Paper:
Hess, S, Stathopoulos, A and Daly, AJ *Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies.* Transportation . (in press).

http://dx.doi.org/10.1007/s11116-011-9365-6

Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies

Stephane Hess* Amanda Stathopoulos† Andrew Daly‡

July 22, 2011

Abstract

The study of respondent heterogeneity is one of the main areas of research in the field of choice modelling. The general emphasis is on variations across respondents in relative taste parameters while maintaining the assumption of homogeneous utility maximising decision rules. While recent work has allowed for differences in the utility specification across respondents in the context of looking at heterogeneous information processing strategies, the underlying assumption that all respondents employ the same choice paradigm remains. This is despite evidence in the literature that different paradigms work differently well on given datasets. In this paper, we argue that such differences may in fact extend to respondents within a single dataset. We accommodate these differences in a latent class model, where individual classes make use of different underlying paradigms. We present four applications using three different datasets, showing mixtures between "standard" random utility maximisation models and lexicography based models, models with multiple reference points, elimination by aspects models and random regret minimisation models. In each of the case studies, the behavioural mixing model obtains significant gains in fit over the base structure where all respondents are hypothesised to use the same rule. The findings offer important further insights into the behavioural patterns of respondents. There is also evidence that what is retrieved as taste heterogeneity in standard models may in fact be heterogeneity in decision rules.

Keywords: random utility; behavioural mixing; taste heterogeneity; elimination by aspects; lexicography; reference-dependence; latent class; random regret

*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611, Fax: +44 (0)113 343 5334

†Department of Economics, Business, Mathematics and Statistics (DEAMS), University of Trieste, Italy, amandairini.blombergstathopoulos@phd.units.it

‡Institute for Transport Studies, University of Leeds, and RAND Europe, daly@rand.org

1 Introduction

A significant part of the recent research effort in the field of choice modelling has been dedicated to the study of respondent heterogeneity, with a particular focus on variations in the parameters of the utility functions. Such heterogeneity has been introduced either through deterministic interactions or through random coefficients, in continuous mixture models (cf. Revelt and Train, 1998; Hensher and Greene, 2003) or finite latent class methods (see e.g. Gopinath, 1995; Greene and Hensher, 2003).

While these departures from a taste homogeneity model generally lead to significant gains in model fit, there is no recognition that what may be causing the heterogeneity are not in fact simply variations in marginal sensitivities but actual differences in the choice process by individual respondents. Indeed, these models are based on the assumption that the underlying behavioural process is the same across respondents.

Recent research on information processing has moved on from this, by allowing the actual utility specification to vary across respondents, for example with some respondents ignoring certain attributes, where in the most appropriate specifications, a probabilistic approach based on latent class structure is used (see e.g. Hess and Rose, 2007; Hensher and Greene, 2010).

Even in the work on heterogeneous information processing strategies, however, the underlying behavioural paradigm remains the same across respondents, namely that of maximisation of utility by individuals, with only the specification of utility varying. There is however evidence in the literature that alternative choice paradigms may fit better on certain datasets (see e.g. recent discussions on happiness by Abou-Zeid and Ben-Akiva 2010 and regret by Chorus et al. 2008).

In the present paper, we highlight the fact that the actual behavioural process used in making a choice may in fact vary across respondents within a single dataset. We discuss how this can be accommodated in a latent class framework, and illustrate the approach in four case studies each concerned with an alternative decision paradigm. While this issue has received some attention in marketing and health economics (see e.g. Gilbride and Allenby, 2004; Araña et al., 2008), it has been largely ignored in a transport context. Moreover, earlier work has focussed on a narrow set of decision paradigms, typically some version of unordered elimination conjuntive/disjunctive rules (Jedidi and Kohli, 2005; Gilbride and Allenby, 2004) or lexicography (Kamel and Rajeev, 2008). Instead, the approach presented here is sufficiently flexible to explore a wide array of decision paradigms. We present a rigourous comparison across different modelling specifications of the relative impact of taste heterogeneity and heterogeneity in decision paradigms. Particularly, we ask the question whether the heterogeneity in relative sensitivities retrieved with standard approaches may in fact be due precisely to such heterogeneity in behavioural process. This extension to multi-paradigm models is very timely, given the renewed interest in alternative paradigms (e.g. Abou-Zeid and Ben-Akiva, 2010; Chorus et al., 2008).

The applications illustrate the improvements gained when allowing for mixtures between "standard" random utility maximisation models and alternative paradigms, namely lexicography based models, models with multiple reference points, elimination by aspects models, and random regret minimisation models. The four studies use three different datasets that are particularly well suited for this analysis. In each of our case studies, the behavioural mixing model obtains significant gains in fit, and further insights into behavioural patterns.

The remainder of this paper is organised as follows. The next section presents the general modelling approach. This is followed by four case studies making use of this approach. Finally, we summarise the findings of the work and present our conclusions.

2 General methodology

The approach used in this paper is based on a latent class (LC) structure. Rather than allowing simply for differences in the utility parameters (or even the utility specification) across classes, we allow for differences across classes in the actual behavioural process. Such a flexible approach allows for the study of a wider array of decision paradigms.

Let us assume that we have N decision makers, where decision maker n is faced with T_n separate choices. Let $P_n\left(\beta_{(m)}, m\right)$ give the probability of that sequence of choices, conditional on using a choice model identified as m, where this model m uses a vector of parameters $\beta_{(m)}$. If as an example m equates to a RUM structure, we would have that:

$$P_n\left(\beta_{(Umax)}, Umax\right) = \prod_{t=1}^{T_n} P\left(U_{j^*_{nt}} \geq U_{j_{nt}}, \forall j \in J\right) \quad (1)$$

where j^*_{nt} is the alternative (out of J) chosen by respondent n in choice situation t, and $U_{j_{nt}}$ is the utility of alternative j as faced by respondent n in choice situation t; the dependency of utilities on estimated parameters and explanatory variables is not explicitly shown here.

We hypothesise that a number of different behavioural processes are used in the data, and thus allow for M different models, each based on its own vector of parameters and with potentially very different model structures. The choice of decision rule for a given respondent is not observed and is thus treated as a latent component. The probability for the sequence of choices observed for respondent n is now given by:

$$P_n = \sum_{m=1}^{M} \pi_{n,m} P_n\left(\beta_{(m)}, m\right) \text{ where } \sum_{m=1}^{M} \pi_{n,m} = 1 \text{ and } 0 \leq \pi_m \leq 1 \; \forall m, \tag{2}$$

where we use different behavioural processes in different classes, i.e. the difference across classes lies not just in the use of different parameters (as is typically the case with LC models), but also in different underlying models.

With this model, we need to estimate parameters of the choice models in the individual classes ($\beta_m, \; m = 1, \ldots, M$), as well as the probabilities for each class ($\pi_m, \; m = 1, \ldots, M$). By performing the averaging at the level of sequences of choices for the same respondent, we take into account the repeated choice nature of panel data, allowing for inter-respondent differences, but maintaining the same model across choices for the same respondent. A possible extension not pursued here is to link class allocation to respondent characteristics, by formulating a class allocation model.

For the present paper, all models were coded and estimated in Ox 4.2 (Doornik, 2001). The estimation of the class probabilities and within class models was performed simultaneously. To deal with the issue of local optima, each time we launched multiple estimation runs of our models with different random sets of starting parameters. In each case, we chose the model leading to the best likelihood. Overall, the solutions obtained with different starting values were rather stable.

3 Case studies

In this section, we present findings from four separate case studies, each comparing a standard model to a structure based on a mixture between models allowing for different behavioural rules to coexist. In the first case study, we investigate a mixture between a standard random utility maximisation (RUM) model and a lexicography based model. In the second case study, we make use of our proposed approach in the context of looking at heterogeneous reference points. The third case study revisits one of the first models for representing choice behaviour, namely the elimination by aspects (EBA) model. Finally, the fourth case study uses a mixture between a standard RUM structure and a random regret minimisation (RRM) model. The modelling approach is tested on three different datasets, where each paradigm-dataset combination reflects the suitability of the data for identifying the decision rule. The precise rationale for the data used in each case study is illustrated in the relevant sections.

The key tests are the generalisations of models to include additional parameters and in these cases formal χ^2 tests can be made. However, in one case study we need to compare models with different numbers of estimated parameters that are not nested and here we use the BIC criterion (Schwarz, 1978) which does not permit formal tests but allows a general assessment of the models' relative success.

3.1 Case study I: RUM and lexicography

3.1.1 Behavioural process under investigation

Standard RUM theory is based on the notion of compensatory behaviour which states that gains in one attribute can be traded against losses in another. Lexicographic models (cf. Luce, 1978; Tversky, 1969) are an expression of bounded rationality leading to a simplification of the choice process. Individuals give priority to a single attribute and only when alternatives are equally good on this attribute do they consider a second attribute. The ordering of attributes in terms of importance potentially varies across individuals.

Some authors have argued that actual lexicography, in the sense of sorting on a preferred choice feature, is not consistent with compensatory modelling frameworks (Sælensminde, 2006). However, distinguishing between lexicography and steep indifference curves, the latter being compatible with RUM, is not always possible (cf. Killi et al., 2007). Indeed, it is questionable whether an analyst could ever infer whether a respondent always choosing the cheapest option is indeed behaving lexicographically, or whether the presented incentives were simply not large enough to encourage trading, a point supported by an adaptive experiment by Cairns and van der Pol (2004). For modelling, we would prefer to take the more plausible explanation for any given individual, and this may vary across respondents.

