Can a better model specification avoid the need to move away from random utility maximisation?

Stephane Hess* Matthew Beck† Romain Crastes dit Sourd‡

Abstract

An ever increasing number of applications in choice modelling in transport are looking at moving away from random utility maximisation (RUM) models, lured by the promise of more realistic behaviour in alternative structures. The most prominent example of such a structure in recent years has been the random regret minimisation model (RRM), though there are others. While these alternative structures are behaviourally interesting, researchers seem to at times forget the many reasons why RUM has been the workhorse in choice modelling for several decades, notably its firm grounding in economic theory. The present paper puts forward the idea that a more flexible treatment of heterogeneity in preferences across decision makers may reduce the benefits of moving away from RUM. We illustrate this point on the basis of three typical stated choice datasets from transport studies, offering strong support to our hypotheses.

Keywords: alternative decision rules; heterogeneity; random regret
Word count: $4,001$ *words and 4 tables*

*Institute for Transport Studies and Choice Modelling Centre, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611
†Institute of Transport and Logistics Studies, The University of Sydney, matthew.beck@sydney.edu.au
‡Institute for Transport Studies and Choice Modelling Centre, University of Leeds, R.CrastesditSourd@leeds.ac.uk

1 Introduction

Given that choice modelling is used to inform many policy and expenditure decisions, not only in transport but across a wide range of important economic sectors, it is not surprising that a large body of research has been devoted to understanding how respondents make decisions. This has included a deep engagement with ideas from behavioural economics and mathematical psychology.

Initial efforts went into attempts to accommodate richer behavioural patterns in the specification of utility functions while retaining the framework of random utility maximisation (RUM). Examples of this include incorporating non-linearity into the utility framework [cf. 1], investigating of reference dependence and gains-losses asymmetry [see e.g. 2–4] and models of expected utility theory which use a non-linear utility formation to capture risk attitudes in decisions where there is uncertainty [e.g. 5, 6].

There has also been work which has purposefully sought to move away RUM, with a growing literature seeking to explore differing behavioural paradigms and incorporate them into the modelling of choices. Examples of these include elimination-by-aspects wherein the decision maker eliminates alternatives one at a time based on some criteria or aspects they deem to be important until only one alternative remains [see e.g. 7], satisficing where the decision maker follows a search path acquiring information about alternatives until one is identified that is good enough based on some simple criteria [see e.g. 8], or lexicographic choice where a respondent may rank attributes as being more or less important, but make the choice based solely on the level of the most important attributes [see e.g. 9].

One behavioural paradigm which as captured a great deal of attention in recent years is random regret minimisation, building on the work of Chorus et al. [10] and Chorus [11]. Regret models are based on the notion that an individual is motivated by the desire to avoid a scenario where the chosen alternative is outperformed by one or more non-chosen alternatives on one or more attributes. A now large body of research has compared the performance of regret minimisation to utility maximisation in a range of different contexts, and research has attempted to incorporate a mix of decision rules (RUM, RRM and others) to account for heterogeneity in choice behavior [see e.g. 12, 13]. For an overview of recent RRM applications, see Chorus et al. [14].

The interest in alternative decision rules has without a doubt reinvigorated the diversity of work in choice modelling, and deepened our engagement as a discipline with other fields of behavioural research. However, without wishing to diminish the contributions made by this work, the move away from a RUM environment has at times been taken too lightly. To borrow a notion from prospect theory, one needs to remember that this departure from our RUM-reference may lead to losses as well as gains, and that it is not clear that the latter outweigh the former.

A departure from RUM will often lead to increases in computational complexity, as we again highlight in this paper. To a large extent, this issue can be addressed by ever better estimation approaches and ever more powerful computers. However, what cannot be ignored is that a key reason for the longstanding popularity of RUM is its strong foundation in economic theory and its inherent ability to produce outputs that are of value in economic analysis. This is especially true for welfare analysis - the computation of marginal rates of substitution, such as value of time (VTT) or other willingness to pay (WTP) measures, remain a key interest in choice modelling. Once we move away from the compensatory RUM setting, we quickly face issues in the derivation of such measures, as highlighted in the case of RRM by Dekker [15].

