Utility maximisation and regret minimisation: A mixture of a generalisation

Stephane Hess[a] , Caspar G. Chorus[b]

[a] Institute for Transport Studies, University of Leeds, Leeds, United Kingdom

[b] Transport and Logistics Group, Delft University of Technology, Delft, the Netherlands

Abstract

There is growing interest in decision rule heterogeneity across individual respondents but also across attributes. This paper brings these two issues together by putting forward a latent class approach which not only allows for different decision rules across classes, but also differences in the decision rules used across attributes within a given class. To deal with the issue of the large number of possible combinations this produces, we focus on the specific case of random utility maximisation (RUM) and random regret minimisation (RRM) and put forward the use of a generalised random regret minimisation (G-RRM) model within individual classes. This allows the *optimal* specification in terms of split between RUM and RRM within a given class to be revealed by the data during estimation, rather than needing to be imposed by the analyst. Initial findings on a standard stated choice dataset are promising and show how a rich pattern of taste heterogeneity and decision rule heterogeneity across respondents and attributes can be revealed.

1. Introduction

Recent years have witnessed a rapidly growing number of studies that aim to incorporate decision rules other than the conventional linear-additive random utility maximisation (RUM) rule in discrete choice models of travel behaviour. Motivated by the wish to increase the behavioural realism of travel demand models, as well as their empirical performance, decision rules like symmetric relative advantage maximisation (e.g., Leong & Hensher, in press(a)), reference dependent utility maximisation (e.g., Stathopoulos & Hess, 2012), random regret minimisation (RRM, e.g., Chorus, 2010) and decision tree approaches (e.g., Arentze & Timmermans, 2007) have been proposed and applied in recent travel behaviour studies. Earlier work tended to focus on comparing the properties and empirical performance of models based on these alternative decision rules with those of models based on the conventional linear-additive utility maximisation rule. While such comparisons have generated interesting insights, more recent work has adopted a different perspective, and puts more emphasis on capturing potential heterogeneity in applied decision rules.

Two main strands of this more recent research can be distinguished: first, it is increasingly being acknowledged that it is unrealistic to assume that every *individual* applies the same decision rule; rather it makes more sense to allow for different groups of individuals to use different decision rules. This conceptual idea can be operationalised using mixture models of the type put forward by Hess et al. (2012) – see Hess & Stathopoulos (2013) and Boeri et al. (2014) for other recent examples. These models incorporate different latent classes, each with their own decision rule (and set of taste parameters). Estimation results show large improvements in fit compared to models that assume one and the same decision rule for the entire population, where the work by Hess & Stathopoulos (2013) also provides further insights into who might be making choices in what way.

A second and related body of papers questions the assumption that every *attribute* is processed using the same decision rule. The resulting (often called: hybrid) model structures allow for different attributes to be processed using different decision rules; an example being the hybrid utility-regret model put forward by Chorus et al. (2013) and employed by, amongst others, Leong & Hensher (in press (b)); this model allows for some attributes to be processed in a regret minimisation fashion while others are processed in a utility maximisation fashion.

An obvious next step would be to combine these two approaches; that is, allowing for different *individuals* to use different decision rules, for different *attributes*. In principle, a combined hybrid-discrete mixture model would be able to accommodate for both types of decision rule heterogeneity simultaneously. However, there is an important caveat that so far has hampered progress in this direction: when the number of attributes is non-trivial, a large number of latent classes is needed to capture all possible combinations of decision-rules. For example, hybrid utility maximisation-regret minimisation models for a four-attribute choice context would already result in 16 latent classes. Resulting mixture models are generally not well behaved due to the large number of classes, precluding successful estimation on empirical data.

This paper builds on a recent advance in regret minimisation modelling to circumvent this combinatorial explosion. The recently proposed generalised random regret minimisation model (Chorus, 2014) estimates a regret weight for each attribute; if equal to 1, conventional regret minimisation behaviour – i.e., as in Chorus (2010) – is obtained. If equal to 0, conventional utility maximisation behaviour is obtained. Values between 0 and 1 imply regret minimisation behaviour, but with a smaller degree of non-linearity in the regret function than is the case in conventional regret minimisation models. Since the regret weight can be estimated on empirical data, this provides an opportunity to infer from the data, for each attribute, if it is processed using a utility maximisation or regret minimisation rule. That is, rather than having to estimate all possible hybrid utility-regret combinations to find the optimal constellation, the generalised random regret minimisation model allows one to directly infer the best fitting decision rule for every attribute.

When estimating hybrid mixture models of the type discussed above, the generalised random regret model provides a clear conceptual and operational advantage over previously used hybrid utility-regret models: conceptually, it is much more elegant to estimate attribute-specific decision rules from the data as opposed to these rules being imposed by the researcher. Operationally, the generalised approach limits the number of latent classes, as one no longer needs to allow for every possible combination of attribute-specific decision rules. This results in better behaved and more manageable models. Nevertheless, the model still clearly allows for the possibility of different classes making use of a different combination of RUM/RRM parameters. This paper is the first to combine the generalised random regret minimisation model and the discrete mixture paradigm; as such, it is the first attempt that we know of, to simultaneously allow for heterogeneity in decision rules across individuals and attributes in a discrete choice model of traveller behaviour.

