ITLS

WORKING PAPER

ITLS-WP-08-12

Efficient stated choice experiments for estimating nested logit models

By

Michiel CJ Bliemer*, John M Rose, and David A Hensher

*ITLS/Delft Partnership Visiting Professor

June 2008

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES
The Australian Key Centre in Transport and Logistics Management

The University of Sydney
Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-08-12

TITLE: **Efficient stated choice experiments for estimating nested logit models**

ABSTRACT: The allocation of combinations of attribute levels to choice situations in stated choice (SC) experiments can have a significant influence upon the resulting study outputs once data is collected. Recently, a small but growing stream of research has looked at using what have become known as efficient SC experimental designs to allocate the attribute levels to choice situations in a manner designed to produce better model outcomes. This research stream has shown that the use of efficient SC designs can lead to improvements in the reliability of parameter estimates derived from discrete choice models estimated on SC data for a given sample size. Unlike orthogonal designs, however, efficient SC experiments are generated in such a manner that their efficiency is related to the econometric model that is most likely to be estimated once the choice data is collected. To date, most of the research on efficient SC designs has assumed an MNL model format. In this paper, we generate efficient SC experiments for nested logit models and compare and contrast these with designs specifically generated assuming an MNL model form. We find that the overall efficiency of the design is maximized only when the model assumed in generating the design is the model that is fitted during estimation.

KEY WORDS: ***Stated choice, efficient experimental designs, nested logit, sample size***

AUTHORS: Michiel CJ Bliemer, John M Rose and David A Hensher

CONTACT: Institute of Transport and Logistics Studies **(C37)**

The Australian Key Centre in Transport Management

The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071

Facsimile: +61 9351 0088

E-mail: itlsinfo@itls.usyd.edu.au

Internet: http://www.itls.usyd.edu.au

DATE: June 2008

1. Introduction

Stated choice (SC) data has proven useful in studying many transportation related problems over the past two to three decades. For example, SC data has been used to examine the demand for a cycle-way networks (e.g., Ortúzar *et al.*, 2000), to examine the benefits derived from various calming measures on traffic (e.g., Garrod *et al.*, 2002)., to study the influences on parking choice (e.g., Shiflan and Bard-Eden, 2001; Hensher and King, 2001; van der Waerden *et al.*, 2002) and to establish the Value of Travel-Time Savings (VTTS) of commuters and non-commuters (e.g., Hensher, 2001a,b). Typically, SC experiments present sampled respondents with a number of different choice situations, each consisting of a universal but finite set of alternatives defined on a number of attribute dimensions. Respondents are then asked to specify their preferred alternatives given a specific hypothetical choice context. These responses may then be used by transport modelers to estimate models of choice behavior, which depending on the type of experiment conducted, may allow for the estimation of the direct or cross elasticities (or marginal effects) of the alternatives as well as on the marginal rates of substitution respondents are willing to make in trading between two attributes (i.e., willingness to pay measures, for example, VTTS).

Unlike most data, SC data requires that the analyst designs the experiment in advance by assigning attribute levels to the attributes that define each of the alternatives which respondents are asked to consider. Traditionally, the attribute levels are allocated to 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 situations using randomized, cyclical, Bayesian or foldover procedures (see for example, Bunch *et al.*, 1994; Louviere and Woodworth, 1983; Huber and Zwerina, 1996; Sandor and Wedel, 2001).

Whilst historically, researchers have tended to rely on orthogonal experimental designs (designs in which the attribute levels between different attributes are uncorrelated, see e.g., Louverie *et al.*, 2000) when conducting SC studies, a small but growing number of researchers have called into question this practice (e.g., Bliemer and Rose, 2006; Carlsson and Martinsson, 2003; Ferrini and Scarpa, 2007; Huber and Zwerina, 1996; Kanninen, 2002; Kessels *et al*., 2006; Sándor and Wedel, 2001, 2002, 2005; Rose and Bliemer, 2006). The central argument against the use of orthogonal designs is that the properties of orthogonality in SC data are not aligned with the properties of the discrete choice models typically estimated on SC data. In linear models, such as linear regression, orthogonality is important in that it avoids problems with multicollinearity in the estimated model, but more importantly, also results in the elements of the models variance-covariance matrix being minimized. It is this second point which is of primary importance. By minimizing the elements of the variance-covariance matrix of the model, the standard errors of the parameter estimates are also minimized, which in turn ensures that the *t*-ratios of the model are maximized.

Unfortunately, discrete choice models are not linear models and the variance-covariance matrices of the parameters of such models are obtained very differently to the variance-covariance matrices of linear regression models. McFadden (1974) showed that the asymptotic variance-covariance (AVC) matrix of the multinomial logit (MNL) model can be derived from the second derivatives of the log-likelihood function of the model. The same also holds for more advanced discrete choice models. Given that (i) the log-likelihood function and second derivatives of discrete choice models are dependent on the choice probabilities obtained from choice data, and (ii) that only differences in the utility of the chosen and non-chosen alternatives matter, the orthogonality of a SC design says little about the expected AVC matrix of the design.