As an additional complexity, while apparent lexicographic behaviour is easy to spot in the case of surveys with only two attributes, this becomes significantly more difficult with a larger number of attributes. Indeed, many different rules will become possible, involving different orderings as well as numbers of levels, and it may not be possible to fully identify the rule leading to a given choice when the design is not conceived to carry out such tests. The fact that there may be uncertainty as to which rule was used argues for the use of a probabilistic approach such as suggested here.

3.1.2 Data & model specification

The analysis in this section makes use of data from the Danish Valuation of Travel Time (VTT) study (see e.g. Fosgerau 2006). This part of the survey presented a binary unlabelled choice between car commute trips, characterised only by different travel time and cost to a sample of drivers. In the present analysis, we employ a sample of $1,676$ respondents, who each faced 8 meaningful choice tasks. The specific reason for making use of this dataset in the present case study is that the use of only two attributes facilitates the identification of lexicography.

An initial analysis of the data showed that 13.66% of respondents always choose the cheaper of the two options, while 5.97% of respondents always choose the faster one. A multitude of different reasons for this type of behaviour arise, as discussed for example by Hess et al. (2010) in the context of this dataset. Lexicography may be a strategy to deal with choice complexity (Sælensminde, 2002), an effect of boredom and disengagement (cf. Bradley and Daly, 1994) or indeed a result of a lack of incentives to trade among attributes, i.e. the presence of strong sensitivities (Ryan and Farrar, 1994). The behaviour may arguably also be limited to the context of the survey at hand. Removing these respondents from the data arbitrarily assumes that they are behaving in a manner that is inconsistent with our analytical framework (cf. Lancsar and Louviere, 2006; Hess et al., 2010). However, their simple inclusion in the models, without treatment, arguably biases findings, especially in terms of heterogeneity, as the model will attempt to explain their non-trading behaviour by allowing for extreme sensitivities.

We first estimate a simple MNL model, attempting no treatment of the potential lexicographic behaviour or any random heterogeneity in sensitivities. Our second model (MNL & LEX) is a LC structure with three classes. Class 1 is a simple MNL model (M_1), using the same specification as in the base model. The remaining two classes are modelled by means of (deterministic) lexicography based rules. Here, model M_2 represents lexicography on travel time (TT), and model M_3 lexicography on travel cost (TC); we note that the data design precludes the existence of ties. With this, we have that:

$$P_n\left(M_2\right) = \prod_{t=1}^{T_n} I_{TT_{j_{nt}}}, \tag{3}$$

where $I_{TT_{j_{nt}}}$ is equal to 1 if the travel time for the alternative chosen by respondent n in choice set t is less than that of the competing alternative. We also have that

$$P_n\left(M_3\right) = \prod_{t=1}^{T_n} I_{TC_{j_{nt}}}, \tag{4}$$

where $I_{TC_{j_{nt}}}$ is defined analogously to $I_{TT_{j_{nt}}}$. Equations 3 and 4 are conditional only on the data and not on any parameters, given the deterministic nature of these two models.

The probability under a given lexicographic rule will be equal to 1 only if every single choice for that respondent can be explained by the specific rule. In other words, only a respondent whose observed choices exhibit *apparent* lexicographic behaviour is *eligible* to be captured by these classes. In this model, the apparent lexicography is accommodated solely through special classes, with no attempts to explain it through taste heterogeneity. As a result, the shares for these two classes will be equal to the sample population shares for this type of behaviour, and estimates for the trading class will be equivalent to what would be obtained if we simply removed lexicographic respondents from the sample.

Our third model once again uses only a single class, given by a Mixed Multinomial Logit (MMNL) model, employing a multivariate Lognormal distribution across respondents. In particular, we have that:

$$P_n(\Omega, M_1) = \int_\beta \prod_{t=1}^{T_n} \frac{e^{V_{j^*_{nt}}(\beta)}}{\sum_{j=1}^{J} e^{V_{j_{nt}}(\beta)}} f(\beta \mid \Omega)\, \mathrm{d}\beta, \tag{5}$$

where Ω is a vector of parameters (to be estimated) of the multivariate distribution $f(\beta \mid \Omega)$. This structure thus offers no special treatment of lexicography, with any non-trading behaviour explained solely through taste heterogeneity.

Finally, our fourth model combines the two approaches, where we once again make use of three classes, with class 1 modelled by Equation 5, and classes 2 and 3 by Equation 3 and Equation 4 respectively. This model thus includes special classes for lexicography while also allowing for random heterogeneity. Our expectation is that this will allow the model to accommodate some of the non-trading behaviour on the basis of reasonably heightened time and cost sensitivities, with any respondents whose behaviour would lead to extreme sensitivities being captured by the two additional classes.

3.1.3 Estimation results

The estimation results for the first case study are summarised in Table 1. We observe very significant gains in model fit when moving from the MNL model to the MNL & LEX model (895.79 units for 2 additional parameters), while the gains when moving from MNL to MMNL are even more substantial ($1,565.27$ units for 2 additional parameters). Finally, the combined MMNL & LEX model comprehensively outperforms the simple MNL model and MNL & LEX models (1575.56 units for 4 parameters and 679.77 units for 2 parameters respectively), while the improvement by 10.29 units over the MMNL model is also statistically significant, coming at the cost of 2 additional parameters.

Three parameters are common across models, namely the mean estimates for the two marginal utility coefficients (μ_{TC} and μ_{TT}), and a constant for the first alternative (δ_1). In the two models incorporating a treatment of random taste heterogeneity, s_{11}, s_{21}, and s_{22} give the elements of the Cholesky matrix, where, with ξ_1 and ξ_2 giving two independently distributed standard normal variates, we have that $\beta_{\mathsf{TC}} = \mu_{\mathsf{TC}} + s_{11}\xi_1$, and $\beta_{\mathsf{TT}} = \mu_{\mathsf{TT}} + s_{21}\xi_1 + s_{22}\xi_2$. Furthemore, γ_2 and γ_3 are used in the class

Table 1: Estimation results for first case study (VTT measures relating to trading classes only in MNL & LEX and MMNL & LEX models)

	MNL		LC			MMNL - LN		LC-MMNL		
Observations	13,408		13,408			13,408		13,408		
Log-likelihood	-8,925.89		-8,030.10			-7,360.62		-7,350.33		
par.	3		5			5		7		
adj. ρ^2	0.0393		0.1354			0.2075		0.2084		

				t-rat.					t-rat.	
	est.	t-rat.	est.	*vs* 0	*vs* $1/3$	est.	t-rat.	est.	*vs* 0	*vs* $1/3$
δ_1	0.1700	9.56	0.2141	10.70	-	0.3080	12.64	0.3054	12.57	-
μ_{TC}	-0.0536	-20.64	-0.0743	-22.80	-	-1.0770	-13.39	-1.2959	-15.94	-
μ_{TT}	-0.0368	-14.18	-0.0556	-17.87	-	-1.8364	-24.53	-1.8464	-25.73	-
s_{11}	-		-		-	2.4043	22.02	-2.0980	-20.34	-
s_{21}	-		-		-	1.5333	14.16	-1.3398	-13.37	-
s_{22}	-		-		-	1.2446	52.57	-0.9242	-25.70	-
γ_2	-		-1.8016	-24.61	-	-		-2.5709	-14.54	-
γ_3	-		-2.6321	-24.75	-	-		-3.5363	-13.35	-
μ_{VTTS} (DKK*/hour)	41.19	24.17	44.89	31.84	-	89.01	11.99	70.69	15.12	-
$\widehat{\mu_{0.5}}_{\mathsf{VTTS}}$ (DKK/hour)	41.19	24.17	44.89	31.84	-	28.08	19.24	34.60	17.25	-
σ_{VTTS} (DKK/hour)	0.00	-	0.00	-	-	267.76	7.38	125.94	8.61	-
cv_{VTTS}	0.00	-	0.00	-	-	3.01	16.88	1.78	15.04	-
cv_{TT}	0.00	-	0.00	-	-	6.96	5.96	3.63	6.66	-
cv_{TC}	0.00	-	0.00	-	-	17.97	3.80	8.98	4.56	-
π_{trading}	100%	-	80.84%	82.83	48.68	100%	-	90.45%	65.46	41.34
$\pi_{\mathsf{lex\text{-}cost}}$	0%	-	13.34%	15.84	-23.74	0%	-	6.92%	6.12	-23.39
$\pi_{\mathsf{lex\text{-}time}}$	0%	-	5.81%	10.04	-47.49	0%	-	2.63%	3.91	-45.58

* DKK 1 is approximately equivalent to UK $£0.12$

allocation model, where the probabilities for the different classes are obtained as:

$$\pi_{\text{trading}} = \frac{1}{1+e^{\gamma_2}+e^{\gamma_3}}, \quad \pi_{\text{lex-cost}} = \frac{e^{\gamma_2}}{1+e^{\gamma_2}+e^{\gamma_3}}, \quad \pi_{\text{lex-time}} = \frac{e^{\gamma_3}}{1+e^{\gamma_2}+e^{\gamma_3}} \tag{6}$$

Table 1 also shows the mean (μ_{VTT}) and median ($\widehat{\mu_{0.5}}_{\text{VTT}}$) VTT measures, the standard deviation for the VTT (σ_{VTT}), and the coefficient of variation (cv) for the VTT and the two marginal utility coefficients. For the MNL & LEX and MMNL & LEX models, the VTT measures only relate to the trading classes; with in effect VTT measures of zero and plus infinity applying in the lex-cost and lex-time classes respectively.