The remainder of this paper is structured as follows. Section 2 presents the model structure put forward in our paper. Section 3 presents empirical analysis based on a stated route choice dataset. Conclusions and directions for further research are put forward in section 4.

2. Model structures

This structure gives an overview of the various model structures used in this paper. We first discuss the generalised random regret minimisation model, which can simplify to both standard RUM and RRM structures, before discussing mixture models.

2.1. The Generalised Random Regret Minimisation model

This section draws heavily from the recently published paper (Chorus, 2014) which puts forward the Generalised Random Regret Minimisation (or: G-RRM) model. For reasons of space limitations and to avoid repetition, we do not present and discuss the conventional RRM model and its properties; for further information on the RRM model, the interested reader is referred to Chorus (2012), with applications for example in Boeri et al. (2012, 2013), Kaplan & Prato (2012) and Prator (2014). The same applies for the standard RUM model, which is well covered elsewhere. Instead, we directly start with introducing the G-RRM model; the G-RRM model assumes that discrete choice behaviour is driven by minimisation of the following objective function:

$$RR_i = R_i + \upsilon_i = \sum_{j \neq i} \sum_m \ln\left(\gamma_m + \exp\left[\beta_m \cdot \left(x_{jm} - x_{im}\right)\right]\right) + \upsilon_i, \quad (1)$$

where:

RR_i denotes the random (or: total) regret associated with a considered alternative i

R_i denotes the 'deterministic' regret associated with i

υ_i denotes the 'unobserved' regret associated with i, its negative being distributed i.i.d. Extreme Value Type I with variance $\pi^2/6$

β_m denotes the estimable taste parameter associated with attribute x_m

x_{im}, x_{jm} denote the values associated with attribute x_m for, respectively, the considered alternative i and another alternative j

γ_m denotes the regret-weight for attribute x_m.

The probability of choosing alternative *i* out of *J* is then given by:

$$P_i = \frac{-R_i}{\sum_{j=1}^{J} -R_j}, \quad (2)$$

i.e. the minimisation of the deterministic regret component.

To illustrate the role of regret-weight γ_m, consider the so-called attribute-regret function $\ln\left(\gamma_m + \exp\left[\beta_m \cdot \left(x_{jm} - x_{im}\right)\right]\right)$, which gives the regret that is associated with comparing considered alternative i with another alternative j, in terms of attribute x_m. Note that when γ_m equals one, the conventional attribute regret function as put forward in Chorus (2010) is obtained. By varying γ_m from 0 to 1, and plotting the resulting attribute regret function for $\left(x_{jm} - x_{im}\right)$ ranging from minus 5 to 5 (keeping β_m fixed at unity), the role of the regret-weight becomes immediately clear; the left hand panel of Figure 1 shows the effect on the attribute regret function of a step-wise variation in γ, and the right hand panel shows the effect of a continuous change in γ.

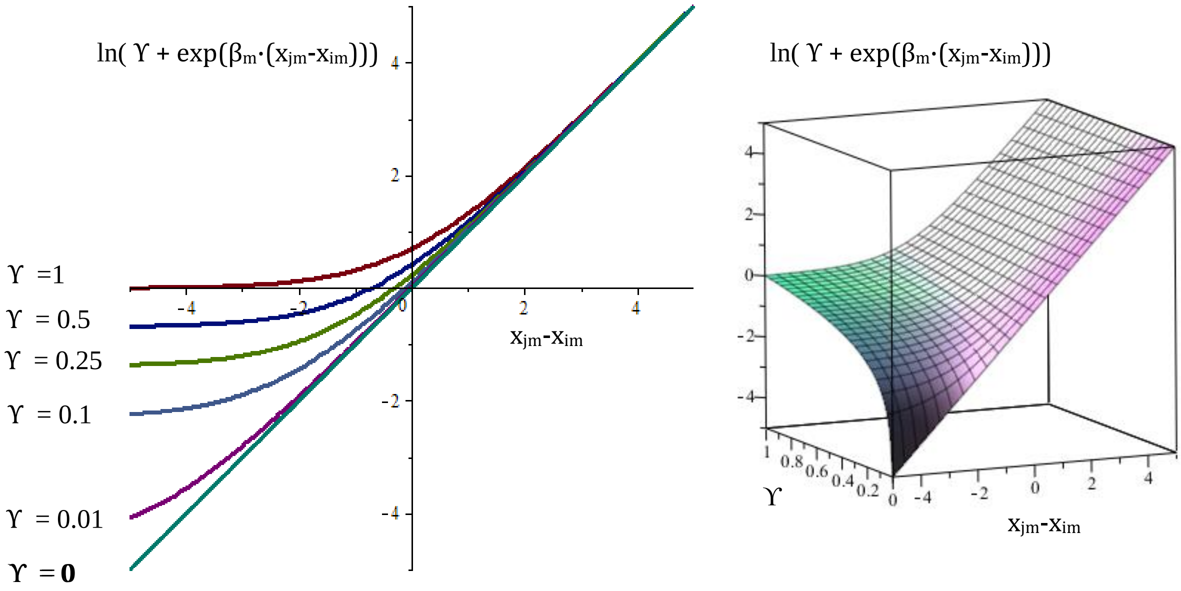

Figure 1: Impact of variation in regret-weight ($\boldsymbol{\gamma}$) on attribute regret ($\boldsymbol{\beta_m}$=1)