Acknowledgement of this fact has resulted in a small but growing stream of research into experimental designs generated specifically to minimize the elements of the AVC matrices for discrete choice models. Such designs are known as efficient designs (see Bliemer and Rose, 2006 for a review of such designs). To date, most research on efficient designs have assumed an MNL model form (see Ferrini and Scarpa, 2007 and Sándor and Wedel, 2002, 2005 for the sole exceptions). In this paper, we examine the generation of efficient SC experimental designs to the nested logit (NL) model form, which has become a popular tool in estimating models based on SC data in the transportation area (e.g., Bhat and Castelar, 2002; Brownstone and Small, 2005; Cherchi and Ortúzar, 2002; Hess and Polak, 2006a,b; Hess *et al.*, 2006; Polydoropoulou and Ben-Akiva, 2001; Yao and Takayuki, 2005). Given the wide scale use of the nested logit model in SC related transportation studies, understanding how better to generate the SC designs for this model is an important issue, particularly given that the AVC matrix of the nested logit model is very different to that of an MNL model.

The NL model is a significant extension to the traditional MNL model. The primary motivation to switch from the MNL model to the NL model is the restrictive MNL assumption of independent and identically distributed (IID) error terms (and the related behavioral assumption - the Independence of Irrelevant Alternatives (IIA) assumption). A particularly important behavioral consequence of IID and IIA is that all pairs of alternatives are equally similar or dissimilar in terms of their unobserved influences (see for example, Ben-Akiva and Lerman, 1985; Louviere *et al.*, 2000; Hensher and Greene, 2002; Koppelman and Wen, 1998a,b; Hensher *et al.*, 2005). This has implications for the treatment of any attributes not observed.

In practice, it is often the case that different subsets of alternatives will share similar unobserved information content, which may translate into correlation between these unobserved influences amongst pairs of alternatives (i.e., non-zero and varying covariances for pairs of alternatives). Differences in error variance and non-zero covariances represent violations of the IID and IIA assumption. By relaxing the IID (and IIA) assumption(s) of the MNL model, the NL model overcomes these problems by allowing for different treatments of the error (co)variances across subsets of the alternatives contained within the model, hence negating the problems often associated with the MNL model.

The main contributions of the paper are two-fold. First, the research presented in this paper generalizes the current state-of-the-art of efficient SC designs towards the NL model, of which the MNL model is a special case. In Section 4, we show that unlike the MNL model, dependence on the choice observations is an issue in NL models, thus making the derivations more complex. To overcome this, it is necessary to rely on analytical approximations. Secondly, through the use of case studies, we demonstrate that the choice of model type (i.e., MNL or NL in this case) and also the nesting structure during the design generation process is important for the efficiency of the choice data at the time of estimation.

The remainder of the paper is organized as follows. In Section 2, we derive the NL model as necessary background before Section 3 discusses the theory on generating efficient experimental designs. In Section 4, we derive the AVC matrix for the NL model. Section 5 presents a case study in which we generate and compare SC experimental designs, and also illustrate losses in efficiency if a different model type or nesting is used for estimation than the design is generated for.

2. Nested logit model

Adopting the definitions used in Hensher and Greene (2002)[1], the elements in the NL model have a tree structure in which the top-level alternatives are referred to as branches, and the alternatives residing at the bottom of the tree structure as elemental alternatives. An example is shown in Figure 1. Typically, the elemental alternatives represent the alternatives individuals are directly faced with in choice situations. The branches of the model then define groupings of elemental alternatives that are assumed to be more or less similar in terms of having the same error variances in their utility functions.

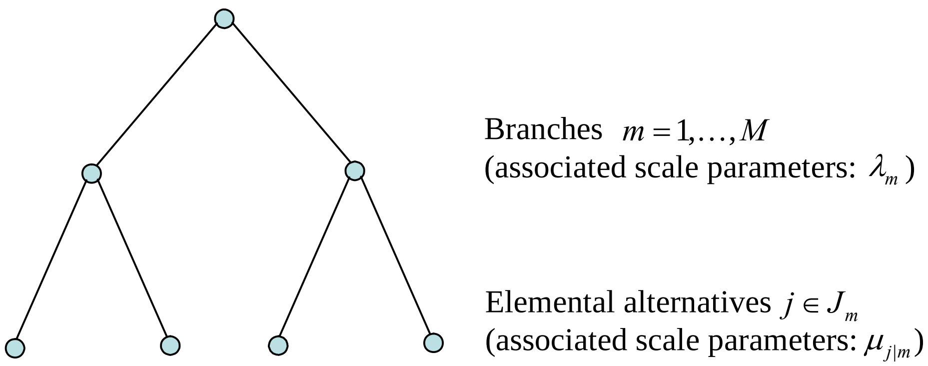

Figure 1: Two level nested logit tree

Within the NL model, each branch and elemental alternative will have an associated scale parameter. Let λ_m denote the scale parameter of branch *m*, and let $\mu_{j|m}$ denote the scale parameter of elemental alternative *j* within branch *m*. Let J_m denote the set of all elemental alternatives belonging to branch *m*. By definition, all scale parameters of the elemental alternatives in this set J_m have the same scale parameter. That is, $\mu_{j|m} = \bar{\mu}_m$ for all $j \in J_m$ for some value $\bar{\mu}_m$, while the scale parameters of elemental alternatives below different branches need not be the same. If $\bar{\mu}_m$ is the same for each branch, *m*, then the NL model will collapse to an MNL model.

