A mixed random utility - Random regret model linking the choice of decision rule to latent character traits

Stephane Hess and Amanda Stathopoulos

Institute for Transport Studies
FACULTY OF ENVIRONMENT

UNIVERSITY OF LEEDS

Presentation outline

Presentation outline

RUM

- Random utility maximisation has been the key paradigm in modelling individual decisions for four decades
- Idea is that decision makers quantify the appeal of an alternative as a utility and choose the alternative which maximises their utility or minimises their disutility
- An inherent characteristic of this framework is the notion of trading between attributes, with poor performance on one attribute (e.g. cost) being compensated for by good performance on another attribute (e.g. time)

Alternatives to RUM

- RUM is in very widespread use, with a range of different models, going way beyond MNL in terms of complexity
- However, RUM has always been criticised, especially by behaviour economists and mathematical psychologists, as giving a simplified representation of reality
- Various alternative models exist, for example elimination by aspects, but these structures have disadvantages in application to large scale problems with large numbers of attributes and alternatives, and also in terms of producing outputs that can be used in economic analysis
- Most recently, there has been growing interest in modelling regret as an alternative to utility

RRM I

- Regret is a negative emotion experienced when we imagine that a present situation would have been better if we had made a different decision
- Several adaptations are needed to allow regret models to be estimated with the typical, multi-alternative, multi-attribute data obtained in choice experiments
- Initial work by by Chorus et *al*. defines 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 ($i \neq j$)

RRM II

- In this specification, regret is computed only considering the best forgone alternative

$$R_{i,n,t} = max_{i \neq j} \left\{ \sum_{k=1..K} max \left\{ 0, \beta_k (x_{j,k,n,t} - x_{i,k,n,t}) \right\} \right\}$$

- Choice probability for alternative *i* with *iid* type 1 extreme value errors is written $P_{i,n,t} = \frac{exp(-R_{i,n,t})}{\sum exp(-R_{j,n,t})}$

RRM III

- RRM model was later revised to move away from just comparison with the best foregone alternative on each attribute
- Incorporate comparisons across all pairs of alternatives

$$R_{i,n,t} = \sum_{j \neq i} \sum_{k=1..K} \ln\left(1 + e^{\beta_k \left(x_{j,k,n,t} - x_{i,k,n,t}\right)}\right)$$

Contrasts I

- Recent work by Chorus and colleagues in this context also shows some evidence that on some datasets, RRM works better than RUM, with the opposite being the case on other datasets
- Also some insights by allowing jointly for the two paradigms within a single model
- The two models are inherently different in terms of interpretation
- Whereas a RUM based analysis derives the sensitivity to attributes, RRM estimates the potential contribution to regret feelings of each attribute.
- Due to this, the comparison across models is not straightforward

Contrasts II

- RRM seems to often favour compromise alternatives, i.e. it is a suitable mechanism for a choice situation where the decision maker wants to avoid poor performance on one or more attributes
- Behaviourally, this could make a lot of sense especially in choices of (long term) products as opposed to (short term) services
- Thus far, it is not possible to easily use RRM for welfare analysis, e.g. computing WTP measures
- Questionable whether this is possible at all given the way regret depends also on performance of other alternatives
- However, it is likely to be a very suitable approach in forecasting scenarios, and can be estimated with standard MNL software

Introduction

- Haven't really looked at what might drive the choice of an appropriate decision rule at the level of a person
- Can allow for different rules for different people in a sample
- Can explain, especially in SP, how this relates to their real world situation

Methodology I

- In line with work by Hess et al. (2012), a general specification of a model allowing for different decision rules within a latent class framework is given by:

$$LC_n\left(\beta_1, \ldots, \beta_S, \pi_1, \ldots, \pi_S\right) = \sum_{s=1}^{S} \pi_s LC_{n,s}\left(\beta_s\right)$$

  where LC_n is the contribution to the likelihood function of the observed choices for respondent n (out of N)

- Two key shortcomings:
  - Risk of confounding between heterogeneity in sensitivities and heterogeneity in decision rules
  - Limited insight into the factors determining the choice of decision rule

Methodology II

- Hess et al. (2012) deal with confounding by including additional random heterogeneity in a continuous manner, i.e. $\beta_s \sim f(\beta_s \mid \Omega_s)$:

$$LC_n(\Omega_1, \ldots, \Omega_S, \pi_1, \ldots, \pi_S) = \sum_{s=1}^{S} \pi_s \left[\int_{\beta_s} LC_{n,s}(\beta_s) f(\beta_s \mid \Omega_s) \, \mathrm{d}\beta_s \right],$$

