lead to assuming that time can be traded against other features of the commute, without incurring a penalty for losses.

The contrasting asymmetry and decreasing sensitivity for the daily fare, however, suggests a more complex picture when ratios of time and cost are considered. Indeed, respondents display a pronounced un-willingness to accept increases in travel time in exchange for fare compensation. Importantly several dimensions, such as the slope, base-line and marginally changing sensitivity for different fare levels contribute to the complex differences between upward and downward shifts in the cost attribute. If we concentrate on the asymmetry, standard policy advice can be formulated, such as the warning that increases from the reference level generate steeper disutility than equal-magnitude gains. More innovative policy guidance can be drawn from the findings concerning marginal substitution. Indeed, when evaluating the ideal and acceptable reference-levels, we find that the law of diminishing returns applies differently to good and bad decision consequences. In particular, we can enrich the finding that gains have a flatter impact on utility, by also noting that utility for further improvements is quickly extinguished. On the contrary, when considering a loss with regard to the ideal fare level, the commuter experiences the same disutility for each marginal increase. From the point of view of a local public transportation authority such sensitivities will prompt a policy that carefully compensates each fare increase with visible improvements in service quality. We can further speculate that with experience ideal values will acquire a similar behaviour to current ones. In this case, long-term implications of a change in fare levels is the stabilisation around a more tolerant reaction where respondents will assimilate changes. A further dimension to consider is the identification of which reference point is envisioned by people when they evaluate options that result in shifts of service features. The findings within this survey suggests that this is highly relevant to understand stated reactions.

The framework proposed in this paper incorporates a set of issues that require further attention. Aside from the single data-source and stated preference nature of the data, calling for further applications, the current findings prompt several further explorations. On the side of validation the criteria of model-fit should be supplemented with analysis to corroborate the effective contributions of reference-dependent formulations. Future research needs to extend these analyses to encompass a wider variety of situations characterised by habitual and novel choices to understand the time dynamics of reference-dependence, such as the updating of reference points. The applicability of the findings would benefit from controlling for a wider set of factors such as personal features, attitudes, task-perception and other context effects, as well as incorporating inter-respondent heterogeneity in sensitivities. Further work should also explore latent variable approaches to improve the modelling of stated indications of reference points.

Acknowledgements

The present work was carried out during a visit by the first author to the Institute for Transport Studies at the University of Leeds early 2010. The authors are grateful to Andrew Daly for discussions on the modelling structure and to Marco Alderighi and Edoardo Marcucci and five anonymous referees at TRB's annual meeting 2011 and referees of Transportation Research Part A for comments on an earlier draft that materially improved the paper. The first author acknowledges support from Trieste University in the form of a *M.U.R. scholarship- Progetto Giovani Ricercatori*. The second author acknowledges the support of the *Leverhulme Trust* in the form of a *Leverhulme Early Career Fellowship*.

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.

Bates, J., Polak, J., Jones, P., and Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, 37(2-3):191–229.

Batley, R., Dargay, J., and Wardman, M. (2011). The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*, 47(1):61–72.

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis*. MIT Press Cambridge, Mass.

Bierlaire, M. (2008). *Estimation of discrete choice models with BIOGEME 1.8, biogeme.epfl.ch*.

Daly, A. (2010). Cost damping in travel demand models: Report of a study for the department for transport. RAND TR-717-DFT.

Daly, A. and Hess, S. (2011). Simple approaches for random utility modelling with panel data. In *Paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C.*

Danielis, R. and Marcucci, E. (2007). Attribute cut-offs in freight service selection. *Transportation Research Part E: Logistics and Transportation Review*, 43(5):506–515.

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.

Delle Site, P. and Filippi, F. (2011). Stochastic user equilibrium and value-of-time analysis with reference-dependent route choice. *European Journal of Transport and Infrastructure Research*, 11(2):194–218.

DfT (2009). Values of time and operating costs, tag unit 3.5.6. http://www.dft.gov.uk/webtag/documents/expert/unit3.5.6.php, 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(3):342–363.

Gunn, H. F. (1981). Travel budgets—a review of evidence and modeling implications. *Transportation Research Part A: Policy and Practice*, 15(1):7–23.

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: Policy and Practice*, 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.

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.

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

Knetsch, J. L. (2007). Biased valuations, damage assessments, and policy choices: The choice of measure matters. *Ecological Economics*, 63(4):684–689.

Koppelman, F. (1981). Non-linear utility functions in models of travel choice behaviour. *Transportation*, 10:127–146.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121:1133–1166.

