Incorporating model uncertainty into the generation of efficient stated choice experiments: A model averaging approach

John M. Rose, Riccardo Scarpa and Michiel C.J. Bliemer

December 2008

John M. Rose*
The University of Sydney
johnr@itls.usyd.edu.au

Riccardo Scarpa
University of Waikato
rscarpa@mngt.waikat.ac.nz

Michiel C.J. Bliemer
Delft University of Technology
m.bliemer@ct.tudelft.nl
Adjunct Professor
The University of Sydney

Abstract

Stated choice (SC) studies typically rely on the use of an underlying experimental design to construct the hypothetical choice situations shown to respondents. These designs are constructed by the analyst, with several different ways of constructing these designs having been proposed in the past. Recently, there has been a move from so-called orthogonal designs to more efficient designs. Efficient designs optimize the design such that the data will lead to more reliable parameter estimates for the model under consideration. The literature dealing with the generation of efficient designs has examined and largely solved the issue of a requirement for *a prior* knowledge of the parameter estimates that will be obtained post data collection. Nevertheless, problems related to the fact that the efficiency of a SC experiment is related to the variance-covariance matrix of the model to be estimated and that different econometric models will have different variance-covariance matrix, thus resulting in different levels of efficiency for the same design, has yet to be addressed. In this paper, we propose the use of a model averaging process over different econometric models to solve this problem. Via the use of a case study, we show that designs generated using the model averaging process prove robust to different model estimation as well as provide decent levels of protection against biased parameter estimates relative to designs generated specifically for a given model type.

1. Introduction

The generation of experimental designs specifically for the purpose of stated choice (SC) surveys has attracted much attention of late. Unlike most data, SC data requires that the analyst design the data in advance by assigning levels to the attributes that define each of the alternatives which respondents are asked to consider. Traditionally, the attribute levels are allocated to the each of the alternatives according to some generated experimental design, with the most common approach being to use a fractional factorial design to generate a series of single alternatives which are then allocated to choice sets using randomised, cyclical, Bayesian or foldover procedures (see e.g., Bunch et al., 1994; Huber and Zwerina, 1996; Kanninen, 2002; Louviere and Woodworth, 1983; Sándor and Wedel, 2001, 2002, 2005). A significant amount of research effort has recently been devoted to how better to assign the attribute levels to alternatives and in turn, the resulting alternatives to choice situations. By and large, these efforts have concentrated on methods to promote greater gains in the statistical efficiency of SC experiments (e.g., Bunch et al., 1994; Carlsson and Martinsson, 2002; Huber and Zwerina, 1996; Kanninen, 2002; Sándor and Wedel, 2001, 2002, 2005). What this stream of research has shown is that how the attribute levels of a design are distributed over the course of the experiment, via the underlying experimental design, will impact to a greater or lesser extent upon whether or not an independent assessment of each attributes contribution to the choices observed to have been made by sampled respondents can be determined.

Primarily, these research efforts have concentrated on the concept of improving the *statistical efficiency* of experimental designs generated for SC studies. In doing so, researchers have defined statistical efficiency in terms of increased precision of the parameter estimates for a fixed sample size. In taking such a definition, statistical efficiency within the literature has therefore been linked to the standard errors likely to be obtained from the experiment, with designs that can be expected to i) yield lower standard errors for a given sample size, or ii) the same standard errors given a smaller sample size, being deemed more statistically efficient. In order to calculate the statistical efficiency of a design, Bunch, et al. (1994), Huber and Zwerina (1996), Sándor and Wedel (2001) and Kanninen (2002), amongst others, have shown that the common use of logit models to analyze discrete choice data requires *a priori* information about the parameter estimates, as well as the final econometric model form to be estimated.

Information on the expected parameter estimates is required in order to calculate the expected utilities for each of the alternatives present within the design. Once known, the expected utilities may in turn be used to calculate the likely choice probabilities. Hence, given knowledge of the attribute levels (the design), expected parameter estimate values and choice probabilities, it becomes a straightforward exercise to calculate the asymptotic variance-covariance (AVC) matrix for the design, from which the expected standard errors can be obtained. By manipulating the attribute levels of the alternatives, for known (assumed) parameter values, the analyst is able to minimize the elements within the AVC matrix, which in the case of the diagonals means lower standard errors and hence greater reliability in the estimates at a fixed sample size, or even at a reduced sample size. The linking of the experimental design generation process to attempts to reduce the asymptotic standard errors of the parameter estimates has resulted in a class of designs known as '*efficient designs*' where a design that, when used in practice, is expected to produce smaller asymptotic standard errors for a given sample size is thought of as being more 'efficient'.

Two main criticisms of the efficient design approach resonate within the literature. Firstly, in order to calculate the AVC matrix for a SC design, the analyst requires *a priori* knowledge of the utility functions for data that will be collected using that design. This is because the values of the AVC matrix map directly to both the attribute levels and the choice probabilities of the alternatives contained within the design. The choice probabilities for a given design are in turn a function of the attribute levels as well as the parameter estimates associated with the attributes. The attribute levels are represented by the design itself, and are determined during the design construction phase. We note that the requirement of determining the attribute levels is independent of the type of experimental design implemented (e.g., if an orthogonal design is used the attribute levels still need to be determined), and hence this issue is not specific to the generation of efficient designs. Unfortunately, the exact parameter values are unlikely to be known at the design construction phase, and as such, the researcher may have to make certain assumptions as to what values (termed priors) these will be in order to generate an efficient design. To overcome this problem, three approaches have been used in the past.

The first approach involves researchers making the strong assumption that all parameter priors are simultaneously equal to zero (e.g., Burgess and Street, 2003; Huber and Zwerina, 1996; Street and Burgess, 2004; Street et al., 2005). This approach has the advantage of not requiring any advanced knowledge of the parameter estimates, however, such designs are likely to be truly optimal in practice only when the parameter estimates are zero. That is, such an approach is likely to require larger sample sizes in practice to achieve given values for the standard errors of the design than otherwise would have been the case had reasonable non-zero priors been used at the time of design construction (see e.g., Bliemer and Rose, 2006, 2009). The second approach that has sometimes been used in the past has been to assume that the parameter priors are non-zero and known with exact certainty (e.g., Carlsson and Martinsson, 2003; Rose and Bliemer, 2005; Scarpa and Rose, 2008). In such an approach, fixed non-zero values are assumed for the prior parameter estimates. This approach is relatively simple to implement, however questions remain as to the robustness of the designs should the assumed parameter priors be misspecified.

Sándor and Wedel (2001) introduced a third approach by relaxing the assumption of perfect *a priori* knowledge of the parameter priors through adopting a Bayesian like approach to the design generation process. Rather than assume a single fixed value for each parameter prior, the efficiency of the design is calculated over a number of simulated draws taken from *prior parameter distributions* assumed by the analyst. Different distributions may be associated with different population moments representing different levels of uncertainty with regards to the true parameter values. In this way, by optimizing the efficiency of the design over a range of possible parameter prior values (drawn from the assumed parameter prior distributions), the design is made as robust as possible, at least for the range assumed[1].

The second criticism often leveled at this design approach is the need to know in advance the precise econometric model that will be estimated once data has been collected. There unfortunately exist many forms of possible discrete choice models that analysts may wish to estimate once SC data has been collected (e.g., MNL, NL, GEV, MMNL). Unfortunately, the log-likelihood functions of different model types are typically different from each other and

---

[1] For example, one would expect for most SC studies that a price parameter will be negative. If the expected parameter value is not known with any level of accuracy, then the analyst may assume say a uniform distribution ranging from -0.1 to -1.0 (or any other negative values depending on the magnitude of the expected parameter), thus optimizing the design over this range of possible population parameter values.

given that the AVC matrix for such models is mathematically given as the inverse of the second derivatives of the models log-likelihood function, the AVC matrix for each type of model will also therefore be different. As such, the construction of efficient designs requires not only an assumption as to the parameter priors assumed, but also what AVC matrix the analyst is attempting to optimize.

Given differences in the AVC matrices of different discrete choice models, attempts at minimising the elements of the AVC matrix assuming one model, may not necessary minimize the elements of the AVC matrix of another model. Similar to the problem involving knowledge of the parameter estimates, the analyst is unlikely to know precisely what model is likely to be estimated in advance, the problem becomes one of having to select the most likely model that will be estimated once data has been collected. In the past, little attention has been given to this issue, with only a small amount of research examining the issue in any depth (see e.g., Bliemer et al., 2009).

In this paper, we propose the use of a model averaging approach to solve this latter problem. This solution involves calculating the AVC matrix for all possible models that might be estimating post data collection and minimising a weighted average measure of efficiency across these possible models. This approach has the advantage of not requiring precise knowledge of the exact econometric model to be estimated whilst allowing for statistical efficiency across a broad spectrum of possible model types. The model averaging approach may also be conducted in conjunction with the use of Bayesian prior parameters, thus allowing for both prior parameter and model estimation uncertainty during the design generation phase. The approach does come at a cost however, with the generated design unlikely to be as efficient a design when compared to a design generated specifically for the particular model type used post data collection.

In this paper, we explore three different possible models in generating SC designs. These include the multinomial logit (MNL), mixed MNL (MMNL, referred to as the random parameters model in some literature) and the error components (EC) models. In discussing the MMNL and EC models, further distinctions are made between the cross sectional and panel formulations of these models. To demonstrate the robustness of this design approach, we construct a case study where we separately construct designs specifically for each model type and compare and contrast these designs with a number of designs constructed using the model averaging approach. In generating the designs, we use combinations of fixed and Bayesian priors, thus demonstrating the complete flexibility of the approach. In constructing the case study, we further compare the effect of using different weights for each model type as well as the impact of using a traditionally generated orthogonal design on the statistical efficiency of the expected results.

The remainder of the paper is organised as follows. In section 2 we discuss the different model types used throughout the paper in detail. In doing so, we discuss differences in the utility specifications, probabilities, log-likelihood functions, and AVC matrices of each of the models. In Section 3, we discuss the process of generating efficient designs as well as outlining statistical measures of the efficiency of designs. Section 3 also discusses the design generation process and statistical efficiency measures when using the model averaging approach, the focus of this paper. In Section 4, we present a case study, the resulting designs of which are presented in Section 5. In generating the designs, we generate different designs that are optimised specifically for different model types which we compare and contrast to designs generated using the model averaging approach as well as to orthogonal designs. Section 6 presents results examining tests conducted on the robustness of the designs

generated, after which Section 7 provides a discussion and conclusion of the general findings presented within the paper.

2. Methodology

In this section, we outline the differences between the MNL, MMNL, and EC models. We begin by examining differences in the utility specifications of these models, after which we discuss the assumptions underlying each of the models log-likelihood functions. We next describe how the various models produce different choice probabilities. Finally, we conclude with an examination of differences in the (co)variance matrices of these models. In discussing the MMNL and EC models, we distinguish between different conceptualizations of these models based on different assumptions related to the interdependence of choice observations within the data.

