On the robustness of efficient experimental designs towards the underlying decision rule

Sander van Cranenburgh[a], John M. Rose[b], Caspar G. Chorus[a]

[a]*Transport and Logistics Group, Delft University of Technology*

[b]*Business Intelligence and Data Analytics, University of Technology Sydney*

Abstract

We present a methodology to derive efficient designs for Stated Choice (SC) experiments based on Random Regret Minimization (RRM) behavioural assumptions. This complements earlier work on the design of efficient SC experiments based on Random Utility Maximization (RUM) models. Capitalizing on this methodology, and using both analytical derivations and empirical data, we investigate the importance of the analyst's assumption regarding the underlying decision rule used to generate the efficient experimental design. We find that conventional RUM-efficient designs can be statistically highly inefficient in cases where RRM is the better representation of the actual choice behaviour, and vice versa. Furthermore, we present a methodology to construct efficient designs that are robust towards the uncertainty on the side of the analyst regarding the underlying decision rule.

1 Introduction

Stated Choice (SC) experiments are widely used to acquire understanding of travel behaviour (Louviere et al. 2000). SC experiments involve respondents being exposed to hypothetical scenarios involving two or more alternatives, at least one of which is described by a set of attributes and attribute levels. Respondents are then asked to review these scenarios and indicate their preferences for the alternatives shown based on the attributes and attribute levels describing each of the alternatives. It is therefore necessary for the analyst to assign the levels describing the attributes prior to writing the survey. The allocation of the attribute levels to the survey task typically occurs via an experimental design, although random allocation is not uncommon in practice. Many different approaches for generating experimental designs have appeared within the literature. Nowadays, the most common approach to generate designs for SC experiments involves what are known as efficient designs. Efficient designs aim to maximise the information

obtained from SC data, resulting in more reliable parameter estimates for a given number of observations (Rose and Bliemer 2009; Kessels et al. 2011).

Early research on efficient design theory mainly concentrated on the Multinomial Logit (MNL) model (e.g., Bunch et al. 1996; Huber and Zwerina 1996), while more recent research focussed on extending the design theory to encompass more advanced choice models, including the NL model (e.g., Bliemer et al. 2009; Goos et al. 2010), the cross-sectional version of Mixed Logit (ML) model (Sándor and Wedel 2002; Yu et al. 2009), and the panel version of the ML model (Bliemer and Rose 2010; Yu et al. 2011). A substantial share of this research has been expended on examining the issue of parameter priors on design efficiency. Two types of assumptions have been used when defining parameter priors. The first type involves designs being derived under the assumption of fixed prior parameters. The prior parameters can be either zero (the resulting design is said to be a utility neutral design, e.g., Huber and Zwerina 1996), or non-zero (e.g., Carlsson and Martinsson 2003). In either case, the design is referred to as a local optimal design. An alternative to locally optimal designs can be derived under the assumption that the prior parameters are drawn from some distribution with a known probability density which reflects uncertainty (by the analyst) about the value of the true population parameters. When such priors are assumed, the resulting design is known as a semi-Bayesian efficient design (Yu et al. 2009).

However, to date, research into experimental design theory for SC experiments has exclusively been based on the (often implicit) assumption that decision-makers make choices using (linear-additive) Random Utility Maximization (RUM) rules. This is despite compelling evidence that decision-makers use a wide range of decision rules when making choices (Hess et al. 2012; Boeri et al. 2014) and despite the rapidly growing interest in the travel behaviour community into alternative decision rules (Leong and Hensher 2012; Ramos et al. 2014; Guevara and Fukushi 2016; Sun et al. 2016). Some related studies have looked into the impacts on statistical efficiency of misspecifications of utility functions and prior parameters (e.g. Ferrini and Scarpa 2007; Yu et al. 2008). But, these studies are fully embedded within the RUM modelling framework.

This paper contributes to the literature by shedding light on the importance of the assumption concerning the underlying decision rule for efficient experimental design. We do so, in the context of a particular non-RUM model – in casu: a Random Regret Minimization (RRM) model (Chorus 2010). The main reason why we use the RRM model for our analyses is because RRM models are among the more widely used non-RUM models (see e.g. Hensher et al. 2016; Boeri

and Longo 2017; Sharma et al. 2017; van Cranenburgh and Chorus (accepted) for recent applications). Moreover, the specific RRM model we use – the P-RRM model (Van Cranenburgh et al. 2015a) – is equally parsimonious as the canonical linear-additive RUM model, and – as we will show below – has very convenient mathematical properties for constructing efficient designs.

In this paper, we first analytically investigate the importance of the design decision rule on statistical efficiency[1]. Specifically, we consider two cases: (1) experimental designs that are optimised for linear-additive RUM while the true Data Generating Process (DGP) is P-RRM, and (2) experimental designs that are optimised for P-RRM while the true DGP is linear-additive RUM. After that, we present a methodology to construct efficient designs that are robust towards decision rule uncertainty. Finally, we use empirical data (specifically collected for this study) to explore statistical properties of RUM, P-RRM and robust efficient designs in the context of empirical data.

The methodological contributions of this paper to the experimental design literature are threefold. Firstly, we enrich the choice modeller's toolbox for designing efficient experimental designs by showing how efficient designs can relatively easily be constructed for the P-RRM-MNL model. Because the P-RRM model has a piecewise linear form, its Asymptotic Variance Covariance matrix (AVC) – which is needed to construct efficient designs – can be determined analytically, just like for the linear-additive RUM-MNL model[2,3]. Secondly, we show that decision rule misspecification may have severe consequences for the efficiency of resulting designs. Thirdly, we present a methodology to construct efficient designs that are robust towards the uncertainty on the side of the analyst on the underlying decision rule.

The remainder of this paper is structured as follows. Section 2 briefly revisits the essentials of efficient design theory, and derives a design methodology to construct efficient for RRM models. Section 3 analytically explores the effect of misspecification of the design decision rule on

[1] In the remainder of this paper, we use the term 'design decision rule' to refer to the decision rule of the model under consideration when constructing the experimental design. The term 'estimation decision rule' refers to the decision rule embedded in the choice model which is estimated on the data collected based on the efficient design. In the analytical part of our work, the term 'DGP decision rule' refers to the decision rule which underlies actual choice behaviour.

[2] Note that we consistently refer to the 'linear-additive' RUM-MNL model. We do so because under the assumption of linear-additivity the AVC matrix can be relatively straightforwardly derived, which is otherwise typically not the case. For this reason, experimental design software packages, such as NGENE, only allow the user to specify linear-additive utility functions.

[3] In this paper we refer to RUM and RRM decision rules as well as to RUM and RRM choice models. To highlight the difference between decision rules and choice models, in case we refer to the latter we explicitly add "-MNL" (note that all choice models in our study are in MNL form) while in case we refer to the former "-MNL" is omitted.

statistical efficiency, and presents a methodology to generate efficient designs that are robust towards decision rule uncertainty. Section 4 presents an empirical case study which highlights the impact of decision rule misspecification. Finally, section 5 reports conclusions and presents a discussion of the robustness of efficient experimental designs.

2 Efficient designs for Random Regret Minimization models

This section presents the methodology to construct efficient designs for RRM models. Specifically, it derives the AVC matrix for the P-RRM model. The AVC matrix is a key mathematical ingredient to construct efficient designs. The P-RRM model is a recently proposed type of RRM model, which imposes very strong regret minimisation behaviour (as compared to other types of RRM models). In this section we first revisit essentials of efficient design theory (section 2.1) and regret minimisation models (section 2.2). Informed readers on these topics may skip these parts. Section 2.3 derives the AVC matrix for the P-RRM model.

2.1 Efficient design theory

Designing an SC experiment involves making a number of decisions, such as how many choice sets are presented to each respondent, how many alternatives per choice set, what are the attributes considered, and what are their levels. After having determined this design set-up, the next step is to select a design strategy. Two design strategies are predominantly used in the literature: 1) those based on the principle of orthogonality, and 2) those based on statistical efficiency. The latter of these two strategies has become the most widely used in recent years, and is the topic of this paper. The aim of an efficient design strategy is to generate designs that maximise the collected information in the data. Depending on the assumed optimality criteria, the premise underlying statically efficient designs is that by doing so, more reliable parameter estimates can be attained with an equal, or lower number of observations than when orthogonal designs would have been used (Rose and Bliemer 2009). In other words, efficient designs aim to minimise the (asymptotic) standard errors, or alternatively, maximise the *t*-ratios, of the parameters estimates of the model under consideration.