In all four models, we note some reading-left-to-right effects in the estimate for δ_1. The mean estimates for the two marginal utility coefficients are negative and significant in all models, while the two random coefficients models retrieve significant levels of unexplained inter-respondent heterogeneity. As expected, the simple MNL & LEX model produces weights for the two lexicography classes in line with sample population shares, while, in the combined MMNL & LEX model, the shares are lower as some of the behaviour is captured by the tail of the Lognormal distribution.

In terms of VTT measures, we observe a small increase when moving from MNL to MNL & LEX. This comes as a result of the larger share of respondents always choosing the cheaper option compared to respondents always choosing the faster option. The mean VTT measures in the two mixture models are substantially higher, with the value for MMNL being the highest. On the other hand, the median VTT measures are lower than in the MNL and MNL & LEX model, where they are lowest in the MMNL model. Most crucially however, we observe a major reduction in the level of random taste heterogeneity when moving from MMNL to MMNL & LEX, where the degree of reduction is very substantial compared to the *retrieved* rates of apparent lexicography. This confirms our hypothesis that even a relatively small share of respondents can have an undue influence on our findings in terms of random heterogeneity if their non-trading behaviour can only be explained by very extreme sensitivities.

Returning to the above point about the MNL & LEX and MMNL & LEX results relating solely to the trading class, it is possible to compute a median (though not mean, given infinite values in class 2) for the overall model, which is 32.28 DKK/hour. This value is thus only slightly lower than the value for the trading class only. It is however higher than the median VTT for the MMNL model of 28.21 DKK/hour, reflecting the fact that this model attempts to accommodate the lexicographic behaviour through an extreme tail on the distribution of the cost coefficient. This finding highlights the importance of accommodating possible lexicography to avoid potential bias in willingness to pay estimates.

The findings from this section are clearly specific to the data at hand, and are also potentially influenced by the distributional assumptions. Indeed, the Lognormal distribution is well suited for accommodating outlying sensitivities and different shares would have been obtained with alternative assumptions. Ideally, the experiment reported here should be repeated with non-parametric distributions (cf. Fosgerau, 2006). The advantage of the model used here is that it allows the non-trading behaviour to be captured by the tails of the distribution but only up to the point where the resulting shape would unduly affect the capability of the model to accommodate the trading part of the sample population.

3.2 Case study II: heterogeneous reference points

3.2.1 Behavioural process under investigation

The theory of reference-dependent choice postulates that in making decisions, individuals identify a reference point and judge possible outcomes in terms of gains and losses relative to this (see e.g. Kahnemann and Tversky, 1979; van Osch et al., 2006). This is in contrast with the standard RUM focus on absolute attribute sensitivity, which in essence equates to the reference point being equal to zero.

A small number of tests have been carried out in a choice experiment setting to control for asymmetric evaluation of multiple attributes in a transport context (see e.g. Hess et al., 2008; De Borger and Fosgerau, 2008; Masiero and Hensher, 2010). An important question arises as to the determination of the reference point. A common approach is to assume that any attribute's reference point coincides with its status quo value, e.g. the current travel time. This approach is especially popular when dealing with datasets that include an explicit reference alternative. Findings in a wide range of situations however indicate that the real reference point can be past states (see e.g. Kahneman et al., 1991), beliefs about future states (Koszegi and Rabin, 2006), aspirations compared to a reference group (Stutzer, 2004), or even an arbitrary anchor with no relation to the choice at hand (Ariely et al., 2003). In a transport setting there have scarcely been any empirical explorations of variations in reference points across respondents, with a notable exception being the work of Masiero and Hensher (2011), where both current and shifted reference points are used. However, these reference points are *presented* to respondents, whereas our work accounts for the fact that reference points *used* by respondents may well be different from those *presented*. Additionally, we use three different possible reference points.

3.2.2 Data & model specification

In line with the above observations, we make use of data that permits the study of the formation of *different* reference points. In particular, we take observations from a survey looking at commuting by rail and bus, collected through an online panel in the United Kingdom in early 2010. The survey presents respondents with choices between three alternatives described by six attributes each: travel time, fare, frequency of seat availability, frequency of delays, extent of delays and the availability of a text message (SMS) delay alert service. The first alternative corresponds to a typical trip for that respondent, while the remaining two alternatives are symmetrically pivoted around the current conditions. The scenarios presented absolute values to respondents to facilitate comparisons. Each of the 360 respondents used in the current sample was presented with 10 such choice scenarios.

To explore the use of reference-dependence with regard to points other than current trip conditions, information on two additional values was collected from respondents, equating to an *acceptable* level and an *ideal* level for each attribute. In defining these points respondents were explicitly instructed to consider technical constraints and the high usage rate of the transit network. A previous paper on this dataset has shown evidence of asymmetrical preference formation around either of these three reference points (Stathopoulos and Hess, 2010). The results from this earlier work also highlight the importance of applying a logarithmic transform for the fare attribute.

Five different models were estimated on this sample. The first model makes use of symmetrical coefficients for all attributes. This is followed by three specifications that allow for asymmetrical preference formation for travel time and fare sensitivities, since previous work indicated symmetric sensitivities for the remaining attributes. The contribution to the utility of alternative i contains the following components relating to travel time, with a corresponding approach applying for the log of the fare attribute:

$$V_{i_{nt}} = ... + \beta_{\mathsf{TT,inc}} \max\left(0, \mathsf{TT}_i - \mathsf{TT}_{\mathsf{ref}}\right) + \beta_{\mathsf{TT,dec}} \max\left(0, \mathsf{TT}_{\mathsf{ref}} - \mathsf{TT}_i\right) + ... \quad (7)$$

where $\mathrm{TT}_{\mathrm{ref}}$ gives the reference point for the travel time attribute, and where $\beta_{\mathrm{TT,inc}}$ and $\beta_{\mathrm{TT,dec}}$ are the coefficients for increases and decreases respectively.

Our first departure from the base model uses the *current* values for travel time and fare as the reference points, which are identical to the values used for the first alternative. This is followed by a model using the respondent-stated *acceptable* values for the travel time and fare attributes, and a model making use of the respondent-stated *ideal* values for travel time and fare. We acknowledge possible issues with endogeneity when using respondent reported reference points, but argue it should be placed in the context of seeking to avoid bias caused by not accounting for asymmetrical preference formation around such points. Finally, the LC model makes use of four different classes, incorporating the specification from the four separate models discussed above. In this model, the coefficients used in the three asymmetric classes are generic, only the definition of the reference point changes. No additional models were estimated that allow for random taste heterogeneity.

3.2.3 Estimation results

The estimation results for the second case study are summarised in Table 2. The specification used for this dataset estimates constants for the first two alternatives (δ_1 and δ_2), along with marginal utility coefficients for travel time (TT), the logarithm of fare (L-FARE), the rate of delays (trips out of 10), the expected delay (rate multiplied by average delay for affected trips), the rate of having to stand (trips out of 10), and the provision of a charged or free delay SMS alert system (dummy coded for the two levels). In addition, Table 2 reports estimates for the three parameters used in the class allocation model (γ_2, γ_3 and γ_4), along with the resulting class allocation probabilities. In the base model, only the linear time and log-fare coefficients are estimated, while only their asymmetrical counterparts are estimated in the three reference dependent models. Finally, both sets of coefficients are used in the LC model.

The estimation results show that all three asymmetrical specifications lead to modest but statistically significant gains in fit over the base model, by 2.83, 20.20 and 27.21 units respectively, each at the cost of 2 additional parameters. The degree of asymmetry is small for travel time, but is very noticeable for log-fare. Of the three specifications, the best performance is obtained when making use of the *ideal* values as reference points, followed by the model making use of the *acceptable* values. This finding alone

Table 2: Estimation results for second case study

	Symmetrical	Ref *base*	Ref *acceptable*	Ref *ideal*	LC
Observations	3,680	3,680	3,680	3,680	3,680
Log-likelihood	-3,369.23	-3,366.40	-3,349.03	-3,342.02	-3,073.30
par.	9	11	11	11	16
adj. ρ^2	0.1644	0.1646	0.1689	0.1706	0.2359

	Symmetrical		Ref *base*		Ref *acceptable*		Ref *ideal*		LC		
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat. *vs* 0	t-rat. *vs* $1/4$
δ_1	0.3887	7.81	0.2680	3.51	0.3350	6.47	0.3386	6.67	0.2261	3.48	-
δ_2	0.1725	3.56	0.1734	3.56	0.1763	3.62	0.1780	3.65	0.2473	4.51	-
β_{TT}	-0.0467	-13.18	-		-		-		-0.0475	-8.08	-
$\beta_{\text{TT,inc}}$	-		-0.0509	-6.47	-0.0485	-10.91	-0.0498	-12.56	-0.1150	-11.31	-
$\beta_{\text{TT,dec}}$	-		0.0452	6.50	0.0451	5.55	0.0358	3.52	0.0321	2.20	-
$\beta_{\text{L-FARE}}$	-5.9654	-28.61	-		-		-		-2.2608	-6.35	-
$\beta_{\text{L-FARE,inc}}$	-		-7.2110	-12.47	-6.7566	-27.18	-6.5516	-28.72	-22.8580	-16.99	-
$\beta_{\text{L-FARE,dec}}$	-		5.2163	13.66	3.5876	8.44	1.8877	3.13	2.5760	3.46	-
$\beta_{\text{rate of delays}}$	-0.1689	-6.79	-0.1679	-6.74	-0.1721	-6.89	-0.1726	-6.91	-0.2703	-9.35	-
$\beta_{\text{expected delay}}$	-0.0989	-6.42	-0.0940	-6.06	-0.0972	-6.29	-0.0976	-6.32	-0.0909	-5.23	-
$\beta_{\text{rate of having to stand}}$	-0.2230	-10.73	-0.2203	-10.57	-0.2242	-10.73	-0.2257	-10.79	-0.2857	-12.04	-
$\beta_{\text{charged SMS}}$	-0.0753	-1.14	-0.0611	-0.93	-0.0694	-1.05	-0.0790	-1.20	-0.0700	-0.94	-
$\beta_{\text{free SMS}}$	0.3315	6.67	0.3364	6.77	0.3399	6.82	0.3247	6.51	0.4405	7.56	-
γ_2	-		-		-		-		-0.5156	-2.39	-
γ_3	-		-		-		-		-0.6920	-2.51	-
γ_4	-		-		-		-		-1.0424	-3.30	-
$\pi_{\text{symm.}}$	100.00%		-		-		-		40.81%	11.43	4.43
$\pi_{\text{ref base}}$	-		100.00%		-		-		24.37%	6.42	-0.17
$\pi_{\text{ref acc}}$	-		-		100.00%		-		20.43%	4.08	-0.91
$\pi_{\text{ref ideal}}$	-		-		-		100.00%		14.39%	3.42	-2.52