2.1 Utility specification

Let U_{nsj} denote the utility of alternative *j* perceived by respondent *n* in choice situation *s*. U_{nsj} may be partitioned into three separate components, an observed component of utility, V_{nsj}, an unobserved (or un-modeled) component of utility, η_{nsj}, and an unobserved (and un-modeled) component, ε_{nsj}, such that

$$U_{nsj} = V_{nsj} + \eta_{nsj} + \varepsilon_{nsj}. \tag{1}$$

The observed component of utility is typically assumed to be a linear relationship of observed attribute levels of each alternative, *x*, and their corresponding weights (parameters), β. In the MNL model, the parameter weights for each attribute are invariant over respondents, such that the observed component of utility may be represented as

$$V_{nsj} = \sum_{k=1}^{K} \beta_{jk} x_{nsjk}. \tag{2}$$

Unlike the MNL model, some or all of the parameter weights of the MMNL model are assumed to vary with density $f(\beta \mid \Omega)$ over the sampled population. Assumptions as to how these parameter weights vary over the population have in the past resulted in two different formulations of the MMNL model. One version of the model, known as the cross sectional MMNL formulation, assumes that the parameter weights vary with density over both *n* and *s* suggesting that preference heterogeneity exists both within and between individuals, even when the same individual is observed to make *s* choices within a similar choice context. The second version of the model, known as the panel MMNL formulation, assumes that preferences vary between individuals but not within. The assumption that preferences vary between and not within respondents accounts for the pseudo panel nature of SP data (Ortúzar and Willumsen, 2001; Revelt and Train, 1998; Train, 2003). Equations (3a) and (3b) represent the observed components of utility under both the cross sectional and panel formulations of the MMNL model specifications.

$$V_{nsj} = \sum_{k=1}^{K} \beta_{nsk} x_{nsjk}, \tag{3a}$$

$$V_{nsj} = \sum_{k=1}^{K} \beta_{nk} x_{nsjk}.$$ (3b)

Like the MMNL model, the EC model involves estimation of one or more random parameters. Unlike the MMNL model however, the random parameter estimates of the EC model are associated with alternatives, *j*, not attributes, *x*. To estimate the model, the analyst first specifies a set of dummy variables, with each dummy variable able to appear in the utility specifications of up to J-1 alternatives. Next, generic normally distributed random parameters with means normalised to zero, represented as η_{nsj} in Equation (1), are estimated for each of the defined dummy variables. By associating each η_{nsj} with different subsets of alternatives, the parameters (which represent standard deviations set around a mean of zero) capture different common error variances associated with those alternatives for which they are estimated for. Note that utility specifications with alternative specific constants and alternative specific error components will be equivalent to a MMNL model with normally distributed random constant terms. Also, as with the MMNL model, the random parameters of the EC model may be estimated with density over both *n* and *s* (cross sectional EC model) or only over *n* (panel EC model).

Assuming the analyst fails to specify error components as part of the utility functions of the model, then Equation (1) will collapse to

$$U_{nsj} = V_{nsj} + \varepsilon_{nsj},$$ (4)

which represents the most common form of utility representation within the literature.

Finally, for all logit type models, the second unobserved component of utility, ε_{nsj}, are assumed to be identically and independently extreme value type 1 (EV1) distributed.

2.2 Model Probabilities

Depending on the assumptions made about the utility specifications as outlined above, different functional forms of the logit model will be arrived at. We now outline in turn how the assumptions made about the different models influence the choice probabilities derived for each of the models.

2.2.1 The MNL Model

The choice probabilities of the MNL model are derived from a number of assumptions about the choice behaviour of respondents. In particular, aside from the assumption that ε_{nsj} are IID EV1, the MNL model assumes that the marginal utilities for the attributes and variables specified within the system of utility equations are fixed for the sampled population and that $\eta_{nsj} = 0$. Under these assumptions, the probability, P_{nsj}, that respondent *n* chooses alternative *j* in choice situation *s* is given by

$$P_{nsj} = \frac{\exp\left(V_{nsj}\right)}{\sum_{i \in J_{ns}} \exp\left(V_{nsi}\right)}.$$ (5)

2.2.2 MMNL and Error Components Models

Both the MMNL and EC models differ from the MNL model in that we now assume that (some of) the parameters (or error components) are random, following a certain probability distribution. The choice probabilities of the MMNL model therefore depend on the random parameters. Both models utilize the MNL probabilities given in Equation (5), however rather than calculate a single probability for each alternative, both models calculate the choice probabilities for each random draw taken from the assumed probability distribution(s). In this way, multiple choice probabilities are obtained for each alternative, as opposed to a single set of probabilities as obtained from the MNL model. It is the expectation of these probabilities over the random draws which are calculated and used in the model estimation process. The expected choice probabilities for the MMNL logit and EC models are given in Equations (6a) and (6b) respectively.

$$E\left[P_{nsj}\right] = \int_{\beta} \frac{\exp\left(V_{nsj}\right)}{\sum_{i \in J_{ns}} \exp\left(V_{nsi}\right)} f\left(\beta \mid \theta\right) d\beta, \tag{6a}$$

$$E\left[P_{nsj}\right] = \int_{\eta} \frac{\exp\left(V_{nsj}\right)}{\sum_{i \in J_{ns}} \exp\left(V_{nsi}\right)} f\left(\eta \mid \theta\right) d\eta. \tag{6b}$$

Equations (6a) and (6b) provide the choice probabilities at the level of the alternatives. In the cross sectional formulations of the MMNL and EC models, it is these probabilities that are used directly in model estimation. In the panel formulations of the MMNL and EC models, the choice probabilities given in Equations (6a) and (6b), whilst calculated, are not of direct interest. Rather, what are of interest are the probabilities of observing the sequence of choices made by each respondent, not the probabilities that specific alternatives will be observed to be chosen. To this end, we define the probability P_n^* that a certain respondent n has made a certain sequence of choices $\{j \mid y_{nsj} = 1\}_{s \in S_n}$ with respect to the set of choice situations, S_n, by

$$P_n^* = \int_{\beta} \prod_{s \in S_n} \prod_{j \in J_{ns}} \left(P_{nsj}\right)^{y_{nsj}} f\left(\beta \mid \theta\right) d\beta, \tag{7a}$$

$$P_n^* = \int_{\eta} \prod_{s \in S_n} \prod_{j \in J_{ns}} \left(P_{nsj}\right)^{y_{nsj}} f\left(\eta \mid \theta\right) d\eta, \tag{7b}$$

for the MMNL and EC models respectively.

2.3 Model Log-Likelihood Functions

Typically, the parameters β contained within each V_{nsi} are unknown and must be estimated from data. Let y_{nsj} equal one if j is the chosen alternative in choice situation s shown to respondent n, and zero otherwise. Then the parameters can be estimated by maximizing the likelihood function L,

$$L = \prod_{n=1}^{N} \prod_{s \in S_n} \prod_{j \in J_{ns}} \left(P_{nsj}\right)^{y_{nsj}}. \tag{8}$$

where N denotes the total number of respondents and S_n is the set of choice situations faced by respondent n.

Rather than maximize the likelihood function, it is more common to maximize the log of the likelihood function instead. This is because taking the product of a series of probabilities will typically produce values that are extremely small and which most computing software packages will be unable to adequately handle. By taking the logs of the probabilities first, large negative values will result, which when multiplied, produce even larger negative values. As such, the log-likelihood function of the model, shown below, is typically preferred.

$$LL = \ln\left[\prod_{n=1}^{N}\prod_{s \in S_n}\prod_{j \in J_{ns}}\left(P_{nsj}\right)^{y_{nsj}}\right]. \tag{9}$$

In the sections that follow, we attempt to differentiate between the log-likelihood functions of the various models explored throughout this paper.

2.3.1 The MNL Model

In order to derive the log-likelihood function of the MNL model, an assumption is made that all choice observations are independent of each other. That is, even in data where the same individual is observed to make multiple choices, the log-likelihood function of the MNL model treats the data as if the observed choices have been made by separate pseudo individuals. Using the mathematical properties $\ln\left(n_1 n_2\right) = \ln(n_1) + \ln(n_2)$ and $\ln(n_1)^{y_{njs}} = y_{njs}\ln(n_1)$, and applying the same mathematical rules to choice tasks, s, and alternatives, j, this independence of choice observations assumption results in Equation (9) being rewritten in the more commonly known form of

$$LL = \sum_{n=1}^{N}\sum_{s \in S_n}\sum_{j \in J_{ns}} y_{nsj}\ln\left(P_{nsj}\right). \tag{10}$$

The Log-likelihood function of the MNL model given in Equation (10) will be globally concave for linear in the parameters utility specifications (see McFadden 1974) suggesting that there should exist a single set of parameter estimates that will maximise this function.

2.3.2 Cross Sectional MMNL and Error Components Models

The log-likelihood functions of the cross sectional MMNL and EC models are derived under the same assumptions of choice observation independence as made with the MNL model. The difference between these two models and the MNL model however is that the choice probabilities used for the MNL are replaced with the expected choice probabilities given in Equations (6a) and (6b). Using the same mathematical rules used to derive the MNL model log-likelihood function, and noting additionally that $E\left(n_1 n_2\right) = E(n_1)E(n_2)$, the log-likelihood functions of the cross sectional MMNL and EC models may be represented as

$$LL = \sum_{n=1}^{N}\sum_{s \in S_n}\sum_{j \in J_{ns}} y_{nsj}\ln\left[E\left(P_{nsj}\right)\right]. \tag{11}$$

2.3.3 Panel MMNL and Error Components Models

8

The derivation of the log-likelihood functions of the panel formulations of the MMNL and EC models differ to those of their equivalent cross sectional forms, as well as to that of the MNL model, in that the choice observations are no longer assumed to be independent within each respondent (although the independence across respondents assumption is maintained). Mathematically, this means that $E(s_1 s_2) \neq E(s_1)E(s_2)$, and hence we are no longer able to invoke the mathematical rule $\ln(s_1 s_2) = \ln(s_1) + \ln(s_2)$. Given this, the log-likelihood functions of the panel MMNL and EC models may respectively be represented as

$$LL = \sum_{n=1}^{N} \ln \int_{\beta} \prod_{s \in S_n} \prod_{j \in J_{ns}} \left(P_{nsj}\right)^{y_{nsj}} f\left(\beta \mid \theta\right) d\beta, \tag{12a}$$

$$LL = \sum_{n=1}^{N} \ln \int_{\eta} \prod_{s \in S_n} \prod_{j \in J_{ns}} \left(P_{nsj}\right)^{y_{nsj}} f\left(\eta \mid \theta\right) d\beta, \tag{12b}$$

or

$$LL = \sum_{n=1}^{N} \ln \left(P_n^*\right) \tag{12c}$$

In the next section we outline the AVC matrices of each of the model types considered within this paper.

2.4 Model Variance-Covariance Matrices

The generation of efficient SC experiments requires first an estimation of the AVC matrix of the design, Ω_N. The AVC matrix Ω_N can be determined as the inverse of the Fisher information matrix, I_N, which in turn can be computed using the second derivatives of the log-likelihood function of the discrete choice model to be estimated (see Train, 2003). Mathematically, the AVC matrix for the MNL may be represented as

$$\Omega_N = I_N^{-1}, \quad \text{with } I_N = -E_N\left(\frac{\partial^2 \log LL}{\partial \beta \partial \beta'}\right), \tag{13a}$$

whilst the AVC matrix of the MMNL and EC models becomes

$$\Omega_N = I_N^{-1}, \quad \text{with } I_N = -E_N\left(\frac{\partial^2 \log E(LL)}{\partial \theta \partial \theta'}\right), \tag{13b}$$

where $E_N(\cdot)$ is used to express the large sample population mean. Hence, the AVC matrix can be determined by calculating the Hessian matrix of the log-likelihood function for the specific model.

As was seen in Section 2.3, different discrete choice models have different log-likelihood functions. Given that the AVC matrix of a discrete choice model is calculated as the inverse of the second derivatives of the log-likelihood function of that model, it is clear that each model will also yield a different AVC matrix. In this section, we reproduce the second derivatives of the log-likelihood functions for each of the models discussed as part of this paper.