Each elemental alternative is assumed to have a corresponding utility function. Given that an individual has chosen an elemental alternative in branch *m*, the utility $U_{jt|m}$ of elemental alternative *j* and individual *t* consists of an observed utility component $V_{jt|m}$ and the unobserved random component $\varepsilon_{jt|m}$,

[1] We will restrict ourselves to NL models with two levels, being the most common. However, the theory can be extended to include more than two levels.

$$U_{jt|m} = V_{jt|m} + \varepsilon_{jt|m}, \quad \forall j \in J_m; m = 1,\ldots,M; t = 1,\ldots,T. \tag{1}$$

The observed utility is assumed to be a linear combination of attribute values $x_{jkt|m}$ (the explanatory variables) with associated weights β_k for each attribute *k*,

$$V_{jt|m} = \mu_{j|m} \sum_{k=1}^{K} \beta_k x_{jkt|m}, \quad \forall j \in J_m; m = 1,\ldots,M; t = 1,\ldots,T. \tag{2}$$

If the same parameter appears in multiple elemental alternatives, then the parameter is called generic across these alternatives, otherwise it is called alternative-specific.[2] Note that we have added scale parameter $\mu_{j|m}$ to the observed utility component to account for differences in scale parameters. The unobserved random components $\varepsilon_{jt|m}$ for all $j \in J_m$ and all decision makers, *t*, are assumed to be independently and identically extreme value type I distributed. The probability of choosing elemental alternative *j* given that branch *m* was chosen can therefore be seen as a simple MNL model, yielding the following conditional probability (McFadden, 1974):

$$P_{jt|m} = \frac{\exp(V_{jt|m})}{\sum_{i \in J_m} \exp(V_{it|m})}, \quad \forall j \in J_m; m = 1,\ldots,M; t = 1,\ldots,T. \tag{3}$$

The utility of a branch *m* is in the literature called an inclusive value (IV) variable. Multiplying this IV variable with the IV parameter (that is, the branch scale parameter) yields this combined observed utility for branch *m* (see Ben-Akiva and Lerman, 1985; Hensher and Greene, 2002; Carrasco and Ortúzar, 2002):

$$V_{mt} = \frac{\lambda_m}{\bar{\mu}_m} \log\left(\sum_{j \in J_m} \exp(V_{jt|m}) \right), \quad \forall m = 1,\ldots,M; t = 1,\ldots,T. \tag{4}$$

The probability of choosing branch *m* is given by

$$P_{mt} = \frac{\exp(V_{mt})}{\sum_{n=1}^{M} \exp(V_{nt})}, \quad \forall m = 1,\ldots,M; t = 1,\ldots,T. \tag{5}$$

---

[2] If attribute *k* does not appear in alternative *j* in branch *m*, then $x_{jkt|m}$ is simply set to 0.

Combining equations (3)–(5) yields the unconditional probability that a decision maker, t, chooses elemental alternative j:

$$P_{jt} = P_{mt} P_{jt|m} = \frac{\left(\sum_{i \in J_m} \exp(V_{it|m}) \right)^{\lambda_m / \bar{\mu}_m}}{\sum_{n=1}^{M} \left(\sum_{i \in J_m} \exp(V_{it|n}) \right)^{\lambda_n / \bar{\mu}_n}} \cdot \frac{\exp(V_{jt|m})}{\sum_{i \in J_m} \exp(V_{it|m})}, \quad \forall j \in J_m;\, m = 1, \ldots, M;\, t = 1, \ldots, T. \quad (6)$$

Unfortunately, the NL model is over-identified and to be able to estimate the parameters (including scale parameters), and as commonly done we normalize all scale parameters of the lower level, i.e. $\mu_{j|m} = \bar{\mu}_m = 1$ for all $j \in J_m$ and all m. For more details on the nested logit model, see for example, Ben-Akiva and Lerman (1985) or Hensher *et al.* (2005). The next section will describe how to construct an efficient design for estimating the set of parameters.

3. Efficient stated choice experiments

The literature on generation of efficient designs for stated choice (SC) experiments state as their basis, the seminal work by McFadden (1974) and described in detail in Ben-Akiva and Lerman (1985) and Louviere *et al*. (2000). Following on from this, the majority of work conducted on the generation of efficient SC experiments has relied on this model (e.g., Bliemer and Rose, 2005; Carlsson and Martinsson, 2003; Huber and Zwerina, 1996; Kanninen, 2002; Kuhfeld *et al.*, 1994; Rose and Bliemer, 2005; Sándor and Wedel, 2001). In this section we will design SC experiments allowing for different error variances across the alternatives (i.e., for the NL model).

Stated choice experiments present sampled respondents with a number of different choice situations, each consisting of a universal but finite set of alternatives defined on a number of attribute dimensions. Respondents are then asked to specify their preferred alternatives given a specific hypothetical choice context. SC data requires that the analyst designs the experiment in advance by assigning attribute levels to the attributes that define each of the alternatives which respondents are asked to consider.