Both graphs show that when the regret-weight becomes smaller and starts to approach zero, the convexity of the regret-function and the resulting reference dependent asymmetry (or: non-linearity) in preferences vanishes. When $\gamma = 0$, there is no asymmetry anymore, implying that the impact on regret of a change in an alternative's attribute is no longer dependent on the alternative's initial performance in terms of the attribute, relative to its competition. Intuitively, the resulting regret function with symmetric preferences (i.e., with γ = 0) looks like a function generated by a linear-in-parameters RUM model, and indeed it can be shown (Chorus, 2014) that if γ = 0 for a particular attribute, the G-RRM model generates the same choice behaviour as does a linear-in-parameter RUM model, for that attribute. As noted by Chorus (2014), the coefficient values obtained when γ = 0 for a particular attribute are scaled down by a factor of *J* compared to the estimation of a RUM model, which is immediately obvious from Equation (1) which shows that the coefficient is used *J* times per G-RRM function, instead of just once as in a RUM function.

Clearly, depending on the values of γ_m for different attributes, different choice models arise; as such, the G-RRM model can be seen as a generic formulation which nests various types of choice models: the conventional RRM model is a special case which is obtained when $\gamma_m = \gamma = 1\ \forall m$, and the conventional linear-in-parameters RUM model is a special case which is obtained when $\gamma_m = \gamma = 0\ \forall m$. Furthermore, hybrid RUM-RRM models of the type proposed in Chorus et al. (2013) are obtained when $\gamma_m \in \{0,1\}\ \forall m$.

A more subtle model structure arises when $\gamma_m \in]0,1[$ for one or more attributes: as can be seen when inspecting Figure 1, for these in-between values of γ, the regret/utility function is still convex, but the degree of non-linearity is not as high as for a conventional RRM model (or: G-RRM with γ = 1). Nonetheless, given the presence of some asymmetry and reference-dependency in the regret function for values of $\gamma \in]0,1[$, the resulting or implied behaviour should be considered as regret-minimisation as opposed to linear-in-parameters utility maximisation behaviour. As a consequence, for the G-RRM model with $\gamma \in]0,1[$, previously derived properties of the conventional RRM model hold, but these properties are less pronounced than for the conventional RRM model.

As an illustration of how values of γ influence the salience of key properties of the RRM model, Figure 2 presents a numerical simulation in the context of the G-RRM model (with two different regret-weights). The numerical example refers to the existence of a compromise effect, which states that consumers tend to prefer alternatives with a reasonable performance on each attribute, as opposed to alternatives with a strong performance on some, and a poor performance on other attributes. The fact that the RRM model predicts such a compromise effect has been reported in various theoretical (Chorus, 2010) and empirical (Chorus & Bierlaire, 2013) studies. As these papers show, it is the convexity of the regret function in the RRM model that generates the compromise effect: deterioration of an attribute with an already poor performance generates a lot of additional regret, and this cannot be compensated by the relatively small decreases in regret associated with further improvement of an already strong attribute – this is easily verified when inspecting the regret curve for $\gamma = 1$ in Figure 1's left panel. As a consequence of this semi-compensatory behaviour implied by the convexity of the regret function, alternatives with a reasonable performance on all attributes gain a market share bonus in the context of

RRM models. Note that, given the symmetric and reference independent treatment of attributes in linear-in-parameter RUM-models, the latter do not feature the compromise effect (as for example highlighted and empirically verified in Chorus & Bierlaire, 2013).

Since RRM's accommodation of the compromise effect is a consequence of the convexity of its regret function, the expectation is that for a G-RRM model with $0 < \gamma < 1$ this compromise effect, while still present, will be less pronounced than in a conventional RRM model (i.e., a G-RRM model with $\gamma = 1$). The following numerical example – which has been used in previous publications to illustrate how RRM generates compromise effects, and as such forms a good benchmark – serves to illustrate and verify this expectation. Assume a choice situation between three alternatives, each defined in terms of two quality attributes (x_1 and x_2) that are equally important to the decision-maker (higher values are preferred over lower ones; β = 1 for each of the two attributes): A = {1, 3}, B = {2+Δ, 2–Δ}, C = {3,1}. In words: where A and C take on relatively extreme positions on the two attributes, B is a compromise alternative to the extent that Δ is close to 0. Figure 2 plots P(A), P(B) and P(C) as a function of Δ, for G-RRM models with $\gamma = 1$ for both attributes, and for G-RRM models with $\gamma = 0.1$ for both attributes, respectively (the former in solid lines, the latter in dotted lines).[1]