2.4.1 The MNL Model

The second derivatives of the log-likelihood function of the MNL depend on whether the parameter estimates are generic or alternative specific (see Bliemer and Rose, 2005). Let x_{nsj}^* and x_{nsj} represent attributes for which generic, given as β^*, and alternative specific, represented by β_j, parameters are to be estimated for respectively. Assuming that all respondents face the same choice situations, s, the second derivatives of the MNL log-likelihood function yields the following expressions (see Rose and Bliemer, 2005)

$$\frac{\partial^2 LogL}{\partial \beta_{k_1}^* \partial \beta_{k_2}^*} = -\sum_{n=1}^{N} \sum_{s \in S_n} \sum_{j \in J_{ns}} x_{nsjk_1}^* P_{nsj} \left(x_{nsjk_2}^* - \sum_{i=1}^{J} P_{nsi} x_{nsik_2}^* \right) \tag{14a}$$

$$\frac{\partial^2 LogL}{\partial \beta_{j_1 k_1} \partial \beta_{k_2}^*} = -\sum_{n=1}^{N} \sum_{s \in S_n} \sum_{j \in J_{ns}} x_{j_1 k_1 s} P_{j_1 s} \left(x_{j_1 k_2 s}^* - \sum_{i=1}^{J} x_{i k_2 s}^* P_{is} \right) \tag{14b}$$

$$\frac{\partial^2 LogL}{\partial \beta_{j_1 k_1} \partial \beta_{j_2 k_2}} = \begin{cases} \sum_{n=1}^{N} \sum_{s \in S_n} \sum_{j \in J_{ns}} x_{j_1 k_1 s} x_{j_2 k_2 s} P_{j_1 s} P_{j_2 s}, & \text{if } j_1 \neq j_2; \\ -\sum_{n=1}^{N} \sum_{s \in S_n} \sum_{j \in J_{ns}} x_{j_1 k_1 s} x_{j_2 k_2 s} P_{j_1 s} \left(1 - P_{j_2 s} \right), & \text{if } j_1 = j_2. \end{cases} \tag{14c}$$

Note that the choice index, y_{njs}, drops out of the second derivatives of the MNL log-likelihood function, with only the design, x, and choice probabilities remaining. Given this result, it is not necessary to know *a prior* what alternatives will be chosen in the sample data in order to calculate the expected AVC matrix of the model. All the analyst requires to know is the design, and the choice probabilities. Given that the choice probabilities are a function of the design as well as the parameter estimates (see Equation (5)), in generating an efficient design, the analyst is required to make certain assumptions regarding the parameter estimates in advance. We discuss this further in Section 3.

2.4.2 Cross Sectional MMNL and Error Components Models

The AVC matrix of the MMNL and EC models are somewhat more complicated than those of the MNL model given that the parameter and error component estimates are now assume to be randomly distributed. Let M_k represent a vector of parameters related to the probability distributions of the k (either random or error component) parameters, β_k, denoted by $\theta_k = [\theta_{km}]$, where $m = 1, \ldots, M_k$. The second derivatives of this model is given as

$$\frac{\partial^2 \log E(L)}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} = \sum_{n=1}^{N} \sum_{s \in S_n} \sum_{j \in J_{ns}} y_{nsj} \left(\frac{1}{E(P_{nsj})} E\left(\frac{\partial^2 P_{nsj}}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} \right) - \frac{1}{\left(E(P_{nsj}) \right)^2} E\left(\frac{\partial P_{nsj}}{\partial \theta_{k_1 m_1}} \right) E\left(\frac{\partial P_{nsj}}{\partial \theta_{k_2 m_2}} \right) \right). \tag{15}$$

Unfortunately, unlike the MNL model, the choice index, y_{njs}, does not drop out when taking the second derivatives of the log-likelihood function of this model. Thus, in order to derive Equation (15), we are forced to rely on asymptotic theory and substitute $E_N(y_{nsj}) = E(P_{nsj})$, where $E_N(\cdot)$ is again the large sample mean. In this way, Equation (15) becomes equivalent to that given in Sándor and Wedel (2002).

2.4.3 Panel MMNL and Error Components Models

Relative to the other models explored herein, the second derivatives of the log-likelihood functions of the panel MMNL and EC models are far more complex to compute as a result of the product terms resident in Equations (12a) to (12b). Nevertheless, such derivations are possible. Bliemer and Rose (2008) show that the second derivatives of Equation (12c) is

$$
\frac{\partial^2 \log E(L)}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} = \sum_{n=1}^{N} \left(\frac{1}{E\left(P_n^*\right)} \frac{\partial^2 E\left(P_n^*\right)}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} - \frac{1}{\left(E\left(P_n^*\right)\right)^2} \frac{\partial E\left(P_n^*\right)}{\partial \theta_{k_1 m_1}} \frac{\partial E\left(P_n^*\right)}{\partial \theta_{k_2 m_2}} \right)
$$
$$
= \sum_{n=1}^{N} \left(\frac{1}{E\left(P_n^*\right)} E\left(\frac{\partial^2 P_n^*}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} \right) - \frac{1}{\left(E\left(P_n^*\right)\right)^2} E\left(\frac{\partial P_n^*}{\partial \theta_{k_1 m_1}} \right) E\left(\frac{\partial P_n^*}{\partial \theta_{k_2 m_2}} \right) \right),
$$

(16)

where

$$
\frac{\partial P_n^*}{\partial \theta_{km}} = P_n^* \frac{\partial \beta_k}{\partial \theta_{km}} \sum_{s \in S_n} \sum_{j \in J_{ns}} \frac{y_{nsj}}{P_{nsj}} \frac{\partial P_{nsj}}{\partial \beta_k},
$$

(17)

and

$$
\frac{\partial^2 P_n^*}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} = \frac{1}{P_n^*} \frac{\partial P_n^*}{\partial \theta_{k_1 m_1}} \frac{\partial P_n^*}{\partial \theta_{k_2 m_2}} - P_n^* \frac{\partial \beta_{k_1}}{\partial \theta_{k_1 m_1}} \frac{\partial \beta_{k_1}}{\partial \theta_{k_2 m_2}} \sum_{s \in S_n} \sum_{j \in J_{ns}} \frac{\partial P_{nsj}}{\partial \beta_{k_2}} x_{nsjk_1}
$$
$$
+ P_n^* \frac{\partial^2 \beta_{k_1}}{\partial \theta_{k_1 m_1} \partial \theta_{k_2 m_2}} \sum_{s \in S_n} \sum_{j \in J_{ns}} \frac{y_{nsj}}{P_{nsj}} \frac{\partial P_{nsj}}{\partial \beta_{k_1}},
$$

(18)

and where $\partial P_{njs} / \partial \beta_k$ is the first derivative of the multinomial logit probability,

$$
\frac{\partial P_{nsj}}{\partial \beta_k} = P_{nsj} \left(x_{nsjk} - \sum_{i \in J_{ns}} P_{nsi} x_{nsik} \right).
$$

(19)

As with the cross sectional MMNL and EC models, the choice index, y_{njs}, does not drop out when taking the second derivatives of the log-likelihood function of this model. Nevertheless, it is possible once more to for the choice outcomes to be replaced by probabilities, since $E_N(y_{nsj}) = P_{nsj}$ (y follows a multinomial distribution). However, $E_N\left(P_n^*\right)$ cannot be approximated that easily, as it describes a generalized multinomial distribution (Beaulieu, 1991). It is therefore necessary, unlike for designs generated specifically for the MNL and cross sectional MMNL and EC models, to simulate a sample based on the design x in order to calculate the second derivatives of the model. To do this, for each respondent n, we first draw a random parameter β_k from each given parameter distribution, then determine the observed utility V_{nsj} for each choice situation s based on design x. Next we separately draw random values for the unobserved component ε_{nsj} for each alternative in each choice situation, and determine y_{nsj} by selecting the alternative with the highest utility

in each choice situation. Note that the same random draw for β_k is used over all choice situations for each respondent, representing the panel formulation.

3. Generation of Efficient Designs

In this section, we discuss the general approach used in constructing efficient designs before moving onto discussion specifically related to the generation of the efficient designs assuming multiple possible econometric model types.
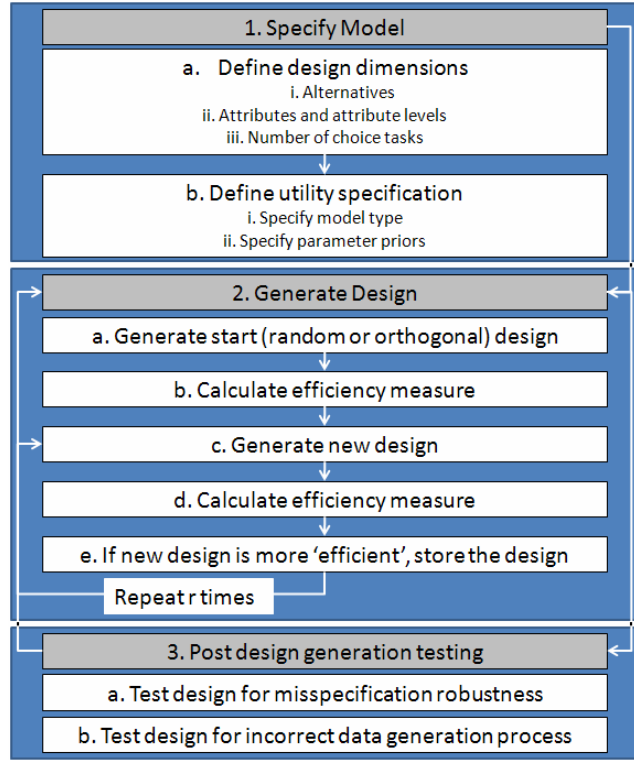
3.1 Efficient design generation process

Figure 1 outlines the generation process typically associated with the construction of efficient designs. The first stage in generating an efficient design is to specify the alternatives, attributes, and attribute levels to be used in the experiment as well as the number of choice situations that are to be shown to each respondent. The experimental design itself, once constructed, describes which hypothetical choice situations the respondents will be presented with over the course of a SC experiment. Typically, an experimental design represents nothing more than a matrix *X* of numbers in which each row represents a choice situation. The numbers in the table correspond to the *attribute levels* for each attribute (e.g., -1, 1) associated with the alternatives and are replaced by their actual attribute levels later on in the questionnaire (e.g., $1, $1.50). Different coding schemes can be used for representing the attribute levels in the experimental design. The most common ones used in practice are design coding (0, 1, 2, 3, etc.), orthogonal coding ({-1,1} for two levels, {-1,0,1} for three levels, {-3,-1,1,3} for four levels, etc.), or coding according to the actual attribute level values. Note that the alternatives, attributes and attribute levels must be defined prior to the design construction process and may be determined by use of secondary research such as focus groups or in-depth interviews.

The number of choice tasks to be shown to respondents is determined by the analyst, however, pilot surveys may offer insight into the maximum number that any one respondent may reasonably be able to handle. The theoretical minimum number of choice tasks, *S*, required for a design is determined by the number of parameters to be obtained from that design and is calculated simply as the number of parameter estimates, not including constants, plus one. This lower bound exists as any number of choice tasks lower than this value will not allow for the Fisher information matrix, I_N, to be inverted, and hence the AVC matrix to be calculated for the design. Factors such as attribute level balance (where each attribute level appears an equal number of times across the experiment) may also influence the minimum number of choice situations that may be required for a given design. Note that with blocking or random assignment of choice tasks to subsets of respondents, the minimum number of choice tasks required for a given design need not be equal to the number of choice tasks shown to each respondent.