Orthogonal designs are widely used in SC experiments; however, this class of designs may not be statistically 'efficient', as they do not take the SC model specification into account.[3] 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. These efforts have concentrated on methods to promote greater gains in the statistical efficiency of SC experiments (e.g., Anderson and Wiley, 1992; Bunch *et al.*, 1994; Carlsson and Martinsson 2003; Huber and Zwerina, 1996; Kanninen, 2002; Laziri and Anderson, 1994; Sándor and Wedel, 2001). Common amongst all of these efforts is minimization of the elements of the AVC matrix of the models to be fitted to SC data.

[3] Originally, orthogonal designs were employed in conjunction with linear regression models, for which they are efficient. However, for other model types such as discrete choice models, orthogonal designs are no longer efficient which has been largely unrecognized by many researchers.

In order to estimate the likely AVC matrix of a SC experiment, the analyst is required to assume a set of prior parameter estimates (Huber and Zwerina, 1996; Sándor and Wedel, 2001).[4] These priors are used to calculate the expected utilities as well as choice probabilities of each of the alternatives. From these choice probabilities, it is possible to calculate the AVC matrix of the model to be estimated. Through manipulation of the design, 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.

Important issues that need to be addressed are how to determine the priors and how to determine the best nesting structure. Often, pilot studies with a small number of respondents are conducted before generating the final experimental design to be given to a large number of respondents. A simple orthogonal design could be used for the pilot study and a range of model types and nesting structures can be estimated. This will give indications of the most suitable nesting structure and corresponding parameters. These parameters can then be used as priors for generating the efficient design.

Let Σ_T denote the AVC matrix of all parameter estimates using a sample size of *T*. There are *K* parameters to estimate and there are *M* scale parameters to estimate for each branch, such that the total number of parameters to estimate is $K+M$. Hence, Σ_T is a symmetric matrix of dimension $(K+M)\times(K+M)$. In the literature, several measures for determining the efficiency of a design have been proposed, such as the D-error, A-error, etc. (see e.g., Huber and Zwerina, 1996). The A-error aims to minimize the average asymptotic variances (of which the roots are the asymptotic standard errors of the parameter estimates). This measure is sensitive to the scale of the parameters. The D-error does not exhibit this sensitivity and as such has become the mostly widely used measure of efficiency. Thus, we adopt this measure for the remainder of the paper.[5] The D-error is the determinant of this matrix (assuming a single respondent) with a certain scaling taking the number of parameters into account:

$$\text{D-error} = \det(\Sigma_1)^{1/(K+M)}. \tag{7}$$

The lower this D-error, the higher the efficiency of the design and therefore the greater the asymptotic efficiency of the parameter estimates. Instead of assuming fixed prior parameters, one could also assume prior parameter distributions and computed the expected D-error over these probability distributions. This will lead to so-called Bayesian efficient designs, see e.g., Sándor and Wedel (2005), Bliemer *et al.* (2006). Algorithms to find efficient designs are briefly discussed in Appendix A.

---

[4] One reviewer questioned the use of prior parameter estimates in generating SC experiments and the impact such priors may have on final model results. The need for using prior parameters need not be a major concern as one can always choose to use zero prior values, which will yield designs with an efficiency equivalent to that of an orthogonal design. However, using information on priors, even if one just knows the sign, one can improve on the efficiency of the design. If priors do not exist in the literature, they can always be obtained from a small pilot study. Misspecification of priors may decrease the efficiency of the design, but the efficiency will in general still be better than assuming zero priors. We would argue that the purpose of research is to recover the population parameters, and that any design, even one randomly generated will recover the true population parameters in large enough sample sizes. The purpose of efficient designs is to recover these at much smaller sample sizes, and even with misspecification of the prior parameters, efficient designs tend to perform much better than other design types (see Bliemer and Rose, 2006).

[5] Note that using a different efficiency measure may lead to a different (optimal) efficient design.

The (expected) D-error will be used in this paper as a criterion to determine efficient designs. The next section will be devoted to the calculation of the AVC matrix.

4. Deriving the asymptotic variance-covariance matrix

In this section, we demonstrate that the AVC matrix can be computed using the design attribute values and prior parameter values. The following important result, proven by McFadden (1974) for the MNL model, will be used to establish the AVC matrix for the NL model. Let $(\bar{\beta},\bar{\lambda})$ denote the true parameter values of the NL model. Furthermore, suppose that $(\hat{\beta}_T,\hat{\lambda}_T)$ are the maximum likelihood estimators of the model for a sample size of *T*. It then holds that $(\hat{\beta}_T,\hat{\lambda}_T)$ is asymptotically normal as $T \to \infty$ with mean $(\bar{\beta},\bar{\lambda})$ and variance-covariance matrix $\Sigma_T = \Omega_T^{-1}$, where Ω_T is the Fisher information matrix consisting of the negative expected second derivatives of the log-likelihood function. Hence, if $L_T(\beta,\lambda)$ is the log-likelihood function (assuming *T* respondents), then it follows that:

$$\Omega_T = -E\left(\frac{\partial^2 L_T(\beta,\lambda)}{\partial(\beta,\lambda)\partial(\beta,\lambda)'}\right). \tag{8}$$