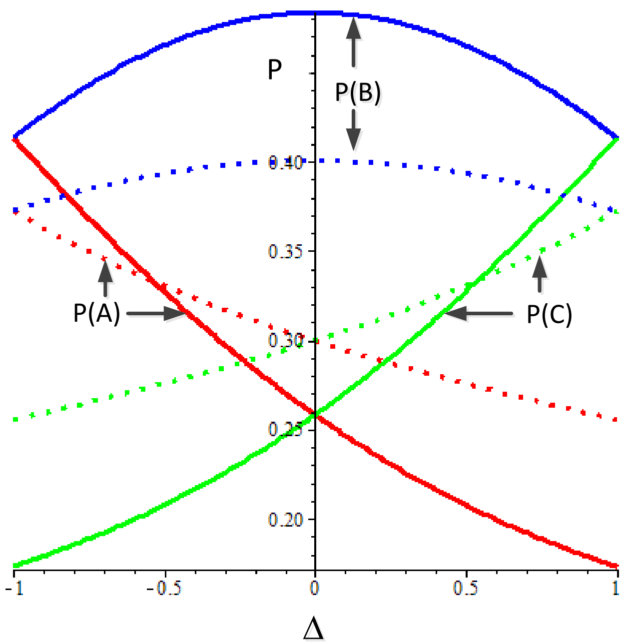

Figure 2: Compromise effect for the conventional RRM-model and for G-RRM with regret-weights 0.1 (the latter in dotted lines)

[1] Note that choice probabilities generated by a RUM's linear-in parameters Logit-model would of course be insensitive to changes in Δ as the model would assign equal choice probabilities of 0.33… for all three alternatives, irrespective of the value for Δ; these probabilities are hence not plotted, for clarity of communication.

Figure 2 clearly shows the presence of a compromise effect for alternative B: as long as its attributes remain close to 2 (i.e., Δ close to 0), it receives a choice probability bonus at the cost of the two extreme alternatives. Secondly, and this is more relevant in the context of this paper, the compromise effect is still present for the case where $\gamma = 0.1$, but less pronounced than for the case where $\gamma = 1$ (i.e., the conventional RRM model). This decreasing salience of the compromise effect for the G-RRM model with the lower regret weight follows directly from the decreasing difference between the sensitivity to attribute changes in the domain of poor performance versus in the domain of strong performance. This decreasing difference in turn, follows directly from the decreased level of non-linearity of the regret curve, which is a direct consequence of the lower regret weight. In sum: regret weights with values between 0 (linear-in-parameters RUM) and 1 (conventional RRM model) still generate regret minimisation behaviour and as such still result in key properties of the conventional RRM model, but with a reduced salience.

For estimation purposes, it is pragmatic to parameterise the regret-weight in terms of a binary logit function: $\gamma_m = \exp(\delta_m)/(1 + \exp(\delta_m))$. For (large) negative values of δ_m, a RUM specification is approached for the attribute, and for (large) positive values, a RRM specification is approached. When δ_m is estimated to lie in-between these two extremes, e.g. when it is estimated to be insignificantly different from 0, implying $\gamma_m = 0.5$, regret minimisation behaviour is obtained but with less emphasis on regret than is the case for the conventional RRM model.

2.2. Mixture model

The mixture model introduced by Hess et al. (2012) is a simple generalisation of a latent class structure where the differences between classes is not just in parameters of the same underlying model structures but also in the use of different model structures in different classes.

A general specification of a model allowing for different decision rules within a latent class framework is given by:

$$LC_n = \sum_{s=1}^{S} \pi_{n,s} \, LC_{n,s}, \qquad (3)$$

where LC_n is the contribution to the likelihood function of the observed choices for respondent *n*. This probability of observed choices is given by a weighted average over *S* different types of models, where LC_{ns} is the probability of the observed sequence of choices for person *n* if model *s* is used, and $\pi_{n,s}$ is the weight attached to model *s* (representing a specific decision process), where $\sum_{s=1}^{S} \pi_{n,s} \quad \forall n$. The mixing of models is performed at the level of individual respondents rather than individual choices.

In existing work, the above specification has been used to combine models such as RUM, RRM and elimination by aspects (EBA). In the present paper, we extend on this by making the individual classes use G-RRM models, thus allowing not just for mixtures between pure

RUM and RRM classes, but also mixed RUM-RRM classes and classes with intermediate specifications for individual parameters.

3. Empirical analysis

3.1. Data

The data used in the present paper is the same as that in Chorus & Bierlaire(2013). The data collection effort focused on route choice behaviour among commuters who travel from home to work by car. A total of 550 people were sampled from an internet panel maintained by IntoMart in April 2011. Sampled individuals were at least 18 years old, owned a car, and were employed. The sample was representative of Dutch commuters in terms of gender, age and education level. Of the 550 sampled individuals, 390 filled out the survey (implying a response rate of 71%).