Once the alternatives, attributes and attribute levels have been determined, the next stage is to define the utility specification in full for the design. This involves determining (i) what parameters will be generic and alternative specific; (ii) whether attributes will enter the utility function as dummy/effects codes or some other format; (iii) whether main effects only or interaction terms will be estimated; (iv) the values of the parameter estimates likely to be obtained once the model is estimated; and (v) the precise econometric model that is likely to be estimated from data collected using the experimental design. Points (i) to (iii) impact directly upon the design matrix *X*, whereas point (iv) influences the AVC matrix via the choice probabilities and point (v) via the choice probabilities as well as influencing the

dimensionality of the AVC matrix itself. Points (i) to (iii) also impact upon the minimum number of choice situations required for the design.

Figure 1: Design Generation Process

Given the design dimensions, the next stage is to generate an initial design. The initial design can be either orthogonal or simply randomly generated, however it is worth noting that it might not always be possible to construct an orthogonal design as such designs only exist for a subset of cases. The next step is to evaluate the statistical efficiency of the design. Many efficiency measures have been proposed in the literature in order to calculate an efficiency value based on the AVC matrix for the assumed model type. Typically these measures are expressed as an efficiency 'error' (i.e., a measure for the *in*efficiency), with the objective then to locate a design that minimizes this efficiency error. The most widely used measure is called the *D-error*, which takes the determinant of the AVC matrix Ω_1, assuming only a single respondent.[2] Other measures exist, such as the *A*-error, which takes the trace (sum of the diagonal elements) of the AVC matrix, however, in contrast to the *D*-error, the *A*-error is sensitive to scaling of the parameters and attributes, hence here only the *D*-error will be discussed.

The *D*-errors are a function of the experimental design X and the prior values (or prior probability distributions) β, and can be mathematically formulated as:

$$D_z\text{-error} = \det\left(\Omega_1(X,0)\right)^{1/K}, \tag{20}$$

$$D_p\text{-error} = \det\left(\Omega_1(X,\beta)\right)^{1/K}, \tag{21}$$

$$D_b\text{-error} = \int_\beta \det\left(\Omega_1(X,\beta)\right)^{1/K}\phi(\beta\,|\,\theta)d\beta. \tag{22}$$

---

[2] The assumption of single respondent is just for convenience and comparison reasons and does not have any further implications. Any other sample size could have been used, but it is common in the literature to normalize it to a single respondent.

where K is the number of parameters to be estimated. Within the literature, designs which are optimized without any information on the priors (i.e., assuming β =0) are referred to as D_z–efficient designs (Equation (21), whereas designs optimized for specific fixed (non-zero) prior parameters are referred to as D_p–efficient designs (Equation (21)). In (Bayesian) D_b–efficient designs (Equation (22)), the priors β are assumed to be random variables with a joint probability density function $\phi(\cdot)$ with given parameters θ.

The next stage in the design generation process involves changing the location of some or all of the attribute levels in the design matrix, X, and recalculating the efficiency measure for the new design, using the same parameter priors. This step is repeated R number of times, each time recording the designs relative level of statistical efficiency and storing the design matrix with the best relative efficiency measure. By changing the design R number of times, the analyst is in effect able to compare the efficiency of each of the R different design matrices. It is important to note that only for designs with relatively small numbers of alternatives, attributes and attribute levels, will it be possible to search the full enumeration of possible attribute level combinations that exist. As such, it is often necessary to turn to algorithms to examine as many different designs as is possible in a given time frame. In order to achieve this, a number of algorithms have been proposed and implemented within the literature for determining how best to change the attribute levels in locating efficiency designs, however a detailed discussion of these algorithms is beyond the scope of the current paper. Interested readers are directed to, for example, Bliemer and Rose (2006), Kessels et al. (2006) or Rose and Bliemer (2008) for a detailed discussion of such algorithms.

Once an efficient design has been generated, it is common practice to test the expected design performance prior to its use in practice. Such tests have typically taken one of two forms in the past. Firstly, some researchers use a similar approach to the design generation process, however rather than fixing the parameter priors (or distributions), changing the design and then testing the efficiency of each new design, they fix the design, change the parameter priors and then test the efficiency of the design under the new set of assumed parameters (see e.g., Rose and Bliemer 2008b). By taking this approach, the analyst is able to observe the robustness of the design to parameter prior misspecification, particularly if the parameter values explored are outside of the range considered in the design generation process.

Other researchers have employed a different approach to testing generated designs in the past, by taking the design and applying Monte Carlo simulations to test i) whether a correct data generation process has been used in constructing the design and ii) the accuracy of the parameter estimates likely to be obtained from the design at various sample sizes should that design be used in practice (see e.g.,. Kessels et al. 2006 or Ferrini and Scarpa 2007). Figure 2 demonstrates the Monte Carlo simulation process often employed within the literature. Starting with the generated experimental design, data based on a sample of respondents (or choice observations) is first simulated, including a choice variable, y_{njs}. The choice variable is created by first calculating the utilities for each choice observation based on the design attributes and a set of known prior parameter estimates which when added to an additional simulated random error term, produces different utility values for each of the alternatives. Once the utilities for each alternative are calculated, the alternative with the highest utility is assumed to be the one selected. Once a sample has been simulated, discrete choice models are estimated on the simulated data with the resulting outputs (e.g., parameter values, standard errors) then compared to the known input values (e.g.,

parameter priors). By simulating different random error terms, different samples may be generated, and the process repeated over each new simulated sample, a total of R times.
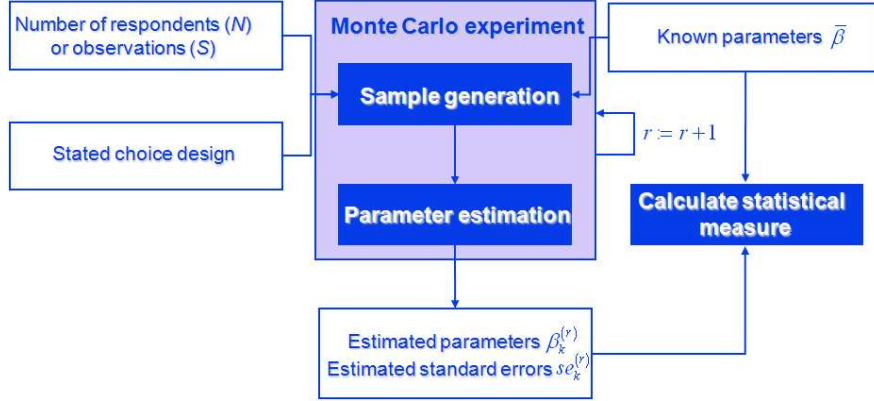
Figure 2: Monte Carlo Simulation Process

A number of different measures have been proposed to compare the input parameters to those obtained from the Monte Carlo simulation process. In this paper, we rely on three, the first two being the mean square error (MSE) and the relative absolute error (RAE). Equations for the calculating the MSE and RAE statistics are given in Equations (23) and (24) respectively for the parameter estimates. Similar equations may be used for other model outputs, such as the standard errors, or even D-error values if so desired.

$$ MSE = \frac{1}{R}\sum_{r=1}^{R}\left(\beta_k^{(r)} - \overline{\beta}\right)^2, \tag{23} $$

$$ RAE = \frac{1}{R}\sum_{r=1}^{R}\left(\beta_k^{(r)} - \overline{\beta}\right)\Big/\overline{\beta}, \tag{24} $$

where $\beta_k^{(r)}$ is the parameter estimate obtained at sample iteration r, and $\overline{\beta}$ is the known prior parameter estimate used in constructing the Monte Carlo simulation.

The final measure we use is the expected mean square error of the parameter estimates (EMSE). Unlike the MSE and RAE which capture statistical evidence of any potential biases in the individual parameters, the EMSE provides a single summary statistic of the overall bias and variances across all parameter estimates which may be used easily to compare different designs. The EMSE is given as Equation (25).

$$ EMSE = \frac{1}{R}\sum_{r=1}^{R}\left(\beta_k^{(r)} - \overline{\beta}\right)'\left(\beta_k^{(r)} - \overline{\beta}\right). \tag{25} $$

3.3 Uncertainty in the model type: Model averaging

The statistical efficiency of a design as discussed in Section 3.1 relates to the AVC matrix for a single assumed model type. As discussed in Section 2.4 however, different discrete choice models will result in different AVC matrices and as such, it follows that the efficiency of a given design will relate to the specific model assumed in its construction. Given this situation, it is likely that a design generated for a specific model type will be less (or possibly more) efficient if applied to a different model type for which it was generated for. As with the parameter priors, however, it is possible to take a somewhat Bayesian approach

to the problem, by constructing multiple AVC matrices for a given design assuming different model types and calculating an efficiency measure for each AVC matrix thus derived. Note that it is also possible to assume different prior parameter estimates for the different models assumed in the design construction process.

Assuming that the same efficiency measure is used for each AVC matrix (e.g., D-error), the analyst may then apply some form of weight to each measure to arrive at a composite weighted average measure of efficiency over all model types. In taking this approach, the relative weight applied to each efficiency measure would reflect the likelihood that the model to which the efficiency measure belongs would be the one that is estimated on the final data collected using that design. In the case where the likelihood that the final model type to be estimated is unknown, it is possible that each efficiency measure be given an equal weight in the averaging process. If however one expects that for example, a panel MMNL model is more likely to be estimated on the final data set collected, then the efficiency measure for the AVC matrix of this model type can be given a greater weight than for other model types assumed when generating the composite efficiency measure. Independent of the weights applied, the design construction process proceeds as previously described, the only difference being that the weighted average efficiency measure is used to compare different designs as opposed to the individual model type efficiency measures.

Problems arise however if different efficiency measures (e.g., *D*-error and *A*-error) are used for the different model types assumed in the weighted averaging of the efficiency measure. This is because the different efficiency measures are measured on different metric scales and are likely to produce values with widely different relative magnitudes. For example, a design might produce a *D*-error value of 0.2 and an *A*-error value of 100. In such a case, the analyst would need to at a minimum normalize in some manner the different measures to some form of common metric. Unfortunately, to do this would require advanced knowledge of the likely magnitudes of the different efficiency measures, which is unlikely to be known prior to design generation. As such, it is recommended that a common efficiency measure type be applied to all model types when using the model averaging approach.

3.4 A note on the relationship between statistical efficiency and sample size

Although within the confines of this paper, we use *D*-error as our sole criteria of efficiency, it is worth highlighting the precise link that exists between the statistical efficiency of a design and the sample size required for that design. As suggested by Equations (13a) and (13b), the AVC matrix Ω_N can be computed for any sample size N. Indeed, mathematically, Bliemer and Rose (2009) argue that $I_N = N \cdot I_1$, and hence $\Omega_N = \left(I_N^{-1} \right) = \frac{1}{N} I_N^{-1} = \frac{1}{N} \Omega_1$. Given such a relationship, it follows that the standard error for the kth attribute of a design may be represented as

$$se_{kN} = \frac{se_k}{\sqrt{N}},$$

(25)

and the asymptotic *t*-ratio

$$t_k = \frac{\beta_k}{\left(se_k \middle/ \sqrt{N} \right)}.$$

(26)

Re-arranging Equation (24), we obtain

16

$$N = \left(\frac{se_k.t_k}{\beta_k} \right)^2 , \tag{27}$$

an equation that may be used to compute the theoretical minimum sample size for each parameter of the design.