In a SC design, different combinations of attribute levels are shown to respondents in each choice situation. Let *S* be the total number of choice situations (choice sets) faced by each individual respondent *t*, where *s* denotes each specific choice situation, and let $x_{jkts|m}$ be the attribute level related to the k^{th} attribute associated with the j^{th} alternative (in branch *m*) shown to respondent *t* in choice situation *s*. As is commonly the case, it is assumed that all respondents face the same choice situations (although more sophisticated approaches exist and the derivations here can be extended), such that the sub-index *t* can be dropped from the attribute variables and the probabilities. The log-likelihood function can thus be written as

$$L_T(\beta,\lambda) = \sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{m=1}^{M}\sum_{j\in J_m} y_{jts}\log P_{js} = \sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{m=1}^{M} y_{mts}\sum_{j\in J_m} y_{jts|m}\log P_{js|m}P_{ms}, \tag{9}$$

where y_{jts} is the vector of outcomes of the SC experiment. This indicator value is equal to one if elemental alternative *j* is chosen in choice situation *s* by respondent *t*, and zero otherwise. Furthermore, this vector is decomposed such that $y_{jts} = y_{jts|m}y_{mts}$, where (for all respondents *t* and choice sets *s*) y_{mts} equals one if branch *m* is 'chosen' and zero otherwise, and $y_{jts|m}$ equals one if elemental alternative *j* is chosen given that branch *m* was chosen, and zero otherwise. The second derivatives needed in equation (10) are derived in Appendix B, which after some manipulations finally lead to the following equations:

$$-E\left(\frac{\partial^2 L_T(\beta,\lambda)}{\partial\beta_{k_1}\partial\beta_{k_2}}\right) = -T\sum_{s=1}^{S}\left\{\sum_{m=1}^{M} P_{ms}\left[(\lambda_m - 1)\left(\sum_{i\in J_{mk_1k_2}} x_{ik_1s|m}P_{is|m}x_{ik_2s|m} - \sum_{i\in J_{mk_1}} x_{ik_1s|m}P_{is|m}\sum_{j\in J_{mk_2}} P_{js|m}x_{jk_2s|m}\right)\right]\right.$$
$$-\left(\sum_{m=1}^{M}\lambda_m P_{ms}\left[\left(\lambda_m\sum_{i\in J_{mk_2}} P_{is|m}x_{ik_2s|m} - \sum_{n=1}^{M}\lambda_n P_{ns}\sum_{i\in J_{nk_2}} P_{is|n}x_{ik_2s|n}\right)\sum_{i\in J_{mk_1}} x_{ik_1s|m}P_{is|m}\right.\right.$$
$$\left.\left.\left. + \sum_{i\in J_{mk_1k_2}} x_{ik_1s|m}P_{is|m}x_{ik_2s|m} - \sum_{i\in J_{mk_1}} x_{ik_1s|m}P_{is|m}\sum_{j\in J_{mk_2}} P_{js|m}x_{jk_2s|m}\right]\right)\right\} \tag{10a}$$

$$-E\left(\frac{\partial^2 L_T(\beta,\lambda)}{\partial\lambda_{m_1}\partial\beta_{k_2}}\right) = T\sum_{s=1}^{S}\left[\log\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)P_{m_1s}\left(\lambda_{m_1}\sum_{i\in J_{m_1k_2}} P_{is|m_1}x_{ik_2s|m_1} - \sum_{n=1}^{M}\lambda_n P_{ns}\sum_{i\in J_{nk_2}} P_{is|n}x_{ik_2s|n}\right)\right] \tag{10b}$$

$$-E\left(\frac{\partial^2 L_T(\beta,\lambda)}{\partial\lambda_{m_1}\partial\lambda_{m_2}}\right) = \begin{cases} T\sum_{s=1}^{S} P_{m_1s}\left(1-P_{m_1s}\right)\left(\log\left[\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)\right]\right)^2, & \text{if } m_1 = m_2 \\ -T\sum_{s=1}^{S} P_{m_1s}P_{m_2s}\log\left[\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)\right]\log\left[\sum_{i\in J_{m_2}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_2}\right)\right], & \text{if } m_1 \neq m_2 \end{cases} \tag{10c}$$

It should be pointed out that, unlike in the MNL model, in the NL model the second derivatives are not independent of observed choices, y; see Appendix B. However, for large T (as we are interested in the asymptotic properties), we can use substitutions of probabilities as shown in Appendix A, such that equations (10a,b,c) are independent again of y. In case $\lambda_m = 1$ for all branches m, the Fisher information matrix of the NL model in equation (10a) collapse to that of the MNL model as reported in Bliemer and Rose (2005), namely

$$-E\left(\frac{\partial^2 L_T(\beta)}{\partial\beta_{k_1}\partial\beta_{k_2}}\right) = T\sum_{s=1}^{S}\sum_{j\in J_{k_1}} P_{js}x_{jk_1s}\left(x_{jk_2s} - \sum_{i\in J_{k_2}} P_{is}x_{ik_2s}\right). \tag{11}$$

Using equations (12a,b,c), the Fisher information matrix, Ω_T, can be calculated and hence, the AVC matrix, Σ_T, can be computed as well (for each sample size T).

The only remaining unknowns are the parameter values. The true parameter values $(\bar{\beta},\bar{\lambda})$ are to be estimated from the model. In constructing efficient designs, it is common to assume prior parameter values $(\tilde{\beta},\tilde{\lambda})$ and optimize the design based on these priors. Considerable effort should be committed to identifying reasonable prior parameter values, as doing so will result in significant pay-offs in terms of requiring (much) smaller sample sizes in the actual SC experiment.