Respondents to the survey were asked to imagine the hypothetical situation where they were planning a new commute from home to work (either because they had recently moved, or because their employer had recently moved, or because they had started a new job). They were asked to choose between three different routes that differed in terms of the following four attributes, with three levels each: average door-to-door travel time (45, 60, 75 minutes), percentage of travel time spent in traffic jams (10%, 25%, 40%), travel time variability (±5, ±15, ± 25 minutes), and total costs (€5.5, €9, €12.5).

Using the Ngene-software package (ChoiceMetrics, 2009), a so-called 'optimal orthogonal in the differences'-design of choice sets was created to ensure a statistically efficient data collection. This design resulted in nine choice tasks per respondent and 3,510 choice observations in total. Figure 3 shows one of these tasks.

1	Route A	Route B	Route C
Average travel time (minutes)	45	60	75
Percentage of travel time in congestion (%)	10%	25%	40%
Travel time variability (minutes)	±5	±15	±25
Travel costs (Euros)	€12,5	€9	€5,5
YOUR CHOICE	□	□	□

Figure 3: An example route choice-task

3.2. Results

A large number of different models were estimated for this study, gradually increasing the number of classes within the latent class structure discussed in Section 2.2, with each class making use of a G-RRM model, where, in varying classes, a priori structures were either imposed (e.g. pure RUM, pure RRM) or the γ parameters were estimated. All models were coded and estimated in Ox (Doornik, 2001), making use of multiple runs with different starting values to reduce the risk of inferior local maxima.

The models presented here are generally those with any additional constraints on γ imposed which arose from the modelling work, where this relates to cases where δ became either very small or very large. The models are compared in terms of the Bayesian Information Criterion (BIC), which incurs a higher penalty for increases in the number of parameters than traditional likelihood ratio or adjusted ρ^2 measures.

We begin with Table 1, which shows three models with a single class each, i.e. *S*=1. In models 1 and 2, we a priori impose a RUM and RRM structure respectively, i.e. constraining γ=0 in model 1, and γ=1 in model 2, across attributes. The results show that the pure RRM structure (model 2) is preferred to the pure RUM structure (model 1), with both producing four significant negative coefficient estimates.

Table 1: single class models

	Model 1		**Model 2**		**Model 3**	
Log-likelihood	-2,613.45		-2,604.96		-2,603.26	
parameters	4		4		6	
adj ρ^2	0.3212		0.3234		0.3233	
BIC	5,252.20		5,235.22		5,244.47	
	est.	rob. t-rat.	est.	rob. t-rat.	est.	rob. t-rat.
β(travel time)	-0.0224	-22.17	-0.0469	-20.57	-0.0330	-3.91
β(perc. in congestion)	-0.0091	-15.21	-0.0181	-14.44	-0.0188	-14.78
β(travel time variability)	-0.0105	-11.83	-0.0210	-11.80	-0.0147	-0.68
β(travel cost)	-0.0576	-16.83	-0.1128	-15.59	-0.1166	-16.66
δ(travel time)	- inf (fixed a priori)		+ inf (fixed a priori)		-0.3622	-0.27
δ(perc. in congestion)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
δ(travel time variability)	- inf (fixed a priori)		+ inf (fixed a priori)		-0.4656	-0.05
δ(travel cost)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
γ(travel time)	0		1		0.41	
γ(perc. in congestion)	0		1		1	
γ(travel time variability)	0		1		0.39	
γ(travel cost)	0		1		1	

In model 3, we initially estimated all four δ parameters freely, but the values for δ(perc. in congestion) and δ(travel cost) reached very large positive values which indicated that γ=1 for these two attributes, meaning a pure RRM treatment. For the remaining two parameters, the δ estimates are negative albeit not significantly different from zero, indicating a value for $\gamma<0.5$. As can be seen from Figure 1, this is still closer to full RRM treatment than to full RUM treatment, a finding that is also in line with the small difference in fit between model 2 and model 3, with the BIC measure indicating that the additional two parameters in model 3 do not justify the improvement in log-likelihood. The findings from the first three models hence point towards a pure RRM treatment for all four parameters. Even though the G-RRM model in the end thus 'collapses' to a simpler RRM model, it highlights the advantage of the structure as a diagnostic tool, rather than having to estimate all possible combinations of RUM and RRM treatment for the different parameters.

We next move to four separate models using two classes each in Table 2. This includes one model (model 4) with two pure RUM classes, one (model 5) with two pure RRM classes, one with one pure RUM class and one pure RRM class (model 6) and one with two G-RRM classes (model 7). All four models show major improvements over the single class models in Table 1. They also all show a split into one class with a roughly 1/3 probability, and one with a roughly 2/3 probability, where a very consistent pattern emerges across the four models. In the second class, the relative importance of travel time and travel time variability is reduced substantially, especially the former, while the importance of congestion and to a lesser extent travel cost is increased. Across the four models, the impact of congestion also only attains low levels of significance in the first class.