Statistical efficiency can be measured in several ways including, but not limited to, A-efficiency, C-efficiency, D-efficiency, and S-efficiency (Kanninen 1993a; Kanninen 1993b; Rose and Bliemer 2013, and see Kessels et al. 2006 for a comparison). Regardless of the exact measure that

is being used, efficient designs exploit the Asymptotic Variance Covariance (AVC) matrix, denoted Ω, of the parameter estimates to determine the efficiency. The AVC matrix is a function of the design itself, the (prior) parameter estimates, and the model specification. The AVC matrix is equal to the inverse of the expected Fisher Information matrix *I*, see equation 1, where *X* denotes the attributes levels, *Y* denotes the choices, and *β* denotes the model parameters. In turn, the Fisher Information matrix can be computed by taking the second-order derivatives of the Log-likelihood (LL) function w.r.t. the model parameters and taking the expectation w.r.t. *Y*.

$$\Omega(\beta \mid X) = \left(I(\beta \mid X)\right)^{-1} \quad where \quad I(\beta \mid X) = -E_Y\left(\frac{\partial^2 \log L(\beta \mid X, Y)}{\partial\beta\partial\beta}\right) \qquad (1)$$

The Log-likelihood function for a variety of discrete choice models, such as MNL, NL, and cross-sectional versions of Mixed Logit and Probit, is given by equation 2. In equation 2, y_{sj} denotes an indicator of the choice, *s* denotes the choice set number, and *j* denotes the alternative. Note that we assume that all respondents face the same design – as is usually the case in SC experiments. Therefore, only one design – consisting of *S* choice sets – needs to be evaluated to be able to maximise the statistical efficiency of a design.

$$LL(\beta) = \sum_{s=1}^{S}\sum_{j=1}^{J} y_{sj} \ln\left(P_{sj}\right) \qquad (2)$$

In the context of this research we focus on one measure of statistical efficiency: the D-error statistic (D-efficiency). The D-error statistic is the most commonly used measure of efficiency in experimental design practice, and is calculated by taking the determinant of the AVC matrix, see equation 3. Accordingly, the optimal efficient design is that design which attains the lowest D-error. Note that to compute the D-error while accounting for the number of parameters, the determinant is sometimes scaled by the power of 1/*k*, where *k* denotes the number of parameters. However, as we do not vary the number of parameters in this study we do not apply this scaling factor.

$$D\text{-error} = \det\left[\Omega\left(X, \beta\right)\right] \qquad (3)$$

2.2 The P-RRM model

RRM models are based on the premise that, when choosing, the decision maker *n* minimises regret. Regret is experienced when a competitor alternative *j* outperforms the considered alternative *i* with regard to attribute *m*. The overall regret of an alternative is conceived to be the sum of all the pairwise regrets that are associated with bilaterally comparing the considered alternative with the other alternatives in the choice set in terms of each of the attributes. The general form of RRM models is given in equation **4**, where RR_{in} denotes the random regret experienced by decision maker *n* considering alternative *i*, R_{in} denotes the observed part of regret, and ε_{in} denotes the unobserved part of regret. In the core of RRM models[4] is the so-called attribute level regret function $r_{ijmn} = f\left(\beta_m, x_{jmn} - x_{imn}\right)$. The attribute level regret function is a convex function that maps the difference between the levels of attribute *m* of the competitor alternatives *j* and the considered alternative *i* onto regret.

$$RR_{in} = R_{in} + \varepsilon_{in} \ where \ R_{in} = \sum_{j \neq i} \sum_{m} r_{ijmn} \qquad (4)$$

There are various ways to specify the attribute level regret function, each leading to different RRM models. A well-known specification defines the attribute regret as $\ln(1 + \exp\left(\beta_m \left[x_{jm} - x_{im}\right]\right)$ (Chorus 2010); however, several studies – see Hensher et al. (2016) for a recent example – have highlighted that empirical differences (e.g., model fit, choice probability predictions, and elasticities) between this RRM-specification and its RUM counterpart are rather small, limiting the added value of this RRM-specification as an alternative choice model. Recent work (Van Cranenburgh et al. 2015a) has shed light on the reasons underlying these small differences, which turn out to be caused by a particular mathematical property of the attribute regret function used in the RRM model proposed by Chorus (2010). A more generic μRRM model is proposed in that latter paper, and this model has been found to result in very considerable differences in model fit, predictions and elasticities when compared with RUM-counterparts (e.g., Van Cranenburgh et al. 2015a; Boeri and Longo 2017; Sharma et al. 2017).

The P-RRM model, which we use in the remainder of our paper, is a special case of the μRRM model. Its attribute level regret function is given by $r_{ijmn} = \max\left(0, \beta_m \left[x_{jmn} - x_{imn}\right]\right)$. The P-

[4] One exception to this form is the RRM specification proposed in Chorus et al. (2008) and used in, for example, Jang et al. (2016)

RRM model is an extreme case of the generic and flexible μRRM model, and has a cornerstone interpretation within the RRM modelling paradigm (see Van Cranenburgh and Prato 2016 for an overview of RRM models): since it postulates no 'rejoice' (this is how the opposite of regret is called in the decision sciences), it yields the strongest regret minimization behaviour, i.e., the highest level of regret aversion which is possible within the RRM framework. At first sight, the max operator in the attribute level regret function may seem undesirable from a numerical optimisation perspective. However, as shown by Van Cranenburgh et al. (2015a), when the signs of the taste parameters are a priori known to the researcher – as is usually the case in a travel behaviour setting – the P-RRM model becomes linear-additive (equation **5**).

Setting priors is a challenge in generating efficient designs. This is the case in general, but especially so in the context of this study, in which we explore the relative efficiencies of RUM and P-RRM designs. To set the prior parameters consistently across the RUM and P-RRM contexts, in equation **5** a choice set size correction factor of $2/J_n$ is presented.[5] This correction factor is originally developed to estimate RRM models on a data set in which the number of alternatives that are available to a decision-maker varies across choice observations (Van Cranenburgh et al. 2015b). However, in the context of this study the correction factor is used to eliminate the effect of the choice set size on the sizes of the RRM model parameters and enables using the same priors (in terms of size) for the RUM and RRM model.

$$R_{in}^{P-RRM} = \sum_m \beta_m \tilde{x}_{imn}^{P-RRM} \qquad where\, \tilde{x}_{imn}^{P-RRM} = \begin{cases} \frac{2}{J_n} \sum_{j \neq i} \max\left(0, x_{jmn} - x_{imn}\right) \; if\, \beta_m > 0 \\ \frac{2}{J_n} \sum_{j \neq i} \min\left(0, x_{jmn} - x_{imn}\right) \; if\, \beta_m < 0 \end{cases} \tag{5}$$

Finally, inspired by RUM-MNL modelling practice, we assume the negative of the error term to be i.i.d. type I Extreme Value distributed with a normalised variance of $\pi^2/6$, resulting in the well-known and convenient closed-form logit formula for choice probabilities (equation **6**). Note that since RRM models assume minimization of regret (rather than maximising of utility), a minus sign is placed in front of the observed regret.

[5] Note that using $2/J_n$ we presume the choice set size is constant across all observations of the same decision-maker *n*. In case the choice set size varies across observations of the same decision-maker, the correction factor would be $2/J_{ns}$, where subscribe *s* refers to the observation number of decision-maker *n*.

$$P_{in} = \frac{\exp\left(-R_{in}\right)}{\sum_{l=1..J} \exp\left(-R_{ln}\right)}$$	6

2.3 The AVC matrix for the P-RRM-MNL model

The second-order derivative of the Log-likelihood function of the P-RRM-MNL model w.r.t. its generic taste parameters is given in equation **7**. For reasons of brevity the derivation is provided in Appendix A. This equation is very similar to that of its linear-additive RUM-MNL counterpart. In fact, the only difference is the replacement of the attribute levels x_{jms}, by the P-RRM model's attribute levels $\tilde{x}_{jms}^{P-RRM}$. Combining equations 1 and **7** allow us to compute the AVC matrix, which in turn enables us to compute the D-error, or any other efficiency measure, for P-RRM-MNL models.

The second-order derivative of the Log-likelihood function of the P-RRM-MNL (equation **7**) is not a function of the actual choices. Therefore, the AVC matrix straightforwardly be derived analytically, just like for linear-additive RUM-MNL models (Huber and Zwerina 1996) and linear-additive RUM-NL models (Bliemer et al. 2009). From a pragmatic viewpoint, this is an important feature, considering the fact that in order to find an efficient design typically many thousands candidate designs need to be evaluated.

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} P_{js} \left(\tilde{x}_{jm_2s}^{P-RRM} - \sum_{i=1}^{J} \tilde{x}_{im_2s}^{P-RRM} P_{is} \right)$$	7

3 Robustness of efficient designs towards misspecification of the design decision rule

This section explores the robustness of efficient designs towards misspecification of the design decision rule. In particular, we 1) assess the efficiency of linear-additive RUM efficient designs when the DGP decision rule is P-RRM, and vice versa; 2) present a methodology to construct designs that are robust towards decision rule uncertainty; and 3) assess the potential of this methodology.

To measure robustness in this context, we assess what we call *Efficiency loss.* Efficiency loss is defined as the decrease in the D-efficiency due to using a misspecified design decision rule, relative to the efficiency that could have been obtained had the correct design decision rule been used. Equation **8** gives the loss function, where $\bar{D}\text{-error}$ is the D-error statistic obtained using the *incorrect* design decision rule, and $D\text{-error}$ is the optimal D-error statistic obtained using the *correctly* specified design decision rule. An efficiency loss close to 1 indicates that the design obtained using the incorrectly specified decision rule is statistically highly inefficient (given the DGP decision rule), whereas an efficiency loss close to 0 indicates that the design obtained using the incorrectly specified decision rule is almost as statistically efficient as the design that would have been obtained when using the correctly specified design decision rule.