As indicated by equations (10a,b,c), the Fisher information matrix is proportional to the sample size *T*, hence, the AVC matrix will be proportional to $1/T$. The square roots of the diagonals of the AVC matrix represent the asymptotic standard errors of the parameter estimates, and hence, these standard errors (*se*) are proportional to $1/\sqrt{T}$. In other words, $se_T(\beta_k) = se_1(\beta_k)/\sqrt{T}$, where $se_T(\beta_k)$ is the asymptotic standard error of parameter β_k assuming a sample size of *T*, while $se_1(\beta_k)$ is the asymptotic standard error considering only a single respondent. As such, similar to the MNL model, the NL model will exhibit diminishing increases in reliability (as measured by lower asymptotic standard errors) as we increase the sample size.

Interestingly, having the prior parameter values $\tilde{\beta}_k$ and also the asymptotic standard errors for each sample size *T*, one is also able to compute the asymptotic *t*-values. In order to test parameters being statistically significant from zero, or scale parameters from one (i.e. testing whether that part of the tree structure is actually an MNL model), the following asymptotic *t*-values can be computed:

$$t_T = \frac{\beta_k}{se_1(\tilde{\beta}_k)/\sqrt{T}}, \qquad \text{and} \qquad t_T = \frac{\lambda_m}{se_1(\tilde{\lambda}_m)/\sqrt{T}}. \tag{12}$$

Re-arranging this term and assuming a certain significance level (e.g., $t_T \geq 1.96$) yields the following theoretical minimum sample sizes T_k^* and T_m^* for obtaining a statistically significant parameter estimate for parameter β_k and λ_m, respectively:

$$T_k^* \geq \left(\frac{t_T se_1(\tilde{\beta}_k)}{\tilde{\beta}_k} \right)^2, \qquad \text{and} \qquad T_m^* \geq \left(\frac{t_T se_1(\tilde{\lambda}_k)}{\tilde{\lambda}_k - 1} \right)^2. \tag{13}$$

These (asymptotic) theoretical minimum sample sizes can be used to assess the efficiency of a design for each parameter estimate separately, instead of for all parameters combined when using the D-error measure. The minimum sample sizes derived from equation (13) hold only under the asymptotic assumption and therefore should be interpreted carefully as representing the lower bounds for the sample sizes. In practice, it is likely that (much) larger sample sizes are required. However, these theoretical sample sizes can give indications about which parameters are likely to be difficult to estimate once data is collected using the design.

It should be stressed that the above derivations of the AVC matrix hold under the assumption of independent choice situations. In the case where respondents face multiple choice situations, as with SC data, the assumption of independence typically does not hold and therefore may bias the derived AVC matrix. Basically all literature on generating efficient designs assumes independent choice situations, thereby intrinsically assuming independent choice situations. In this paper, we recognize this problem, but do not aim to resolve the issue. The only way to correct for dependent choice situations is to switch to a mixed logit model formulation in which the dependency is treated as in panel data estimation. To our best knowledge, the only paper dealing with this topic of dependent choice situations is Bliemer and Rose (2008). In that paper, they show that it is possible in theory to generate designs for the panel version of the mixed logit model, but in practice it adds great computational complexity to the design generation process. Some preliminary results show, however, that experimental design results for the MNL model are much closer to those of the panel mixed logit model than for the cross-sectional mixed logit model (e.g., considered in Sándor and Wedel, 2002). The MNL model being a special case of the NL model, this may suggest that creating designs assuming a nested logit form is not far from results obtained for a panel version. A compromise would be to generate an efficient design based on the NL model, and then assess if the design assuming a panel formulation also has a good efficiency.

5. Case studies

In this section, we generate a number of experimental designs for several models in which we allow for differences in the error variances across subsets of alternatives. In Section 5.1 a typical example of a nested logit model in transportation will be formulated and will serve as the basis of most of the analyses in the subsequent sections. Different (orthogonal and/or efficient) designs will be compared in Section 5.2. Potential effects of misspecification of prior parameters (in particular, the scale parameters) are discussed in Section 5.3. Also in this section, a Bayesian efficient design is analyzed, which is less sensitive to prior parameter misspecification. Section 5.4 examines the impact of model misspecification, which shows that generating a design for an MNL model while the true model is an NL model (or the other way around) will lead to a loss in efficiency.

5.1 Model formulation and design dimensions

The example we use involves an experiment with four elemental alternatives; ‘car on toll road’ (cart), ‘car on non-tolled road’ (carnt), ‘bus’, and ‘train’. The observed components of these elemental alternatives are represented by

$$V_s^{\text{cart}} = \beta_0^{\text{cart}} + \beta_1^{\text{car}}\text{TT}_s^{\text{cart}} + \beta_2^{\text{car}}\text{RC}_s^{\text{cart}} + \beta_3^{\text{cart}}\text{TOLL}_s^{\text{cart}}, \tag{14}$$

$$V_s^{\text{carnt}} = \beta_1^{\text{car}}\text{TT}_s^{\text{carnt}} + \beta_2^{\text{car}}\text{RC}_s^{\text{carnt}}, \tag{15}$$