In terms of model structure, we first note that model 5 outperforms model 4, i.e. a structure with two pure RRM classes is preferred to a structure with two pure RUM classes. However, model 6, which a priori imposes one pure RUM class and one pure RRM class, obtains an even better BIC measure, suggesting heterogeneity across respondents not just in sensitivities but also in decision rules. Model 7 relaxes the assumption in model 6 about within class homogeneity of the decision rule for different attributes. The findings for the second class remain the same as the imposition of a pure RRM treatment for all attributes in model 6. After additional constraints for δ_1(% cong.), δ_1(tt var) and δ_1(cost), the picture in the first class is slightly different from model 6. We still see a pure RUM treatment for congestion and travel time variability. However, for travel time, we see a γ value of around one third, indicating a treatment that is not pure RUM and actually suggests a fairly regret based treatment, while, for cost, the treatment is purely regret based. Despite these differences, the log-likelihood for models 6 and 7 is essentially the same, where the one additional parameter (δ1(trav. time)) in model 7 leads to a higher (worse) BIC. The recommendations from the two class structures would thus be a further constrained version of model 7, using the G-RRM model once again as a specification search tool, leading to a mixed RUM-RRM treatment in one class, and a pure RRM treatment in the second class.

Table 2: two class models

	Model 4		Model 5		Model 6		Model 7	
Log-likelihood	-2,431.59		-2,416.78		-2,412.92		-2,412.83	
parameters	9		9		9		10	
adj ρ2	0.3671		0.3709		0.3719		0.3717	
BIC	4,920.11		4,890.49		4,882.77		4,888.91	
	est.	rob. t-rat.	est.	rob. t-rat.	est.	rob. t-rat.	est.	rob. t-rat.
β_1(trav. time)	-0.0559	-10.15	-0.1582	-5.44	-0.0559	-9.71	-0.0558	-9.77
β_1(% cong.)	-0.0025	-1.44	-0.0052	-1.35	-0.0030	-1.65	-0.0034	-1.91
β_1(tt var)	-0.0261	-6.33	-0.0510	-4.54	-0.0260	-6.20	-0.0259	-6.22
β_1(cost)	-0.0437	-4.63	-0.0864	-4.36	-0.0404	-4.16	-0.0808	-4.16
β_2(trav. time)	-0.0146	-12.81	-0.0314	-11.23	-0.0310	-12.23	-0.0309	-12.19
β_2(% cong.)	-0.0131	-13.38	-0.0266	-12.37	-0.0275	-12.74	-0.0276	-12.71
β_2(tt var)	-0.0088	-7.92	-0.0182	-8.27	-0.0180	-7.88	-0.0180	-7.87
β_2(cost)	-0.0725	-15.72	-0.1451	-14.05	-0.1495	-13.75	-0.1496	-13.76
δ_1(trav. time)	- inf (fixed a priori)		+ inf (fixed a priori)		- inf (fixed a priori)		-0.6918	-4.10
δ_1(% cong.)	- inf (fixed a priori)		+ inf (fixed a priori)		- inf (fixed a priori)		- inf (fixed)	
δ_1(tt var)	- inf (fixed a priori)		+ inf (fixed a priori)		- inf (fixed a priori)		- inf (fixed)	
δ_1(cost)	- inf (fixed a priori)		+ inf (fixed a priori)		- inf (fixed a priori)		+ inf (fixed)	
γ_1(trav. time)	0		1		0		0.33	
γ_1(% cong.)	0		1		0		0.00	
γ_1(tt var)	0		1		0		0.00	
γ_1(cost)	0		1		0		1.00	
δ_2(trav. time)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
δ_2(% cong.)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
δ_2(tt var)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
δ_2(cost)	- inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed a priori)		+ inf (fixed)	
γ_2(trav. time)	0		1		1		1.00	
γ_2(% cong.)	0		1		1		1.00	
γ_2(tt var)	0		1		1		1.00	
γ_2(cost)	0		1		1		1.00	
δ_{s1}	-0.6917	-4.18	-0.77489	-3.88	-0.6956	-4.14	-0.6918	-4.10
δ_{s2}	0	-	0	-	0	-	0	-
π_1	33.36%		31.54%		33.28%		33.36%	
π_2	66.64%		68.46%		66.72%		66.64%	

Table 3 contains our final set of four models, where models with more classes than those presented in Table 3 gave a worse BIC. To allow for a generic format in the presentation of results, Table 3 makes use of six classes throughout, where the first two are pure RUM classes, followed by two pure RRM classes, with the fifth and sixth being freely estimated G-RRM models (subject to additional constraints). Not each model makes use of each class, as detailed in the table.

Model 8 combines a pure RUM class (class 1) with a pure RRM class (class 3) and a G-RRM class (class 5). This model outperforms the best two class models from Table 2, where the findings suggest that, in the third class, all parameters except travel time variability should have a pure RRM treatment, where the estimate for travel time variability is however also not statistically significant in this class. This third class obtains around one quarter of the overall weight in the model, with a slightly higher probability for the other (a priori) pure RRM class than for the pure RUM class. These findings are overall compatible with those from model 7, albeit that congestion and travel time now have a pure RRM treatment outside the pure RUM class.

