Towards more behaviourally robust travel demand forecasts:
Catering to utility maximisers and regret minimisers

Thijs Dekker*, Sander van Cranenburgh, Caspar G. chorus
Transport and Logistics Group, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, Netherlands. E: t.dekker@tudelft.nl; s.vancranenburgh@tudelft.nl; c.g.chorus@tudelft.nl
T: +31 15 278 2015

* Corresponding author

Version: 30/08/2013

ABSTRACT

Choice probabilities and related outputs of discrete choice models form a critical input to many travel demand forecasting and transport project evaluation studies. The decision rule underlying a discrete choice model describes how individuals make their decisions and thereby co-determines the choice probabilities. Uncertainty from the side of the analyst regarding the underlying decision rule(s) may therefore translate into alternative predictions regarding the behavioural response to changing travel conditions. In this paper, we contrast the well-known Random Utility Maximization framework, on which most travel demand forecasts are based, with its Random Regret Minimization counterpart. Based on a review of the existing empirical comparisons between the two frameworks we discuss the connections and dissimilarities between both model types and the associated implications for travel demand forecasting. The empirical comparisons reveal that both models perform about equally well in terms of model fit and external validation, which makes it hard to identify one model as a superior specification for forecasting. Despite these small differences in overall model fit, choice probabilities and elasticities can differ substantially (and predictably) in specific choice-contexts. One such example is the compromise effect where the Random Regret Minimization framework predicts a market share bonus for 'in-between' alternatives. The paper discusses model averaging techniques to generate predictions when a clear winning model cannot be identified. Finally, the paper puts these considerations in the context of a regret-based Dutch National Model, which is currently under construction.

1. INTRODUCTION

It is generally acknowledged that uncertainty in travel demand forecasts is one of the most prominent causes underlying uncertainties in transport infrastructure evaluation (e.g. Rasouli and Timmermans, 2012). Forecasting travel behaviour is intrinsically linked to the extent to which travellers respond to a changing environment as represented by travel times, costs and other related travel conditions. For example, the Dutch National Model is a marginal model predicting changes in travel behaviour relative to a baseline scenario (e.g., Daly & Sillaparcharn, 2008; Willigers & de Bok, 2009). Like many other macro-models, the model is based on the Random Utility Maximization (RUM) framework (McFadden, 1974) and assumes travellers maximise their underlying utility functions and select e.g. the mode or destination associated with the highest level of overall utility. The functional form of the utility function has direct implications for the extent to which respondents are willing to make trade-offs between e.g. travel time and travel costs (or other travel attributes) and simultaneously their inclination to alter their mode or route choice under changing conditions.

Transportation modellers and policy analysts deal with inherent uncertainties underlying descriptive models of travel behaviour by performing sensitivity analyses using, for example, alternative values for the implicit trade-offs (e.g., values of time) across attributes. Similarly, demographic trends can also be varied systematically in sample enumeration procedures to provide additional insight into the likely ranges (confidence intervals) around best estimates of relevant model outcomes, such as market shares under varying policy packages. In this paper, we point out that sensitivity tests so far have ignored the possibility of variations in the underlying decision rule used by travellers when making decisions. Hess et al. (2012) point out that the applied decision rule is associated with a similar degree of uncertainty to the analyst, while significant evidence of the presence of heterogeneous decision heuristics has been found within the transport literature (Leong and Henser, 2012; Chorus, 2013).

In this paper, we specifically focus on the Random Regret Minimization (RRM) framework as an alternative decision heuristic (Chorus, 2010). Conceptually, RRM's decision-rule differs from its RUM-counterpart by postulating that choices are determined by the desire to minimize anticipated regret. Within the RRM model, regret is conceptualized as the emotion that is felt when an alternative performs worse than another alternative on a specific attribute. The fact that regret depends on the performance (at the attribute level) relative to other available alternatives introduces a context dependency not present in the conventional RUM framework which relies on linear-in-parameters specifications of utility functions. Where this linear-in-parameters RUM model assumes fully compensatory behaviour across attributes, the RRM model is associated with a specific form of semi-compensatory behaviour. An important implication of this property is that the RRM-model predicts a market share bonus for so-called compromise alternatives with an 'in-between' performance on all attributes (e.g., Chorus and Bierlaire, 2013).

Chorus et al. (2013a) point out that, at the aggregate level and for most datasets, the RUM and RRM model appear to perform similarly in terms of model fit and predictive ability on hold out data. Interestingly enough, it is found that – notwithstanding the small differences between the two paradigms in terms of model fit and predictive ability – RUM and RRM do have the potential to generate markedly different predictions regarding relative attribute importance, willingness to pay, elasticities, and market share forecasts. Hensher et al. (2011) is one of several empirical examples confirming that the differences in functional form between the RUM and RRM model translate into different behavioural implications such as elasticities. It is exactly those elasticities and associated choice probabilities that determine the responsiveness of individuals to changing circumstances, and they explain why specific segments of the population make alternative decisions. Moreover, this may cause diverging patterns in forecasting methods. Accordingly, the key question is how RRM models can be put to use effectively for supporting travel demand forecasts.

We will argue that – given the small differences between the two model types in terms of model fit and predictive ability – one fruitful approach to use RRM for supporting travel demand forecasts would be to simultaneously employ the two choice model types (RUM and RRM/Hybrid RUM-RRM) for the analysis and prediction of choice behaviour. The outcomes of the different models (in terms of, for example, elasticities and/or market share forecasts) may then be used to obtain what may be called 'behavioural confidence intervals' and/or to perform what may be called 'behavioural sensitivity analyses'. That is, to the extent that RUM and RRM generate different outcomes, it makes sense to consider each model type (and associated outcome) a possible scenario. Confronted with these different behavioural scenarios, policy-makers and planners may then apply conventional techniques for dealing with multiple scenarios with the aim of developing 'behaviourally robust' policies i.e., policies that are likely to turn out effective, irrespective of which behavioural scenario (RUM or RRM/Hybrid RUM-RRM) in the end turns out to be the most correct one. An alternative approach based on model averaging techniques (Hoeting et al. 2007; Wagenmakers and Farrell, 2004) is also suggested. Such approaches allow for increasing the robustness of predictions while taking uncertainties about which is the correct model into account.