$$V_s^{\text{bus}} = \beta_0^{\text{bus}} + \beta_1^{\text{bus}}\text{TT}_s^{\text{bus}} + \beta_2^{\text{bus}}\text{FARE}_s^{\text{bus}}, \tag{16}$$

$$V_s^{\text{train}} = \beta_1^{\text{train}}\text{TT}_s^{\text{train}} + \beta_2^{\text{train}}\text{FARE}_s^{\text{train}}. \tag{17}$$

where TT represents travel time, RC running costs, TOLL toll costs, and FARE the fare costs. The elemental alternatives are nested into two groups: car = {cart, carnt} for car alternatives and pt = {bus, train} for public transport alternatives. The IV variables of these nests are given by

$$V_s^{\text{car}} = \lambda^{\text{car}} \log\left(\exp(V_s^{\text{cart}}) + \exp(V_s^{\text{carnt}})\right), \tag{18}$$

$$V_s^{\text{pt}} = \lambda^{\text{pt}} \log\left(\exp(V_s^{\text{bus}}) + \exp(V_s^{\text{train}})\right). \tag{19}$$

There are in total 11 parameters to be estimated, of which nine at the elemental level (of which two generic parameters, β_1^{car} and β_2^{car}, and seven alternative-specific parameters), and two scale parameters. There are nine attributes for which different combinations of attribute levels have to be determined in the experimental design. The attribute levels and priors to be used in generating the design are listed in Table 1. The priors were chosen based on previous study results as well as to preserve realistic estimates of VTTS for the car and public transport alternatives. Following common practice, we limit the experiments to attribute level balanced designs (although such a constraint may result in the generation of a sub-optimal design).

Table 1: Prior parameter values and attribute levels for case study 1

Prior parameter values:

β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
-0.4	-0.5	-0.9	-1.3	-0.4	-0.4	-1.5	-0.45	-1.6	1	0.6

Attribute levels:

TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{\text{cart}}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{\text{bus}}$ (A$)	TT_s^{train} (min.)	$FARE_s^{\text{train}}$ (A$)
10	1	2	20	2	40	1	30	2
20	2	3	30	3	50	2	40	3
30	3	4	40	4	60	3	50	4

5.2 Efficiency of different designs

As our focus in the case study will be to compare the efficiency of different designs, *ceteris paribus* (given a certain model type and prior parameter values), we keep all design dimensions fixed (i.e., the number of alternatives, the number of attributes, the number of choice situations, the number of attribute levels, the attribute level range, etc.), while only changing the order that the attribute levels appear within the design. Later in the paper, we investigate losses of efficiency under prior parameter and model type misspecification using the same design dimensions. The number choice situations chosen (i.e., 12) was selected such that both attribute level balance and orthogonality can be achieved. Whilst we could have chosen different design dimensions, the purpose of the paper is not to test the impact of these dimensions on the design efficiency. Rather, the aim of the paper is to illustrate the methodology and demonstrate the importance of selecting more efficient SC designs for the particular model type that is likely to be estimated, once data has been collected.

We generate three different (attribute level balanced) designs with 12 choice situations assuming the above NL model: an efficient (non-orthogonal) design, and two orthogonal designs. Orthogonal designs, in which the attribute levels of different attributes are all uncorrelated, are still the mainstream design type to use. For this reason we include orthogonal designs in our comparison. As it is common practice when undertaking SC studies to only generate a single orthogonal design, and in doing so, not test the overall efficiency of the design, it is possible (indeed probable) that researchers may obtain sub-optimal designs. The two orthogonal designs we generate consist of the most efficient one and the least efficient one, which allows for an examination of the likely range of impact upon the reliability of the parameter estimates obtained from different orthogonal designs of the same dimensions.

Table 2 shows the three generated designs. Shown in the table are the attribute level combinations for each of the final designs constructed, as well as the D-error values for the designs computed from the AVC matrices derived using equations (10a,b,c) (see Table 3). In calculating the D-error values, we remove the rows and columns from the AVC matrix related to the two alternative-specific constants.[6]

To generate the designs, we used Ngene 0.7,[7] which evaluates in an intelligent way the D-error of a large number of possible designs, and stores the most efficient design (with or without the restriction of orthogonality). For the non-orthogonal designs, the algorithm employed a simple swapping procedure similar to that discussed in Appendix B or Huber and Zwerina (1996) and Sándor and Wedel (2001).

Table 2: Experimental designs for case study 1

Efficient design for NL model D-error = 0.1421

s	TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{cart}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{bus}$ (A$)	TT_s^{train} (min.)	$FARE_s^{train}$ (A$)
1	10	1	4	20	4	50	2	40	3
2	20	1	2	40	3	40	1	30	4
3	30	2	3	30	2	50	3	50	2
4	20	1	2	20	4	60	2	40	4
5	30	2	3	30	2	60	1	50	2
6	10	3	4	40	3	40	3	30	4
7	20	1	4	30	4	60	1	50	3
8	30	2	3	30	4	60	3	50	4
9	10	3	2	20	2	50	2	30	3
10	10	3	4	20	2	50	2	40	3
11	20	2	2	40	3	40	3	40	2
12	30	3	3	40	3	40	1	30	2

[6] These rows and columns are generated from the design, however, they are simply ignored in calculating the D-error measure. For many SC studies, it is the attribute parameters (or ratios thereof) which are of prime importance, and not the constants (Hensher *et al.*, 2005). As the D-error measure is a global measure of statistical efficiency, minimization of this measure often results in trade-offs having to be made between elements of the AVC matrix of designs. Inclusion of constant terms may therefore result in lower efficiency levels being achieved (i.e., higher standard errors) for the design attributes which are often considered to be of more importance.