The remaining three models all make use of two RUM classes and two RRM classes, where models 10 and 11 add in one, respectively two, additional G-RRM classes. While model 9 provides an improvement over previous structures, this is solely a result of allowing for taste heterogeneity within the RRM segment of the models (i.e. classes 3&4) as the additional RUM class (class 1) obtains a near zero weight (1.92%) with no significant parameter estimates. The fact that further improvements in fit are obtained in models 10 and 11 highlights the value of the G-RRM approach in capturing attribute specific treatments of decision rules. Firstly, by allowing for these additional classes, we are able to capture some of the within decision rule heterogeneity for the RUM classes, where we now see a more even split in probability across RUM classes, along with some significant effects in both classes. The majority of the weight remains with a RRM treatment of attributes, whether in the pure RRM classes or the RRM treatment within G-RRM classes. Additionally however, we see that more pure RRM classes are needed (given that one additional G-RRM class in model 11 becomes pure RRM) before we can recover a class with a stronger mixture between RUM and RRM, namely class 5 in model 11.

Table 3: three to six class models

	Model 8		Model 9		Model 10		Model 11	
Log-likelihood	-2,347.19		-2,322.86		-2,301.42		-2,284.89	
parameters	15		19		25		30	
adj ρ2	0.3874		0.3927		0.3967		0.3997	
BIC	4,789.26		4,765.90		4,760.98		4,759.54	
	est.	rob. t-rat.	est.	rob. t-rat.	est.	rob. t-rat.	est.	rob. t-rat.
β_1(trav. time)	-0.0531	-11.35	-0.5172	0.00	-0.0223	-2.47	-0.0040	-1.57
β_1(% cong.)	-0.0043	-2.18	-0.2440	0.00	-0.0347	-3.92	-0.0119	-2.54
β_1(tt var)	-0.0250	-7.09	-0.1769	-0.35	-0.0129	-1.84	-0.0104	-3.73
β_1(cost)	-0.0354	-4.29	-4.0859	-0.41	-0.0630	-2.16	-0.0222	-1.73
β_2(trav. time)		-	-0.0548	-10.78	-0.0598	-7.90	-0.0363	-2.38
β_2(% cong.)		-	-0.0041	-2.06	-0.0063	-1.34	-0.0277	-4.24
β_2(tt var)		-	-0.0261	-6.85	-0.0135	-2.67	-0.0085	-0.62
β_2(cost)		-	-0.0341	-4.05	-0.0370	-1.68	-0.0442	-0.78
β_3(trav. time)	-0.0654	-3.75	-0.0151	-4.07	-0.0982	-2.07	-0.6426	-0.88
β_3(% cong.)	-0.0306	-8.00	-0.0271	-8.66	-0.0135	-0.39	-0.0653	-0.61
β_3(tt var)	-0.0158	-4.54	-0.0211	-6.08	-0.0989	-4.70	-0.0359	-1.36

β_3(cost)	-0.3049	-4.79	-0.0810	-6.20	-0.1308	-1.46	-1.0933	-0.62
β_4(trav. time)	-		-0.1007	-4.91	-0.0783	-2.77	-0.1529	-4.37
β_4(% cong.)	-		-0.0319	-7.35	-0.0257	-5.26	-0.0288	-1.93
β_4(tt var)	-		-0.0167	-4.27	-0.0150	-3.72	-0.0226	-3.19
β_4(cost)	-		-0.3873	-6.04	-0.3980	-4.32	-0.5812	-6.04
β_5(trav. time)	-0.0115	-1.91	-		-0.0084	-1.71	-0.0520	-1.98
β_5(% cong.)	-0.0270	-6.71	-		-0.0157	-2.50	-0.0036	-0.80
β_5(tt var)	-0.0134	-0.26	-		-0.0108	-0.17	-0.0400	-4.24
β_5(cost)	-0.0652	-2.83	-		-0.0692	-1.91	-0.0953	-3.03
β_6(trav. time)	-		-		-		-0.0243	-2.69
β_6(% cong.)	-		-		-		-0.0282	-1.68
β_6(tt var)	-		-		-		-0.0128	-1.70
β_6(cost)	-		-		-		-0.2781	-3.78
γ_1(generic)	0 (fixed to pure RUM)		0 (fixed to pure RUM)		0 (fixed to pure RUM)		0 (fixed to pure RUM)	
γ_2(generic)	-		0 (fixed to pure RUM)		0 (fixed to pure RUM)		0 (fixed to pure RUM)	
γ_3(generic)	1 (fixed to pure RRM)		1 (fixed to pure RRM)		1 (fixed to pure RRM)		1 (fixed to pure RRM)	
γ_4(generic)	-		1 (fixed to pure RRM)		1 (fixed to pure RRM)		1 (fixed to pure RRM)	
δ_5(trav. time)	+ inf (fixed)		-		+ inf (fixed)		-1.9107	-0.49
δ_5(% cong.)	+ inf (fixed)		-		+ inf (fixed)		+ inf (fixed)	
δ_5(tt var)	-1.2144	-0.05	-		-2.1617	-0.03	+ inf (fixed)	
δ_5(cost)	+ inf (fixed)		-		+ inf (fixed)		+ inf (fixed)	
γ_5(trav. time)	1.00		-		1.00		0.13	
γ_5(% cong.)	1.00		-		1.00		0.00	
γ_5(tt var)	0.23		-		0.10		0.00	
γ_5(cost)	1.00		-		1.00		1.00	
δ_6(trav. time)	-		-		-		+ inf (fixed)	
δ_6(% cong.)	-		-		-		+ inf (fixed)	
δ_6(tt var)	-		-		-		+ inf (fixed)	
δ_6(cost)	-		-		-		+ inf (fixed)	
γ_6(trav. time)	-		-		-		1.00	
γ_6(% cong.)	-		-		-		1.00	
γ_6(tt var)	-		-		-		1.00	
γ_6(cost)	-		-		-		1.00	
δs_1	0.3460	0.97	-2.8410	-6.95	-0.2268	-0.40	0.2737	0.35
δs_2	-		0.0226	0.11	0.3144	0.64	0.1060	0.13
δs_3	0.4757	0.99	-0.0383	-0.16	-0.4835	-0.78	0.1725	0.30
δs_4	-		0	-	0.4541	0.81	0.4091	0.67
δs_5	0	-	-		0	-	0.1591	0.27
δs_6	-		-		-		0	-
π_1	35.14%		1.92%		14.88%		18.03%	
π_2	-		33.61%		25.56%		15.25%	
π_3	40.00%		31.62%		11.51%		16.29%	
π_4	-		32.86%		29.39%		20.64%	
π_5	24.86%		-		18.66%		16.08%	
π_6	-		-		-		13.71%	