To further clarify equation **8**, suppose a RUM optimal design is used to collect data, but the true DGP is RRM. Then, the design is chosen suboptimal. The loss in efficiency due to the use of a suboptimal design is computed by comparing the D-error of the design that was chosen based on the incorrect assumption on the side of the analyst on the underlying decision rule ($\bar{D}\text{-error}$) – in this case the RUM optimal design – with the D-error of the best design that the analyst could have chosen ($D\text{-error}$) – in this case the RRM optimal design.

Furthermore, note that the measure in equation **8** is closely related to some previously proposed measures in the efficient design literature (see e.g. Scarpa and Rose 2008; Danthurebandara et al. 2011). The fundamental difference between the measure proposed by Scarpa and Rose (2008) and equation **8** is the aim of the measure. The measure of Scarpa and Rose (2008) aims to capture the efficiency loss caused by the fact that there is a difference between the priors that are used to construct the design and the actual parameter estimates obtained after having collected the data. In contrast, our measure aims to capture the efficiency loss due to the use of a misspecified design decision rule.

$$L = 1 - \frac{D\text{-error}}{\bar{D}\text{-error}}$$	8

3.1 Experimental design set-ups

We consider two relatively small experimental design set-ups.[6] By doing so we are able to determine the optimal efficient designs, i.e., the most efficient design of all possible designs. This is crucial in this context as we want to avoid presenting potentially misguided results due to the possibility that the search algorithm (which otherwise would have to be used) may not have been able to find the optimal solution. In fact, in the context of large solution spaces search algorithms are only able to assess a very small part of the total solution space (Palhazi Cuervo et al. 2014).

Table 1 presents our experimental design set-ups. The two design set-ups both comprise of three alternatives; each alternative is defined by two generic attributes (e.g., cost and time). We consider three alternatives as this is the minimum number of alternatives to be able distinguish between RUM behaviour and RRM behaviour. Furthermore, the alternatives comprise of two attributes, as this is the minimum number of attributes needed to analyse trade-offs in discrete choice models. Design set-ups 1 and 2 differ from one another in terms of the number of levels per attribute and the number of choice sets per design. In design set-up 1 four attribute levels are used, and four choice sets per respondent, while in design set-up 2 five attribute levels and three choice sets per respondent are used.

Table 1: Design set-ups

	Design set-up 1	**Design set-up 2**
Number of alternatives	3	3
Number of generic attributes	2	2
Attribute levels	1,2,3,4	1,2,3,4,5
Number of choice sets per respondent	4	3
Number of unique choice sets, without duplicate or dominant alternatives	16	100
Number of designs based on choice sets without duplicate or dominant alternatives	$\binom{16}{4} = 1{,}820$	$\binom{100}{3} = 161{,}700$

For set-ups 1 and 2 respectively, we can construct 560 and 2,300 unique choice sets. In turn, respectively 4 billion and 2 billion possible designs can be constructed for design set-ups 1 and 2.

[6] The term 'design set-up' refers to the constraints for the experimental design. This should not be confused with the design itself, nor with the notion of a choice set, which is confronted by the participant to the experiment.

However, we can strongly limit the solution space by removing choice sets containing one or more dominant alternatives.[7] The first and most important reason for doing so is that dominant alternatives are known to bias parameters estimates (Huber et al. 1982; Bliemer and Rose 2011; Bliemer et al. 2017). Therefore, it is common practice to remove choice sets containing dominant alternatives (Hensher et al. 1988; Kouwenhoven et al. 2014), although sometimes a dominant choice set is added to see whether respondents understood the choice set (see e.g. Bradley and Daly 1994). A second reason relates to the fact that respondents may get agitated by choice sets containing dominant alternatives, especially in the case of relatively simple SC experiments which comprise of just two or three attributes, such as ours. Removing those choice sets containing dominant alternatives reduces the number of unique choice sets to a mere 16 in design set-up 1 and to 100 choice sets in design set-up 2. As a consequence, there are just 1,820 and 161,700 possible designs for design set-up 1 and 2, respectively. Given these relatively small solution spaces, we are able to identify the globally optimal designs, and to analytically explore the robustness of efficient designs with respect to the design decision rule.

As a final note before presenting results, we use 'perfect priors'. That is, the priors used to construct the efficient designs are equal to the parameters used in the DGP. As such, we do not look into the effect of misspecified parameter priors on statistical efficiency. Moreover, without loss of general applicability, RUM and P-RRM parameters are set to the same values. While this presumption is not a theoretical necessity (other than in binary choice situations, see Chorus 2010), our considerable experience with estimating (P-)RRM models suggests that – after correcting for the choice set size (see section 2.2) – this is a very reasonable assumption.

3.2 Results

Figure 1 and Figure 2 show the efficiency loss for respectively design set-ups 1 and 2, as a function of the model parameters β_1 and β_2. The left-hand plots show the efficiency loss of P-RRM models when applied on RUM efficient designs; the right-hand plots show the efficiency loss of the linear-additive RUM model when applied on P-RRM efficient designs. To explore the effect of the size of the parameters, we created a two-dimensional grid space. We let β_1 and β_2 range from 0.025 to 3 with 119 equal intervals of 0.025. Next, we compute the efficiency loss for each of the 119 × 119 = 14,161 combinations of β_1 and β_2, by taking the following steps:

[7] An alternative A is said to dominate alternative B if alternative A performs at least as well as alternative B in terms of all attributes, and performs better in at least one attribute.

1. We search for the best (i.e. globally optimal) designs. Specifically, we find the design that minimises the RUM D-error (conditional on β_1 and β_2), and the design that minimises the RRM D-error (conditional on β_1 and β_2). To do this, we compute the D-errors of across all feasible designs (1,820 for set-up 1 and 161,700 for set-up 2) and find their lowest instances. We store the best RUM design and the best RRM design as well as their associated D-errors. To compute the D-errors, we use Equation 3.
2. We compute the D-error obtained using the *incorrect* decision rule: $\overline{D}\text{-error}$. To do this, we compute the RRM D-error (conditional on β_1 and β_2) for the best RUM design (which was stored at step 1), and we compute the RUM D-error (conditional on β_1 and β_2) using the best RRM design (which was also stored at step 1).
3. We compute the efficiency loss using Equation 8, based on the results of step 1 and 2.

Based on Figure 1 and Figure 2 a number of observations can be made. Firstly, there are parameter spaces where RUM efficient designs are statistically highly inefficient when the DGP decision rule is P-RRM, and there are parameter spaces where P-RRM efficient designs are highly inefficient when the DGP decision rule is RUM. Secondly, RUM and P-RRM designs appear to be by and large equally sensitive (i.e., non-robust) towards misspecification of the decision rule. Thirdly, in both design set-ups we see that the statistical efficiency of the design which is misspecified in terms of the underlying decision rule is highly dependent on the exact combination of parameter sizes. For several combinations of β_1 and β_2 the efficiency loss changes very abruptly when parameters change only slightly.

All in all, these results suggest that – in terms of statistical efficiency – efficient experimental designs are not robust with respect to misspecification of the decision rule. That is, efficient experimental designs can be statistically highly inefficient when the design decision rule does not match the DGP decision rule. However, it is worthwhile to notice that the efficiency loss as presented in Figure 1 and Figure 2 may partly be attributed to the adopted parameterisation. In the absence of a theoretical link between RUM and RRM model parameters, they are imposed to be equal across models (after correction for choice set size, see section 2.2). A substantial body of empirical studies indicates that this is a reasonable approach. Nonetheless, it may somewhat inflate the presented losses.