There are of course uses of choice modelling other than valuation, especially forecasting, but these are generally of more interest in large scale applications where the computational complexity of alternatives to RUM can again cause problems. Another argument put forward in the literature on RRM is that non-compensatory or semi-non-compensatory models are better able to deal with behavioural phenomena such as compromise effects [16]. However, the comparison [e.g. in 14] has generally been against a simple linear in attributes specification of a RUM model not accounting for heterogeneity across decision makers, even in the more recent work of van Cranenburgh et al. [17]. Many of the behavioural complexities of decision making can in fact be accommodated within a RUM framework, as demonstrated for example in the case of reference dependence [see e.g. 2–4]. The issue is of course one of determining what these complexities are, and finding an appropriate functional form.

In the present paper, we return to arguments in McFadden and Train [18] that the Mixed Multinomial Logit (MMNL) model is able to approximate any RUM model arbitrarily closely. We exploit this notion to gain further insights into whether a treatment of heterogeneity reduces the "requirement" for a departure from RUM. We are partly motivated to do this following the finding in Hess and Stathopoulos [19], where allowing for heterogeneity in sensitivities to individual attributes reduced the benefit of allowing for heterogeneity in decision rules. Our focus is specifically on the case of RRM given its current popularity and due to the convenient testing framework that arises from a recent development by van Cranenburgh et al. [17].

The remainder of this paper is organised as follows. Section 2 presents our empirical testing framework. This is followed in Section 3 by our empirical application. Section 4 summarises the findings of the paper.

2 Framework for our tests

Our empirical testing makes use of three base models and their random coefficient counterparts.

The base structure is a standard Multinomial Logit (MNL) model, where we have that the utility that respondent n obtains from alternative i (out of J) in choice task t is given by $U_{int} = V_{int} + \varepsilon_{int}$, where V_{int} and ε_{int} are the deterministic and random components of utility, respectively. We specify $V_{int} = \delta_i + \beta x_{int}$ where we normalise the alternative specific constant (ASC) δ_i to 0 for one alternative. The assumption of a type I extreme value distribution for ε_{int} then gives us the typical MNL choice probabilities:

$$P_{\text{MNL},int} = \frac{e^{V_{int}}}{\sum_{j=1}^{J} e^{V_{jnt}}} \tag{1}$$

Our second model is the typical random regret minimisation (RRM) model, where, following Chorus [11], the deterministic regret for alternative i for respondent n in choice task t is given:

$$R_{int} = \delta_{\text{RRM},i,k} + \sum_{k=1}^{K} \sum_{j \neq i} \ln\left(1 + e^{\beta_k\left(x_{jntk} - x_{intk}\right)}\right), \tag{2}$$

where $k = 1, \ldots, K$ is an index across attributes. The regret is informed by all the pairwise comparisons, where the regret for alternative i increases whenever an alternative $j \neq i$ performs better than i on a given attribute. Working again under the assumption of type I extreme value errors, the probability of respondent n choosing alternative i in choice task t is now simply given by the analogue of a MNL formula as:

$$P_{\text{RRM},int} = \frac{e^{-R_{int}}}{\sum_{j=1}^{J} e^{-R_{jnt}}}, \tag{3}$$

where the negative signs relate to minimising rather than maximising regret.

The third model is the μ-RRM model of van Cranenburgh et al. [17]. This new specification allows the variance of the error term (i. e. the scale parameter), μ, to be estimated. We rewrite Equation 2 as:

$$R_{int,\mu} = \delta_{\text{RRM},i,k} + \sum_{k=1}^{K} \sum_{j \neq i} \ln \left(1 + e^{\frac{\beta_k}{\mu} \left(x_{jntk} - x_{intk} \right)} \right), \tag{4}$$

and then rewrite the choice probabilities in Equation 3 as:

$$P_{\mu\text{-RRM},int} = \frac{e^{-\mu R_{int,\mu}}}{\sum_{j=1}^{J} e^{-\mu R_{jnt,\mu}}}, \tag{5}$$

where μ must be positive.