- Imposes substantial demands in terms of computational complexity as well as empirical identification
- We put forward the use of an additional layer of latent classes:

$$LC_n\left(\beta^{(1)}, \ldots, \beta^{(S)}, \pi, \varpi^{(1)}, \ldots, \varpi^{(S)}\right) = \sum_{s=1}^{S} \pi_s \sum_{k=1}^{K_s} \varpi_{k,s} LC_{n,s}(\beta_{s,k})$$

- Model now uses K_s different classes for model s
- Averaging across classes is performed at the level of individual respondents, recognising the repeated choice nature of the data

Methodology III

- Other shortcoming is the lack of explanation as to what drives the likelihood of a given paradigm being more appropriate for one specific respondent than another
- Rather than linking class allocation probabilities π to respondent specific characteristics, we link it to unobserved character traits
- Working with a single such trait for the sake of exposition, let us refer to it as α_n for respondent n
- Simplifying our overall structure further to the case of just two decision paradigms, we now write:

$$\pi_{n,1} = \frac{1}{1 + e^{\delta_{\pi,2} + \tau \alpha_n}} \tag{1}$$

$$\pi_{n,2} = \frac{e^{\delta_{\pi,2} + \tau \alpha_n}}{1 + e^{\delta_{\pi,2} + \tau \alpha_n}}. \tag{2}$$

where α_n is a latent component specific to respondent n

Methodology IV

- Model the value of α_n as:

$$\alpha_n = \sum_{l=1}^{L} h_l \left(z_{n,l}, \gamma_l \right) + \eta_n$$

  where η_n is standard Normal across individuals

- Likely difficult to find meaningful socio-demographic explanators for underlying character traits
- More likely to be intrinsic to a person and shaped by experience and lifestyle, either of which are difficult to capture in data

Methodology V

- Thus far, simply allows for random (through η_n) and deterministic (through $\gamma' z_n$) variations in the class allocation probabilities
- Estimates relationship between latent character traits and likely decision rules only on the basis of data on choices
- Only provide a snapshot of preferences in a very controlled settings at a particular point in time
- Arguably do not permit us to make the full link to what we regard as person specific character traits which are constant over a longer time horizon
- Make use of additional information relating to other manifestations of these character traits

Methodology VI

- Assume data contains additional variables which we hypothesise to be a function of the same latent character traits
- Identification of such variables is a difficult task and could encompass a range of different formats, be it answers to questions on attitudes and perceptions, or descriptors of lifestyle and past experiences
- Of crucial importance within the behavioural concept at the heart of our approach is that they need to relate to long term traits rather than short term feelings
- Model values of these *indicators*, say $I_{n,1}$ to $I_{n,M}$ grouped together into a vector I_n as:

$$I_{n,m} = \delta_{I,m} + \zeta_m \alpha_n + \upsilon_{n,m}$$

- i.e. they are linked to the same underlying character traits α_n

Methodology VI

- In estimation, we now jointly maximise the likelihood of the observed choices and the observed values of the indicators
- Enable the model to create a link between the behaviour in the short term context (i.e. stated choice) and the longer term character traits
- First hybrid model allowing long term character traits to explain decision rule heterogeneity in choice data

$$L_n = \int_{\eta_n} \left[\sum_{s=1}^{S} \pi_{n,s}\left(\alpha_n\right) \sum_{k=1}^{K_s} \varpi_{n,s,k} LC_{n,s}\left(\beta_{s,k}\right) \right] \left[\prod_{m=1}^{M} LI_{n,m}\left(\alpha_n\right) \right] \phi\left(\eta_n\right) \mathrm{d}\eta_n$$

Data I

- Online survey conducted on rail and bus commuters in the UK in 2010
- Route choice between reference trip and two hypothetical alternatives, with six attributes (travel time, fare, rate of crowded trips, rate of delays, average length of delays, provision of a delay information service)
- Collected data concerning acceptable and ideal conditions for each attribute of their reference trip
- Hypothesis: a declared shortfall from ideal reference values makes it less likely that the respondent chose that real world commute trip by minimising regret
- Assume that travellers have a certain amount of influence on their commute journey and over time align it with their aspirations
- Test whether size of gap between ideal/acceptable and current values is related to the predisposition to use a regret-minimising decision rule