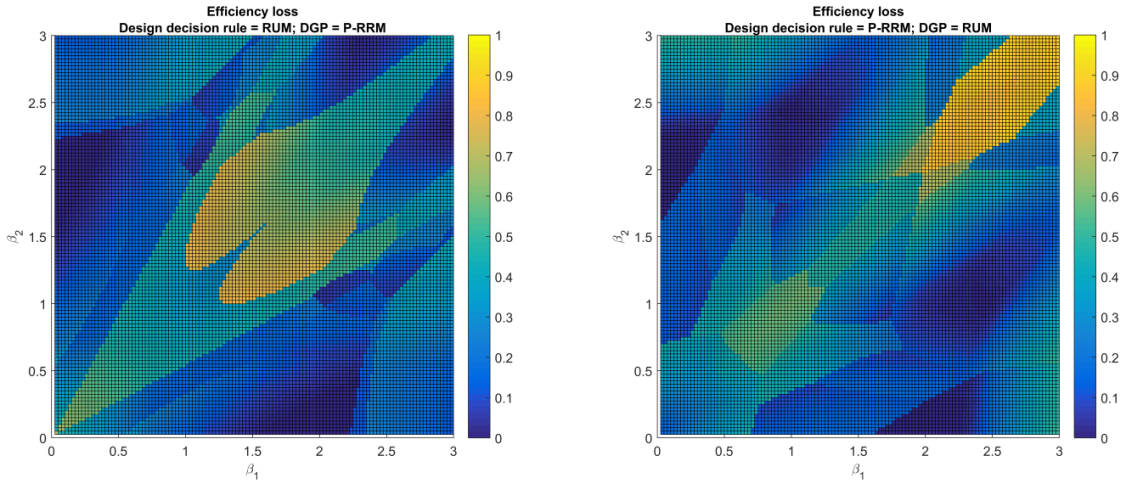

Figure 1 Efficiency loss design set-up 1: RUM and P-RRM design

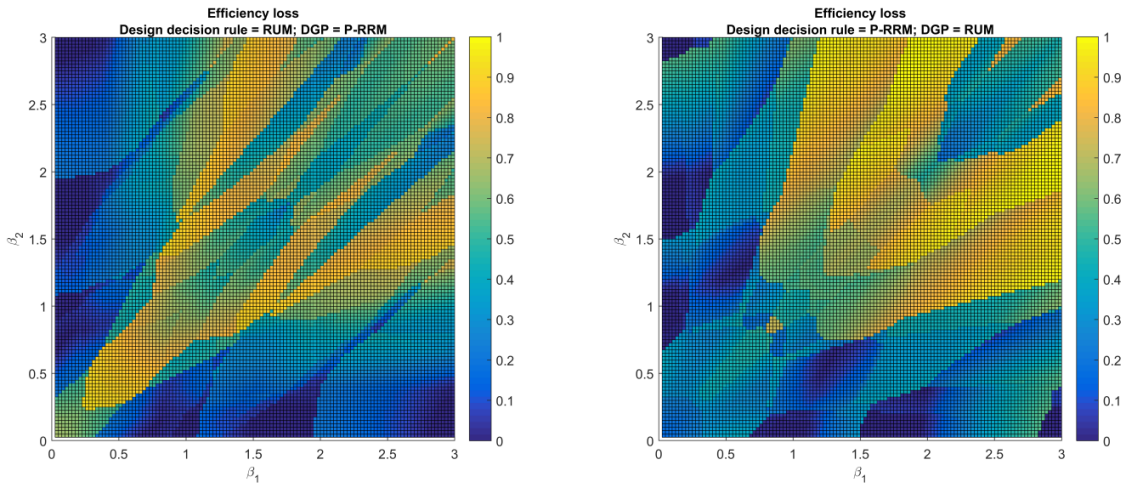

Figure 2 Efficiency loss design set-up 2: RUM and P-RRM design

3.3 Generating designs that are robust towards decision rule uncertainty

Motivated by the results presented in section 3.2, we develop a methodology to construct designs that are robust towards the uncertainty on the side of the analyst on the underlying decision rule. In particular, to construct such designs we minimise a composite, rather than a single, efficiency measure, see equation **9**. This composite measure takes into account the probabilities of each decision rule being the DGP decision rule. It does so by weighing the contribution of the D-errors associated with each decision rule to the composite D-error. Basically, this composite D-error is thus to be seen as a discrete mixture of D-errors (one for each decision rule).

The weights w_r in equation **9** are set by the analyst, and may reflect his or her prior believes on what is the most likely decision rule. Alternatively, in the case where decision rule heterogeneity is explicitly taken into account by the analyst e.g. using Latent Class (LC) models, it may reflect his or her believes concerning the 'market shares' of the decision rules within the target population. In practice, in many situations the analyst may not have prior indications for one decision rule being more likely than another, nor prior believes on the market shares of decision rules within the target population. In that case, the analyst may simply opt to set all weights to be equal. A more informed way to learn about the distribution of decision rules in the target population however is to estimate Latent class (LC) models based on data from a small pilot study, which then needs to be conducted prior to deploying the full survey. In the LC model the individual classes can be employed to embed different decision rules (see e.g. Hess et al. 2012). The insights from such a pilot study can then be used to set the weights and design the experiment for the full survey in a statistically optimal way.

$$D_{composite} = \sum_{r=1\ldots R} w_r \cdot D_r\text{-error} \quad where \sum_{r=1\ldots R} w_r = 1 \qquad (9)$$

To assess the efficacy of this methodology in terms of generating experimental designs that are robust with respect to decision rule misspecification, we applied the methodology in the context of design set-ups 1 and 2 and repeated the analyses presented in the previous section. Specifically, we identified those designs that minimised the composite D-error in equation **10** for each combination of β_1 and β_2 and assessed the loss of efficiency (equation 8). As can be seen, we used equal weights (i.e., $w_{RUM} = 0.5, w_{P-RRM} = 0.5$), reflecting the situation in which the analyst does not have a prior beliefs concerning which is the most prevalent decision rule in the target population.

$$D_{composite} = \frac{1}{2} D_{RUM} + \frac{1}{2} D_{P-RRM} \qquad (10)$$

Figure 3 and Figure 4 show the efficiency loss for the optimal discrete mixture design for, respectively, design set-ups 1 and 2 as a function of the model parameters β_1 and β_2. The efficiency loss is relative to the best design that could have been constructed in case the correct design decision rule was known and used (rather than a mixture of decision rules). The left-hand plots show the efficiency loss of P-RRM models when applied on discrete mixture efficient

designs; the right-hand plots show the efficiency loss of the linear-additive RUM model when applied on discrete mixture efficient designs. Again, we explore the effect of the size of the parameters using the same range as in our analysis in section 3.2.

Comparing Figure 3 with Figure 1 and Figure 4 with Figure 2 reveals that the discrete mixture efficient designs strongly improve the robustness towards the analyst's uncertainty associated with the underlying decision rule. In Figure 3 and Figure 4 the loss in efficiency is generally small and rarely exceeds 0.5. In contrast, designs optimal for a single decision rule (see above) show large parameters spaces in which the loss in efficiency is very substantial, exceeding 0.9.

We also conducted the same analyses using different weights: $w_{RUM} = 0.75, w_{P-RRM} = 0.25$ and $w_{RUM} = 0.25, w_{P-RRM} = 0.75$ (results not presented here for reasons of succinctness). In line with intuition, these show that taking different decision rules into account when creating the efficient design will decrease the statistical efficiency as compared to when the design and model decision rule perfectly match, but doing so will increase the robustness of the design towards decision rule misspecification (i.e. the case where the design is optimised for decision rule A, but decision rule B is estimated).

In sum, we believe the presented methodology to account for the fact that the underlying decision rule is unknown to the analyst is promising. An alternative approach to create such robust designs would be to employ Bayesian efficient designs in combination with the μRRM model (which generalises both the P-RRM and the linear-additive RUM model). However, although pursuing this idea would be an interesting avenue for further research, the above methodology is much easier to implement.