According to van Cranenburgh et al. [17], the strength of the μ-RRM model is its ability to estimate a profundity of regret. In practice, this allows the model to accommodate different behavioural processes, depending of the value taken by μ, and this is our motivation for applying the model in our example. We also feel that profundity of regret is but one of the interpretations for the advantages of the model - the fact that the model allows for a different degree of non-linearity compared to a standard RRM model clearly gives it another advantage (and a reason to outperform a simple MNL model, as shown in [17]).

van Cranenburgh et al. [17] list three cases for μ and the profundity of regret:

- When μ is equal to one (or insignificantly different from one), the model corresponds to a classical RRM;
- When μ is arbitrarily large, the attribute level regret functions are linear (or very close to linear) and $\frac{\beta}{\mu}$ is very small. In such cases, the difference between the utility one gets from a gain and the regret one gets from a loss is close to zero. When μ tends to infinity, the μ-RRM model collapses to linear RUM behaviour; and
- When μ is arbitrarily small, the difference between the utility one gets from a gain and the regret one gets from a loss is very strong. In this case, the μ-RRM model takes the form of the P-RRM model [cf. 17]

Alongside these three base specifications, we estimate their mixed counter-parts, where we allow for random variation across individual respondents (assuming within-respondent homogeneity) in the values of β. Let $P_{j_{nt}}(\beta)$ give the probability of the observed choice for respondent n in task t (j_{nt}), conditional on β. We assume that β follows a random distribution across respondents, with $\beta \sim f(\beta \mid \Omega)$, where Ω is an estimated vector of parameters. We then have that the probability of the sequence of T choices for respondent n, conditional on Ω is given by:

$$L_n(\Omega) = \int_{\beta} \prod_{t=1}^{T} P_{j_{nt}}(\beta) f(\beta \mid \Omega) \, \mathrm{d}\beta, \tag{6}$$

where $P_{j_{nt}}(\beta)$ can take the form from Equation 1, 3 or 5. For the first structure, this thus leads to a Mixed Multinomial Logit (MMNL) model, while for the other two, we obtain mixed counterparts to a RRM and a μ-RRM model. From McFadden and Train [18], we know that the most flexible specification of a MMNL model arises when we allow for all elements of β to be random, and with a full covariance matrix estimated between them. In the present work, we limit ourselves to using fixed ASCs across respondents, along with univariate distributions for individual elements in β. With the aim of providing an illustrative application, we also rely on Normal distributions only as we do not seek to reach policy conclusions or estimated WTP measures.

3 Empirical application

We now discuss the data used in our empirical example before presenting the results. All models were coded in R [20], using 250 MLHS draws in estimation [21], with the same seeds used across models.

3.1 Data

We use three typical transport stated choice (SC) datasets in our study. The first two datasets come from Australian toll road studies. Both datasets present each respondent with 16 choice tasks involving a choice between an invariant reference trip and two hypothetical alternatives, described by free flow time, slowed down time, crawl time (only in the second dataset), travel time variability, running costs and tolls. The first dataset includes data from 243 respondents, while the second dataset includes data from 304 respondents. The two datasets are described in more detail in Hess et al. [22] and the references therein.

The third dataset was collected from a sample of 368 public transport commuters in the UK, who were presented with 10 choice tasks involving a choice between an invariant reference trip and two hypothetical alternatives, described by travel time, fare, the rate of crowded trips (out of ten trips), the rate of delays (out of ten trips), the average length of delays (across delayed trips), and the provision of a delay information service with different pricing. The last of these attributes proved not to have a statistically significant impact, while we allowed the length of delays to enter the model as an interaction with the rate of delays, leading to the expected average delay. For more details on this dataset, see Hess and Stathopoulos [19].

3.2 Results for first Australian dataset

The results for the first Australian dataset are presented in Table 1. Along with overall model fit statistics and individual parameter estimates, we show a normalised estimation time, which is set to 1 for the base MNL model. This highlights a jump in estimation time when moving from RUM to RRM and especially to μ-RRM, where this is even more substantial in the versions with random coefficients.

In this first dataset, the coefficients in the base RUM and RRM models have the expected signs and are statistically significant except for β_{VAR}. The MNL model obtains a better fit than the RRM model, and this finding is confirmed by the μ-RRM model where the large value for μ is moving us towards a RUM model. The inclusion of random coefficients leads to very substantial gains in model fit for all three structures, as would be expected, and we see high levels of statistical significance for the standard deviation terms which capture the heterogeneity across respondents. What is most relevant for the present paper is that the difference in model fit between the RUM and RRM models actually increases when we incorporate random heterogeneity, and albeit that the μ-RRM model now shows a slightly better model fit than the simple MMNL model, the estimate of μ increases further compared to the closed form model, again pointing towards a RUM specification.