In generating a design however, the values of β_k are not estimated but rather are assumed inputs in the form of the parameter priors. Similarly, the values of t_k are not estimated, but must be pre-specified by the analyst, with a logical value being 1.96 or greater to ensure that the parameter will be statistically significant with at least 95 percent certainty. In taking this approach, each parameter of the design will have a different theoretical minimum sample with the theoretical minimum sample size for the overall design being the value of the largest calculated N, thus ensuring that all parameters of the design are likely to be found statistically significant post data collection. Note that the sample sizes calculated in this manner represent a theoretical minimum sample size as other factors, such as parameter stability may require larger sample sizes than suggested by Equation (27).

4. Case Study

In order to illustrate the theory of efficient designs and the use of model averaging in the construction of such designs, we now consider a case study examining a specific discrete choice problem. In doing so, we generate 10 separate designs for purposes of comparison. The first two designs we generate are orthogonal designs. We construct two different orthogonal designs firstly to highlight that different orthogonal designs often exist for the same problem, and secondly to show that different orthogonal designs can and will display different levels of statistical efficiency. Thirdly, the construction of two separate orthogonal designs allows for a better comparison of the predominant method of experimental design construction used within the literature to date. At present, it is common practice to randomly generate a single orthogonal design, without reference to the statistical efficiency of the design, and as such, the results obtained from the design, in terms of statistical efficiency, may also be considered to be random. Next we generate designs specifically optimized for each of the model types discussed earlier; MNL, MMNL panel and cross sectional and EC panel and cross sectional. Finally we generate three separate designs applying the model averaging approach, allowing for different weights for each design generated.

Before we discuss the results obtained for each of the designs generated, we first outline the specific design dimensions (i.e., number of alternatives, attributes, etc.) as well as the utility specifications for each of the model types, as utilized for the purposes of the case study. In Section 5, we discuss the resulting designs before demonstrating the statistical tests that may be performed on each design post construction in Section 6.

4.1 Case study design dimensions

Consider a choice experiment involving three alternatives, the first two of which are described by four attributes, and the last representing a no choice or status quo alternative and hence having no associated attributes. For simplicity we assume that all parameters are generic, although the theory and application is easily extended to alternative specific parameter estimates. In generating the design, we assume that each respondent will view 16 choice situations each.

Within the SC experiment, the eight attributes (four per each alternative) can take on different levels over the different choice situations shown to respondents. Let us assume that the first three attributes can take on one of four levels such that the first attribute for each alternative can take the values $\{5, 10, 15, 20\}$, and the second and third attributes can take the levels $\{0, 1, 2, 3\}$. Assume that the last attribute can take two levels, $\{0, 1\}$. These values were chosen for demonstrative purposes only, and any values could have been selected for the case study. Following common practice, we constrain ourselves to balanced designs (although such a constraint may result in the generation of a sub-optimal design).

Equation (28) shows the utility specification set up used for the case study.

$$U(A) = \beta_{0A} + \beta_1 x_{1A\{5,10,15,20\}} + \beta_2 x_{2A\{0,1,2,3\}} + \beta_3 x_{3A\{0,1,2,3\}} + \beta_4 x_{4A\{0,1\}} + \eta_{A,B},$$

$$U(B) = \beta_{0B} + \beta_1 x_{1B\{5,10,15,20\}} + \beta_2 x_{2B\{0,1,2,3\}} + \beta_3 x_{3B\{0,1,2,3\}} + \beta_4 x_{4B\{0,1\}} + \eta_{A,B}, \qquad (28)$$

$$U(C) = 0,$$

were $\eta_{A,B} \sim N(0, \sigma^2)$, representing the error components applicable only for the EC cross sectional and panel models.

4.2 Case study prior parameters

Table 1 shows the prior parameter and error component estimates associated with each model used as part of the case study. In specifying the priors for generating the case study designs, we have assumed different parameter values for each of the models. In each case we have assumed alternative specific constants for the two non-status quo alternatives and in the case of the EC models we have assumed a generic error component across these same two alternatives. For the MNL model, we have assumed Bayesian prior parameter distributions for the first two parameter estimates, representing uncertainty as to the precise parameter values these estimates will take once data has been collected. For the same model, we have assumed fixed prior parameters for the remaining two estimates. For the EC models, we have employed Bayesian prior parameter distributions for the first parameter, but fixed parameter priors for the remaining three estimates. For both the panel and cross-sectional EC designs, we have also utilized the Bayesian approach to design generation for the error component priors, drawing the assumed standard deviation parameters from uniform distributions.

For the panel and cross sectional MMNL models, we have assumed the first two parameters to be random parameters drawn from Normal distributions, thus representing preference heterogeneity over the sampled population for these two attributes. In specifying the prior parameter values for these distributions, we have again assumed a Bayesian design generation approach for the first random parameter, with the mean of the random parameter distribution being drawn from a Normal distribution and the standard deviation prior from a uniform distribution. In this way, we have incorporated uncertainty as to the precise values both population moments will take post data collection for this random parameter. For the second random parameter distribution, we have assumed fixed or known mean and standard deviation parameter priors for the design generation process. To compute the AVC matrices for the panel designs, it is necessary to simulate a sample data set. For the present study, we generate samples of 2,500 respondents (40,000 choice observations). Although not reported here, we tested the impact that different simulated sample sizes have upon the efficiency measures and found that the efficiency measures were stable at levels far lower than the 2,500 respondents used. Finally, taking the draws

for the MMNL and EC models, we used Gaussian quadrature with four points each for the random parameter and Bayesian prior parameter draws.

Table 1: Case study parameter priors

		MNL	MMNL	EC	MMNL	EC
			Panel		Cross Sectional	
		MNL	MMNL	EC	MMNL	EC
Parameter Priors						
β_{0A}	Mean	-3.2	-2.4	-3.0	-3.2	-3.3
β_{0B}	Mean	-3.4	-2.2	-2.8	-3.0	-3.2
β_1	Mean	$N(-0.07,0.03)$	$N(-0.08,0.01)$	$N(-0.06,0.02)$	$N(-0.02,0.01)$	$N(-0.05,0.02)$
	Std Dev.	-	$U(0.02,0.04)$	-	$U(0.01,0.03)$	-
β_2	Mean	$N(1.2,0.2)$	1.2	1	1.4	1.1
	Std Dev.	-	0.4	-	0.3	-
β_3	Mean	1.8	1.2	1.5	1	1.2
β_4	Mean	-0.6	-0.7	-0.5	-0.6	-0.4
Error Component Priors						
$\eta_{1A,B}$	Std Dev.	-	-	$U(1.0,2.0)$	-	$U(1.5,2.5)$

In order to explore the impact different weights have on the model averaging process during design generation, we generate three different model average designs using different sets of assumed weights. In the first two cases, we apply larger weights to the MMNL and EC panel models than for the cross-sectional MMNL and EC models, to reflect the greater preponderance within the literature for estimating such models. The final weights we apply assume an equal chance of each model being estimated once data is collected using the design. The weights used are summarized in Table 2.

Table 2: Model average prior weights

Model	Weight 1	Weight 2	Weight 3
MNL	0.2	0.125	0.2
MMNL (Pan.)	0.2	0.25	0.3
EC (Pan.)	0.2	0.25	0.3
MMNL (C.S.)	0.2	0.125	0.1
EC (C.S.)	0.2	0.25	0.1

5. Case Study Designs

Table 3 presents the 10 designs generated as part of the case study. Each design was generated using Ngene, an experimental design software package currently under development by ChoiceMetrics. For each choice set *s*, we omit the final alternative (*j* = 3) given that this alternative represents a no choice option and as such has no attribute levels. Aside from the orthogonal designs, all efficient designs reported where the best designs located after evaluating a number of potential candidates using a randomize-and-swap algorithm. The average time per design evaluation depended upon the model type being evaluated. Based on a notebook computer running Windows XP with a 2.0Ghz Pentium processor and 2GB RAM, the average time per design evaluation for each model type is given in Table 4. Also reported is the number of potential designs explored by model type during the design construction process.

Table 3: Case study designs

S	J	Orthogonal 1				Orthogonal 2				MNL Design				MMNL (pan.)				MMNL (C.S.)				EC (pan.)				EC (C.S.)				Model Average 1				Model Average 2				Model Average 3			
		x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	1	5	0	0	1	5	0	0	1	10	1	0	0	5	3	0	1	5	1	3	1	5	1	3	1	5	3	2	0	10	1	0	0	5	1	3	1	20	2	2	0
1	2	5	0	0	1	5	0	0	1	15	2	0	1	20	0	3	0	20	2	0	0	20	2	0	0	20	2	2	1	15	2	0	1	20	3	2	0	5	3	1	1
2	1	15	1	3	1	15	1	3	1	10	1	1	1	5	0	3	0	20	2	2	0	20	2	2	0	10	2	0	1	10	2	0	1	5	3	1	0	5	3	1	0
2	2	20	0	2	1	20	0	2	1	15	1	0	0	15	2	1	1	15	0	3	1	15	0	3	1	15	0	0	0	15	0	2	0	20	1	3	1	20	1	3	1
3	1	10	1	2	0	20	1	2	0	15	2	2	0	15	2	0	1	5	3	1	0	5	3	1	0	10	1	1	0	15	2	2	1	20	2	2	0	10	1	0	0
3	2	5	1	2	0	5	1	2	0	10	0	3	1	10	1	0	0	10	2	1	1	10	2	1	1	15	1	0	1	10	2	0	0	5	3	2	1	15	2	0	1
4	1	20	0	1	0	10	0	1	0	20	0	3	0	15	3	1	0	15	2	0	1	15	2	0	1	20	2	3	1	10	0	1	1	10	1	0	0	15	2	0	1
4	2	20	1	0	0	20	1	0	0	5	3	1	1	10	0	2	1	10	0	1	0	10	0	1	0	5	3	1	0	15	2	0	0	15	2	0	1	10	1	0	0
5	1	10	3	1	1	10	3	1	1	5	0	3	0	20	3	2	1	5	2	0	0	5	2	0	0	15	1	1	1	20	3	1	0	15	0	1	1	15	3	1	0
5	2	15	0	3	0	15	0	3	0	20	3	2	1	5	1	2	0	15	1	1	1	15	1	1	1	10	0	0	0	5	0	3	1	15	2	0	0	5	0	3	1
6	1	20	2	2	1	20	2	2	1	20	3	2	1	20	2	2	0	20	0	3	1	20	0	3	1	5	0	3	0	20	2	2	0	20	1	3	1	10	0	3	0
6	2	10	0	1	0	10	0	1	0	5	1	2	0	5	3	1	1	5	0	1	0	5	0	1	0	20	3	1	1	5	3	1	1	5	3	1	0	15	2	1	1
7	1	5	2	3	0	15	2	3	0	15	2	0	0	10	1	1	0	15	1	2	1	15	1	2	1	15	0	1	1	15	2	0	1	15	3	1	0	15	0	1	0
7	2	15	1	1	1	15	1	1	1	10	2	0	1	15	3	2	1	20	1	3	0	20	1	3	0	10	2	0	0	10	1	0	0	5	0	3	1	10	0	1	1
8	1	15	3	0	0	5	3	0	0	5	3	1	1	20	2	2	0	15	0	3	0	15	0	3	0	20	2	2	0	5	3	1	0	10	0	3	0	20	1	2	0
8	2	10	1	3	1	10	1	3	1	20	1	3	0	5	2	2	1	10	2	0	1	10	2	0	1	5	2	3	1	20	1	3	1	15	2	1	1	5	1	3	1
9	1	15	2	2	0	5	2	2	0	20	1	3	1	20	1	3	0	20	3	1	0	20	3	1	0	15	0	0	0	15	1	3	1	20	3	2	1	10	0	1	1
9	2	15	2	2	0	15	2	2	0	5	3	1	0	5	3	1	1	5	3	2	1	5	3	2	1	15	0	2	1	5	3	1	0	10	1	1	0	15	2	0	0
10	1	20	1	3	0	10	1	3	0	10	0	3	1	10	0	3	1	5	0	3	0	5	0	3	0	20	3	2	1	5	1	3	1	10	0	1	0	20	3	2	1
10	2	5	2	1	1	5	2	1	1	10	3	1	0	20	3	0	0	15	3	0	1	15	3	0	1	5	1	2	0	20	3	2	0	10	0	1	1	10	1	2	0
11	1	15	3	1	1	15	3	1	1	15	0	2	1	5	3	0	0	20	3	2	1	20	3	2	1	10	1	2	1	5	0	3	0	5	0	3	0	5	1	3	1
11	2	5	3	0	0	5	3	0	0	10	2	0	0	20	1	3	1	5	3	0	0	5	3	0	0	15	2	3	0	15	2	1	1	20	3	2	1	20	3	2	0
12	1	20	0	0	1	20	0	0	1	10	2	0	1	15	0	3	1	10	0	0	0	10	0	0	0	5	0	3	1	10	0	3	0	20	1	2	0	20	1	3	1
12	2	15	3	3	1	15	3	3	1	15	0	2	0	10	2	0	0	20	0	2	1	20	0	2	1	20	3	1	0	20	3	2	1	5	1	3	1	5	3	1	0
13	1	10	2	1	0	20	2	1	0	15	2	0	0	10	2	0	1	10	2	1	0	10	2	1	0	5	3	1	1	15	0	1	0	10	2	0	1	5	3	0	1
13	2	10	2	2	1	10	2	2	1	15	1	1	1	15	0	3	0	10	3	2	1	10	3	2	1	20	1	3	0	10	0	1	1	15	0	2	0	20	0	3	0
14	1	5	1	0	0	15	1	0	0	5	3	1	1	5	1	2	0	10	1	1	1	10	1	1	1	10	3	0	0	20	1	2	0	15	2	2	1	10	2	0	1
14	2	20	2	1	0	20	2	1	0	20	0	3	0	20	2	3	1	5	1	3	0	5	1	3	0	10	1	3	1	5	1	3	1	10	2	0	0	15	0	2	0
15	1	10	3	2	1	10	3	2	1	5	1	2	0	10	1	1	1	10	3	0	1	10	3	0	1	20	2	3	0	5	3	0	1	15	2	0	1	5	0	3	0
15	2	20	3	0	1	20	3	0	1	20	2	2	1	10	1	0	0	15	1	3	0	15	1	3	0	5	3	1	1	20	0	3	0	10	1	0	0	20	3	2	1
16	1	5	0	3	1	5	0	3	1	20	3	1	0	15	0	1	1	15	1	2	1	15	1	2	1	15	1	0	0	20	3	2	1	5	3	0	1	15	2	2	1
16	2	10	3	3	0	10	3	3	0	5	0	3	1	15	1	0	0	20	2	2	0	20	2	2	0	10	0	2	1	10	1	2	0	20	0	3	0	10	2	0	0