This paper starts by introducing both the linear-in-parameters RUM model and the RRM model (Section 2). Section 3 provides an overview of recently performed empirical comparisons between the two models in terms of model fit, predictive ability on hold out data, and policy-relevant output. Observed differences and similarities are further illustrated using simulated data. Section 4 then pays attention to potential implications for forecasting exercises when both the RUM and RRM model provide a roughly similar fit, but generate alternative predictions. Section 5 provides an update of a project that is being undertaken by the authors, in collaboration with Significance consultancy and the Dutch Ministry of Infrastructure and the Environment: the aim of the project is to build an RRM-based counterpart of the existing RUM-based Dutch National Model and to use the outcomes of the two National Models jointly as input for a behavioural sensitivity analysis along the lines suggested above. Section 6 concludes and discusses potentially fruitful avenues for further research.

2. ALTERNATIVE DECISION HEURISTICS: RUM VS. RRM

Discrete choice models form a key element in many travel demand forecasting models. Forecasts can be based on applications of the full model, where sample enumeration methods (see Ben-Akiva & Lerman, 1985 or Train 2009) are applied to derive choice probabilities over the available alternatives for specific groups of individuals and combined with weights to forecast aggregate market shares. Such applications can be present in the trip generation stage, but also in the stages where the choice for mode of transport, route and time-of-day are modelled. More generally, Daly (2013) describes the expected demand Q_i for alternative i by Equation (1) where w_k represents the number of individuals of type k, $P_{ik}(x_k)$ the probability of choosing alternative i given x_k, $q_{ik}(x_k)$ the quantity demanded when i is chosen. Note that the latter is generally normalized to unity in discrete choice models. Finally, x_k summarizes the characteristics of the individuals of type k, but also incorporates product characteristics of the available alternatives as perceived by the individual.

$$Q_i = \sum_k w_k P_{ik}\left(x_k\right) q_{ik}\left(x_k\right) \tag{1}$$

Other applications of discrete choice models in forecasting are based on the parameter outcomes of the models and rely more on the willingness-to-pay estimates or elasticities as behavioural output of the choice model. Overall, the dependence of these forecasting methods on the underlying choice probabilities make them heavily dependent on assumptions from the side of the analyst concerning the applied decision rule and associated mathematical model of behaviour. These assumptions will be the focus of this paper.

Equation (2) describes the standard logit choice probability associated with the Random Utility Maximization (RUM) framework, where P_{ik} represents the probability that type k will select alternative i when presented with alternatives $j=1,\ldots, J$.

$$P_{ik} = \frac{\exp\left(V_{ik}\right)}{\sum_{j=1}^{J} \exp\left(V_{jk}\right)} \tag{2}$$

Deterministic utility V_{ik} is commonly assumed to be a linear function of characteristics x_{ik} describing the alternative, i.e. $V_{ik} = x_{ik}\beta$. The linear specification implies that individuals are willing to make trade-offs between any two characteristics at a constant rate, as represented by the ratio of marginal utilies - i.e the β's. The Random Regret Minimization (RRM) model works with choice probabilities of a similar form (see Equation (3)), but the deterministic regret function R_{ik} exhibits several essential differences from V_{ik}.

$$P_{ik} = \frac{\exp\left(-R_{ik}\right)}{\sum_{j=1}^{J} \exp\left(-R_{jk}\right)} \tag{3}$$

Where in the RUM model utility only depends on the characteristics of alternative i, regret R_{ik} depends on the performance of the alternative on each characteristic m relative to the performance of other alternatives on that same characteristic.[1] The deterministic regret function is described in Equation (4), where regret arises when alternative i performs worse than alternative j on characteristic m. The regret of alternative i is increasing with the number of attributes on which alternative i is outperformed, the importance of those attributes (as denoted by the associated parameter), and with the number of alternatives by which alternative i is outperformed (as denoted by the summation over $j \neq i$). This introduces a clear context or choice set dependency not present in the linear-in-parameters RUM model.

$$R_{ik} = \sum_{j \neq i} \sum_{m} \ln\left(1 + \exp\left(\beta_m \left(x_{jm} - x_{im}\right)\right)\right) \tag{4}$$

The functional form also implies that performing increasingly better on an attribute (relative to the competition) does not lead to constant decreases in regret of alternative i. In fact, marginal binary regret rapidly approaches zero when $(x_{jm}-x_{im})<0$ and converges to β_m when $(x_{jm}-x_{im})$ is sufficiently large (when $\beta_m>0$) (Chorus, 2010).[2] In other words, when an alternative already has a strong performance on a particular attribute, relative to a competing alternative, then further improving its performance on that attribute leads to (very) small decreases in regret. However, when that alternative already has a poor performance on the respective attribute, then further deteriorating its performance on that attribute leads to large increases in regret. Contrary to the linear-additive RUM model, the implied marginal rates of substitution across attributes are therefore also no longer constant. They depend to a large extent on the composition of the choice set and imply a particular form of semi-compensatory behaviour in the sense that further deterioration of an attribute on which the alternative already performs poorly is very difficult to compensate by means of an improvement of an attribute on which the alternative already performs strongly. Willingness to pay for an improvement of a quality-

[1] It is important to note that the independence of irrelevant alternatives (IIA) axiom no longer holds in the RRM framework, also when – as is the case in Logit-specifications – error components are i.i.d., since attribute levels of all other alternatives enter the regret function of alternative *i*.
[2] Note that the subscripts *k* is removed from $(x_{jm}-x_{im})$ for notational convenience.

attribute therefore crucially depends on the performance of the alternative *and* its competition in terms of price and the quality-attribute relative to other alternatives in the choice set (Chorus, 2012a).