already justifies the interest in looking at departures from the typical approach of using current values as the reference points. It should also be noted that the *ideal* reference point has the most extreme values in the data, and as such it is arguably not surprising that the use of that reference point in the model gives the best performance, possibly due to providing greater flexibility.

Moving from a hypothesised population-wide use of a single reference point to a LC model which probabilistically accommodates different reference points offers very substantial gains in model fit. We note an improvement over the MNL model by 295.93 units for 5 additional parameters, and improvements over the three reference dependent models by 293.10, 275.73 and 268.72 units respectively, each at the cost of 3 additional parameters. The size of these gains is very significant when compared to moving from symmetrical MNL to asymmetrical MNL with a common reference point. This suggests that it is important to allow for heterogeneity in reference points across respondents, although part of the gains can be explained by the use of a panel specification which recognises that while the reference points vary across respondents, they stay constant across choices for the same respondent. The model shows very high asymmetry in the three asymmetric classes, especially for the log-fare coefficient. As was the case in the three base models, losses are valued more negatively than gains are valued positively. While some of the remaining coefficients retain scales similar to the four other models, an increased sensitivity is noted for rate of delays, the rate of having to stand, and the provision of a free delay information service. Finally, while for the base models, the best fit was obtained with the *ideal* values for the reference point, followed by the *acceptable* and *current* values, the opposite ordering applies to the class allocation probabilities, while overall, a bigger combined weight is given to the three asymmetric classes than to the base class (59.19% *vs* 40.81%). Here, it should be noted that none of the weights of the three asymmetry classes is statistically different from $1/4$.

Table 3 shows willingness-to-pay (WTP) measures for improvements in services, as well as the cost reductions required to accept a lower quality of service (i.e. willingness-to-accept, WTA). All measures are computed for a journey costing $\pounds 3$. For the symmetrical model, the two types of measures are clearly equivalent to one another. For the three simple asymmetrical models, we compute WTP measures and WTA measures separately, on the basis of the appropriate marginal coefficients. Finally, for the LC model, we compute both types of measures on the basis of the symmetrical as well as asymmetrical coefficients,

Table 3: Analysis of results for second case study

Willingness-to-pay for improved quality of service (at base cost of $£3$)

	Symmetrical		Ref *base*		Ref *acceptable*		Ref *ideal*		LC	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
travel time reduction ($£$/hr)	1.41	13.62	1.13	6.08	1.20	5.60	0.98	3.55	1.69	5.43
one fewer train delayed out of 10 ($£$)	0.08	6.89	0.07	6.05	0.08	6.96	0.08	7.00	0.17	5.66
expected delay reduction ($£$/hr)	2.99	6.41	2.35	5.26	2.59	6.23	2.68	6.29	3.38	4.23
standing in one fewer train out of 10 ($£$)	0.11	10.80	0.09	8.19	0.10	10.70	0.10	10.87	0.18	6.14
free delay information system ($£$)	0.17	6.62	0.14	6.13	0.15	6.75	0.15	6.45	0.27	5.19

Cost reductions required to accept lower quality of service (at base cost of $£3$)

	Symmetrical		Ref *base*		Ref *acceptable*		Ref *ideal*		LC	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
travel time increase ($£$/hr)	1.41	13.62	1.76	5.93	2.43	6.99	4.74	3.09	6.30	4.41
one more train delayed out of 10 ($£$)	0.08	6.89	0.10	6.35	0.14	5.51	0.27	2.89	0.33	5.17
expected delay increase ($£$/hr)	2.99	6.41	3.24	5.92	4.87	5.20	9.30	2.85	6.72	3.94
standing in one more train out of 10 ($£$)	0.11	10.80	0.13	9.07	0.19	6.92	0.36	3.04	0.35	5.47
no delay information system ($£$)	0.17	6.62	0.19	6.09	0.28	5.34	0.52	2.86	0.54	4.72

and use the class allocation weights to produce a weighted average[1].

We observe that, where separate measures are applicable, the cost reductions required to accept a lower quality of service are higher than the WTP for improved service. This is in line with the strong asymmetry in the fare sensitivity. For the WTP measures, the results remain roughly comparable across the three asymmetrical models, but are lower than in the symmetrical model. For the WTA measures, we get higher values in the asymmetrical models, especially the model making use of the *ideal* values as reference points. In all four base models, we observe low measures for the WTP for travel time reductions, where these are however in line with the low average journey cost in this dataset. What is somewhat more surprising is the low valuation for changes in the rate of delay and the rate of standing. Here, and also for the WTP for travel time reductions, higher and arguably more realistic values are obtained by the LC model. The findings concerning the WTP measures could indicate that commuters carry out trade-offs in a consistent manner across different reference points when dealing with improved trip conditions in return for a higher fare. However, an analysis of the WTA measures reveals great variations depending on which reference point is used in the models. A further notable fact is that the LC model

[1] As an example, the willingness to pay for travel time reductions is obtained as $\left(\pi_{\text{symm.}} \frac{\beta_{\text{TT}}}{\beta_{\text{L-FARE}}} + (\pi_{\text{ref base}} + \pi_{\text{ref acc}} + \pi_{\text{ref ideal}}) \frac{\beta_{\text{TT,dec}}}{\beta_{\text{L-FARE,inc}}}\right) \cdot fare$, where we use $fare = £3$.

points towards a possibly more realistic (smaller) ratio between WTA and WTP than the model based on the *ideal* reference point would imply.

3.3 Case study III: RUM & EBA

3.3.1 Behavioural process under investigation

Elimination By Aspects (EBA) is a paradigm originally proposed by Tversky (1972a,b). It represents choice as a process of eliminating alternatives successively, on the basis of their failure to possess certain attributes (or fulfill certain criteria), referred to as *aspects*, until a single alternative remains. The key driver is the order in which the attributes are considered. The ordering used by a given respondent is unobserved, and the model thus selects attributes randomly, with probabilities proportional to weights, the most important attributes having larger weights, thus giving the process its random character.

In particular, in the context of the example presented in this section, we have five different aspects, with weights $w_1, \ldots, w_5$. With five aspects, we obtain 120 different orderings of attributes, where, as an example, the probability of the first ordering $\langle 1, 2, 3, 4, 5 \rangle$ is given by

$$p_{1,2,3,4,5} = \frac{w_1}{\sum_{j=1}^{5} w_j} \frac{w_2}{\sum_{j=2}^{5} w_j} \frac{w_3}{\sum_{j=3}^{5} w_j} \frac{w_4}{\sum_{j=4}^{5} w_j}, \tag{8}$$

where $w_j > 0,\ j = 1, \ldots, 5$.

In any given choice scenario t for respondent n, we first remove any alternative that does not possess aspect 1. If more than one alternative remains, we move on to aspect 2, and so on, until just a single alternative remains. The probability $P_{nt}\left(1, 2, 3, 4, 5\right)$ of the actual observed choice under this given rule is equal to 1 if the remaining alternative is equal to the chosen alternative, and 0 otherwise (if $K > 1$ alternatives remain, their probabilities are $\frac{1}{K}$ each). The probability of the actual sequence of choices for respondent n under a given rule is equal to $P_n\left(1, 2, 3, 4, 5\right) = \prod_{t=1}^{T} P_{nt}\left(1, 2, 3, 4, 5\right)$, and the unconditional probability is then given by a weighted average across the different possible orderings:

$$P_n = \sum_{a=1}^{5} \sum_{b \neq a} \sum_{c \neq a,b} \sum_{d \neq a,b,c} \sum_{e \neq a,b,c,d} p_{a,b,c,d,e} \prod_{t=1}^{T} P_{nt}\left(a, b, c, d, e\right). \tag{9}$$

The only parameters to be estimated for this model are the different weight parameters, with an appropriate normalisation, e.g. setting one weight parameter to a value of 1. The location of the product across tasks inside the weighted summation means that the EBA model accommodates the panel structure of the data.

EBA represents a process fundamentally different from RUM. Nevertheless, Batley and Daly (2003) show that, by appropriate selection of the weights, hierarchical EBA models can be made exactly equivalent to GEV models of the tree form, in the context of models with dummy coefficients only (i.e. with no continuous attributes). Whatever form of EBA and RUM models are compared with each other, it is clear that the coefficients of the RUM model and the weights of the EBA model are not directly related. In the simple tree example given by Batley and Daly (2003), RUM coefficients are equal to logarithms of EBA weights, but this cannot be extended to more general model forms since no precise equivalence exists. Weight ratios, or even ratios of logs of EBA weights have no interpretation as values.