20

Table 4: Average design evaluation time by assumed model type

Model	Evaluation Time (seconds)	# of design replications tested
MNL	0.0015	100,000
MMNL (Pan.)	174	10,000
EC (Pan.)	50	50,000
MMNL (C.S.)	0.125	10,000
EC (C.S.)	0.04	50,000
Model average	400	5,000

Care is required in comparing the time differences required in computing the efficiency of each model type given that each of the different design model types had different number of Bayesian parameters from which simulated draws were required to be taken. Also, the panel MMNL and EC models, and by implication the weighted average designs, required the generation of simulated sample data in order to compute the AVC matrix for the design, and hence require additional computation time.

Table 5 reports the efficiency results for the 10 generated designs calculated as if the designs were used to estimate different model forms. In calculating the efficiency measures for each design by and model type, we have employed the prior parameter estimates given in Table 1. Within Table 5, efficiency results are presented for the three weighted average measures (Table 2) as well as for the unweighted efficiency measures. The bold values represent the D_b-errors for the model type that a specific design was optimised for. In each case, designs generated for a specific model type were found to be more statistically efficient than designs generated for another model type. For example, if the estimated model is of the MNL type, then the MNL design will perform best, with a D_b-error of 0.0897 (which is smaller than the D_b-errors of the other nine designs). Similarly, the panel MMNL design will perform best when estimating a panel MMNL model (with a D_b-error of 0.0752), and the panel EC design and will perform best for estimating the panel EC model (with a D-error of 0.1539). Likewise, the cross-sectional MMNL and EC designs would be expected to perform best when estimated using the same model for which they were optimised for. Examination of the two orthogonal designs suggest that these designs are likely to perform poorly in terms of their efficiency levels independent of the model type estimated when compared to all other remaining designs.

At the base of Table 5 is the minimum sample sizes and number of choice observations rounded to the nearest respondent required for each design to obtain significant asymptotic *t*-ratios for all parameters at the 95 percent confidence level. Reading across the table, bolded values represent the smallest sample size required from each of the 10 designs based on the five different model estimations assumed. As is to be expected, the design specifically generated with a particular model form in mind tends to produce the lowest expected sample size requirement when estimated using that model type. Nevertheless, it is worth noting that with the exception of the MMNL and EC cross sectional models, the MNL model design appears to perform extremely well in terms of the expected minimum sample size required independent of the model type estimated. Of note is the poor performance of the two orthogonal designs in terms of expected minimum sample size requirements with both orthogonal designs expected to require substantially more respondents when applied to all model types when compared to all of the other designs generated. Such an observation is not surprising given that other researchers have noted similar findings elsewhere (notably Bliemer and Rose, 2008, 2009, Scarpa and Rose, 2008 and Rose and Bliemer, 2009). One possible explanation of this is that the orthogonality of a design suggests nothing about the choice probabilities obtained from the design, and given that the choice probabilities are

21

Table 5: Case study D-error Results

	Model Assumed for Design									
	Orthogonal 1	Orthogonal 2	MNL	MMNL (Pan.)	EC (Pan.)	MMNL (C.S.)	EC (C.S.)	Model Av. 1	Model Av. 2	Model Av. 3
	D-error Value (non-weighted)									
MNL	0.2263	0.1994	**0.0897**	0.1124	0.0903	0.1599	0.1177	0.0960	0.0969	0.0938
MMNL (Pan.)	0.1200	0.1136	0.0804	**0.0752**	0.0799	0.0905	0.0821	0.0792	0.0812	0.0791
EC (Pan.)	0.2772	0.2601	0.1553	0.1726	**0.1539**	0.2142	0.1731	0.1588	0.1588	0.1565
MMNL (C.S.)	0.2853	0.2709	0.2533	0.2246	0.2333	**0.1616**	0.2333	0.2132	0.2175	0.2199
EC (C.S.)	0.4508	0.4154	0.3354	0.2935	0.3318	0.3313	**0.2463**	0.2761	0.2711	0.2772
Model Av.	1.3596	1.2594	0.9141	0.8783	0.8892	0.9574	0.8525	**0.8233**	0.8255	0.8265
	D-error Value (Model Average Weighting 1)									
MNL	0.0453	0.0399	**0.0179**	0.0225	0.0181	0.0320	0.0235	0.0192	0.0194	0.0188
MMNL (Pan.)	0.0240	0.0227	0.0161	**0.0150**	0.0160	0.0181	0.0164	0.0158	0.0162	0.0158
EC (Pan.)	0.0554	0.0520	0.0311	0.0345	**0.0308**	0.0428	0.0346	0.0318	0.0318	0.0313
MMNL (C.S.)	0.0571	0.0542	0.0507	0.0449	0.0467	**0.0323**	0.0467	0.0426	0.0435	0.0440
EC (C.S.)	0.0902	0.0831	0.0671	0.0587	0.0664	0.0663	**0.0493**	0.0552	0.0542	0.0554
Model Av.	0.2719	0.2519	0.1828	0.1757	0.1778	0.1915	0.1705	**0.1647**	0.1651	0.1653
	D-error Value (Model Average Weighting 2)									
MNL	0.0283	0.0249	**0.0112**	0.0140	0.0113	0.0200	0.0147	0.0120	0.0121	0.0117
MMNL (Pan.)	0.0300	0.0284	0.0201	**0.0188**	0.0200	0.0226	0.0205	0.0198	0.0203	0.0198
EC (Pan.)	0.0693	0.0650	0.0388	0.0431	**0.0385**	0.0536	0.0433	0.0397	0.0397	0.0391
MMNL (C.S.)	0.0357	0.0339	0.0317	0.0281	0.0292	**0.0202**	0.0292	0.0267	0.0272	0.0275
EC (C.S.)	0.1127	0.1039	0.0839	0.0734	0.0829	0.0828	**0.0616**	0.0690	0.0678	0.0693
Model Av.	0.2759	0.2561	0.1856	0.1774	0.1819	0.1992	0.1692	0.1672	**0.1671**	0.1674
	D-error Value (Model Average Weighting 3)									
MNL	0.0453	0.0399	**0.0179**	0.0225	0.0181	0.0320	0.0235	0.0192	0.0194	0.0188
MMNL (Pan.)	0.0360	0.0341	0.0241	**0.0226**	0.0240	0.0271	0.0246	0.0238	0.0244	0.0237
EC (Pan.)	0.0832	0.0780	0.0466	0.0518	**0.0462**	0.0643	0.0519	0.0476	0.0476	0.0470
MMNL (C.S.)	0.0285	0.0271	0.0253	0.0225	0.0233	**0.0162**	0.0233	0.0213	0.0218	0.0220
EC (C.S.)	0.0451	0.0415	0.0335	0.0294	0.0332	0.0331	**0.0246**	0.0276	0.0271	0.0277
Model Av.	0.2380	0.2206	0.1475	0.1486	0.1447	0.1727	0.1481	0.1395	0.1402	**0.1392**
	S-error rounded to nearest N (N.S)									
MNL	15 (240)	15 (240)	**5 (80)**	8 (128)	6 (96)	9 (144)	8 (128)	7 (112)	7 (112)	6 (96)
MMNL (Pan.)	8 (128)	8 (128)	**5 (80)**	**5 (80)**	**5 (80)**	6 (96)	**5 (80)**	6 (96)	6 (96)	**5 (80)**
EC (Pan.)	29 (464)	24 (384)	**12 (192)**	16 (256)	**12 (192)**	18 (288)	14 (224)	13 (208)	13 (208)	**12 (192)**
MMNL (C.S.)	262 (4192)	233 (3728)	205 (3280)	201 (3216)	187 (2992)	**147 (2352)**	173 (2768)	194 (3104)	217 (3472)	218 (218)
EC (C.S.)	221 (3536)	158 (2528)	167 (2672)	132 (2112)	148 (2368)	121 (1936)	**80 (1280)**	96 (1536)	87 (1392)	99 (1584)

instrumental in calculating the AVC matrix, an orthogonal design would not be expected to result in low standard error values when compared to designs specifically generated for this purpose.