These differences in the marginal rate of substitution between the RUM and RRM model have direct implications for the distribution of choice probabilities across alternatives and associated elasticities. For example, RRM-models predict that it is more effective (in terms of avoiding regret) to select a compromise alternative with an intermediate performance on most or all attributes. Even when such a compromise alternative fails to have a strong performance on any of the attributes (relative to the other alternatives in the set), RRM-models predict that it will still only generate modest levels of regret as long as it does not have a particularly poor performance on any of the attributes. This is a direct consequence of the model property that good characteristics do not compensate for the large amounts of regret associated with bad characteristics (Chorus and Bierlaire, 2013). Hensher et al. (2011) derive the elasticities for the RRM model, which are remarkably different from those implied by the RUM model (see Train, 2009, pp. 59).[3]

Similar to accounting for heterogeneity in preferences across individuals, it is unlikely that all individuals base their decisions on the same decision rule. Intermediate specifications of the RUM and RRM model have been developed recently, i.e. so-called Hybrid RUM-RRM models, but these only account for variations in decision rules at the level of travel characteristics (e.g. Chorus et al. 2013c). That is, some characteristics are treated in the standard RUM approach, while others are treated based on the RRM approach. Overall, a uniform decision rule is still imposed across respondents in these hybrid models. Hess et al. (2012) is one of the few papers allowing for heterogeneity in the decision rule using a latent class approach. Due to the limited set of empirical studies allowing for such heterogeneity and given that Hess et al. (2012) use an older RRM-specification which differs from the one we focus on in this paper, we will only be able to compare the empirical performance of RRM, RUM and hybrid RUM-RRM at the sample level and not across individuals. In the remaining sections, we are particularly interested in the predictions of each of these overall models in terms of choice probabilities and elasticities associated with specific choice tasks.

3. EMPIRICAL PERFORMANCE OF THE RRM AND RUM MODEL

Since its recent introduction, the Random Regret Minimization (RRM) approach to discrete (travel) choice modelling has gained attention among a yet small but growing group of choice modellers, leading to a growing body of theoretical and empirical literature on RRM. The majority of empirical RRM-studies feature comparisons with linear-additive RUM models in the context of one or more stated or revealed choice datasets (e.g., Boeri et al., 2013; Beck et al., 2013; Hensher et al., 2011; Kaplan and Prato, 2012; Thiene et al., 2012). In Chorus et al. (2013a), we summarize a set of 19 journal papers which empirically contrasted the RUM and RRM model. These studies are co-authored

[3] Note that in a binary choice setting, the RUM and RRM model generate equivalent choice probabilities and therefore also embody equivalent elasticities and welfare measures.

by a total of 28 scholars and published in fields such as transportation, urban planning, environmental economics, and health economics. Most studies are associated with multiple empirical comparisons translating in to an overall number of 33 datasets for which the RUM and RRM model have been contrasted in the international peer-reviewed literature. Comparisons between RRM and RUM are generally made in terms of dimensions as diverse as model fit, predictive ability, willingness to pay, elasticities, and (or) market share forecasts.

3.1 Model fit and external validity

Chorus et al. (2013a) stipulate that in terms of model fit the RRM model or a hybrid RRM/RUM generally performs better than the conventional RUM model, but that differences are generally small (although statistically significant). Similar results are obtained when contrasting the external validity of both models. Results are somewhat ambiguous, however, since some studies report a higher model fit for one of the two models, in combination with a worse out-of-sample predictive ability. These results are, however, not surprising due to the close connection between the RUM and RRM model. First, both models are usually estimated in the same econometric form (i.e., MNL or Mixed Logit specification). Second, they have the same number of degrees of freedom, i.e. the same number of parameters describing the same set of characteristics, which are applied to exactly the same data. Third, in a binary choice setting both models provide identical model outcomes. Chorus (2010) established that utility differences in this case are identical to the regret differences between the two available alternatives. The RUM and RRM model start to diverge when additional alternatives are added to the choice sets. In the RUM model, utility levels of the current set of alternatives are not affected by adding new alternatives to the set. In the RRM model, regret changes when the choice set composition changes (including the situation where new alternatives are added to the set), as is clearly seen upon inspecting Equation (4).

A close connection thus exist between the RUM and RRM model, where the levels of β_m and x largely determine the extent to which similar results are obtained between both models in terms of overall model fit: both models attempt to describe the same set of choices over the same choice sets and using an equivalent number of parameters. Especially since the β's are optimized during the estimation process, similar levels of the overall log-likelihood can be expected. In some choice tasks the RRM model will perform better and in other choice tasks the RUM model will. The same conclusions can be drawn with respect to model validation on hold-out samples. As long as the decision process (and experimental design) between the estimation and hold out sample are comparable, the RUM and RRM model are likely to generate a comparable aggregate fit (i.e., averaged over all observations), where the context dependence in the RRM model may again translate into a better or worse performance for particular observations. There are not many *a priori* expectations on the direction of such differences at the aggregate level, which warrants additional future research on the implications of varying experimental designs and varying choice set sizes on the relative performance of the RUM and RRM model.[4] One finding however, which appears to be rather robust (e.g. Chorus & Bierlaire, 2013) is that when an outspoken compromise alternative is present in a particular choice situation, the RUM-model tends to underestimate its popularity and the RRM-model is likely to provide a better fit with choice observations made in the context of that choice situation, also at the aggregate level.

To further illustrate the close connection between the two models, we simulate two synthetic datasets. In each choice task every individual is presented with $J=3$ alternatives, which are described by $M=3$ attributes. The x values for all alternatives and characteristics are drawn from a uniform distribution between [0,1] and the model parameters are set to $\beta_m=2$. In the first dataset, generated choices are based on the standard RUM model, and in the second dataset the RRM rule is applied. Both datasets are estimated using both the RUM and RRM model providing a 2x2 experimental set-up. Table 1 presents the log-likelihood values for a dataset comprising 5,000 observations. Differences in model fit are in line with intuition (RUM model fitting better on RUM data, etc.); they are rather small and have a maximum of 26 LL points in the case of the RRM data.