4. Summary and conclusions

There is growing interest in the notion that different decision rules may work differently well in explaining the choices observed in data used for travel behaviour analysis. Going further, there is now growing evidence that different decision makers may well be making their choices based on different rules. Finally, there are also results that suggest that the *optimal* decision rule may in fact vary across attributes within a given dataset. The present paper has brought these different notions together by putting forward a latent class approach which not only allows for different decision rules across classes, but also differences in the decision rules used across attributes within a given class. This would clearly lead to a very large number of different possible combinations of rules across classes, and, within the context of two popular paradigms, RUM and RRM, we have proposed to address this through the use of a G-RRM model within individual classes. This allows the *optimal* specification in terms of split between RUM and RRM within a given class to be revealed by the data during estimation, rather than needing to be imposed by the analyst. Initial findings on a standard stated choice dataset are promising and show how a rich pattern of taste heterogeneity, and decision rule heterogeneity across respondents and attributes can be revealed. It has also highlighted that, while in many cases, the G-RRM models collapse to either RUM or RRM for individual coefficients, this provides a very useful diagnostic approach rather than needing to estimate each possible combination of RUM and RRM separately.

References

Arentze, T.A., Timmermans, H.J.P., 2007. Parametric action trees: incorporating continuous attribute variables into rule-based models of discrete choice. *Transportation Research Part B*, 41(7), 772-783

Boeri, M., Longo, A., Doherty, E., Hynes, S., 2012. Site choices in recreational demand: A matter of utility maximization or regret minimization? Journal of Environmental Economics and Policy, 1(1), 32-47

Boeri, M., Longo, A., Grisolia, J.M., Hutchinson, W.G., Kee, F., 2013. The role of regret minimization in lifestyle choices affecting the risk of coronary heart disease. Journal of Health Economics, 32(1), 253-260

Boeri, M., Scarpa, R., Chorus, C.G., 2014. Stated choices and benefit estimates in the context of traffic calming schemes: Utility maximisation, regret minimisation, or both? *Transportation Research Part A*, 61, 121-135

Chorus, C.G., 2010. A new model of Random Regret Minimisation. *European Journal of Transport and Infrastructure Research*, 10(2), pp. 181-196

Chorus, C.G., 2012. Random regret-based discrete choice modeling: A tutorial. Springer, Heidelberg, Germany

Chorus, C.G., in press. A generalised random regret minimisation model. *Transportation Research Part B*

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

Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision-rules in discrete choice models: An approach and four case-studies. *Transportation*, 39(3), 565-591

Hess, S., Stathopoulos, A., 2013. A mixed random utility - random regret model linking the choice of decision rule to latent character traits, *Journal of Choice Modelling*, 9, 27-38.

Kaplan, S., Prato, C.G., 2012. The application of the random regret minimization model to drivers' choice of crash avoidance manoeuvres. Transportation Research Part F, 15(6), 699-709

Leong, W., Hensher, D.A., in press (a). Relative advantage maximisation as a model of context dependence for binary choice data, *Journal of Choice Modelling*

Leong, W., Hensher, D.A., in press (b). Contrasts of relative advantage maximisation with random utility maximisation and regret minimisation, *Journal of Transport Economics and Policy*

Prato, C.G., 2014. Estimating Random Regret Minimization models in the route choice context. Transportation, 41(2), 351-375

Stathopoulos, A., Hess, S., 2012. Revisiting reference point formation, gain-loss asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. *Transportation Research Part A*, 46, 1673-1689