3.3.2 Data & model specification

For this case study, we make use of data from a survey looking at rail travel behaviour, collected through an online panel in the United Kingdom in early 2010. In particular, we rely on a sample of $7,968$ observations collected from 996 respondents, each faced with 8 scenarios involving a choice between three alternatives, where the attributes were pivoted around those of a reported trip (but without including a reference alternative). The alternatives were described on the basis of travel time, fare, the guarantee of a reserved seat, the provision of free wifi, and whether the ticket offered flexibility (in terms of rescheduling). The last three were described in terms of presence/absence, making them ideally suitable for the present analysis in an EBA framework.

Two different RUM specifications were used as the compensatory model. Firstly, a simple MNL structure was used, using a logarithmic transform for fare, with alternative-specific constants for the first two alternatives. Secondly, we used a MMNL model, with random taste heterogeneity in the travel time, seat, wifi and flexibility coefficients, using a Weibull distribution (with estimated parameters b and c, where $\beta = -b\left(-\ln U\right)^{\frac{1}{c}}$, where U is a uniform draw, with $b \geq 0$ and $c > 0$). No significant additional random heterogeneity was found for the cost coefficient after making use of the log transform.

While the final three attributes are ideally suited for use in an EBA framework, given their presence/absence nature, this is not the case for the travel time and fare attributes, where a transformation is required to determine whether an alternative is eliminated or not when that specific attribute is used as a determinant. The practical use of EBA models with mixed attributes in the context of different data sets remains an intriguing field of exploration for future applications; here, we made use of four different straightforward specifications:

EBA$_1$ eliminates the worst (for the considered attribute) of any remaining alternatives at a given stage;

EBA$_2$ eliminates all but the best (for the considered attribute), equating to a dominance based approach;

EBA$_3$ eliminates all options that are 10 minutes slower than the reference trip, or $£0.50$ more expensive when using fare (depending on which attribute is used[2]); and

EBA$_4$ eliminates an alternative if the time or fare is worse than that for the reference trip (again depending on which attribute is used).

Independently of which of the four EBA approaches is used, it is not possible to estimate a stand-alone EBA model on this dataset, as there are choices that cannot be explained by such an approach, leading to a notionally minus infinity contribution to the log-likelihood function. Rather, the EBA model is only ever used in conjunction with an RUM model, using a two class specification. In the EBA part, weights are estimated for the first four attributes, with the weight for flexibility being normalised to a value of 1.

3.3.3 Estimation results

The estimation results for the fixed coefficient models are summarised in Table 4. In addition to the alternative specific constants and the five marginal utility coefficients, we report the constant used in the class allocation model (γ_2) and the weights used in the EBA classes. In addition, the first part of Table 6 reports the WTP measures calculated from the coefficients estimated for the compensatory class only in the different models, with no obvious interpretion for the parameters of the EBA class.

[2] As pointed out by an anonymous referee, an alternative way of specifying EBA$_3$ would have been to work with percentage differences rather than absolute differences.

Table 4: Estimation results for third case study: models without additional random taste heterogeneity

	MNL		MNL & EBA_1			MNL & EBA_2			MNL & EBA_3			MNL & EBA_4		
Observations	7,968		7,968			7,968			7,968			7,968		
Log-likelihood	-6,288.87		-6,166.07			-5,767.27			-6,170.22			-6,122.78		
par.	7		12			12			12			12		
adj. ρ^2	0.2808		0.2942			0.3398			0.2938			0.2992		
				t-rat.			t-rat.			t-rat.			t-rat.	
	est.	t-rat.	est.	*vs* 0	*vs* $1/2$	est.	*vs* 0	*vs* $1/2$	est.	*vs* 0	*vs* $1/2$	est.	*vs* 0	*vs* $1/2$
δ_1	0.1075	3.16	0.0905	2.54	-	0.0834	2.11	-	0.0945	2.65	-	0.1019	2.83	-
δ_2	0.0017	0.05	-0.0146	-0.41	-	-0.0536	-1.33	-	-0.0082	-0.23	-	-0.0184	-0.51	-
β_{TT}	-0.0207	-29.46	-0.0216	-29.32	-	-0.0168	-20.33	-	-0.0208	-27.50	-	-0.0203	-26.71	-
$\beta_{\text{L-FARE}}$	-5.8199	-33.33	-6.0428	-32.66	-	-4.9725	-23.84	-	-5.9180	-31.74	-	-6.0089	-31.57	-
$\beta_{\text{reserved seat}}$	0.9864	27.22	0.9222	23.65	-	0.9906	22.51	-	0.9800	24.89	-	0.9983	25.25	-
β_{wifi}	0.2679	6.82	0.2291	5.53	-	0.1682	3.66	-	0.2038	4.91	-	0.2118	5.05	-
$\beta_{\text{flexible ticket}}$	0.2452	4.09	0.1876	2.94	-	0.0460	0.65	-	0.1659	2.60	-	0.1625	2.53	-
γ_2	-	-	-2.5495	-16.99	-	-0.9091	-11.40	-	-2.5770	-15.47	-	-2.3742	-16.17	-
w_{TT}	-	-	0.5321	2.97	-	4.9818	5.28	-	1.6930	2.75	-	3.2625	3.03	-
$w_{\text{L-FARE}}$	-	-	0.9722	3.45	-	5.6043	5.31	-	1.3631	2.84	-	1.3772	2.53	-
$w_{\text{reserved seat}}$	-	-	5.2627	3.27	-	4.9675	4.85	-	5.7797	2.80	-	5.6305	2.91	-
w_{wifi}	-	-	0.4932	2.88	-	0.6231	4.14	-	1.2130	2.65	-	1.7628	2.98	-
$w_{\text{flexible ticket}}$	-	-	1	-	-	1	-	-	1	-	-	1	-	-
π_{MNL}	100.00%	-	92.75%	91.98	42.40	71.28%	43.68	13.04	92.94%	84.96	39.25	91.48%	79.96	36.26
π_{EBA}	0.00%	-	7.25%	7.19	-42.40	28.72%	17.60	-13.04	7.06%	6.46	-39.25	8.52%	7.44	-36.26

In the base model, we observe negative effects of increases in time (β_{TT}) and log-fare ($\beta_{\text{L-FARE}}$), with positive utilities for the provision of a reserved seat, free wifi, or ticket flexibility. There is also some evidence of left-to-right reading effects.

We next look at the model combining a MNL structure with the first of the EBA specifications. We observe an improvement in model fit by 122.80 units, which is highly significant at the cost of just five additional parameters. We note that the MNL part of the model still accounts for over ninety percent of the class probabilities, but nevertheless observe quite substantial reductions in WTP for the three quality of service attributes in that class. In the EBA model, the biggest weight by far is obtained by the seat reservation attribute.

The second LC model obtains even larger improvements in model fit over the base model (521.60 units for 5 parameters), and a much greater share (almost thirty percent) for the EBA class. We also note that the coefficient for ticket flexibility is no longer significant in the trading model, and that while the VTT and the WTP for wifi is reduced, that for seat reservation increases by over 17% in comparison with the base model. Relatively equal weights are obtained for travel time, fare, and seat reservation in the EBA component, with lower weight for wifi and ticket flexibility.

Our third LC model obtains the smallest (albeit still significant) of the four improvements over the base model (118.65 units for 5 parameters), and gives a share of seven percent to the EBA component. The largest weight in the EBA component is once again given to seat reservation, with relatively equal weights for travel time, fare and wifi, followed by ticket flexibility. While the VTT and the WTP for seat reservation are almost identical to the overall MNL results, we see a major reduction in the WTP for the provision of wifi and for ticket flexibility.

Our final LC model once again significantly outperforms the base model (166.09 units for 5 parameters) and offers the second best performance of the four LC structures, along with the second highest share for the EBA component of the model. The highest weights in the EBA component are obtained for seat reservation and travel time, followed by wifi, fare, and ticket flexibility. With this model, the biggest impact on the MNL component is once again the decrease in the WTP for wifi and ticket flexibility.

