Transportation Research Part A 109 (2018) 50–64

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

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

A B S T R A C T

We present a methodology to derive efficient designs for Stated Choice (SC) experiments based on Random Regret Minimisation (RRM) behavioural assumptions. This complements earlier work on the design of efficient SC experiments based on Random Utility Maximisation (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

* Corresponding author.
E-mail address: s.vancranenburgh@tudelft.nl (S. van Cranenburgh).

https://doi.org/10.1016/j.tra.2018.01.001
Received 25 March 2017; Received in revised form 13 December 2017; Accepted 3 January 2018
0965-8564/

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 Maximisation (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 Minimisation (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 (in press) 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 (software to create efficient designs for RRM can be downloaded from www.advancedRRMmodels.com). 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 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 minimisation 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] 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.