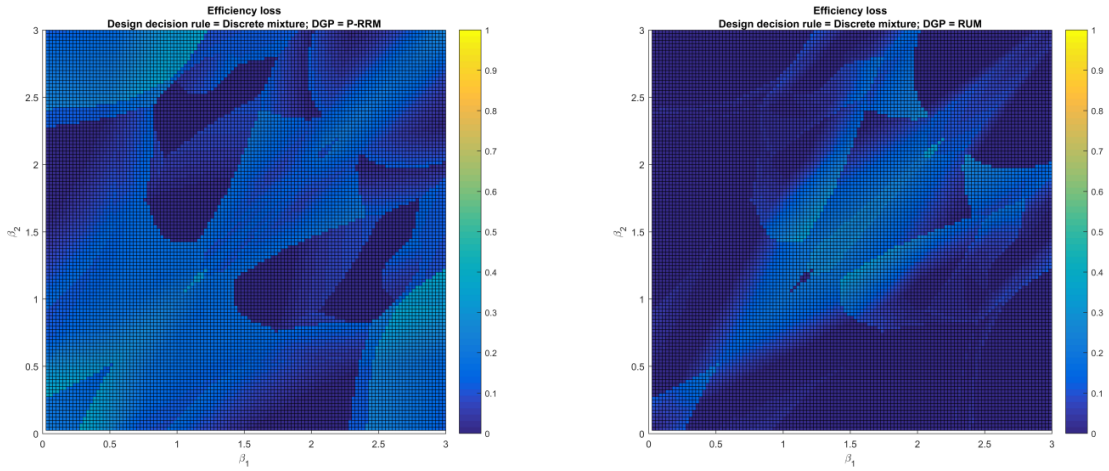

Figure 3 Efficiency loss design set-up 1: Discrete mixture design

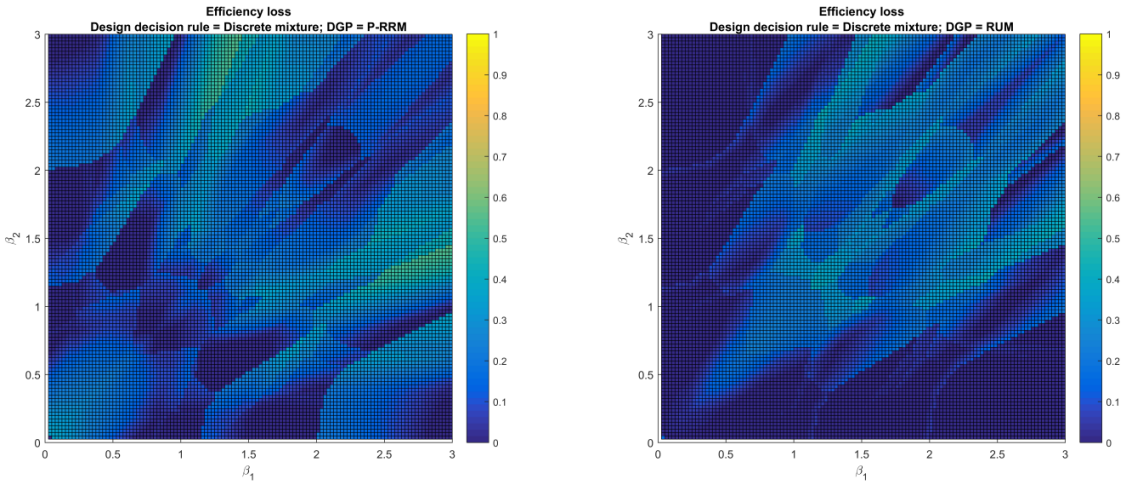

Figure 4 Efficiency loss design set-up 2: Discrete mixture design

4 Empirical case study: Value-of-Travel time

To empirically investigate the importance of the design decision rule on the statistical efficiency of efficient designs we conducted a case study using real data. More specifically, we conducted a Value-of-Travel Time (VoT) SC experiment. SC experiments are widely used to infer the VoT, which, in turn, plays a crucial role in transport infrastructure evaluations (Hess et al. 2005; Börjesson and Eliasson 2014; Kouwenhoven et al. 2014; Ojeda-Cabral et al. 2016a; Wardman et al. 2016). In addition, VoT SC experimental designs are typically relatively small in terms of the

number of alternatives and attributes. This makes VoT SC designs particularly suitable in this context and congruent with our analytical analyses in section 3.

To ensure that our results regarding the (non-)robustness of efficient designs remain tractable and easily interpretable we needed to strongly constrain our design set-up. In particular, we used a relatively small number of choice sets and attribute levels, and we used locally optimal designs (e.g., rather than more complex Bayesian designs). This 'simple' approach enables us to perform an unambiguous analysis of the main determinants of design (non-)robustness.

4.1 The experimental design

Table 2 shows the SC design set-up. Note that this design set-up is very similar to Set-up 1 of Section 3. To infer the VoT we present respondents route choices. Each route choice set consists of 3 unlabelled alternatives. Furthermore, we use just two attributes: travel time and travel cost. Attribute levels are selected as follows: the range of the travel times was chosen such that they are in consonance with the range of the travel times presented in previous European VoT SC-experiments, which is usually in the order of 10 to 15 minutes. The minimum travel time is set at 23 minutes, and the maximum at 35 minutes, with equally spaced 4 minutes intervals. By using 4 minutes intervals we avoid unresolved issues relating to the valuation of small travel time savings (Welch and Williams 1997; Daly et al. 2014). As shown in section 3.1, for this design set-up in total 16 choice sets can be created without dominant or duplicate alternatives. Given that we also use 4 choice sets per respondent, this implies a total of $\binom{16}{4}$ =1,820 possible experimental designs.

Table 2: Experimental design set-up VoT

Number of alternatives	3
Number of generic attributes	2
Number of choice sets per respondent	4
Attribute levels Time [minutes]	23, 27, 31, 35
Attribute levels Cost [euro]	3, 4, 5, 6

To construct the efficient designs, we need priors for β_{Time} and β_{Cost}. In the most recent Dutch VoT study (Kouwenhoven et al. 2014), a VoT of €9 per hour (€0.15/minute) was found for car drivers in the commute. However, not only the ratio of the parameters, but also the scale (thus the amount

of observed versus the amount of unobserved utility) matters for efficient experimental designs. For linear-additive RUM-MNL models, any combination of β_{Time} and β_{Cost} that satisfies $\beta_{Time}/\beta_{Cost} = 0.15$ yields a VoT of €9.[8] In other words, we need to set one of the βs for normalization purposes. Based on a pre-test with under graduate students, we set β_{Time} to -0.15, implying β_{Cost} = -1.00, both for the RUM and P-RRM designs.

To find the optimal D-efficient RUM, P-RRM, and discrete mixture designs we computed the associated D-error statistics across all 1,820 possible designs, given the prior parameters. The designs with the lowest D-error were selected. For the mixture design, we selected the design with the lowest composite D-error. In the absence of information on the market shares of decision rules in the target population and in consonance with the analyses presented in section 3.3, we use equal weights (i.e., $w_{RUM} = 0.5, w_{P-RRM} = 0.5$). Furthermore, as we focus exclusively on statistical efficiency, no attention is paid to attribute level balance, utility balance, and other potential design considerations. It goes without saying that when constructing SC experiments to infer the VoT numerous such other design considerations would also need to be considered (see e.g. Hess et al. 2017).

Table 3 shows the obtained optimal RUM (left), P-RRM (middle) and discrete mixture (right) designs. The choice sets of these three designs are combined into one SC experiment. Note that the optimal RUM, P-RRM and discrete mixture designs have one choice set in common, being choice set number 2. The discrete mixture design also has one other choice set in common with the RUM design and another choice set in common with the P-RRM design. Hence, all together, eight unique choice sets are needed to construct all three optimal designs. Finally, note that by inspecting the choice sets it is difficult (if not impossible) to tell which design is efficient given either presupposed decision rule. For the interested reader, Appendix B shows the choice probabilities associated with each optimal design.

In line with the findings in Section 3, the D-error statistics reveal that the design optimised for RUM is statistically relatively inefficient for P-RRM, and vice versa. Specifically, the loss of efficiency in case the design decision rule is RUM, while the DGP decision rule is P-RRM is $L = 1 - 0.0110/0.0194 = 0.57$; and, the loss of efficiency in case the design decision rule is P-RRM while the DGP decision rule is RUM is $L = 1 - 0.0178/0.0317 = 0.44$. The discrete mixture

[8] In the context of RRM VoT is not univocally defined in a welfare theoretic sense, see Dekker (2014) for more details.

design, in contrast, performs quite well in terms of statistical efficiency, regardless of the DGP decision rule. When the DGP decision rule is P-RRM, the loss of efficiency of the discrete mixture design is $L = 1 - 0.0110/0.0133 = 0.17$; when the DGP decision rule is RUM the loss of efficiency is just $L = 1 - 0.0178/0.0184 = 0.03$.

Table 3: Experimental designs

	RUM design						P-RRM design						Discrete mixture design					
D_{RUM}	0.0178						0.0317						0.0184					
D_{RRM}	0.0194						0.011						0.0133					
Choice task	Route A		Route B		Route C		Route A		Route B		Route C		Route A		Route B		Route C	
	TT	TC	TT	TC	TT	TC	TT	TC	TT	TC	TT	TC	TT	TC	TT	TC	TT	TC
1	23	5	31	4	35	3	23	5	27	4	35	3	23	6	27	4	35	3
2	23	6	27	4	35	3	23	6	27	4	35	3	23	6	27	5	31	3
3	23	6	31	5	35	3	23	6	27	5	35	3	23	6	27	5	35	4
4	27	6	31	5	35	3	23	6	27	5	35	4	23	6	31	5	35	3

TT = Travel Time; TC = Travel Cost

To further explore the optimal designs, Figure 5 visualizes the choice sets associated with the three optimal designs. The left plot shows the 4 choice sets belonging to the optimal RUM design; the middle plot shows the 4 choice sets belonging to the optimal P-RRM design; and the right plot shows that the 4 choice sets belonging to the optimal discrete mixture design. The black points in these plots represent alternatives. The blue lines connect the alternatives within a single choice set. Note that all blue lines have a negative slope. This is a result of the fact that we removed choice sets containing dominant alternatives. Figure 5 reveals that the RUM and P-RRM designs are in fact quite different. Although they do have one choice set in common, it is noteworthy to see that in the P-RRM design all four blue lines – connecting the alternatives within a choice set – have a convex shape. In contrast, in the RUM design, just one line is convex, the other three being concave. Furthermore, the discrete mixture design appears to be symmetric along the diagonal. However, it is unclear how these visual observations relate to statistical efficiency and to the behavioural properties of the underlying decision rules.