For the MNL & EBA specifications, the best performance is thus obtained by a model in which a dominance rule is used for time and fare in the EBA model. As a next step, we look at the models

Table 5: Estimation results for third case study: models with additional random taste heterogeneity

	MMNL	MMNL & EBA$_2$
Observations	7,968	7,968
Log-likelihood	-5,453.85	-5,357.85
par.	11	16
adj. ρ^2	0.3757	0.3861

	MMNL		MMNL & EBA$_2$		
				asy. t-rat.	
	est.	t-rat.	est.	*vs* 0	*vs* $1/4$
δ_1	0.1944	4.19	0.2104	4.33	-
δ_2	-0.0285	-0.62	-0.0115	-0.24	-
b_{TT}	0.0507	20.22	0.0475	17.33	-
$\beta_{\text{L-FARE}}$	-9.5498	-34.72	-8.0603	-28.52	-
$b_{\text{reserved seat}}$	1.6362	15.45	1.6863	14.48	-
b_{wifi}	0.2938	3.94	0.2934	2.45	-
$b_{\text{flexible ticket}}$	0.4481	5.02	0.3929	4.34	-
c_{TT}	1.0513	18.98	1.1399	16.36	-
$c_{\text{reserved seat}}$	0.8014	7.40	0.8837	10.01	-
c_{wifi}	0.4880	8.00	0.4935	3.83	-
$c_{\text{flexible ticket}}$	0.5350	12.09	0.5318	10.63	-
γ_2	-		-1.9980	-11.22	-
w_{TT}	-		9.0729	1.80	-
$w_{\text{L-FARE}}$	-		36.9050	1.79	-
$w_{\text{reserved seat}}$	-		3.3226	1.06	-
w_{wifi}	-		0.2774	1.03	-
$w_{\text{flexible ticket}}$	-		1	-	-
π_{MMNL}	100.00%	-	88.06%	46.60	20.14
π_{EBA}	0.00%	-	11.94%	6.32	-20.14

incorporating additional random taste heterogeneity, with results summarised in Table 5, and WTP measures for the compensatory part of the model shown in the second half of Table 6. A first observation to be made is that with the exception of the model using a dominance rule for time and fare, the remaining MMNL & EBA models collapsed back to the MMNL model. This would suggest that any heterogeneity in behaviour that would be captured by mixing the two decision rules can be adequately modelled in the MMNL model alone. Even for the MMNL & EBA$_2$ model, we see a drop in the EBA share from 28.72% to 11.94%. This thus suggests that almost two thirds of the heterogeneity captured by making use of a MNL & EBA mixture can be captured by the MMNL component alone in the MMNL & EBA mixture.

The MMNL & EBA$_2$ model gives us a highly significant improvement in LL over the simple MMNL

Table 6: WTP measures for trading component of model, at base fare of £40

Models without additional random taste heterogeneity

	MNL		MNL & EBA$_1$			MNL & EBA$_2$			MNL & EBA$_3$			MNL & EBA$_4$		
	est.	t-rat.	est.	t-rat.	vs MNL	est.	t-rat.	vs MNL	est.	t-rat.	vs MNL	est.	t-rat.	vs MNL
time (£/hr)	8.55	37.78	8.60	37.56	0.56%	8.13	25.35	-4.90%	8.42	34.04	-1.46%	8.11	32.68	-5.16%
reserved seat (£)	6.78	18.52	6.10	16.86	-9.96%	7.97	14.42	17.54%	6.62	17.40	-2.30%	6.65	17.48	-1.97%
wifi (£)	1.84	7.55	1.52	6.01	-17.63%	1.35	3.94	-26.51%	1.38	5.30	-25.20%	1.41	5.47	-23.42%
flexible ticket (£)	1.69	4.44	1.24	3.12	-26.34%	0.37	0.66	-78.04%	1.12	2.74	-33.48%	1.08	2.67	-35.82%

Models with additional random taste heterogeneity

	MMNL						MMNL & EBA$_2$						change		
	mean	t-rat.	std.dev.	t-rat.	c.v.	t-rat.	mean	t-rat.	std.dev.	t-rat.	c.v.	t-rat.	mean	std.dev.	c.v.
time (£/hr)	12.50	11.08	11.89	32.16	0.95	13.99	13.50	7.95	11.87	28.21	0.88	9.25	8.05%	-0.15%	-7.59%
reserved seat (£)	7.76	6.72	9.76	2.85	1.26	4.86	8.90	7.08	10.09	4.17	1.13	9.87	14.70%	3.40%	-9.85%
wifi (£)	2.58	6.22	5.96	3.87	2.31	2.78	2.98	4.73	6.80	1.72	2.28	1.35	15.79%	13.94%	-1.60%
flexible ticket (£)	3.34	7.99	6.80	8.08	2.04	5.15	3.50	6.87	7.20	7.66	2.05	4.60	4.92%	5.78%	0.82%

model by 96 units for 5 additional parameters. Similarly, the MMNL and MMNL & EBA$_2$ models outperform their MNL and MNL & EBA$_2$ counterparts, with improvements in LL by 835.02 and 409.42 units respectively, for four additional parameters. Moreover, the MMNL model outperforms all of the MNL & EBA models, as it has a better likelihood with fewer degrees of freedom. Two interesting differences arise between the MMNL & EBA$_2$ model and its MNL & EBA$_2$ counterpart. Firstly, while the ticket flexibility coefficient in the MNL & EBA$_2$ model was not statistically significant, both the mean and standard deviation in the MMNL & EBA$_2$ model are statistically significant. Secondly, while the MNL & EBA$_2$ showed relatively similar weights for travel time, fare, and seat reservation in the EBA component, this is no longer the case in the MMNL & EBA$_2$ model, where fare dominates, followed by travel time.

Both models show significant heterogeneity in the four randomly distributed coefficients. However, some interesting differences arise, as highlighted in the WTP findings in Table 6. Here, we can see that when incorporating the mixing between MMNL and EBA$_2$, the degree of heterogeneity in the compensatory part, expressed as the coefficient of variation, is reduced for travel time and seat reservation, with a smaller reduction for wifi provision, and a very small increase for ticket flexibility. For wifi, the standard deviation of the WTP measure has a high associated standard error. The mean values for all four WTP measures are increased in comparison with the simple MMNL model, by between five and sixteen percent. This points to the ability of the EBA component of the MMNL & EBA_2 model to absorb a portion of the heterogeneity previously assigned to random taste variance in the simple MMNL model. What is more, substantial increases in WTP are observed when comparing these results to their taste homogeneity counterparts (MNL and MNL & EBA_2).

3.4 Case study IV: RUM & random regret

3.4.1 Behavioural process under investigation

Regret is a negative emotion experienced when we imagine that a present situation would have been better had we made a different decision (cf. Simonson, 1992). Early intuitions by economists argue that people base decisions on a 'minimax regret' rule (cf. Savage, 1951), which holds that the maximum of possible regret is calculated for each option, and the option that minimises potential regret is chosen. A formal theory of regret was developed independently by Bell (1982) and Loomes and Sugden (1982). The fundamental assumption in regret theory is that final utility depends not merely on the realised outcome but also on what could have been obtained by choosing a different course of action.

Chorus et al. (2008) define a Random Regret Minimisation (RRM) model where regret is equal to the largest among the binary regrets based on pairwise comparisons of the considered (i) and remaining alternatives ($i \neq j$). While we concentrate on the binary comparisons for the sake of simplicity, more recent developments of the RRM framework (cf. Chorus, 2010) have looked at a calculation of regret with regard to all available alternatives; such an extension within a RUM-RRM mixture is straightforward.

What is estimated is really the weights that denote the performance of each attribute k in the binary regret computation.

$$R_i = max_{i \neq j} \left\{ \sum_{k=1..K} max \left\{ 0, \beta_k (x_{jk} - x_{ik}) \right\} \right\} \tag{10}$$

Regret is computed only considering the best forgone alternative. The choice probability for alternative i with iid type 1 extreme value errors is written $P_i = \frac{exp(-R_i)}{\sum exp(-R_j)}$. There are only a few empirical applications of the RRM framework, with Hensher et al. (Forthcoming) being one example. Concerning interpretations of attribute coefficients, whereas a RUM based analysis derives the sensitivity to attributes, RRM estimates the potential contribution to regret feelings of each attribute. For this reason, the comparison across model specifications is not straightforward.

3.4.2 Data & model specification

For the present analysis, we once again use the data from the commuter survey described in the second case study. The presence of strong reference dependence in this data for at least two attributes (cf. Stathopoulos and Hess, 2010) makes it well suited for an application of a RRM framework which similarly entails comparisons of alternatives on individual attributes. For the purpose of being able to use this dataset with a RRM model, the information service attribute was dropped from the model specification. This only had a very small impact on the remaining model parameters.

We once again looked first at models without additional random taste heterogeneity. Here, alongside the MNL model, we estimated a simple RRM model, and two LC models. In the first LC model (MNL & RRM), only the two alternative specific constants were specified to be class specific (i.e. using separate constants for MNL and RRM), while the coefficients in the MNL and RRM classes were specified to be equal to one another. In the second of the LC models (MNL & RRM_{sep}), all parameters were class specific. Allowing for distinct attributes by class gives recognition to the fact that coefficients have an entirely different interpretation across choice paradigms. We next estimated models allowing for additional random taste heterogeneity. Here, significant variations were only observed in the stand-alone RUM model (MMNL) and in the RUM component of a combined model (MMNL & RRM), but not in the stand-alone RRM model or the RRM component of the combined model. The MMNL component of the joint model made use of a Weibull distribution for the four non-fare coefficients, with a fixed coefficient for fare but maintaining the log transform.