3.3 Results for second Australian dataset

The results for the first Australian dataset are presented in Table 2. The increases in estimation times we observe when moving from RUM to RRM and μ-RRM are in line with those from the first Australian dataset, as are the increases from including random heterogeneity. The coefficients in the base RUM and RRM models again all have the expected signs and are statistically significant, except for β_{VAR} which is now also of the wrong sign, highlighting issues with respondent understanding of this attribute. The MNL model again obtains a better fit than the RRM model, and this finding is confirmed by the μ-RRM model where the large value for μ is moving us towards a RUM model.

The gains in fit from the inclusion of random coefficients are somewhat smaller than in the first case study, but remain substantial, and we see high levels of statistical significance for the standard deviation terms. The move towards a model with random coefficients slightly decreases the difference between the RUM and RRM models, while we now see a more sizable benefit for the mixed μ-RRM model. However, the estimate for μ is still very large, pointing towards a RUM structure.

3.4 Results for UK dataset

The results for the UK dataset are presented in Table 3. The increases in estimation times we observe when moving from RUM to RRM and μ-RRM are a little smaller than in the other two case studies, but remain very sizable, especially when including random heterogeneity.

With the UK data, the signs for all marginal utility coefficients are in line with expectations for the simple RUM and RRM models, and the estimates are all statistically significant. With this dataset, we see a final log-likelihood that is better by 22 units for RRM compared to RUM, and the mu-RRM model points towards a value of μ less than 1. On face value, an analyst may use these results to highlight the benefits of a RRM structure. However, the picture changes substantially when moving to the random coefficients versions of the models. The earlier advantage of RRM over RUM is now reversed, and in the mixed versions, the RUM structure obtains a log-likelihood that is 20 units better than that of the RRM counterpart. The mixed version of the μ-RRM model obtains a better fit still, but the value of μ is now larger than 1, moving us away from RRM. While the higher fit for the mixed μ-RRM model compared to the MMNL model may be interpreted as evidence that a regret specification of some form still offers benefit, we would caution against such an interpretation and suggest that simple non-linearity may be at play here.

Some additional work was carried out with the UK data, allowing for random heterogeneity in the μ parameter, where we used a symmetrical Triangular distribution, with lower bound μ_A and a range of μ_B. We observe that for both the model with fixed β and the model with random β, there is no improvement in fit over the models with fixed μ, suggesting that, at least with this data, there is no benefit in allowing for a random μ parameter. To some extent, this is not surprising. A random μ term would move us to a continuous analogue of the behavioural mixing model of Hess et al. [12], however one that keeps the β coefficients the same across model forms. With the large differences we would see between coefficients for RUM and RRM, it is not a complete surprise that such a model performs poorly.

4 Conclusions

The work in this paper was motivated by the simple question whether the differences between models using different behavioural paradigms might change when an analyst allows for heterogeneity across respondents in the parameters of the individual models. This was partly motivated by the findings in Hess and Stathopoulos [19] which showed a reduction in the weight for RRM classes in a RUM-RRM mixture model when allowing for heterogeneity in sensitivities within the individual RUM and RRM models. Additionally, we have noted that the majority of the comparisons between models with different behavioural paradigms in the literature assume homogeneity in model parameters across respondents.

As a first observation, we note the substantial increase in estimation times for RRM structures compared to RUM, even with datasets with only three alternatives, and thus two pairwise comparisons per attribute and alternative. The computational costs of these comparisons becomes much larger in the random coefficients specifications of the models, and this likely plays a role in the fact that we have seen no major work with random coefficients versions of RRM. In real world applications, were the number of alternatives is much larger, this increase in computational cost would be formidable.

Returning to the key interest of our paper, we see that, for two out of the three datasets, the RUM model benefits more than the RRM model from the inclusion of random coefficients, while, in the second Australian dataset, the RRM model benefits slightly more. The most striking result comes in the UK data, where the closed form results would suggest that RRM is the more appropriate structure for the data, while this is reversed when moving towards a random coefficients structure.