The three model average designs generated suggest moderate sample size requirements compared to the other designs generated. Examination of Table 5 suggests that the three model average designs appear to outperform the orthogonal, MNL, panel MMNL and EC designs when applied to the cross sectional MMNL and EC model forms, but tend to perform at least as well or only marginally worse than the other non-orthogonal designs when applied to the other model types. As such, the model averaging process appears to offer some form of robustness in terms of sample size across all the model types explored herein when the exact model type to be used in estimation is unknown at the time of design generation. Table 6 details the percentage of respondents required for each of the three model average designs relative to all of the other designs. Percentage values less than 100 suggest that the model average design would require fewer respondents relative to the comparison design, whilst values greater than 100 percent suggest more respondents would be required. Based on the table, all three model average designs outperform all the other designs in terms of sample size requirements when different model types are used in estimation. Nevertheless, the MNL and EC panel designs appear to offer superior sample size requirements when applied to all model types other than the EC cross sectional model.

Table 6: Percentage of Respondents Comparison with Model Average Designs

		Orth. 1	Orth. 2	MNL	MMNL (Pan.)	EC (Pan.)	MMNL (C.S.)	EC (C.S.)
		Model Average 1						
	MNL	**46.67%**	**46.67%**	140.00%	**87.50%**	116.67%	**77.78%**	**87.50%**
	MMNL (Pan.)	**75.00%**	**75.00%**	120.00%	120.00%	120.00%	100.00%	120.00%
	EC (Pan.)	**44.83%**	**54.17%**	108.33%	**81.25%**	108.33%	**72.22%**	**92.86%**
	MMNL (C.S.)	**74.05%**	**83.26%**	**94.63%**	**96.52%**	103.74%	131.97%	112.14%
	EC (C.S.)	**43.44%**	**60.76%**	**57.49%**	**72.73%**	**64.86%**	**79.34%**	120.00%
		Model Average 2						
	MNL	**46.67%**	**46.67%**	140.00%	**87.50%**	116.67%	**77.78%**	**87.50%**
	MMNL (Pan.)	**75.00%**	**75.00%**	120.00%	120.00%	120.00%	100.00%	120.00%
	EC (Pan.)	**44.83%**	**54.17%**	108.33%	**81.25%**	108.33%	**72.22%**	**92.86%**
	MMNL (C.S.)	**82.82%**	**93.13%**	105.85%	107.96%	116.04%	147.62%	125.43%
	EC (C.S.)	**39.37%**	**55.06%**	**52.10%**	**65.91%**	**58.78%**	**71.90%**	108.75%
		Model Average 3						
	MNL	**40.00%**	**40.00%**	120.00%	**75.00%**	100.00%	**66.67%**	**75.00%**
	MMNL (Pan.)	**62.50%**	**62.50%**	100.00%	100.00%	100.00%	**83.33%**	100.00%
	EC (Pan.)	**41.38%**	**50.00%**	100.00%	**75.00%**	100.00%	**66.67%**	**85.71%**
	MMNL (C.S.)	**83.21%**	**93.56%**	106.34%	108.46%	116.58%	148.30%	126.01%
	EC (C.S.)	**44.80%**	**62.66%**	**59.28%**	**75.00%**	**66.89%**	**81.82%**	123.75%

Model Assumed for Estimation

6. Post design generation Testing

In order to evaluate the robustness of the designs generated, a number of statistical tests are performed. In Sections 6.1 and 6.2 we present the results of tests discussed in Section 3.1. Section 6.1 presents the results of parameter misspecification tests whilst Section 6.2 discusses the results from Monte Carlo simulations designed to examine possible parameter biases that might derive from use of the various generated designs.

6.1 Parameter Prior Misspecification

In constructing each of the non-orthogonal designs, we have assumed that the prior parameter values correspond to the true parameter values held by the population, although we have allowed for some degree of uncertainty via the use of Bayesian prior parameter distributions. Nevertheless, this remains a strong assumption that is unlikely to hold in practice. To test the impact misspecification of the prior parameters has on an experimental design once generated, it is possible to fix the design and apply different sets of priors to it and in doing so recalculate the expected AVC matrix.

Taking the parameter priors given in Table 1 as the true parameter population estimates, we apply each of the 10 generated designs to the five estimation model types and vary the parameter estimates by ±50 percent. For parameters for which Bayesian parameter distributions were employed, we assume the mean of the distributions represent the true population parameter values (e.g., the mean and standard deviation for the first design parameter of the panel MMNL were assumed to be $N(-0.08, 0.01)$ and $U(0.02, 0.04)$ in which case we assume the true population mean and standard deviation parameters to be -0.08 and 0.03 respectively). Figure 3 graphs the impact upon D-error and sample size requirements given misspecification of the β_1 random parameter assuming estimation of a panel MMNL model. Note that the sample sizes shown tend to be larger than those given in Table 5, even when the parameter is assumed not to have been misspecified. This is because the sample sizes calculated for Table 5 were generated assuming the population moments of several parameters were drawn from Bayesian distributions, whereas those shown in Figure 3 where not and hence the two are not directly comparable.
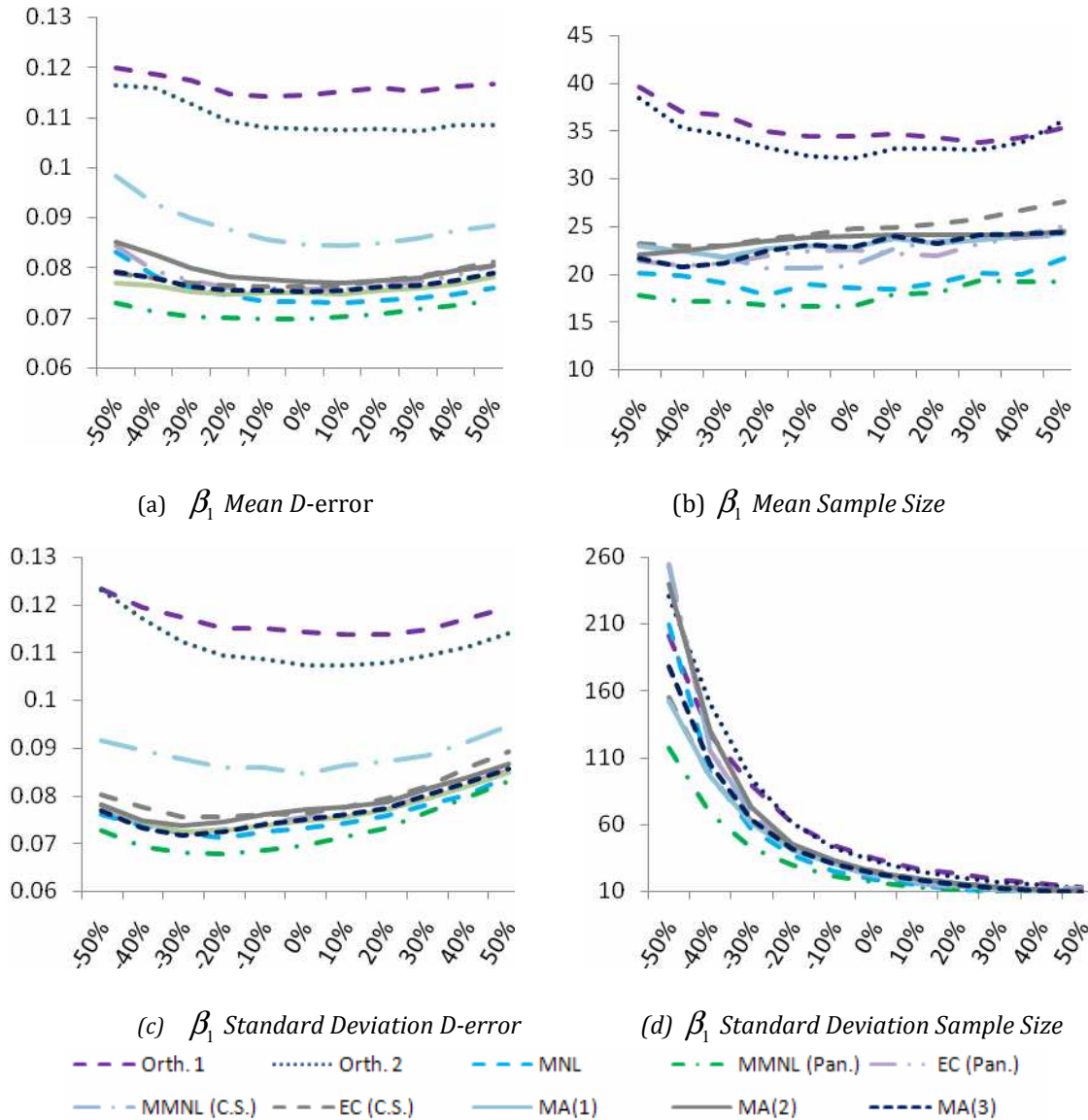
The D-error values for the two orthogonal designs are consistently higher than for all other designs independent of the degree of misspecification for this parameter assuming estimation of a panel MMNL model. As is to be expected, the MMNL panel design appears to perform best in terms of D-error even given parameter prior misspecification. Interestingly, the MNL design also performs well in given the range of prior parameters examined. The D-error values for the three model average designs appear to be somewhere in between those of the MMNL panel and MNL designs, and the two orthogonal designs, across all misspecification values explored.

In terms of expected sample size requirements, the sample sizes required for statistical significance for the mean of the parameter distribution largely mirror the D-error values in terms of the observed results based on the 10 designs generated. Similar results are observed for small values of the standard deviation parameter across all designs however as the magnitude of the parameter increases, all designs converge to a similar ability to detect statistical significance at relatively small sample sizes. Nevertheless, the two orthogonal designs still require slightly larger samples than the other designs for this parameter and model type even given significantly wider population standard deviation.

Similar results are observed for other model types and parameters, both in terms of statistical efficiency results and sample sizes. Rather the present all the results here (there being 360 parameter misspecifications to explore; 36 parameters across 10 design types), the D-error and theoretical minimum sample sizes values may be downloaded from http://www.econ.usyd.edu.au/19129.html in table form. We note that in some cases, misspecification of a prior parameter can results in D-error values increasing (decreasing) whilst the sample size actually decreases (increase). Bliemer and Rose (2008) argue that such a result is possible given that the D-error represents a form of average over all parameters, while the sample size requirements relate only to parameter that is most difficult to estimate. As

such, it is possible that on average, the standard errors for each design are decreasing, but that the largest standard error within the AVC matrix actually decreases.

(a) β_1 *Mean D-error* (b) β_1 *Mean Sample Size*

(c) β_1 *Standard Deviation D-error* (d) β_1 *Standard Deviation Sample Size*

Legend: Orth. 1 — Orth. 2 ⋯ MNL — — MMNL (Pan.) — ·— EC (Pan.) — ·· MMNL (C.S.) — ·— EC (C.S.) — — MA(1) ——— MA(2) ——— MA(3) - - - - -

Figure 3: β_1 **parameter prior misspecification assuming estimation of a MMNL panel model**

Independent of the design type, the results shown here demonstrate that misspecification of the prior parameter values can potentially have a significant impact upon the overall efficiency of different designs and model types. Thus, for design efficiency to be truly translated into estimation efficiency, the parameter priors assumed during the generation process should be as close to possible to the true, but as yet unknown, population level parameters. Bliemer and Rose (2008) offer some suggestions as to how this might be best achieved.

6.2 Parameter Prior Misspecification

As well as test the impact parameter misspecification is likely to have upon the different designs, a number of Monte Carlo simulations were performed on each design assuming the five different model estimation types. A total of 15,000 model simulations were performed with r = 100 iterations per each of the five model estimation types over the 10 design types assuming three different samples sizes (N = 100, 250 and 500). The MSE and RAE results for these simulations may be found at http://www.econ.usyd.edu.au/19129.html. Table 7 presents

the EMSE results for the simulations. Bolded values in the table represent the design type with the smallest EMSE for a given model type at the three different sample sizes.