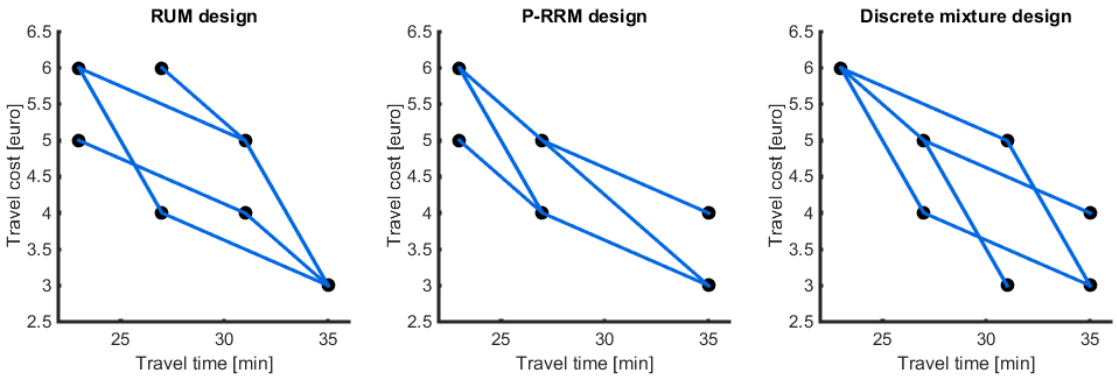

Figure 5: Visualisation of choice sets

4.2 Data collection

The data collection took place in The Netherlands in May 2016. Respondents were recruited using a panel company (TNS NIPO). Only respondents who commuted two or more days per week by car were admitted to the survey. In total 106 respondents completed the full survey. Socio-demographic characteristics of the respondents were provided by the panel company, although for 17 respondents socio-demographic data were missing due to registration errors on the side of the respondents. A relatively balanced sample has been obtained in terms of gender, age, education and income. Sample statistics can be found in Appendix C.

Each respondent is confronted with all eight choice sets that were needed to construct all three optimal designs. The first choice set that is presented to each respondent is choice set 2 (see Table 3), as this choice set is present in all optimal efficient designs. To avoid ordering effects, the remaining choice sets belonging to the RUM, P-RRM and the discrete mixture designs were presented alternately in a random order. The order of the alternatives within each choice set was also shuffled in order to reduce a possible sense of repetition on the side of the respondent. Finally, as our contract with the panel company entitled us to give each respondent 10 choice sets, two additional choice sets were added at the end, but these choice sets are not used in the present analyses.

4.3 Results

Table 4 shows estimation results, split out between data collected using the optimal RUM design, data collected using the optimal P-RRM design, and data collected using the optimal discrete mixture design. On each separate data set we estimated a RUM-MNL and a P-RRM-MNL model.

Based on Table 4 a number of observations can be made. Firstly, in line with expectations, we see that the RUM model attains the lowest empirical D-error when estimated based on the RUM design data, while the P-RRM-MNL model attains the lowest empirical D-error when estimated on the P-RRM design data. More specifically, in case the design decision rule matches the estimation decision rule the D-error is about two times as small as in the case where it does not. Note that here we use the term 'empirical D-error' to refer to the fact that we use the estimated AVC matrix to compute the D-error. Additional analyses (not presented here for reasons of succinctness) show that the empirical and analytically derived D-errors correspond closely, meaning that similar inferences would be made using the analytical D-errors.

Secondly, Table 4 shows that the discrete mixture design is statistically efficient under both a RUM and P-RRM estimation decision rule. The empirical D-errors obtained using the discrete mixture design are only slightly higher than those obtained when using the design decision rule that matches the estimation decision rule. This shows that the methodology presented in section 3.3 to derive robust efficient designs works in practice, at least in the context of these empirical data.

Thirdly, and we consider this a striking finding, results suggest that inferences concerning which model (in casu: RUM or P-RRM) best describes the choice data based on the final Log-likelihood metric is influenced by the design decision rule. More specifically, we see that the RUM model obtains the best model fit when the design decision rule is RUM, while the P-RRM model obtains the best model fit when the design decision rule is P-RRM. In fact, the Ben-Akiva-Swait test (Ben-Akiva and Swait 1986) for non-nested models shows that the obtained model fit differences are highly significant ($p < 0.001$). In addition, the ratios of the parameters of the same model are also very different from one another across the subsets of the data (i.e., the RUM design data and the P-RRM design data). For instance, the ratio of β_{time} over β_{cost} of the RUM model when estimated on RUM design data is 0.276 (implying a VoT of €16.56 per hour), while when estimated on the P-RRM design data it equals 0.221 (implying a VoT of €13.26 per hour). An independent sample *t*-test shows that these values are statistically highly significantly different

from one another ($p < 0.001$). RUM-VoT estimated on the Mixture design data lies in between these two values. These results add to the growing concerns in the travel behaviour research community about potential bias induced by efficient SC experimental designs (e.g. Fosgerau and Börjesson 2015; Ojeda-Cabral et al. 2016b). Further research is needed to further explore these issues. We believe the methodology presented in this study provides a good starting point for such work.

Table 4: Estimation results

DATA SET	**RUM design data**				**P-RRM design data**				**Mixture design data**			
Individuals	106				106				106			
Observations	424				424				424			
LL(0)	-465.8				-465.8				-465.8			
RUM-MNL												
LL(β)	-437.2				-448.9				-451.2			
ρ^2	0.06				0.04				0.03			
Empirical D-error	0.0135				0.0306				0.0149			
	Est	Std err	t-test	p-val	Est	Std err	t-test	p-val	Est	Std err	t-test	p-val
B_time	-0.196	0.026	-7.49	0.00	-0.187	0.033	-5.66	0.00	-0.137	0.027	-5.01	0.00
B_cost	-0.711	0.104	-6.82	0.00	-0.848	0.156	-5.45	0.00	-0.568	0.107	-5.33	0.00
Ratio	0.276	0.016	16.9	0.00	0.221	0.012	17.8	0.00	0.242	0.018	13.2	0.00
P-RRM-MNL												
LL(β)	-448.7				-444.3				-447.0			
ρ^2	0.04				0.05				0.04			
Empirical D-error	0.0134				0.0071				0.0092			
	Est	Std err	t-test	p-val	Est	Std err	t-test	p-val	Est	Std err	t-test	p-val
B_time	-0.137	0.024	-5.74	0.00	-0.098	0.015	-6.35	0.00	-0.107	0.019	-5.57	0.00
B_cost	-0.460	0.093	-4.97	0.00	-0.497	0.085	-5.88	0.00	-0.491	0.083	-5.92	0.00
Ratio	0.298	0.031	9.72	0.00	0.197	0.022	8.91	0.00	0.219	0.023	9.33	0.00

5 Conclusion and discussion

This paper sheds light on the importance of the assumption concerning the underlying decision rule for generating efficient experimental designs. While previous studies investigating efficient designs have been based on the – often implicit – assumption that decision-makers are (linear-additive) utility maximisers, we show that efficient designs can relatively easily be constructed based on the notion that decision-makers minimise regret when making choices. Thereby, we have extended the toolbox for designing efficient experimental designs, opening up the possibility to investigate robustness of efficient designs towards the assumption on the underlying decision

rule. Using this theoretical result, and using a series of analytical and empirical (based on a data set specifically collected for this purpose) tests, we show that conventional RUM efficient designs can be statistically highly inefficient (i.e., non-robust) when the true DGP is based on regret minimization behaviour. This reinforces the notion that the developed regret-based efficient design methodology is a potentially valuable addition to the choice modeller's toolbox. Furthermore, we present a methodology to construct efficient designs that are statistically robust towards the uncertainty on the side of the analyst concerning the underlying decision rule. Based on an analytical and empirical assessment we find that these so-called mixture designs are promising, in the sense that they are robust towards the decision rules that were being considered (in casu: RUM and RRM).

Finally, we would like to point out several limitations to this study, providing avenues for further research. Firstly, regarding our analytical work, our robustness analyses are based on one alternative decision rule: the P-RRM model. These analyses could be extended by investigating possibilities to construct efficient design for other non-RUM models, such as reference dependent models of choice, e.g. based on prospect theory (Kahneman and Tversky 1979). Secondly, to keep the results tractable we used a relatively small number of choice sets per design. Future research may explore efficiency and robustness losses in the context of larger number of choice sets per design. Thirdly, throughout our analytical analyses we have used perfect priors. Future research may consider investigating the case in which priors are misspecified. Currently it is, for instance, unclear to what extent P-RRM efficient designs may be more, or less, sensitive towards prior misspecification than RUM efficient designs. Fourthly, in this research we have focussed on locally optimal designs. Future research may explore the robustness of Bayesian efficient designs towards decision rule misspecification. Fifthly, in the mixture designs in our study we used equal weights. More research is needed to find best strategies to set the weights when using mixture designs. Sixthly, apart from removing choice sets with dominant alternatives, we have not restricted our designs, e.g., by imposing attribute-level balance or other design considerations. In practice, attribute level balance is considered by many analysts a desirable property. Although it is not clear if and to what extent imposing attribute level balance would affect our results, it is worthwhile to further explore this. Seventhly, given that our empirical analyses are based on just one data set it is advisable to repeat these analyses based on multiple other data sets. This will clarify to what extent our empirical findings e.g., regarding the efficacy of the proposed method to construct efficient designs that are robust towards decision rule uncertainty, are more generally valid. Finally, our research raises robustness concerns related to efficient experimental designs.