Table 1: LogLikelihood results for two simulated datasets (RUM-RRM) using two alternative models (RUM-RRM)

	Data	
Model	**RUM**	**RRM**
RUM	-4.345,9	-3.651,5
RRM	-4.351,2	-3.625,4
Obs	5.000	

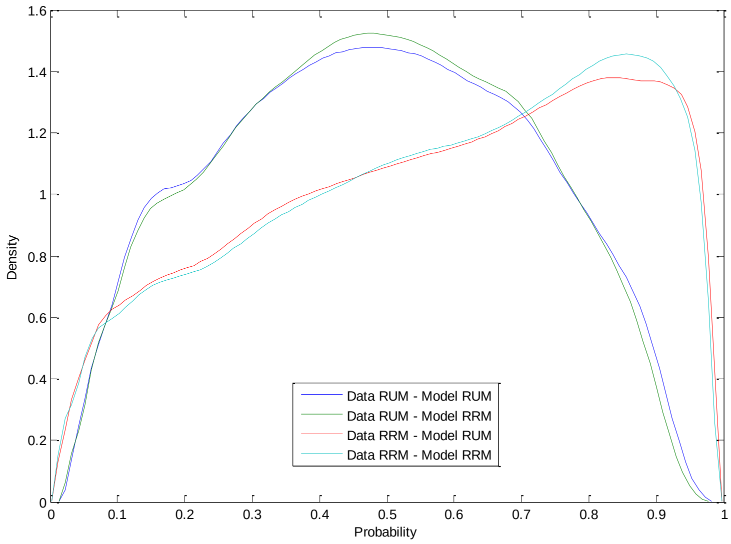
The close connection between the RUM and RRM in aggregate model fit can be further illustrated by plotting the (smoothed) kernel density of the choice probabilities for the *chosen* alternatives based on the four cases (see Figure 1). The higher log-likelihood for the RRM data observed in Table 1 is illustrated by a larger share of high choice probabilities in the two associated models. This higher likelihood of both models on the RRM data is a consequence of the fact that the same parameter size is used for generating RUM and RRM data ($\beta_m=2$), while the RRM model – due to its addition of strictly positive binary regrets – generates smaller parameters than RUM for choice sets containing more than 2 alternatives (see Chorus 2012a).

Most important, the kernel densities point out that *on average* both the RUM and RRM model can approximate the observed choices about equally well translating into a similar model fit. Hensher et al. (2011) provide similar kernel density plots which display some divergence when their data are applied using the two alternative models. In their case, the RRM model provides a better model fit of 12 LL points, and the observed difference in the kernel density plot can be explained by the use of non-synthetic (i.e less well-behaved) data. The fact that on average similar choice probabilities are

[4] For example, Hensher's 'Design-of-designs' approach may turn out useful in this context (Hensher et al. 2006).

assigned to the chosen alternatives does not imply that both models assign similar choice probabilities to particular choice tasks. Furthermore, it is exactly those differences that are important when using the results from discrete choice models in forecasting travel demand based on for example Equation (1). We pay more attention to these differences in the next subsection.

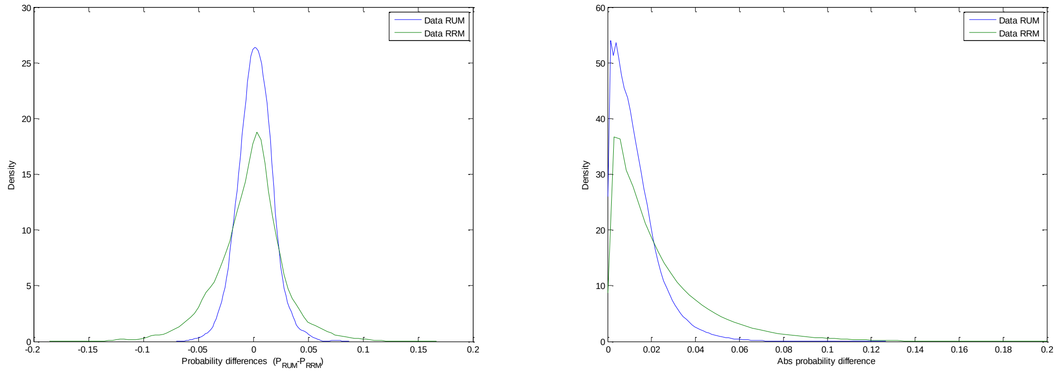
Figure 1: Kernel density plot of the choice probabilities for the chosen alternatives



3.2 Choice probabilities, market shares and elasticities

While the kernel densities in Figure 1 show a close match at the aggregate level when the RUM and RRM data are applied to the same synthetic dataset, Figure 2 highlights that, at the choice task level, differences in the (absolute) choice probability between the RUM and RRM model for the chosen alternative can be as large as 17,2 percentage points for the RRM dataset and 7,9 percentage points for the RUM dataset. Similar differences in choice probabilities are observed by Chorus et al. (2013b) in choices for alternative fuel vehicles (AFVs). In their case, in no less than 7% of all cases, the two models identify different alternatives as the most popular ones. In our synthetic set-up, both models only predict an alternative winner in about 3% of the cases for both datasets. A nice property of the RRM model is that these differences in choice probabilities can easily be related to the composition of the choice task, i.e. the relative performance of alternatives, and the levels of model parameters. The compromise effect is one of these effects, which has been revealed in various papers for specific choice situations. Chorus and Bierlaire (2013) show that, in line with theoretical expectations, the RRM model predicted a substantially higher choice probability than the RUM model (27 vs. 23%) to a ‘compromise’ route in a route choice experiment amongst Dutch commuters. Similar compromise effects are illustrated by Chorus et al. (2013b) in two exemplary choice sets amongst AFVs, and de Bekker-Grob and Chorus (2013) in a health economics case study on osteoporosis drug treatments and human papillomavirus vaccinations.

Figure 2: Probability differences between the RUM and RRM model for the chosen alternative



The implications of these differences in choice probabilities for market shares, or aggregate demand are less clear, since they are weighted across different types of agents and the choice situations they face (see Equation (1)) . First, differences in choice probabilities potentially cancel out over the population of interest, particularly when predictions are based on a range of future scenarios drawn from a stated choice design. Such designs typically describe a broad range of trade-offs to identify marginal sensitivities, but also make it more likely that the predicted choice probabilities vary in different directions between the RUM and RRM model. In real world forecasting exercises, usually a more focussed scenario is applied making it more likely that different predictions between the RUM and RRM model become apparent. These aggregate effects have, however, not been studied in the existing empirical literature. Second, the fact that different types of agents are affected differently by a policy change may be an important source of information to policy makers, even if the overall impacts on aggregate demand are limited.