Utility specification

- Make use of two paradigm specific latent classes to capture additional heterogeneity, such that $K_s = 2, \forall s$
- The deterministic utility for alternative i ($i = 1, \ldots, 3$) for respondent n in choice task t is given:

$$
\begin{aligned}
V_{n,t,i,k} = \; & \delta_{RUM,i,k} \\
& + \beta_{RUM,TT,k} TT_{n,t,i} \\
& + \beta_{RUM,LF,k} \ln \left(FARE_{n,t,i} \right) \\
& + \beta_{RUM,RA,k} RA_{n,t,i} \\
& + \beta_{RUM,RI,k} RI_{n,t,i} \\
& + \beta_{RUM,C,k} C_{n,t,i}
\end{aligned}
$$

Regret specification

- The deterministic regret for alternative i ($i = 1, \ldots, 3$) for respondent n in choice task t is given:

$$
\begin{aligned}
R_{n,t,i,k} = \delta_{RRM,i,k} & \\
& + \sum_{j \neq i} \ln\left(1 + e^{\beta_{RRM,TT,k}\left(TT_{n,t,j} - TT_{n,t,i}\right)}\right) \\
& + \sum_{j \neq i} \ln\left(1 + e^{\beta_{RRM,LF,k}\left(LF_{n,t,j} - LF_{n,t,i}\right)}\right) \\
& + \sum_{j \neq i} \ln\left(1 + e^{\beta_{RRM,RA,k}\left(RA_{n,t,j} - RA_{n,t,i}\right)}\right) \\
& + \sum_{j \neq i} \ln\left(1 + e^{\beta_{RRM,RI,k}\left(RI_{n,t,j} - RI_{n,t,i}\right)}\right) \\
& + \sum_{j \neq i} \ln\left(1 + e^{\beta_{RRM,C,k}\left(C_{n,t,j} - C_{n,t,i}\right)}\right)
\end{aligned}
$$

Measurement equations I

- Did not include a deterministic component for the latent variable structural equation
- Motivated by a desire to not confound the drivers of decision rule heterogeneity with heterogeneity caused by socio-demographic factors
- Focussed on how well a respondent's current commute journey lined up with their aspirations
- Commute journeys evolve over time and can be assumed to be based on informed choices for most travellers
- Hypothesis: a respondent who is more likely to be driven by regret is less likely to have settled on a current commute journey which performs poorly against their desired values

Measurement equations II

$$I_{n,1} = TT_{n,1} - TT_{n,ideal} - \sum_{n=1}^{N} \frac{TT_{n,1} - TT_{n,ideal}}{N}$$

$$I_{n,2} = TT_{n,1} - TT_{n,acc} - \sum_{n=1}^{N} \frac{TT_{n,1} - TT_{n,acc}}{N}$$

$$I_{n,3} = \ln\left(FARE_{n,1}\right) - \ln\left(FARE_{n,ideal}\right) - \sum_{n=1}^{N} \frac{\ln\left(FARE_{n,1}\right) - \ln\left(FARE_{n,ideal}\right)}{N}$$

$$I_{n,4} = \ln\left(FARE_{n,1}\right) - \ln\left(FARE_{n,acc}\right) - \sum_{n=1}^{N} \frac{\ln\left(FARE_{n,1}\right) - \ln\left(FARE_{n,acc}\right)}{N}$$

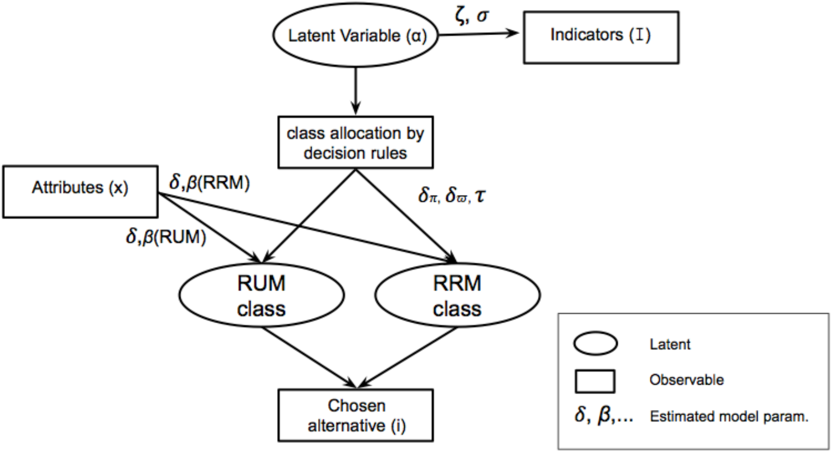
$$I_{n,5} = RA_{n,1} - RA_{n,ideal} - \sum_{n=1}^{N} \frac{RA_{n,1} - RA_{n,ideal}}{N}$$