Lanz, B., Provins, A., Bateman, I., Scarpa, R., Willis, K., and Ozdemiroglu, E. (2010). Investigating willingness to pay-willingness to accept asymmetry in choice experiments. In S, H. and A, D., editors, *Choice Modelling: The State-of-the-Art and the State-of-Practice*. Emerald Publishing, UK.

Loomes, G., Orr, S., and Sugden, R. (2009). Taste uncertainty and status quo effects in consumer choice. *Journal of Risk and Uncertainty*, 39(2):113–135.

Mabit, S. and Fosgerau, M. (2011). Demand for alternative-fuel vehicles when registration taxes are high. *Transportation Research Part D: Transport and Environment*, 16:225–231.

Mahmassani, H., Caplice, C., and Walton, M. (1990). Characteristics of urban commuter behaviour switching propensity and use of information. *Transportation Research Record*, 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*, 44(5):349–358.

Masiero, L. and Hensher, D. A. (2011). Shift of reference point and implications on behavioral reaction to gains and losses. *Transportation*, 38(2):249–271.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in econometrics*. Academic Press, New York.

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

Metz, D. (2010). Saturation of demand for daily travel. *Transport Reviews*, 30(5):659.

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: Policy and Practice*, 38(9/10):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.

Páez, A. and Whalen, K. (2010). Enjoyment of commute: A comparison of different transportation modes. *Transportation Research Part A: Policy and Practice*, 44(7):537–549.

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. and Bliemer, M. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews*, 29:587 – 617.

Rotaris, L., Danielis, R., Sarman, I., and Marcucci, E. (2012). Testing for nonlinearity in the choice of a freight transport service. *European Transport*, 50:1–22.

Scarpa, R., Ferrini, S., and Willis, K. G. (2005). Performance of error component models for status-quo effects in choice experiments. In Scarpa, R. and Alberini, A., editors, *Applications of Simulation Methods in Environmental and Resource Economics*, pages 247–274. Springer.

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.

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

Senbil, M. and Kitamura, R. (2004). Reference points in commuter departure time choice: A prospect theoretic test of alternative decision frames. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 8(1):19–31.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35(10):903–928.

Tapley, N. (2008). *Nonlinearities in discrete choice attributes: a study of transport-related choices*. PhD thesis, The University of Leeds, Institute for Transport Studies.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–1061.

Verplanken, B., Aarts, H., and Knippenberg, A. V. (1997). Habit, information acquisition, and the process of making travel mode choices. *European Journal of Social Psychology*, 27(5):539–560.

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 Part B: Methodological*, 38:215–234.

Table 6: Appendix A: Descriptive statistics for the sample

Attributes	Definition	Mean	St.dev	% rates
Age (years)	Average of mean age within 7 age bands	34.61	10.95	
Income (£)	Average of mean annual income within 9 income bands	25,136	16,143	
Sex	0=male, 1=female	0.61	0.49	
Education reached	1=mandatory school, 2=high school, 3=university	1.81	0.75	40 % university
Information service	0=not available, 1=available at charge, 2=available for free	0.79	0.95	36% free info. service
Car availability	1=no car availability, 2=car availability	1.51	0.50	51% has car
Current tt (min)	Average stated travel time	45.79	26.72	
Current fare (£)	Average stated daily fare	2.86	3.80	
Current delay (freq)	Average stated number of delays in 10 trips	3.41	2.53	
Current delay (min)	Average stated delay across delayed trips	10.07	9.25	
Current crowding (freq)	Average stated number of times having to stand in 10 trips	3.33	3.07	

Table 7: Appendix B. model comparison of intermediary specifications

	Model 1	Model 1.A	Model 1.B	Model 1.C	Model 2
elements added		Mod. 1 + ref. dependence information	Mod. 1 + ref. dependence information + p-w reliability	Mod. 1 + ref. dependence information + p-w crowding	Mod. 1 + ref. dependence information + p-w crowding + p-w reliability = Mod. 2
par.	10	12	14	15	17
LL(est.)	-3360.43	-3357.76	-3352.78	-3341.44	-3336.93
ρ^2	0.1688	0.1695	0.1707	0.1735	0.1746
adj. ρ^2	0.1663	0.1665	0.1672	0.1698	0.1704
LR against Mod. 1		5.34			
LR against Mod. 1.A			9.96	32.64	
LR against Mod. 1.B					31.7
LR against Mod. 1.C					9.02

Notes: Log likelihood ratio statistics illustrate model evolution to justify elements added from model 1 to 2. Critical χ^2 value for 2 degrees of freedom is 5.99 and with 3 degrees is 7.81