Of equal importance in forecasting exercises is how choice probabilities, and market shares, respond to changes in product characteristics or travel conditions. Elasticities are generally used for this purpose and are directly comparable between the RUM and RRM model. Elasticities measure the percentage change in the choice probability of choosing a particular alternative as a result of a one percentage change in the value of one of the characteristics. Due to the dependency of elasticities on the current choice probabilities and levels of the characteristics, the elasticities vary by choice task in *both* the RUM and RRM model (see Train, 2009 and Hensher et al. 2011).[5] Unfortunately, the routine to derive elasticities in NLOGIT takes the average of the elasticities over all observations (weighted or not) and thereby possibly average out important differences between the RRM and RUM model (see the above discussion regarding averaging market shares in stated choice scenarios). Obviously, the weights w_k assigned to specific types of agents in Equation (1) may also affect the differences in elasticities between the two decision rules in forecasting exercises.

[5] Specifically, in the RUM model elasticities are highest when choice probabilities are similar across alternatives, i.e. the choice task is undecided.

The four studies (i.e. Chorus and Bierlaire 2013; Greene et al. 2012; Hensher et al. 2011 and Thiene et al. 2012) which present a comparison of elasticities between the RUM and RRM model do this at the aggregate level, but are rather unclear on the weights applied to derive these measures. Moreover, only the paper by Chorus and Bierlaire (2013) investigates confidence intervals around these elasticities, making it hard to evaluate whether differences in elasticities between the RUM and RRM model that are reported in the other three studies are significant or not. We provide a brief overview of what has been found in these four papers. Thiene et al. (2012) report that, for six of the eight attributes considered in their stated preference study of recreational choice in the Italian Alpes, the RRM model translated into higher elasticities than the RUM model. The size of this difference turns out to be around 10% when examining the ratio of RUM and RRM elasticities, but elasticities are low in general. The latter is also reflected by the fact that when park entrance fees are increased by 15% choice probabilities only increase by approximately 3% in the RUM model and 2% in the RRM model.[6] Greene et al. (2012) find, similar to Thiene et al. (2012), that cost elasticities are smaller in a hybrid-RRM model specification compared to the RUM model. This finding is, however, not confirmed by Hensher et al. (2011) and Chorus and Bierlaire (2013). The former consistently find higher elasticities for the RRM model, although absolute differences with the RUM model are generally small. In Chorus and Bierlaire (2013) the differences in elasticities between the two models are even smaller. The derived confidence intervals even show significant overlap between the RUM and RRM elasticities questioning the existence of these differences. The latter highlights the importance of examining the standard errors of these elasticities. Although the formula for the elasticities is difficult and analytical standard errors are hard to define, the Krinsky and Robb (1986, 1989) method can be applied to simulate the required confidence intervals. Alternatively, Bayesian estimation methods can easily be applied for the same purpose.

We now examine the differences in elasticities at the choice task level for our synthetic dataset, specifically focusing on how choice probabilities of alternatives respond to changes in their *own* characteristics. For the RUM dataset, we find that elasticities range between 0 (no response) and approximately 1.9. this range is comparable when the RRM model is applied to this data. The range increases to approximately 2.9 in the RRM dataset, which is a direct consequence of the chosen parameters (see the discussion presented earlier in the context of higher likelihood for RRM-data), but remains highly comparable between the RUM and RRM model. On average, the RUM model shows elasticities which are only about 0.01-0.02 higher than those for the RRM model. More interesting is again the spread and the density of those deviations across choice tasks, which again reveal that both positive and negative differences are observed between both models. Sometimes the RUM model provides a higher response, while in other cases the RRM model displays higher elasticities. For example, respondents respond more heavily in RUM models than in RRM models when the

[6] The RUM model displays a higher elasticity than the RRM model explaining the higher responsiveness.

alternative under consideration already performs relatively well on that particular characteristic. This can be directly related to the formulation of the regret function, which provides a higher weight to negative performance than to a positive performance. Due to its good performance the change in regret and therefore the elasticity will be only minor in this particular situation.

3.3 Willingness-to-pay and welfare measures

An alternative output of discrete choice models used for predicting demand or the welfare applications of policy changes are willingness-to-pay measures, which describe the extent to which individuals are willing to pay for improvements in the particular characteristics. Similarly, the concept of the logsum (or the associated metric of consumer surplus) can be applied to assess whether individuals are better off in one choice set relative to another (e.g. De Jong et al. 2007a). The RUM model is well established within the welfare economics literature and has convenient properties to analyse such welfare changes, particularly when a linear cost coefficient is specified (implying an assumption of absence of income effects associated with the transport policy). The context dependence of the RRM model complicates welfare analysis and a well-established framework is currently not yet established, but current research initiatives are looking into these issues. Accordingly, only a limited empirical comparison can be discussed in this subsection.

With respect to willingness-to-pay (WTP) measures, the concept of indifference is critical. Chorus et al. (2013c) and Chorus (2012a), for example, develop the marginal rate of substitution between two attributes in the RRM model by assuming that the regret associated with the specific alternative under consideration remains constant. This approach, however, neglects the property of the regret function that the regret of every alternative is affected when an attribute level changes of any alternative. Accordingly, keeping the regret of one alternative constant does not necessarily imply that a considered alternative is equally likely to be chosen in the 'old' and 'new' situation, nor that the individual is indifferent between the choice set in the 'old' and 'new' situation. This is one of the issues that needs to be addressed in future research. In line with results discussed in the previous subsections, Chorus et al. (2013c) show that the RRM model describes a richer range of WTP estimates which are driven by the composition of the choice-set and can directly be related to the semi-compensatory behaviour underlying the RRM model. Similar types of differences (up to 20%) in WTP estimates are found by de Bekker-Grob and Chorus (2013).