$$I_{n,6} = RA_{n,1} - RA_{n,acc} - \sum_{n=1}^{N} \frac{RA_{n,1} - RA_{n,acc}}{N}$$

$$I_{n,7} = C_{n,1} - C_{n,ideal} - \sum_{n=1}^{N} \frac{C_{n,1} - C_{n,ideal}}{N}$$

$$I_{n,8} = C_{n,1} - C_{n,acc} - \sum_{n=1}^{N} \frac{C_{n,1} - C_{n,acc}}{N}$$

Model structure



Results: model fit

	Observations			
	choices	indicators	par	log-likelihood
RUM	3,680	0	7	-3,401.68
RRM	3,680	0	7	-3,402.59
LC ($K_s = 1, \forall s$)	3,680	0	15	-3,171.61
LC ($K_s = 2, \forall s$)	3,680	0	31	-3,025.64
Hybrid ($K_s = 2, \forall s$)	3,680	2,944	48	-9,533.25

Results: RUM choice model component

	base model			hybrid		
	est	rob. t-rat	$\frac{10\beta.}{\beta_{LF}}$	est	rob. t-rat	$\frac{10\beta.}{\beta_{LF}}$
$\delta_{\text{RUM},1,1}$	0.5345	3.32	-0.47	0.3462	1.91	-0.41
$\delta_{\text{RUM},2,1}$	-0.1014	-0.67	0.09	-0.0264	-0.18	0.03
$\beta_{\text{RUM},TT,1}$	-0.1475	-6.44	0.13	-0.1256	-7.30	0.15
$\beta_{\text{RUM},LF,1}$	-11.4640	-11.82	-	-8.5000	-5.71	-
$\beta_{\text{RUM},RD,1}$	0.0981	0.94	-0.09	-0.0907	-1.07	0.11
$\beta_{\text{RUM},ED,1}$	-0.3699	-4.53	0.32	-0.0852	-1.64	0.10
$\beta_{\text{RUM},C,1}$	-0.2653	-2.85	0.23	-0.1692	-1.91	0.20
$\delta_{\text{RUM},1,2}$	0.4063	1.90	-0.12	-0.3114	-1.30	2.38
$\delta_{\text{RUM},2,2}$	0.5492	1.83	-0.17	0.3820	2.83	-2.92
$\beta_{\text{RUM},TT,2}$	-0.0917	-4.12	0.03	-0.0299	-2.69	0.23
$\beta_{\text{RUM},LF,2}$	-32.9400	-4.05	-	-1.3103	-2.84	-
$\beta_{\text{RUM},RD,2}$	-0.5718	-2.40	0.17	-0.3550	-4.29	2.71
$\beta_{\text{RUM},ED,2}$	0.0253	0.51	-0.01	-0.1736	-3.11	1.32
$\beta_{\text{RUM},C,2}$	-0.5787	-2.05	0.18	-0.3365	-4.07	2.57
$\delta_{\pi,\text{RUM}}$	0	-	-	0	-	-
$\delta_{\varpi,\text{RUM},1}$	0.6274	1.97	-	0.3980	1.30	-
τ_{RUM}	-		-	0	-	-

Results: RRM choice model component

	base model			hybrid		
	est	rob. t-rat	$\frac{10\beta.}{\beta_{LF}}$	est	rob. t-rat	$\frac{10\beta.}{\beta_{LF}}$
$\delta_{RRM,1,1}$	-1.5751	-8.16	5.80	-1.6072	-8.30	5.78
$\delta_{RRM,2,1}$	-0.4866	-2.55	1.79	-0.4481	-2.42	1.61
$\beta_{RRM,TT,1}$	-0.0280	-3.49	0.10	-0.0276	-3.22	0.10
$\beta_{RRM,LF,1}$	-2.7138	-2.76	-	-2.7822	-2.97	-
$\beta_{RRM,RD,1}$	-0.2062	-3.05	0.76	-0.1872	-2.68	0.67
$\beta_{RRM,ED,1}$	-0.0577	-2.49	0.21	-0.0891	-1.91	0.32
$\beta_{RRM,C,1}$	-0.2238	-2.98	0.82	-0.2207	-3.24	0.79
$\delta_{RRM,1,2}$	0.3702	1.09	-3.57	-0.1386	-0.75	0.07
$\delta_{RRM,2,2}$	-0.3633	-2.87	3.50	-0.4404	-1.52	0.22
$\beta_{RRM,TT,2}$	-0.0201	-3.06	0.19	-0.0562	-4.89	0.03
$\beta_{RRM,LF,2}$	-1.0381	-3.47	-	-20.0590	-5.83	-
$\beta_{RRM,RD,2}$	-0.2367	-4.31	2.28	-0.1938	-1.60	0.10
$\beta_{RRM,ED,2}$	-0.0384	-1.40	0.37	-0.0682	-0.46	0.03
$\beta_{RRM,C,2}$	-0.1559	-3.15	1.50	-0.2747	-3.36	0.14
$\delta_{\pi,RRM}$	0.0450	0.27	-	-0.1127	-0.37	-
$\delta_{\varpi,RRM,1}$	0.0703	0.16	-	0.1361	0.49	-
τ_{RRM}	-		-	0.7600	2.57	-