3.4.3 Estimation results

The results for the four models without additional taste heterogeneity are summarised in Table 7. We first note the better performance for the MNL model compared to the RRM model, suggesting that overall, RUM fits this dataset better than RRM. Additionally, we can observe a somewhat strong correspondence

Table 7: Estimation results for fourth case study: models without additional random taste heterogeneity

	MNL	RRM	MNL & RRM	MNL & RRM$_{\text{sep}}$
Observations	3,680	3,680	3,680	3,680
Log-likelihood	-3,401.68	-3,476.82	-3,288.28	-3,175.76
par.	7	7	10	15
adj. ρ^2	0.1569	0.1383	0.1842	0.2108

	MNL		RRM		MNL & RRM			MNL & RRM$_{\text{sep}}$		
						t-rat.			t-rat.	
	est.	t-rat.	est.	t-rat.	est.	*vs* 0	*vs* $1/2$	est.	*vs* 0	*vs* $1/2$
$\delta_{1,\text{RUM}}$	0.3996	8.62	-		-0.0327	-0.42	-	0.6146	9.77	-
$\delta_{2,\text{RUM}}$	0.1730	3.58	-		0.1563	2.88	-	0.2310	3.42	-
$\beta_{\text{TT,RUM}}$	-0.0474	-13.47	-		-0.0481	-13.31	-	-0.0551	-10.73	-
$\beta_{\text{L-FARE,RUM}}$	-5.7719	-28.34	-		-5.8875	-27.95	-	-3.1119	-10.70	-
$\beta_{\text{rate of delays,RUM}}$	-0.1628	-6.63	-		-0.1402	-5.45	-	-0.3182	-8.29	-
$\beta_{\text{expected delay,RUM}}$	-0.0994	-6.49	-		-0.1016	-6.11	-	-0.1106	-4.51	-
$\beta_{\text{rate of having to stand,RUM}}$	-0.2221	-10.78	-		-0.1969	-9.34	-	-0.3137	-10.39	-
$\delta_{1,\text{RRM}}$	-		-0.1424	-2.98	-1.9022	-10.50	-	-0.1335	-1.18	-
$\delta_{2,\text{RRM}}$	-		-0.1744	-3.68	-0.4001	-2.02	-	-0.2444	-2.09	-
$\beta_{\text{TT,RRM}}$	-		-0.0400	-10.44	-0.0481*	-13.31*	-	-0.0654	-8.66	-
$\beta_{\text{L-FARE,RRM}}$	-		-5.4082	-25.11	-5.8875*	-27.95*	-	-14.3150	-12.62	-
$\beta_{\text{rate of delays,RRM}}$	-		-0.1261	-3.31	-0.1402*	-5.45*	-	0.0000	0.00	-
$\beta_{\text{expected delay,RRM}}$	-		-0.0816	-3.62	-0.1016*	-6.11*	-	-0.0864	-4.72	-
$\beta_{\text{rate of having to stand,RRM}}$	-		-0.1708	-8.89	-0.1969*	-9.34*	-	-0.1325	-3.00	-
γ_2	-		-		-1.3792	-6.29	-	-0.4491	-3.04	-
π_{RUM}	100%		0%		79.89%	22.68	8.48	61.04%	17.80	3.22
π_{RRM}	0%		100%		20.11%	5.71	-8.48	38.96%	11.36	-3.22

* RRM parameters constrained to estimates from MNL class

in the relative coefficient values in the two base models. Moving to the model that accommodates the two types of decision making but with equal coefficient values, we can observe significant gains in model fit over both base models (113.4 units and 188.54 units respectively, each at the cost of 3 additional parameters), with a $80-20$ split in the weight for the two types of decision making. Moving finally to the model that allows for class specific coefficients, we observe a further improvement in model fit by 112.52 units at the cost of 7 additional parameters. We also observe an increase in the weight for the RRM class to almost forty percent. Most interestingly, while the relative weight of the cost component is reduced in the RUM class, the cost attribute now dominates in the RRM class, suggesting that this class captures respondents with heightened cost sensitivity alongside those respondents for whom a RRM framework is more suitable for explaining their choices. Such findings imply that fare may be particularly relevant in guiding choices away from situations where non-chosen alternatives offer lower fares. The findings also confirm the results from the heterogeneous reference-point model carried out on the same sample where deterioration from the reference fare yields significant gain/loss asymmetry for a large portion of respondents. Further tests of the RRM paradigm are needed to assess the empirical links to other choice processes, including compensatory RUM.

The results for the model incorporating additional heterogeneity in the RUM component are shown in Table 8. Here, only two models were used, a simple MMNL model, and a MMNL & RRM combination, making use of model-specific parameters throughout (labelled MMNL & RRM$_{\text{sep}}$). We observe a highly significant improvement in model fit by 180.95 units when moving from MMNL to MMNL & RRM$_{\text{sep}}$, at the cost of 8 additional parameters. Similarly, we observe improvements over their MNL and MNL & RRM counterparts by 162.44 and 117.47 units respectively, both at the cost of 4 additional parameters.

The MMNL model outperforms the MNL & RRM model but is outperformed by the MNL & RRM$_{\text{sep}}$ model; the likelihoods of these models cannot be compared directly, as they are not generalisations of each other, but calculation of the BIC index gives values of $-3,233.74$, $-3,284.28$ and $-3,168.26$ respectively, maintaining the large differences indicated by the simple log-likelihood values. The share for the RRM component is reduced somewhat in comparison with the MNL & RRM model, dropping from 38.96% to 32.43%, but remains large. Another interesting observation can however be made. In the MNL & RRM$_{\text{sep}}$ model, $\beta_{\text{L-FARE,RRM}}$ dominated, suggesting that this class captured respondents

Table 8: Estimation results for fourth case study: models with additional random taste heterogeneity

	MMNL		MMNL & RRM_{sep}		
Observations	3,680		3,680		
Log-likelihood	-3,239.24		-3,058.29		
par.	11		19		
adj. ρ^2	0.1961		0.2388		
				t-rat.	
	est.	t-rat.	est.	*vs* 0	*vs* $1/2$
$\delta_{1,\text{RUM}}$	0.5502	10.08	0.5210	5.35	-
$\delta_{2,\text{RUM}}$	0.2311	4.07	0.0193	0.19	-
$b_{\text{TT,RUM}}$	0.0626	7.54	0.1315	7.37	-
$\beta_{\text{L-FARE,RUM}}$	-7.1929	-27.55	-14.6240	-16.25	-
$b_{\text{rate of delays,RUM}}$	0.0598	1.61	0.0661	1.22	-
$b_{\text{expected delay,RUM}}$	0.1355	3.49	0.2302	3.48	-
$b_{\text{rate of having to stand,RUM}}$	0.2419	4.66	0.4695	3.82	-
$c_{\text{TT,RUM}}$	0.8852	7.39	0.9378	7.07	-
$c_{\text{rate of delays,RUM}}$	0.3262	5.64	0.2673	5.57	-
$c_{\text{expected delay,RUM}}$	0.5920	5.98	0.6014	7.47	-
$c_{\text{rate of having to stand,RUM}}$	0.4934	6.89	0.4726	8.34	-
$\delta_{1,\text{RRM}}$	-		-0.6879	-6.16	-
$\delta_{2,\text{RRM}}$	-		-0.5256	-5.47	-
$\beta_{\text{TT,RRM}}$	-		-0.0292	-5.76	-
$\beta_{\text{L-FARE,RRM}}$	-		-1.3569	-25.11	-
$\beta_{\text{rate of delays,RRM}}$	-		-0.2634	-10.62	-
$\beta_{\text{expected delay,RRM}}$	-		-0.0578	-2.34	-
$\beta_{\text{rate of having to stand,RRM}}$	-		-0.1131	-5.24	-
γ_2	-		-0.7342	-5.22	-
π_{RUM}	100%		67.57%	26.80	6.66
π_{RRM}	0%		32.43%	13.47	-6.66

who were strongly fare sensitive. However, in the MMNL & RRM_{sep} model, a far more balanced picture emerges, and $\beta_{\text{rate of delays,RRM}}$ is now also significant, while it was essentially zero in the MNL & RRM_{sep} model. The increase of the fare coefficient in the RUM section of the MMNL & RRM_{sep} model could imply that the taste homogeneity counterpart (MNL & RRM_{sep}) may have assigned some of the unmodelled taste heterogeneity to the regret minimisation decision rule. This observation offers further evidence as to the complex distinction between taste and decision paradigm heterogeneity.

Table 9 gives WTP measures for the compensatory model components at a journey cost of $\pounds 3$. In the MNL case, the incorporation of a RRM class leads to major increases in the WTP measures, while,

Table 9: Willingness to pay measures for compensatory model components for fourth case study (at a travel cost of $£3$)

	MNL		MNL & RR$_{\text{sep}}$	
	est.	t-rat	est.	t-rat
travel time reduction (£/hr)	1.48	13.90	3.18	8.06
one fewer train delayed out of 10 (£)	0.08	6.68	0.31	6.32
expected delay reduction (£/hr)	3.10	6.48	6.40	4.27
standing in one fewer train out of 10 (£)	0.12	10.86	0.30	7.69

	MMNL					
	mean	t-rat	std.dev.	t-rat	c.v.	t-rat
travel time reduction (£/hr)	1.66	7.66	1.88	19.48	1.13	10.98
one fewer train delayed out of 10 (£)	0.16	1.98	0.74	1.41	4.55	0.91
expected delay reduction (£/hr)	5.19	5.21	9.29	6.03	1.79	4.65
standing in one fewer train out of 10 (£)	0.21	6.19	0.47	2.48	2.28	2.58

	MMNL & RRM$_{\text{sep}}$					
	mean	t-rat	std.dev.	t-rat	c.v.	t-rat
travel time reduction (£/hr)	1.67	5.53	1.78	12.66	1.07	8.46
one fewer train delayed out of 10 (£)	0.22	1.42	1.56	0.47	7.03	0.39
expected delay reduction (£/hr)	4.25	5.23	7.45	6.47	1.75	5.03
standing in one fewer train out of 10 (£)	0.21	4.99	0.52	2.08	2.42	2.06

in the MMNL context, changes in the mean values are only observed for two of the WTP measures (rate of delays and expected delay), where these changes are far less substantial than was the case for MNL. There is also an increase in the heterogeneity for the WTP for reduced rate of delays. Overall, these findings suggest that while in the MNL case, the RRM class captures those respondents with high cost sensitivity, this is not the case in the MMNL context. Here, the respondents captured by the RRM class may simply be those whose behaviour can be better explained by such a model.