The logsum can be established for both the RUM and RRM model without any of the issues mentioned above. It describes the expected maximum (minimum) utility (regret) of a choice set. Chorus (2012a) shows that in contrast to the RUM model, the RRM logsum does not necessarily improve when an alternative's performance is improved on one attribute. Take for example the situation where a previously very poorly performing alternative improves its performance on an attribute. When, despite this improvement, the choice probability of the alternative remains low, the main effect of this improvement in one of its attributes is that regret levels of the more popular alternatives rise, leading to an increase in expected regret associated with the choice set. An extension of the RRM logsum to the related concept of consumer surplus is, however, not yet supported due to the non-linear cost-coefficient.[7]

4. DECIDING BETWEEN RUM AND RRM IN FORECASTING

The small differences (and sometimes mixed findings) relating to model fit and external validity between the RUM and RRM model often make it hard to decide which of the two models should be applied to predict market shares or to generate other relevant model outputs. This is especially the case when the behaviourally responses predicted by the two models for specific contexts are substantially different. The latter is most likely going to be the case when a choice situation includes one or more outspoken compromise alternatives, or when market shares are predicted for alternatives with a very strong or very poor base performance on relevant attributes. The parsimonious nature of the RRM model, and its compatibility with popular software packages such as Biogeme and NLOGIT, makes it relatively easy for researchers to estimate both models and generate predictions based on both estimated models simultaneously. When prediction is possible, the researcher may apply (arbitrary) selection criteria to pick one of the two models whilst ignoring the predictions of the other.[8] Obviously, this does not take into account the uncertainty associated with which of the two models is actually most likely to represent the true behavioural process.

A second approach would be not too choose either of the two, but to implement them both in the prediction stage, after conducting similar types of sensitivity analyses to both models. This way, policies can be designed that are 'robust from a behavioural perspective', i.e., robust across different decision rules. Indeed, this approach will take into account the analyst's uncertainty associated with both models, but it remains hard to judge the weight that needs to be assigned to either prediction (RUM versus RRM). This brings us to a third, and probably most preferred, approach (especially when combined with the 'behavioural robustness approach'). Wagenmakers and Farrell (2004) cover the topic of model selection and propose the use of Akaike (or Bayesian) Information Criterion based weights. These Akaike weights provide the researcher additional insights into the relative performance when contrasting a range of competing models. The approach is closely related to the Bayesian concept of model averaging (e.g. Hoeting et al. 2007), where the predictive density for the concept of interest based on each considered model is weighted by the posterior model probability. The latter measures the probability that after observing the data a particular model is the correct model. Simple rules of probability, including the marginal likelihood, can be applied to derive these weights. In comparison, the Akaike weights are based on the AIC statistic of a model, which is like the marginal

[7] This could be solved in a Hybrid-RUM-RRM setting where the cost coefficient is treated as a linear-in-parameters utility attribute.

[8] It might very well be the case that analysts prefer to stick to the standard RUM model due to its proven track record. The model needs little introduction and its outcomes can be easily interpreted. The properties of the RRM model are still relatively unknown and its outcomes somewhat harder to interpret. It still needs to earn its reputation.

likelihood based on the model fit combined with a penalty for the number of parameters that are estimated. The benefit of this third approach is that a single point estimate can be provided in the forecasting exercise whilst taking into account the uncertainty associated with the underlying decision rules.

Overall, the RRM model appears to be ready to be applied in travel demand forecasting exercises not directly aimed at deriving welfare implications of changing travel conditions. To date no forecasting applications of the model are known to the authors of this paper. Although this is an obvious research gap, we do note again that at the aggregate level predictions between the RUM and RRM model may not differ that much. This, however, needs to be tested empirically.

5. TOWARDS AN RRM-BASED NATIONAL DEMAND MODEL

In the Netherlands research efforts are currently undertaken to develop an RRM-based counterpart of the RUM-based Dutch National Model (e.g., Daly & Sillaparcharn, 2008; Willigers & de Bok, 2009, Significance, 2012). The Dutch National Model, abbreviated as the LMS, has been developed as a tool for policy evaluation of large transport projects in the 1980s (De Jong et al. 2007b). It produces medium to long-term forecasts on a national scale. Despite the aggregate nature of its outputs, choices associated with transport are modelled at the disaggregate level of the individual or at the level of the households, based on RUM-premises. This disaggregate specification of the LMS provides a behavioural foundation from which to analyse and predict future mobility patterns. In the current version, the LMS 2011, three core models can be identified: the tour-frequency model (by travel purpose), destination/mode/time-of-the-day model (by travel purpose), and a route assignment model (Significance, 2012). Each stage yields input for the next stage and iterative procedures are included to take into account demand-supply interactions, such as congestion. Nine travel purposes are distinguished such as commuting, education, shopping, business and other work-based travel.[9]

In this section we discuss, at the level of the sub-models, some of the challenges that lie ahead when replacing the underlying RUM-based choice models by its RRM counterpart. For reasons of conciseness we limit our discussions to tour-frequency models and the mode-destination-time-of-day model. Similar kinds of challenges lie ahead for the route assignment models as for the tour-frequency and mode/destination/time-of-day models.

5.1 Tour-frequency models

Tour frequency models generate the number of trips for each specific travel purpose. Figure 3 shows how the tour-frequency is modelled: as a series of binary choice situations. The probability that a decision maker will undertake a tour (which is defined as a two-way trip) depends on socio-

[9] To forecast future demand the LMS uses a so-called pivot-point procedure, i.e. the model is geared to predict changes relative to an observed base-year situation. As the base-year situation can usually be rather accurately measured, this so-called pivot point procedure improves the accuracy in the forecasts (see Daly 2005 for a discussion on this method).

demographics (such as age, sex, holding a driving license, etc.), zone characteristics, as well as the accessibility characteristics. For many travel purposes accessibility is found to positively affect the number of tours. The measure of accessibility that enters the tour-frequency models as an explanatory variable is the logsum (e.g. Ben-Akiva and Lerman, 1985; Geurs and van Wee, 2004). This logsum is directly imputed from the underlying mode–destination–time-of-the-day model.

From a methodological perspective, no substantial challenges are expected when replacing the underlying RUM-based tour-frequency models by its RRM counterparts. At first sights, there is even no need at all to replace the RUM-based tour frequency model with RRM-based models. After all, in binary choice situations RRM and RUM yield identical outcomes (Chorus, 2010). However, the regret logsum is fundamentally different from the RUM logsum (Chorus 2012b).[10] Therefore, as the RRM-logsum enters the tour-frequency model, outcomes of RRM-based and RUM-based tour frequency models can be expected to divert as a result of this seemingly subtle difference.