Results: class allocation probabilities

	class allocation probabilities within rules			
	RUM-A	RUM-B	RRM-A	RRM-B
Base model	65.19%	34.81%	51.76%	48.24%
Hybrid 5%	-	-	-	-
Hybrid mean	59.82%	40.18%	53.40%	46.60%
Hybrid 95%	-	-	-	-

	overall					
	RUM-A	RUM-B	RRM-A	RRM-B	RUM	RRM
Base model	31.86%	17.01%	26.46%	24.66%	48.87%	51.13%
Hybrid 5%	12.05%	8.10%	8.95%	7.81%	20.15%	16.77%
Hybrid mean	31.59%	21.22%	25.20%	21.99%	52.82%	47.18%
Hybrid 95%	49.79%	33.44%	42.64%	37.21%	83.23%	79.85%

Results: measurement model

par	est	rob. t-rat
$\zeta_{\text{time ideal}}$	-3.0804	-2.41
$\sigma_{\text{time ideal}}$	15.176	8.71
$\zeta_{\text{time acceptable}}$	-3.7674	-3.15
$\sigma_{\text{time acceptable}}$	13.134	8.48
$\zeta_{\text{log-fare ideal}}$	-0.05814	-2.12
$\sigma_{\text{log-fare ideal}}$	0.39272	10.56
$\zeta_{\text{log-fare acceptable}}$	-0.07216	-2.00
$\sigma_{\text{log-fare acceptable}}$	0.4282	6.70
$\zeta_{\text{crowding ideal}}$	-0.94657	-5.33
$\sigma_{\text{crowding ideal}}$	2.4254	19.96
$\zeta_{\text{crowding acceptable}}$	-0.94933	-5.03
$\sigma_{\text{crowding acceptable}}$	2.2044	19.51
$\zeta_{\text{reliability ideal}}$	-2.2996	-14.80
$\sigma_{\text{reliability ideal}}$	1.0781	12.49
$\zeta_{\text{reliability acceptable}}$	-2.2985	-13.64
$\sigma_{\text{reliability acceptable}}$	0.85885	10.70

Conclusions I

- Results show a link between likely decision rule in SC scenarios and that respondent's stated satisfaction with the real world performance of their current commute journey
- Hypothesis that both outcomes (stated choices and stated satisfaction with real world choices) are influenced by deep rooted character traits
- Findings point towards a link between the tendency for regret minimisation and the effective minimisation of disparity with desired trip features for a respondent's real world commute journey
- Most regret-prone respondents in our sample have, to a larger extent, aligned their reference trip performance to their aspirational values

Conclusions II

- Use of indicators as dependent rather than explanatory variables avoids risk of endogeneity bias, makes model suitable for forecasting and accommodates measurement error in the indicators
- Additionally. the causality link is very clear in our specific context
  - measures relating to satisfaction with the real life commute are meant to relate to the outcome of real world choice processes that are driven by the same character traits as in SC
  - different from assumption that satisfaction with real life commute influences the choice of decision rule in SC
- Also some evidence of higher risk of confounding between heterogeneity in sensitivities and in decision rules in base model
- Reduced in hybrid model as any implied heterogeneity in decision rule also needs to be consistent with the measurement model component

Conclusions III

- A key issue remains the choice of appropriate indicators for the measurement component of the model
- Richer set of indicators opens up possibilities of using multiple latent variables that relate to different character traits
- Also possibilities of linking the choice of decision rules to the values of presented alternatives
- But this moves us away from the notion that the likely decision rule is influenced in particular by underlying character traits
- More work needed on actual implications of results, say in forecasting

Overall conclusions

- Big interest in decision rules in recent years
- Evidence that relative performance varies across datasets or even across people within the same data
- Big role for survey design and especially framing of opt outs
- Major issue remains how to use results from non-RUM models
- Need to remember benefits of RUM and fact that many of the traits might be able to be approximated by RUM

Questions . . .