This is however not to say that the more traditional approach based on orthogonal designs is 'the solution'. Although orthogonal designs have been found to be more robust than efficient designs under certain conditions (Walker et al. 2015), more research is needed to better understand the robustness of orthogonal designs in the context of different decision rules.

Acknowledgements

The first and third authors gratefully acknowledge support from the Netherlands Organisation for Scientific Research (NWO), in the form of VIDI Grant 016-125-305.

Appendix A: Second-order derivative of the Log-likelihood function w.r.t. its parameters for the P-RRM-MNL model

This appendix derives the second-order derivatives of the Log-likelihood function of the P-RRM model w.r.t. its parameters. The second-order derivatives are needed to construct the Fisher information matrix, which in turn is needed to evaluate the statistical efficiency of a design. To derive the second-order derivatives we first derive the first-order derivatives.

First-order derivative

The simplest way to find the first-order derivatives of the Log-likelihood function w.r.t. its parameters for the P-RRM-MNL model is to take advantage of the results for the linear-additive RUM-MNL model, and apply the chain rule:

$$\frac{\partial \log L(\beta \mid X,Y)}{\partial \beta_m} = \frac{\partial \log L}{\partial V_{js}} \cdot \frac{\partial V_{js}}{\partial \beta_m} = \sum_{s=1}^{S}\sum_{j=1}^{J}\left(y_{js} - P_{js}\right)x_{jms}$$

Substitution of $V_{js} = -\tilde{R}_{js}$ and applying the chain rule directly yields the first-order derivative of the Log-likelihood function of the P-RRM-MNL model with respect to (generic) parameter β_m:

$$\frac{\partial LL^{P-RRM}(\beta \mid X,Y)}{\partial \beta_m} = -\sum_{s=1}^{S}\sum_{j=1}^{J}\left(y_{js} - P_{js}\right)\tilde{x}_{jms}^{P-RRM}$$

$$where\, \tilde{x}_{jms}^{P-RRM} = \begin{cases} \dfrac{2}{J_s}\displaystyle\sum_{l \neq j} \max\left(0, x_{lms} - x_{jms}\right) \; if\, \beta_m > 0 \\ \dfrac{2}{J_s}\displaystyle\sum_{l \neq j} \min\left(0, x_{lms} - x_{jms}\right) \; if\, \beta_m < 0 \end{cases}$$

Hence, the only differences between the P-RRM-MNL and the RUM-MNL model in terms of their first-order derivative are the appearance of a minus sign and the replacement of the attribute level x_{jms} with its (choice set size corrected) P-RRM counterpart $\tilde{x}_{jms}^{P-RRM}$. This is in line with intuition given the facts that (1) regret can be conceived as negative utility, and (2) under the P-RRM model's regrets are linear-additive.

Second-order derivative

Below we derive the second-order derivatives for combinations of generic and alternative specific parameters.

Second-order derivative for two generic parameters

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial \beta_{m_1}\partial \beta_{m_2}} = \frac{\partial}{\partial \beta_{m_2}}\frac{\partial LL^{P-RRM}(\beta \mid X,Y)}{\partial \beta_{m_1}} = -\frac{\partial}{\partial \beta_{m_2}}\left(\sum_{s=1}^{S}\sum_{j=1}^{J} y_{js} - P_{js}\right)\tilde{x}_{jm_1s}^{P-RRM}$$

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = \sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} \frac{\partial}{\partial\beta_{m_2}} P_{js}$$

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = \sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} \frac{\partial}{\partial\beta_{m_2}} \left(\frac{e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}}}{\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}} \right)$$

Applying the quotient rule:

$$f = e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}}, \qquad \frac{\partial f}{\partial\beta_{m_2}} = -\tilde{x}_{jm_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}}$$

$$g = \sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}, \qquad \frac{\partial g}{\partial\beta_{m_2}} = -\sum_{i=1}^{J} \tilde{x}_{im_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}$$

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = \sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} \left(\frac{\left[-\tilde{x}_{jm_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}} \right]\left[\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}} \right] - \left[e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}} \right]\left[-\sum_{i=1}^{J} \tilde{x}_{im_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}} \right]}{\left(\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}} \right)^2} \right)$$

Rearranging terms yields:

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = \sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} \left(\frac{-\tilde{x}_{jm_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}}}{\left(\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}} \right)} - \frac{e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{jms}^{P-RRM}}}{\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}} \cdot \frac{-\sum_{i=1}^{J} \tilde{x}_{im_2s}^{P-RRM} \cdot e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}}{\sum_{i=1}^{J} e^{-\sum_{m=1}^{M}\beta_m \tilde{x}_{ims}^{P-RRM}}} \right)$$

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J} \tilde{x}_{jm_1s}^{P-RRM} \left(\tilde{x}_{jm_2s}^{P-RRM} P_{js} - P_{js} \cdot \sum_{i=1}^{J} \tilde{x}_{im_2s}^{P-RRM} P_{is} \right)$$

Finally, we obtain the second-order derivative of the Log-likelihood function w.r.t. the generic model parameters β_{m_1} and β_{m_2}:

$$\frac{\partial^2 LL^{P-RRM}(\beta \mid X,Y)}{\partial\beta_{m_1}\partial\beta_{m_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J}\tilde{x}_{jm_1s}^{P-RRM}P_{js}\left(\tilde{x}_{jm_2s}^{P-RRM} - \sum_{i=1}^{J}\tilde{x}_{im_2s}^{P-RRM}P_{is}\right)$$

<u>Second-order derivative for generic and alternative specific parameters</u>

Likewise, the second-order derivative of the Log-likelihood function w.r.t. the alternative specific parameter $\breve{\beta}$ and generic parameter β can be derived. The derivation and notation is similar as under RUM, see Rose and Bliemer (2005):

$$\frac{\partial^2 LL}{\partial\breve{\beta}_{j_1m_1}\partial\beta_{m_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J}x_{j_1m_1s}P_{j_1s}\left(\tilde{x}_{j_1m_2s}^{P-RRM} - \sum_{i=1}^{J}\tilde{x}_{im_2s}^{P-RRM}P_{is}\right)$$

$$\frac{\partial^2 LL}{\partial\beta_{m_1}\partial\breve{\beta}_{j_2m_2}} = -\sum_{s=1}^{S}x_{j_2m_2s}P_{j2s}\left(\tilde{x}_{j_2m_1s}^{P-RRM} - \sum_{i=1}^{J}\tilde{x}_{im_1s}^{P-RRM}P_{is}\right)$$

<u>Second-order derivative for two alternative specific parameters</u>

Finally, the second-order derivative of the Log-likelihood function w.r.t. two alternative specific parameters $\breve{\beta}_{m_1}$ and $\breve{\beta}_{m_2}$ can be derived. Since treatment of dummy variables is similar under RUM and RRM these derivatives are equal to those for the RUM model. The derivation is can be found in Rose and Bliemer (2005). To compute the second-order derivatives on the diagonal cells, we use:

If $j_1 = j_2$

$$\frac{\partial^2 LL}{\partial\breve{\beta}_{j_1m_1}\partial\breve{\beta}_{j_2m_2}} = -\sum_{s=1}^{S}x_{j_1m_1s}x_{j_2m_2s}P_{j_1s}\left(1-P_{j_2s}\right)$$

and, to compute the off-diagonal cells, we use:

If $j_1 \neq j_2$

$$\frac{\partial^2 LL}{\partial\breve{\beta}_{j_1m_1}\partial\breve{\beta}_{j_2m_2}} = \sum_{s=1}^{S}x_{j_1m_1s}x_{j_2m_2s}P_{j_1s}P_{j_2s}$$

Appendix B: Choice probabilities of efficient designs

	RUM design			P-RRM design			Discrete mixture design		
	P(Y=A)	P(Y=B)	P(Y=C)	P(Y=A)	P(Y=B)	P(Y=C)	P(Y=A)	P(Y=B)	P(Y=C)
RUM MODEL									
1	0.33	0.27	0.40	0.27	0.40	0.33	0.12	0.48	0.40
2	0.12	0.48	0.40	0.12	0.48	0.40	0.12	0.17	0.71
3	0.19	0.16	0.65	0.17	0.26	0.57	0.27	0.40	0.33
4	0.12	0.17	0.71	0.27	0.40	0.33	0.19	0.16	0.65
P-RRM MODEL									
1	0.24	0.41	0.36	0.22	0.56	0.22	0.07	0.67	0.26
2	0.07	0.67	0.26	0.07	0.67	0.26	0.13	0.32	0.55
3	0.18	0.30	0.52	0.18	0.46	0.35	0.22	0.56	0.22
4	0.13	0.32	0.55	0.22	0.56	0.22	0.18	0.30	0.52

Appendix C: Sample statistics

Variable	Sample frequency	Percentage [%] in sample
Gender		
Male	44	42%
Female	45	42%
Missing	17	16%
Age		
18 to 24 yr.	2	2%
25 to 34 yr.	24	23%
35 to 44 yr.	25	24%
45 to 54 yr.	21	20%
55 to 64 yr.	15	14%
65 to 74 yr.	2	2%
Missing	17	16%
Completed education		
No education	0	0%
Elementary school	7	7%
Lower education	5	5%
Middle education	39	37%
Higher education	34	32%
University education	4	4%
Missing	17	16%
Income		
I < €9,400	2	2%
€9,400 ≤ I < €14,700	4	4%
€14,700 ≤ I < €20,600	5	5%
€20,600 ≤ I < €33,500	14	13%
€33,500 ≤ I < €67,000	37	35%
I ≥ €67,000	27	25%
Missing	17	16%

References

Ben-Akiva, M. & Swait, J. (1986). The Akaike Likelihood Ratio Index. *Transportation Science*, *20(2)*, 133-136.