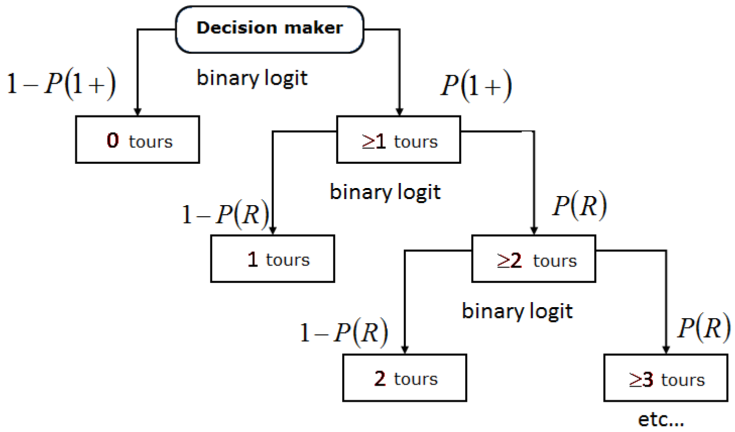


Figure 3: Tour generation in the Dutch National Model

5.2 Destination/mode/time-of-day models

Destination/mode/time-of-day models capture the joint choice of travel mode, destination and time-of-the-day. At this stage the tours generated in the previous stage are assigned to destinations which are reached by specific modes at specific times. The model comprises around 1380 zones (destinations); six different modes of transport, namely: car-driver, car passenger, train, bus/tram/metro, bicycle, and walking; and distinguishes nine periods of time-of-day. The latter results in 45 possible combinations in the day to depart and return.

These choices across travel modes, destinations, and the time-of-day are correlated. For instance, it is generally found that there is higher substitution between alternative time-of-day periods than between alternative modes (Hess et al. 2007). As the multinomial logit model implies proportional substitution across alternatives, the MNL model is not able to accommodate for such substitution patterns. Accordingly, to capture such correlations – while still featuring a closed-form expression for choice probabilities – the LMS uses Nested Logit models. Key when estimating nested

[10] The regret logsum is the expected maximum regret associated with the decision maker's choice set.

logit models is to identify the most appropriate nesting structure (i.e. such that the nesting coefficients do not exceed the theoretical constraint of one when estimated). In the LMS for most travel purposes either one of the following nesting structures is found: mode, above time-of-day, above destination, or mode, above destination, above time-of-day.

It is relatively easy to specify the RRM model in a Nested Logit form. The RRM model is as flexible as the RUM model with regard to error term specifications. Therefore, in principle, no challenges are expected in this regard. However, despite the growing number of studies using RRM models, Nested Logit RRM model specifications are absent in the literature. As such, little is known on estimation of nesting coefficients in RRM models. It seems possible that different nesting structures may be found to be most appropriate under RRM than under RUM – potentially having substantial effects on forecasts.

A more substantial challenge lying ahead when replacing the underlying RUM-based choice models by RRM counterparts stems from the large number of alternatives present in the mode/destination/time-of-day choice model. In the LMS it is assumed that each decision maker considers all available alternatives. As such, the resulting number of alternatives for which choice probabilities need to be computed is considerable. This implies that it is very important that choice probabilities can be computed with relative computational ease. Unfortunately, the computation of choice probabilities using RRM becomes computationally burdensome when choice set sizes are large which is the case for the destination/mode/time-of-day choice. This is a direct consequence of the fact that regret is choice set dependent (see Eq. 4): to compute an alternative's choice probability all attributes differences across all alternatives need to be evaluated since attributes of each alternative enter the regret function. Accordingly, the number of computations required to calculate RRM choice probabilities grows rapidly with the choice set size and the number of attributes per alternative.

Fortunately in this regard, methods for sampling of alternatives in Random Regret Minimization models have been proposed recently (Guevara et al. 2013). These authors analytically show, and illustrate using synthetic data, that consistent, and efficient estimators are obtained when estimating RRM models (with an extended regret function) on sampled choice sets of considerably smaller size than the universal set. Nonetheless, it should be emphasized that estimation on synthetic data is only a first step towards a solid understanding of RRM-based estimation on sampled choice sets. There is a need for follow up research in this regard, using empirical data. More generally speaking, further exploration of sampling of alternatives in RRM models is needed such as relating to the stability of estimates as a function of the particular properties of the applied sampling method.

In all, it can be concluded that a number of methodological challenges lie ahead in the development of an RRM-based national demand model. From a scholarly and methodological viewpoint, this raises the opportunity to further extend the scope of RRM-models, and to study differences between RRM and RUM at a far more aggregate level than has currently been done. Furthermore, since differences in aggregate forecasts between the two National Models (RUM and

RRM) can be expected, tackling these challenges in the process of developing an RRM-based National Model serves an important societal cause as it may ultimately contribute to better-informed policy-making by providing insights on the robustness of travel demand forecasts and transport project evaluation.

6. CONCLUSIONS

In recent years, the Random Regret Minimization framework has evolved rapidly as an alternative decision heuristic to the Random Utility Maximization framework in discrete choice models. The growing range of empirical comparisons between the two behavioural frameworks reveals that in most cases the RRM model is able to describe observed behaviour about equally well as the standard linear-in-parameters RUM model, at least at the aggregate level. For specific choice situations, the two models may however generate remarkably different choice probabilities, elasticities and other related model outputs. In this paper, we have summarized the empirical comparisons and illustrated differences and similarities between the two frameworks using synthetic data. What can be concluded is that the RRM model offers, in terms of estimation, an equally parsimonious alternative to the RUM model, which describes a different range of behaviour based on the premises of regret minimization and semi-compensatory behaviour. A clear progression in understanding the properties of the model relative to the RUM model can be observed and the observed deviations are generally a clear consequence of the context dependency embedded in the RRM model.

The next challenge for the RRM model is to find its way into real world applications, including forecasting exercises based on discrete choice models. Based on the existing literature, it can be established that for specific choice situations the predicted market shares (and elasticities) between the RUM and RRM models can differ substantially. Specifically when compromise alternatives are considered. However, when a range of future scenarios is evaluated across different types of agents differences between the RUM and RRM model may cancel out at the aggregate level. The size and direction of these differences are hard to identify a priori and this is therefore identified as a clear topic for future research. Important in this perspective is also the comparable performance in model fit (and external validity on hold-out samples) making it hard to select one of the two models to predict demand. Selecting either of the two neglects the uncertainty regarding the 'true' underlying decision rule. We have proposed two alternative approaches dealing with this issue. One option is to apply the two models simultaneously and compare range of the predictions between the two models. This will enable policy makers to develop robust policies that accommodate both utility maximisers and regret minimisers. The difficulty is still to assign a degree of importance to both frameworks, particularly when predictions diverge. We propose the use of model averaging techniques as a second option. Akaike weights or Bayesian Model Averiging can be applied to weight the predictions of either model and generate a point estimate whilst taking into account the uncertainty regarding the behavioural model. The readily availability of the RRM model in existing model packages and the relative ease to estimate both models in a Bayesian fashion will open up opportunities to apply the latter strategy.