The evidence in this paper is of course limited to only three datasets and different findings might arise with other data sources. It should also be noted that we have relied on univariate Normal distributions alone, only going someway to the full flexibility discussed by McFadden and Train [18]. The premise that a fully specified MMNL model can approximate any RUM structure arbitrarily closely does of course not on theoretical grounds extend to non-RUM structures, but it seems a reasonable assumption that with a more flexible specification of heterogeneity, the reasons for moving away from RUM may be reduced further. Our work has also relied on a simple linear in attributes specification, and the incorporation of non-linearity in a RUM structure offers further avenues for improvement. In closing, we feel that regret, along with other non-compensatory processes, may be useful if the goal is to gain alternative insight into behaviour, but if the analyst seeks to model choices for the purposes of valuation or forecasting, then a well specified RUM model will likely be able approximate any improvement in model fit resulting from moving to other behavioural paradigms, while maintaining its many desirable properties.

5 Acknowledgements

The authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS.

References

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

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

[3] B. De Borger and M. Fosgerau. 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, 2008.

[4] Amanda Stathopoulos and Stephane Hess. Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. *Transportation Research Part A: Policy and Practice*, 46(10):1673–1689, dec 2012. doi: 10.1016/j.tra.2012.08.005.

[5] Yaron Hollander. Direct versus indirect models for the effects of unreliability. *Transportation Research Part A: Policy and Practice*, 40(9):699–711, nov 2006. doi: 10.1016/j.tra.2005.12.007.

[6] David A. Hensher, William H. Greene, and Zheng Li. Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings. *Transportation Research Part B: Methodological*, 45(7):954–972, aug 2011. doi: 10.1016/j.trb.2011.05.023.

[7] R. P. Batley and A. Daly. On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. *Journal of Mathematical Psychology*, 50(5):456–467, 2006.

[8] Elena Reutskaja, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel. Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926, apr 2011. doi: 10.1257/aer.101.2.900.

[9] Marit Killi, Åse Nossum, and Knut Veisten. Lexicographic answering in travel choice: Insufficient scale extensions and steep indifference curves? *European Journal of Transport and Infrastructure Research*, 7(1):39–62, 2007.

[10] C.G. Chorus, T.A. Arentze, and H.J.P. Timmermans. A random regret minimization model of travel choice. *Transportation Research Part B*, 42(1):1–18, 2008.

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

[12] S. Hess, A. Stathopoulos, and A. J. Daly. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591, 2012.

[13] S. Hess and A. Stathopoulos. Linking the decision process to underlying attitudes and perceptions: a latent variable latent class construct. *paper presented at the* 13^{th} *International Conference on Travel Behaviour Research, Toronto*, 2012.

[14] Caspar Chorus, Sander van Cranenburgh, and Thijs Dekker. Random regret minimization for consumer choice modeling: Assessment of empirical evidence. *Journal of Business Research*, 67(11):2428–2436, nov 2014. doi: 10.1016/j.jbusres.2014.02.010.

[15] Thijs Dekker. Indifference based value of time measures for random regret minimisation models. *Journal of Choice Modelling*, 12:10–20, sep 2014.

[16] Caspar G. Chorus and Michel Bierlaire. An empirical comparison of travel choice models that capture preferences for compromise alternatives. *Transportation*, 40(3):549–562, nov 2012. doi: 10.1007/s11116-012-9444-3.

[17] Sander van Cranenburgh, Cristian Angelo Guevara, and Caspar G. Chorus. New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice*, 74: 91–109, apr 2015. doi: 10.1016/j.tra.2015.01.008.

[18] D. McFadden and K. Train. Mixed MNL Models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000.

[19] Stephane Hess and Amanda Stathopoulos. A mixed random utility — random regret model linking the choice of decision rule to latent character traits. *Journal of Choice Modelling*, 9: 27–38, dec 2013. doi: 10.1016/j.jocm.2013.12.005.

[20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

[21] S. Hess, K. Train, and J. W. Polak. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice. *Transportation Research Part B*, 40(2):147–163, 2006.

[22] Stephane Hess, David Hensher, and Andrew Daly. Not bored yet: revisiting respondent fatigue in stated choice surveys. *Transportation Research Part A*, 46(3):626–644, 2012.