Bliemer, M. C. & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, *44(6)*, 720-734.

Bliemer, M. C., Rose, J. M. & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological*, *43(1)*, 19-35.

Bliemer, M. C. J. & Rose, J. M. (2011). Experimental design influences on stated choice outputs: An empirical study in air travel choice. *Transportation Research Part A: Policy and Practice*, *45(1)*, 63-79.

Bliemer, M. C. J., Rose, J. M. & Chorus, C. G. (2017). Detecting dominancy in stated choice data and accounting for dominancy-based scale differences in logit models. *Transportation Research Part B: Methodological*, *102*, 83-104.

Boeri, M. & Longo, A. (2017). The importance of regret minimization in the choice for renewable energy programmes: Evidence from a discrete choice experiment. *Energy Economics*, *63*, 253-260.

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

Börjesson, M. & Eliasson, J. (2014). Experiences from the Swedish Value of Time study. *Transportation Research Part A: Policy and Practice*, *59*, 144-158.

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

Bunch, D. S., Louviere, J. J. & Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. *University of California Davis Graduate School of Management Working Paper*, 11-96.

Carlsson, F. & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health economics*, *12(4)*, 281-294.

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., Arentze, T. A. & Timmermans, H. J. P. (2008). A random regret-minimization model of travel choice. *Transportation Research Part B-Methodological*, *42(1)*, 1-18.

Daly, A., Tsang, F. & Rohr, C. (2014). The value of small time savings for non-business travel. *Journal of Transport Economics and Policy (JTEP)*, *48(2)*, 205-218.

Danthurebandara, V. M., Yu, J. & Vandebroek, M. (2011). Effect of choice complexity on design efficiency in conjoint choice experiments. *Journal of Statistical Planning and Inference*, *141(7)*, 2276-2286.

Dekker, T. (2014). Indifference based value of time measures for Random Regret Minimisation models. *Journal of Choice Modelling*, *12(0)*, 10-20.

Ferrini, S. & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, *53(3)*, 342-363.

Fosgerau, M. & Börjesson, M. (2015). Manipulating a stated choice experiment. *Journal of Choice Modelling*, *16*, 43-49.

Goos, P., Vermeulen, B. & Vandebroek, M. (2010). -optimal conjoint choice designs with no-choice options for a nested logit model. *Journal of Statistical Planning and Inference*, *140(4)*, 851-861.

Guevara, C. A. & Fukushi, M. (2016). Modeling the decoy effect with context-RUM Models: Diagrammatic analysis and empirical evidence from route choice SP and mode choice RP case studies. *Transportation Research Part B: Methodological*, *93, Part A*, 318-337.

Hensher, D. A., Barnard, P. O. & Truong, T. P. (1988). The role of stated preference methods in studies of travel choice. *Journal of Transport Economics and Policy*, 45-58.

Hensher, D. A., Greene, W. H. & Ho, C. Q. (2016). Random Regret Minimization and Random Utility Maximization in the Presence of Preference Heterogeneity: An Empirical Contrast. *Journal of Transportation Engineering*, 04016009.

Hess, S., Bierlaire, M. & Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, *39(2–3)*, 221-236.

Hess, S., Daly, A., Dekker, T., Cabral, M. O. & Batley, R. (2017). A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological*, *96*, 126-149.

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

Huber, J., Payne, J. W. & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, *9(1)*, 90-98.

Huber, J. & Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research*, *33(3)*, 307-317.

Jang, S., Rasouli, S. & Timmermans, H. (2016). Incorporating psycho-physical mapping into random regret choice models: model specifications and empirical performance assessments. *Transportation*, 1-21.

Kahneman, D. & Tversky, A. (1979). Prospect Theory - Analysis of Decision under Risk. *Econometrica*, *47(2)*, 263-291.

Kanninen, B. J. (1993a). Design of Sequential Experiments for Contingent Valuation Studies. *Journal of Environmental Economics and Management*, *25(1)*, S1-S11.

Kanninen, B. J. (1993b). Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation. *Land Economics*, *69(2)*, 138-146.

Kessels, R., Goos, P. & Vandebroek, M. (2006). A Comparison of Criteria to Design Efficient Choice Experiments. *Journal of Marketing Research*, *43(3)*, 409-419.

Kessels, R., Jones, B., Goos, P. & Vandebroek, M. (2011). The usefulness of Bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry*, *27(3)*, 173-188.

Kouwenhoven, M., de Jong, G. C., Koster, P., van den Berg, V. A. C., Verhoef, E. T., Bates, J. & Warffemius, P. M. J. (2014). New values of time and reliability in passenger transport in The Netherlands. *Research in Transportation Economics*, *47*, 37-49.

Leong, W. & Hensher, D. A. (2012). Embedding Decision Heuristics in Discrete Choice Models: A Review. *Transport Reviews*, *32(3)*, 313-331.

Louviere, J. J., Hensher, D. A. & Swait, J. D. (2000). *Stated choice methods : analysis and applications*. (Cambridge, UK: Cambridge University Press).

Ojeda-Cabral, M., Batley, R. & Hess, S. (2016a). The value of travel time: random utility versus random valuation. *Transportmetrica A: Transport Science*, *12(3)*, 230-248.

Ojeda-Cabral, M., Hess, S. & Batley, R. (2016b). Understanding valuation of travel time changes: are preferences different under different stated choice design settings? *Transportation*, 1-21.

Palhazi Cuervo, D., Goos, P. & Sörensen, K. (2014). Optimal design of large-scale screening experiments: a critical look at the coordinate-exchange algorithm. *Statistics and Computing*, 1-14.

Ramos, G. M., Daamen, W. & Hoogendoorn, S. (2014). A State-of-the-Art Review: Developments in Utility Theory, Prospect Theory and Regret Theory to Investigate Travellers' Behaviour in Situations Involving Travel Time Uncertainty. *Transport Reviews*, *34(1)*, 46-67.

Rose, J. & Bliemer, M. (2005). Constructing Efficient Choice Experiments.

Rose, J. M. & Bliemer, M. C. J. (2009). Constructing Efficient Stated Choice Experimental Designs. *Transport Reviews: A Transnational Transdisciplinary Journal*, *29(5)*, 587-617.

Rose, J. M. & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40(5)*, 1021-1041.

Sándor, Z. & Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, *21(4)*, 455-475.

Scarpa, R. & Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian journal of agricultural and resource economics*, *52(3)*, 253-282.

Sharma, B., Hickman, M. & Nassir, N. (2017). Park-and-ride lot choice model using random utility maximization and random regret minimization. *Transportation*.

Sun, L., Karwan, M. H. & Kwon, C. (2016). Incorporating Driver Behaviors in Network Design Problems: Challenges and Opportunities. *Transport Reviews*, *36(4)*, 454-478.

van Cranenburgh, S. & Chorus, C. G. ((accepted)). Does The Decision Rule Matter For Large-Scale Transport Models? *Transportation Research Part A: Policy and Practice*.

Van Cranenburgh, S., Guevara, C. A. & Chorus, C. G. (2015a). New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice*, *74(0)*, 91-109.

Van Cranenburgh, S. & Prato, C. G. (2016). On the robustness of Random Regret Minimization modelling outcomes towards omitted attributes. *Journal of Choice Modelling*,*(18)*, 51-70.

Van Cranenburgh, S., Prato, C. G. & Chorus, C. (2015b) Accounting for variation in choice set size in Random Regret Minimization models,

Walker, J. L., Wang, Y., Thorhauge, M. & Ben-Akiva, M. (2015). D-efficient of deficient? A robustness analysis of SP experimental design in a VOT estimation context. Presented at the. Annual Meeting of the Transportation Research Board, Washington DC, USA.

Wardman, M., Chintakayala, V. P. K. & de Jong, G. (2016). Values of travel time in Europe: Review and meta-analysis. *Transportation Research Part A: Policy and Practice*, *94*, 93-111.

Welch, M. & Williams, H. (1997). The Sensitivity of Transport Investment Benefits to the Evaluation of Small Travel-Time Savings. *Journal of Transport Economics and Policy*, *31(3)*, 231-254.

Yu, J., Goos, P. & Vandebroek, M. (2008). Model-robust design of conjoint choice experiments. *Communications in Statistics—Simulation and Computation®*, *37(8)*, 1603-1621.

Yu, J., Goos, P. & Vandebroek, M. (2009). Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*, *28(1)*, 122-135.

Yu, J., Goos, P. & Vandebroek, M. (2011). Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*, *28(4)*, 378-388.