Overall, the RRM model seems ready to be applied in forecasting exercises. It is able to generate choice probabilities and other required model outputs such as the logsum. The levels of these statistics may vary from the RUM model, and therefore generate alternative predictions. This is clearly an interesting topic of future research. Therefore, we have illustrated the possible implications and challenges of applying the RRM model to the Dutch National Model in the final section of this paper. An exercise which is on-going work by the authors of this paper. A challenge that still remains is the use of the RRM model for evaluating the welfare implications of changing travel conditions.

REFERENCES

Beck, M.J., Chorus, C.G., Rose, J.M., Hensher, D.A., 2013. Vehicle purchasing behavior of individuals and groups: Regret or reward? *Journal of Transport Economics and Policy,* 47(3), 475-492

de Bekker-Grob, E.W., Chorus, C.G., 2013. Random regret-based discrete choice modelling: An application to health care. *PharmacoEconomics*, 31(7), 623-634

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis: Theory and application to travel demand. MIT Press, Cambridge, MA, USA

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, 253-260.

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

Chorus, C.G. 2012a. Random Regret Minimization: An overview of model properties and empirical evidence. *Transport Reviews,* 32(1), 75-92.

Chorus, C.G. 2012b. Logsums for utility-maximizers and regret-minimizers, and their relation with desirability and satisfaction. *Transportation Research Part A,* 46(7), 1003-1012.

Chorus, C.G., 2013. Capturing alternative decision rules in travel choice models: A critical discussion. *Handbook of Choice Modelling* (forthcoming).

Chorus, C.G., Bierlaire, M., 2013. An empirical comparison of travel choice models that capture preferences for compromise alternatives. *Transportation*, 40(3), 549-562

Chorus, C.G., van Cranenburgh, S., and Dekker, T., 2013a. Random Regret Minimization for consumer choice modeling: assessment of empirical evidence. Paper presented at the *35th ISMS Marketing Science Conference*, 11-13 July, Istanbul, Turkey.

Chorus, C.G., Koetse, M.J., Hoen, A., 2013b. Consumer preferences for alternative fuel vehicles: comparing a utility maximization and a regret minimization model. *Energy Policy,* 61, 901-908

Chorus, C.G., Rose, J.M., Hensher, D.A., 2013c. Regret minimization or utility maximization: it depends on the attribute. *Environment and Planning B: planning and design,* 40, 154-169.

Daly, A. 2013. Forecasting behaviour: with applications to transport. In: Hess, S. and Daly, A. (Eds.), *Choice Modelling: The state of the art and the state of practice*, Edward Elgar, Cheltenham, UK.

Daly, A.J., Fox, J., Tuinenga, J.G, 2005. Pivot-point procedures in practical travel demand forecasting.Paper presented at *European Regional Science Association meeting*, Amsterdam.

Daly, A., and Sillaparcharn, P., 2008. National models. In: Hensher, D.A., Button, K.J. (Eds.) *Handbook of Transport Modellin*g, Pergamon, Amsterdam, The Netherlands.

Guevara, C.A., Chorus, C.G., Ben-Akiva, M.E., 2013. Sampling of alternatives in random regret minimization models, *paper presented at the 2013 TRB conference,* Jan 13-17, Washington D.C.

Hensher, D.A., 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and Resource Economics,* 34, 7-44.

Hensher, D.A., Greene, W.H., Chorus, C.G., 2011. Random Regret Minimization or Random Utility Maximization: An exploratory analysis in the context of automobile fuel choice. *Journal of Advanced Transportation* (forthcoming).

Hess, S., Daly, A., Rohr, C., Hyman, G. 2007, On the development of time period and mode choice models for use in large scale modelling forecasting systems, *Transportation Research Part A* 41(9), 802-826.

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

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T., 2007. Bayesian model averaging: a tutorial. *Statistical Science,* 14(4), 382-417.

de Jong, G., Daly, A., Pieters, M., and van der Hoorn, T., 2007a. The logsum as an evaluation measure: review of the literature and new results. *Transportation Research Part A*, 41(9), 874-889.

de Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R. and Hofman, F., 2007b. Uncertainty in traffic forecasts: literature review and new results for the Netherlands. *Transportation,* 34, 375-395.

Geurs, K., Van Wee, G.P., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2), 127-140.

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

Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 68(4), 715-719.

Krinsky, I., Robb, A.L., 1990. On approximating the statistical properties of elasticities: a correction. *The Review of Economics and Statistics,* 72(1), 189-190.

Leong, W., and Hensher, D., 2012. Embedding decision heuristics in discrete choice models: a review. *Transport reviews*, 32(3), 313-331.

McFadden, D., 1974. The measurement of urban travel demand. *Journal of Public Economics*, 3(4), 303-328.

Rasouli, S., Timmermans, H.J.P., 2012. Uncertainty in travel demand forecasting models: Literature review and research agenda. *Transportation Letters*, 4(1), 55-73.

Significance, 2012. Schattingen van keuzemodellen voor het LMS 2011. Jasper Willigers, Michiel de Bok. Rapport voor de dienst verkeer en schaapvaart van Rijkswaterstaat (in Dutch).

Thiene, M., Boeri, M., Chorus, C.G., 2012. Random Regret Minimization: Exploration of a new choice model for environmental and resource economics. *Environmental and Resource Economics*, 51(3), 413-429.

Train, K.E., 2009. Discrete choice methods with simulation, *Cambridge University Press.*

Wagenmakers, E., and Farrell, S., 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review*, 11(1), 192-196.

Willigers, J., and de Bok, 2009. Updating and extending the disaggregate choice models in the Dutch national model. Proceedings of the *37th European Transport Conference*, Leeuwenhorst, The Netherlands.