Table 1: Results for first Australian dataset

	MNL		**RRM**		**μ-RRM**	
$LL(\beta)$	-3,027.66		-3,039.18		-3,027.94	
pars.	7		7		8	
BIC	6,113.18		6,136.22		6,122.01	
Runtime (normalised)	1.00		2.22		5.79	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.2744	1.82	-0.2155	-1.48	-0.0261	-1.17
δ_2	0.0765	1.33	-0.0650	-1.13	-0.0073	-1.00
β_{FFT}	-0.0688	-9.24	-0.0471	-9.05	-0.0460	-9.23
β_{SDT}	-0.0909	-15.78	-0.0636	-15.27	-0.0610	-15.97
β_{VAR}	-0.0058	-1.06	-0.0045	-1.32	-0.0039	-1.08
β_{RC}	-0.3135	-10.80	-0.2081	-10.94	-0.2093	-10.81
β_{TOLL}	-0.3614	-13.86	-0.2556	-13.33	-0.2427	-13.85
μ	-		-		10.2881	1.57

	mixed RUM		**mixed RRM**		**mixed μ-RRM**	
$LL(\beta)$	-2,368.10		-2,392.83		2,365.25	
pars.	12		12		13	
BIC	4,835.39		4,884.84		4,837.96	
Runtime (normalised)	50.56		326.83		531.11	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.3326	1.44	-0.3799	-1.63	-0.0100	-1.25
δ_2	0.1029	1.15	-0.0730	-0.83	-0.0021	-0.95
μ_{FFT}	-0.1082	-7.20	-0.0785	-8.99	-0.0791	-10.27
μ_{SDT}	-0.1478	-13.07	-0.1088	-12.25	-0.0984	-12.64
μ_{VAR}	-0.0096	-0.99	-0.0169	-1.58	-0.0143	-2.00
μ_{RC}	-0.5849	-11.89	-0.3530	-12.10	-0.3652	-12.52
μ_{TOLL}	-0.6186	-15.53	-0.4704	-13.83	-0.4275	-14.94
σ_{FFT}	0.1272	9.89	0.0918	9.15	0.0842	9.06
σ_{SDT}	0.0925	10.42	0.0725	9.88	0.0646	8.86
σ_{VAR}	0.0933	10.18	0.0597	6.82	0.0624	8.78
σ_{RC}	0.3147	6.06	0.2070	5.85	0.2223	6.41
σ_{TOLL}	0.4582	10.31	0.3012	9.53	0.2754	10.25
μ	-		-		46.7409	1.75

Table 2: Results for second Australian dataset

	MNL		**RRM**		**μ-RRM**	
$LL(\beta)$	-2,668.39		-2,681.47		-2,668.72	
pars.	8		8		9	
BIC	5,404.69		5,430.85		5,413.86	
Runtime (normalised)	1.00		1.87		5.74	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.3886	3.50	-0.3954	-3.64	-0.0350	-1.28
δ_2	-0.0521	-0.88	0.0477	0.81	0.0048	0.77
β_{FFT}	-0.0797	-4.26	-0.0530	-4.36	-0.0533	-4.27
β_{SDT}	-0.1174	-9.34	-0.0761	-8.98	-0.0781	-9.26
β_{CT}	-0.1672	-11.13	-0.1105	-11.08	-0.1119	-11.13
β_{VAR}	0.0312	1.08	0.0242	1.27	0.0212	1.09
β_{RC}	-0.5618	-6.83	-0.4125	-7.28	-0.3777	-6.89
β_{TOLL}	-0.4997	-12.01	-0.3356	-11.95	-0.3344	-12.00
μ	-		-		11.0192	1.45

	mixed RUM		**mixed RRM**		**mixed μ-RRM**	
$LL(\beta)$	-2,202.87		-2,214.42		-2,177.07	
pars.	14		14		15	
BIC	4,524.60		4,547.70		4,481.48	
Runtime (normalised)	53.18		338.04		617.19	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.1423	1.07	-0.1408	-1.08	-0.0103	-0.57
δ_2	-0.0559	-0.64	0.0266	0.30	0.0062	0.54
μ_{FFT}	-0.1524	-6.12	-0.1184	-5.71	-0.1188	-6.91
μ_{SDT}	-0.1913	-11.56	-0.1229	-10.04	-0.1284	-8.80
μ_{CT}	-0.2411	-11.40	-0.1626	-9.95	-0.1955	-11.33
μ_{VAR}	0.0086	0.27	0.0171	0.81	-0.0240	-0.91
μ_{RC}	-0.8230	-7.19	-0.5919	-7.98	-0.5797	-8.84
μ_{TOLL}	-1.1298	-12.59	-0.7684	-10.32	-0.8059	-12.22
σ_{FFT}	0.2447	10.08	0.1752	9.72	0.1799	8.49
σ_{SDT}	0.1165	8.00	0.0713	5.77	0.0539	4.26
σ_{CT}	0.1966	7.76	0.1476	7.10	0.1356	7.57
σ_{VAR}	0.3078	7.31	0.1842	5.63	0.1829	6.31
σ_{RC}	0.5765	3.19	0.3516	2.25	0.5443	6.34
σ_{TOLL}	0.7822	13.44	0.5716	10.02	0.4855	10.20
μ	-		-		7.9383	1.83