Table 7: EMSE Results

			Model Assumed for Estimation			
		N	**MNL**	**MMNL (Pan.)**	**EC (Pan.)**	**MMNL (C.S.)**	**EC (C.S.)**
Orthogonal 1		100	0.00481	0.00438	0.00481	0.00653	0.01425
		250	0.00170	0.00143	0.00167	0.00225	0.01344
		500	0.00073	0.00063	0.00055	0.00100	0.00910
Orthogonal 2		100	0.00607	0.00557	0.00451	0.00590	0.02678
		250	0.00200	0.00146	0.00168	0.00182	0.02331
		500	0.00065	0.00070	0.00057	0.00048	0.02169
MNL		100	0.00302	0.00321	**0.00260**	0.00614	0.01637
		250	0.00109	0.00126	**0.00075**	0.00298	0.01299
		500	0.00040	0.00028	**0.00024**	0.00198	0.01040
MMNL (Pan.)		100	0.00378	**0.00272**	0.00320	0.00408	**0.00507**
		250	0.00138	**0.00106**	0.00106	0.00130	**0.00315**
		500	0.00038	**0.00025**	0.00030	0.00069	**0.00270**
EC (Pan.)		100	0.00217	0.00435	0.00260	**0.00341**	0.03147
		250	**0.00067**	0.00129	0.00083	**0.00117**	0.02448
		500	0.00032	0.00031	0.00032	0.00059	0.02085
MMNL (C.S.)		100	0.00673	0.00446	0.00525	0.00443	0.02155
		250	0.00209	0.00157	0.00171	0.00195	0.01927
		500	0.00068	0.00053	0.00057	0.00099	0.01863
EC (C.S.)		100	0.00317	0.00435	0.00352	0.00660	0.01625
		250	0.00094	0.00143	0.00105	0.00271	0.01362
		500	0.00047	0.00039	0.00048	0.00102	0.01308
Model Av. 1		100	0.00222	0.00470	0.00299	0.00410	0.00990
		250	0.00085	0.00242	0.00107	0.00143	0.00799
		500	0.00029	0.00099	0.00069	0.00048	0.00741
Model Av. 2		100	0.00287	0.00490	0.00295	0.00392	0.01619
		250	0.00084	0.00263	0.00110	0.00168	0.01385
		500	**0.00025**	0.00135	0.00050	0.00050	0.00731
Model Av. 3		100	**0.00202**	0.00425	0.00264	0.00416	0.00812
		250	0.00079	0.00248	0.00098	0.00149	0.00628
		500	0.00029	0.00108	0.00060	**0.00046**	0.00600
Percent of cells in which model average designs have lower EMSE values						
Model Av. 1			95.00%	5.00%	40.00%	80.00%	90.00%
Model Av. 2			95.00%	5.00%	55.00%	85.00%	70.00%
Model Av. 3			100.00%	25.00%	50.00%	85.00%	90.00%

(The leftmost vertical label reads: "Model Assumed for Design")

Not surprisingly, the two orthogonal designs tend to offer the highest amounts bias as represented by the EMSE values, although these designs do offer the least amount of bias for several specific parameter estimates as suggested by the MSE and RAE statistic values reported (e.g., the first orthogonal design produces the lowest MSE and RAE values for the β_2 parameter

estimated using the cross sectional EC model with 500 respondents). However, somewhat of a surprise is the fact that only in the case of the MMNL panel design estimated using that same model type is a design generated for a specific model type found to produce the least amount of bias when estimated on the model type for which it was generated. Indeed, the MMNL panel design appears to outperform the EC cross sectional design at all sample sizes whilst the EC panel design appears to have less bias in the parameter estimates obtained from the design when estimated using the cross sectional MMNL model.

At the base of Table 7, the percentage of times the model average designs are observed to have a lower EMSE value than the other designs are presented for each model estimation type. All three model average designs appear to perform well in comparison to the other designs generated in terms of the biases produced for all model types except for the two panel models. For the panel MMNL model, the first and second model average designs tend to produce larger errors than those observed for 95 percent of the other designs with the third model average design only marginally better with higher EMSE values than only 75 percent of designs. For the panel EC model, the three model average designs produce lower EMSE values than 40, 55 and 50 percent of the other designs over the three sample sizes used in the Monte Carlo simulations. When compared to the MNL and cross sectional MMNL and EC models, the three model average designs appear to perform exceptionally well compared to the other experimental designs.

7. Discussion and Conclusions

This paper addresses the issue of how to generate efficient SC experiments under uncertainty over the final model that will be estimated on data collected using the design. As we have shown, the AVC matrix likely to be obtained from a SC design will differ depending upon the model type that will be estimated once data is collected. To overcome this problem, the paper examines the use of a model averaging process of measures of statistical efficiency in generating experimental designs for SC studies. In our case study, we have also examined the impact of applying different priors on the model averaging process is likely to have. Finally, in exploring the idea of using model averaging techniques to generate efficient designs, we have further examined the use of Bayesian prior parameter distributions to compensate for a lack of a priori understanding of the exact parameter estimates that are also likely to result during application. As such, this paper attempts to handle the two major sources of uncertainty that exists when generated experimental designs for SC studies.

Whilst we have attempted to show via the use of a case study the importance of generating experimental designs specifically for econometric models for which they are likely to be applied to once data has been collected, the model averaging process explained herein provides a degree of robustness if the analyst is unsure as to exactly what model will be estimated post data collection. One short coming of this process however is that the model averaging process requires a much larger computational effort than generating designs specifically for a given model type. Furthermore, designs that are efficient for the cross-sectional MMNL and EC appear to be typically very different in terms of efficiency to other model types making inclusion of such model types difficult in the model averaging process. The results here also suggest that a good starting point in generating an efficient design would appear to be firstly generating a design specifically for the MNL model type and evaluate the efficiency this design under the other model assumptions, as the MNL design appears to perform very well independent of the final model estimated. Whilst Bliemer and Rose (2008) found similar results, further research is required to confirm this finding. Nevertheless, the model averaging process does offer many benefits, appearing to produce lower levels of parameter bias than most other design types, except when applied to panel MMNL and EC model estimation. Whilst

this finding is the result of a single case study, if such a result is confirmed to be more widespread, then one potential solution would be to weight these model types much more in the averaging process than the other model forms.

This paper also suffers from a number of limitations that need to be acknowledged. Firstly, we have employed here only a single case study, with fixed design dimensions. As such, the results presented herein are specific to the case study selected for analysis and does not allow for a detailed analysis of the impact different design dimensions might play in terms of design efficiency for model average type designs. Questions such as does having more or less choice situations impact upon the statistical efficiency of such designs or provide more or less bias or does having a wider attribute level range improve or detract from the efficiency of such designs? Similar questions can be asked as to the number of attributes and alternatives of designs? Whilst these questions have been asked about designs generated for other model types and answers attempted to be given (see e.g., Bliemer and Rose, 2005, and Rose and Bliemer, 2005), we have not attempted to address these issues here.

Secondly, in generating each of the designs, we have remained true to the growing number of experimental design papers by imposing a number of limitations such as attribute level balance within each of the designs. This means that each level of an attribute will appear an equal number of times over the course of the experiment. Whilst there may be reasons for imposing such a restriction, there also exist arguments against such restrictions. The main argument against such an imposition is that attribute level balance acts as a constraint in terms of allocating appropriate attribute levels across a design. For example, a particular choice situation may be improved if one particular level is swapped with that of another choice situation, however the swap may detract from the second choice situation. As such, attribute level balance may cause a number of choice situations to include dominant alternatives, despite the design being overall more efficient. This problem was identified by Toner et al. (1999) and Kanninen (2002) who proposed different methods such as allowing one attribute to be continuous, to provide substantial efficiency gains over designs with fixed attribute levels. Similar research for more advanced models is necessary.

Acknowledgments

References

Beaulieu, N.C. (1991) On the Generalized Multinomial Distribution, Optimal Multinomial Detectors, and Generalized Weighted Partial Decision Detectors, *IEEE Transactions on Communications*, 39(2), 193-194.

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

Bliemer, M.C.J., and J.M. Rose (2008) Construction of experimental designs for mixed logit models allowing for correlation across choice observations, *ITLS working paper: itls-wp-08-13*.

Bliemer, M.C.J., and J.M. Rose (2006) Designing Stated Choice Experiments: State-of-the-Art, *Proceedings of the 11th International Conference on Travel Behaviour Research*, Kyoto, Japan.

Bliemer, M.C.J., and J.M. Rose (2005) Efficiency and Sample Size Requirements for Stated Choice Studies, *ITLS working paper*: *itls-wp-05-08*.

Bunch, D.S., J.J. Louviere and D.A. Anderson (1994) A comparison of experimental design strategies for Multinomial Logit Models: The case of generic attributes, working paper, Graduate School of Management, University of California at Davis.

Burgess, L. and D.J. Street (2005) Optimal designs for choice experiments with asymmetric attributes, *Journal of Statistical Planning and Inference*, 134, 288-301.
Carlsson, F. and P. Martinsson (2002) Design Techniques for Stated Preference Methods in Health Economics, *Health Economics*, 12, 281-294.

Ferrini, S., and R. Scarpa (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.

Huber, J., and K. Zwerina (1996) The Importance of utility Balance and Efficient Choice Designs, *Journal of Marketing Research*, 33, 1996, pp. 307-317.

Kanninen, B.J. (2002) Optimal Design for Multinomial Choice Experiments, *Journal of Marketing Research*, 39(2), 214-217.

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

Louviere, J.J. and G. Woodworth (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data, *Journal of Marketing Research*, 20, 350-367.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour, In Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, New York, 105-142.

Ortúzar, J. de D., and L.G. Willumsen (2001) Modelling Transport. Third edition, John Wiley & Sons Inc., New York.

Revelt, D. and Train, K. (1998) Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level, *Review of Economics and Statistics*, 80(4), 647-657.

Rose, J.M., and M.C.J. Bliemer (2008) Stated preference experimental design strategies, in Handbook of Transport Modelling, ed. D.A Hensher and K.J. Button, Elsevier Oxford, United Kingdom, 151-79.

Rose, J.M., and M.C.J. Bliemer (2005) Constructing Efficient Choice Experiments, *ITLS working paper*: *itls-wp-05-07*.

Rose, J.M. and R. Scarpa (2008) Designs Efficiency for Non-market Valuation with Choice Modelling: How to Measure It, What to Report and Why, University Of Waikato: Working paper in Economics 21/07.

Sándor, Z., and M. Wedel (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, 38, 430-444.

Sándor, Z., and M. Wedel (2002) Profile Construction in Experimental Choice Designs for Mixed Logit Models, *Marketing Science*, 21(4), 455-475.

Sándor, Z., and M. Wedel (2005) Heterogeneous conjoint choice designs, *Journal of Marketing Research*, 42, 210-218.

Street, D.J., and L. Burgess (2004) Optimal and Near Optimal Pairs for the Estimation of Effects in 2-Level Choice Experiments, *Journal of Statistical Planning and Inference*, 118 (1–2), pp. 185–99.

Street, D.J., L. Burgess, and J.J. Louviere (2005) Quick and Easy Choice Sets: Constructing Optimal and Nearly Optimal Stated Choice Experiments, *International Journal of Research in Marketing*, 22, 459-470.

Toner, J.P., S.D., Clark, S.M., Grant-Muller and A.S. Fowkes (1999) Anything you can do, we can do better: a provocative introduction to a new approach to stated preference design, *WCTR Proceedings*, 3, Antwerp, 107-120.

Train, K. (2003) Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.