4 Conclusions

This paper has looked at the benefits of allowing the analyst to use a mixture of different behavioural processes to explain the choices observed in a sample population. The approach uses a latent class structure, where the core distinction with the majority of latent class work lies in the use of a different underlying model structure in different classes. The resulting model is highly flexible and potentially able to accommodate very rich spectra of behavioural heterogeneity, including fundamentally different non-RUM decision protocols.

The paper has presented evidence from four separate case studies, each showing significant improvements in model fit when allowing for heterogeneity in the behavioural processes across respondents, while also offering further insights into actual decision making, and in several instances improving the reasonableness of the willingness-to-pay measures. We acknowledge that part of the gains in fit obtained when comparing a simple MNL model to a mixture between two different models (e.g. MNL and EBA) could be a result of the mixture model capturing correlations between choices for the same respondent[3]. However, aside from it not being clear what those correlations are, if they are not to be related to taste heterogeneity, and how they should be captured, a brief analysis on the second case study (results available on request) showed that while the inclusion of respondent-specific error components (distributed identically but independently across alternatives) led to further gains in fit, there was only a very small impact on the results in terms of the mixing of decision rules or indeed the gains resulting from that approach.

As with any treatment of unobserved model components, we can of course not say with certainty whether the processes that our models allow for actually exist in the data, or are present to the degree indicated by our estimates. But the same clearly applies in models making use of a standard approach for accommodating random taste heterogeneity. Given the repeated evidence in the literature of departures from standard choice paradigms in some datasets, it is clearly conceivable that differences in behavioural processes actually arise between individuals within a single dataset. Accounting for a wider range of behavioural heterogeneity in choice modelling, may, as illustrated in our case studies, also lead to important shifts in willingness-to-pay and accept measures. Indeed, the welfare measures typically calculated from choice modelling results may not apply in groups who are not using RUM consistent decision protocols or will be radically different in scope and interpretation.

The paper has also highlighted the possible risk of confounding between 'standard' taste heterogeneity and heterogeneity in decision making paradigms, with potentially substantially different patterns of heterogeneity emerging. Indeed, the share of the non-RUM classes is reduced when allowing for random heterogeneity in the RUM class. Conversely, we however also see reductions in the degree of random

[3]We would like to thank an anonymous referee for helpful comments on this point.

heterogeneity in the RUM class compared to the simple (one class) MMNL models. This leads to the tentative observation that some of the behaviour that is traditionally perceived as taste heterogeneity in applied research may well be explained by alternative choice paradigms that appear to describe behaviour of a sub-set of respondents particularly well. The conclusion seems to be that it is crucial to account for both types of behaviour at the same time, so as to avoid overstating the weight of the non-RUM class, where, for interpretation reasons, explaining as much behaviour as possible by the RUM component is preferable.

Much work remains to be done, including further investigation into the confounding between taste and process heterogeneity. Additionally, other behavioural processes should be considered, as should mixture models incorporating more than two decision rules; this possibly requires more flexible datasets than were available for this study. The role of experimental design and the degree to which it allows identification of different behavioural processes or even influences the use of certain rules in the first place needs to be explored. To gain insight on these points it would be desirable to extend the empirical work to a wider range of designs and datasets, as we have only tested one dataset per paradigm. Furthermore, the analysis of the results in terms of WTP measures has in the present paper focussed solely on compensatory classes; the interpretation of estimates from the non-compensatory classes (i.e. dominance, EBA, and RRM) remains an area for future work. Finally, the role of mixture distributions in the RUM component of any model allowing for random heterogeneity also needs further attention; different choices of distributions are likely to lead to different shares for the RUM component.

A further issue that is to be resolved is how these models could be used to deal with changes in the variables influencing choice, whether to forecast the impact of transport policy or to calculate expected welfare benefit. The specific difficulties that arise are not a function of the underlying latent class structure but apply to the specific paradigms used. Indeed, a general point that applies to most of these paradigms is that there are threshold values and non-linearities. The impact of a given change will depend to a very large extent on how the population is distributed relative to the threshold points, while a further problem is that the forecast or welfare benefit will depend on the order in which changes are made, i.e. there is path dependence. Nevertheless, an understanding that the population may not only have diverse taste but also behave according to different decision rules should contribute towards formulating more sophisticated transport policy.

Acknowledgements

The first author also acknowledges the support of the Leverhulme Trust in the form of a *Leverhulme Early Career Fellowship*. The second author acknowledges support of the Trieste University scholarship *M.U.R. Progetto Giovani Ricercatori*. The authors are grateful to three anonymous referees whose comments on an earlier version led to significant improvements.

References

Abou-Zeid, M., Ben-Akiva, M., 2010. A model of travel happiness and mode switching. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald Publishing, UK.

Araña, J. E., León, C. J., Hanemann, M. W., 2008. Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly. Journal of Health Economics 27 (3), 753–769.

Ariely, D., Loewenstein, G., Prelec, D., 2003. “coherent arbitrariness”: Stable demand curves without stable preferences. Quarterly Journal of Economics 118 (1), 73–105.

Batley, R. P., Daly, A., 2003. On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. Journal of Mathematical Psychology 50 (5), 456–467.

Bell, D., 1982. Regret in decision making under uncertainty. Operations Research 20, 961–81.

Bradley, M. A., Daly, A., 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21 (2), 167–184.

Cairns, J., van der Pol, M., 2004. Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments. Social Science and Medicine 58 (11), 2211–2218.

Chorus, C., 2010. A new model of random regret minimization. European Journal of Transport and Infrastructure Research 10 (2), 181–196.

Chorus, C., Arentze, T., Timmermans, H., 2008. A random regret minimization model of travel choice. Transportation Research Part B 42 (1), 1–18.

De Borger, B., 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.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

Fosgerau, M., 2006. Investigating the distribution of the value of travel time savings. Transportation Research Part B 40 (8), 688–707.

Gilbride, T. J., Allenby, G. M., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. Marketing Science 23 (3), 391–406.

Gopinath, D., 1995. Modeling heterogeneity in discrete choice processes: Application to travel demand. Ph.D. thesis, MIT, Cambridge, MA.

Greene, W. H., Hensher, D. A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681–698.

Hensher, D., Greene, W., Chorus, C., Forthcoming. Random regret minimization or random utility maximization: An exploratory analysis in the context of automobile fuel choice. Journal of Advanced Transportation.

Hensher, D. A., Greene, W. H., 2003. The Mixed Logit Model: The State of Practice. Transportation 30 (2), 133–176.

Hensher, D. A., Greene, W. H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empirical Economics 39 (2), 413–426.

Hess, S., Rose, J. M., 2007. A latent class approach to recognising respondents' information processing strategies in SP studies. paper presented at the Oslo Workshop on Valuation Methods in Transport Planning, Oslo.

Hess, S., Rose, J. M., 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.

Hess, S., Rose, J. M., Polak, J. W., 2010. Non-trading, lexicographic and inconsistent behaviour in sp choice data. Transportation Research Part D 15 (7), 405–417.

Jedidi, K., Kohli, R., 2005. Probabilistic subset-conjunctive models for heterogeneous consumers. Journal of Marketing Research 42 (4), 483–494.

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

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

Kamel, J., Rajeev, K., 2008. Inferring latent class lexicographic rules from choice data. Journal of Mathematical Psychology 52 (4), 241–249.

Killi, M., Nossum, Å., Veisten, K., 2007. Lexicographic answering in travel choice: Insufficient scale extensions and steep indifference curves? European Journal of Transport and Infrastructure Research 7 (1), 39–62.

Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. The Quarterly Journal of Economics 121 (4), 1133–1165.

Lancsar, E., Louviere, J., 2006. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics 15 (8), 797–811.

Loomes, G., Sugden, R., 1982. Regret theory: An alternative theory of rational choice under uncertainty. Economic Journal 92, 805–24.

Luce, R. D., 1978. Lexicographic tradeoff structures. Theory and Decision 9 (2), 187–193, 10.1007/BF00131773.

Masiero, L., Hensher, D. A., 2010. Analyzing loss aversion and diminishing sensitivity in a freight transport stated choice experiment. Transportation Research Part A: Policy and Practice.

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

Revelt, D., Train, K., 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4), 647–657.

Ryan, M., Farrar, S., 1994. A pilot study using conjoint analysis to establish the views of users in the provision of orthodontic services in grampian. Tech. rep., Discussion Paper 07/94, Health Economics Research, Aberdeen: University.

Sælensminde, K., 2002. The impact of choice inconsistencies in stated choice studies. Environmental and Resource Economics 23 (4), 403–420, 10.1023/A:1021358826808.

Sælensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological Economics 59 (3), 331–340, 10.1016/j.ecolecon.2005.11.001.

Savage, L. J., 1951. The theory of statistical decision. Journal of the American Statistical Association 46, 55–67.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Simonson, I., 1992. The influence of anticipating regret and responsibility on purchase decisions. The Journal of Consumer Research 19 (1), 105–118.

Stathopoulos, A., Hess, S., 2010. Modelling multi-attribute reference dependence in a commuting choice experiment: Does one size fit all? In: Proceedings of the 12th WCTR, July 11-15, 2010 – Lisbon, Portugal.

Stutzer, A., 2004. The role of income aspirations in individual happiness. Journal of Economic Behavior and Organization 54, 89–109.

Tversky, A., 1969. Intransitivity of preferences. Psychological Review 76 (1), 31–48.

Tversky, A., 1972a. Choice by elimination. Journal of Mathematical Psychology 9, 341–367.

Tversky, A., 1972b. Elimination by aspects: a theory of choice. Psychological Review 79, 281–299.

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