Table 3: Results for UK dataset

	MNL		**RRM**		**μ-RRM**	
$LL(\beta)$	-3,721.67		-3,699.49		-3,698.89	
pars.	7		7		8	
BIC	7,500.81		7,456.46		7,463.47	
Runtime (normalised)	1.00		1.58		2.11	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.1978	3.25	-0.1625	-2.68	-0.2083	-2.09
δ_2	0.1395	3.13	-0.1440	-3.23	-0.2080	-2.26
β_{TT}	-0.0376	-7.98	-0.0263	-8.92	-0.0263	-8.97
β_{FARE}	-0.9873	-4.39	-0.7636	-4.42	-0.7670	-4.43
β_{REL}	-0.1224	-4.45	-0.0819	-4.40	-0.0814	-4.42
$\beta_{EXPDELAY}$	-0.0796	-4.50	-0.0561	-4.47	-0.0560	-4.51
β_{CROWD}	-0.1698	-7.34	-0.1148	-7.35	-0.1139	-7.33
μ	-		-		0.6924	3.10

	mixed RUM		**mixed RRM**		**mixed μ-RRM**	
$LL(\beta)$	-3,184.89		-3,205.27		-3,174.96	
pars.	12		12		13	
BIC	6,468.30		6,509.08		6,456.66	
Runtime (normalised)	50.75		316.98		335.69	
	est	rob. t-rat.	est	rob. t-rat.	est	rob. t-rat.
δ_1	0.7403	9.58	-0.4569	-6.13	-0.2085	-2.68
δ_2	0.3045	4.30	-0.3147	-4.46	-0.0989	-1.98
μ_{TT}	-0.0836	-10.73	-0.0566	-10.79	-0.0582	-10.42
μ_{FARE}	-5.9438	-14.64	-3.2270	-11.20	-3.6589	-11.68
μ_{REL}	-0.2647	-5.05	-0.1918	-5.65	-0.1773	-5.07
$\mu_{EXPDELAY}$	-0.1790	-4.24	-0.1242	-5.03	-0.1408	-4.41
μ_{CROWD}	-0.3585	-8.06	-0.2458	-7.69	-0.2764	-7.86
σ_{TT}	0.0658	6.63	0.0446	5.55	0.0497	6.53
σ_{FARE}	4.7719	14.08	4.3979	8.54	4.1441	12.36
σ_{REL}	0.4363	8.01	0.3022	7.47	0.3093	8.25
$\sigma_{EXPDELAY}$	0.1089	2.73	0.0933	3.30	0.0979	4.15
σ_{CROWD}	0.4917	8.37	0.3188	7.86	0.3348	7.67
μ	-		-		3.1552	2.23

Table 4: Results for UK dataset with random μ

	μ-RRM		**mixed μ-RRM**	
$LL(\beta)$	-3,698.89		-3,174.92	
pars.	9		14	
BIC	7,471.68		6,464.79	
Runtime (normalised)	249.07		429.66	
	est	rob. t-rat.	est	rob. t-rat.
δ_1	-0.2063	-1.76	-0.2059	-2.61
δ_2	-0.2059	-1.83	-0.0974	-1.94
μ_{TT}	-0.0263	-8.89	-0.0582	-10.41
μ_{FARE}	-0.7668	-4.39	-3.6591	-11.67
μ_{REL}	-0.0814	-4.42	-0.1774	-5.08
$\mu_{EXPDELAY}$	-0.0560	-4.51	-0.1408	-4.41
μ_{CROWD}	-0.1139	-7.33	-0.2764	-7.86
σ_{TT}			0.0497	6.53
σ_{FARE}			4.1443	12.35
σ_{REL}			0.3093	8.25
$\sigma_{EXPDELAY}$			0.0980	4.15
σ_{CROWD}			0.3349	7.67
μ_A	0.5262	0.39	3.1503	2.16
μ_B	0.3472	0.11	0.1106	1.48