[7] Ngene, developed by Econometric Software Inc., is dedicated software for generating experimental designs for stated choice studies, and currently has prototype status.

Orthogonal design (most efficient) for NL model D-error = 0.2983

s	TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{cart}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{bus}$ (A$)	TT_s^{train} (min.)	$FARE_s^{train}$ (A$)
1	20	3	3	30	2	60	2	50	3
2	30	1	3	20	4	40	2	50	3
3	30	1	2	40	3	60	1	40	2
4	10	2	2	40	4	50	3	30	3
5	30	3	4	40	3	40	3	40	2
6	10	3	2	20	3	40	1	40	2
7	20	1	3	30	2	40	2	30	4
8	20	2	2	30	2	50	3	50	4
9	30	3	3	20	4	60	2	30	4
10	10	2	4	40	4	50	1	50	4
11	10	1	4	20	3	60	3	40	2
12	20	2	4	30	2	50	1	30	3

Orthogonal design (least efficient) for NL model D-error = 1.0477

s	TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{cart}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{bus}$ (A$)	TT_s^{train} (min.)	$FARE_s^{train}$ (A$)
1	20	3	3	30	4	60	2	30	3
2	10	1	3	40	2	40	2	30	3
3	10	1	2	20	3	60	1	40	4
4	30	2	2	20	2	50	3	50	3
5	10	3	4	20	3	40	3	40	4
6	30	3	2	40	3	40	1	40	4
7	20	1	3	30	4	40	2	50	2
8	20	2	2	30	4	50	3	30	2
9	10	3	3	40	2	60	2	50	2
10	30	2	4	20	2	50	1	30	2
11	30	1	4	40	3	60	3	40	4
12	20	2	4	30	4	50	1	50	3

Table 3: Asymptotic variance-covariance (AVC) matrices for case study 1

AVC matrix for efficient design

	β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
β_0^{cart}	**9.04**	0.10	0.41	-2.27	1.38	-0.10	-0.16	-0.09	-0.03	0.18	-0.16
β_1^{car}	0.10	**0.24**	0.36	0.66	-0.03	0.00	-0.02	0.00	-0.02	0.43	-0.04
β_2^{car}	0.41	0.36	**0.84**	0.97	0.16	-0.01	-0.01	-0.01	0.03	0.64	-0.10
β_3^{cart}	-2.27	0.66	0.97	**2.61**	-0.31	0.03	-0.01	0.02	0.01	1.15	-0.08
β_0^{bus}	1.38	-0.03	0.16	-0.31	**28.47**	-0.53	-2.34	-0.13	0.76	0.02	-0.07
β_1^{bus}	-0.10	0.00	-0.01	0.03	-0.53	**0.14**	0.48	0.14	0.51	-0.01	0.19
β_2^{bus}	-0.16	-0.02	-0.01	-0.01	-2.34	0.48	**2.39**	0.50	1.92	-0.06	0.68
β_1^{train}	-0.09	0.00	-0.01	0.02	-0.13	0.14	0.50	**0.15**	0.54	-0.01	0.20
β_2^{train}	-0.03	-0.02	0.03	0.01	0.76	0.51	1.92	0.54	**2.67**	-0.01	0.80
λ^{car}	0.18	0.43	0.64	1.15	0.02	-0.01	-0.06	-0.01	-0.01	**1.04**	0.10
λ^{pt}	-0.16	-0.04	-0.10	-0.08	-0.07	0.19	0.68	0.20	0.80	0.10	**0.40**

AVC matrix for orthogonal design (most efficient)

	β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
β_0^{cart}	**14.76**	1.45	1.23	0.32	-2.70	0.19	2.35	0.24	0.47	1.57	-0.52
β_1^{car}	1.45	**0.58**	0.95	1.49	-0.18	0.02	0.14	0.03	-0.02	0.91	-0.16
β_2^{car}	1.23	0.95	**2.76**	3.08	5.86	-0.08	-1.17	-0.04	0.34	1.77	-0.29
β_3^{cart}	0.32	1.49	3.08	**5.15**	2.40	-0.02	-0.78	-0.01	-0.08	2.46	-0.49
β_0^{bus}	-2.70	-0.18	5.86	2.40	**102.76**	-0.75	-10.4	0.44	9.42	-0.30	0.45
β_1^{bus}	0.19	0.02	-0.08	-0.02	-0.75	**0.18**	0.93	0.20	0.61	0.00	0.25
β_2^{bus}	2.35	0.14	-1.17	-0.78	-10.40	0.93	**7.69**	1.08	2.43	-0.11	1.27
β_1^{train}	0.24	0.03	-0.04	-0.01	0.44	0.20	1.08	**0.25**	0.77	0.01	0.31
β_2^{train}	0.47	-0.02	0.34	-0.08	9.42	0.61	2.43	0.77	**4.28**	-0.25	0.98
λ^{car}	1.57	0.91	1.77	2.46	-0.30	0.00	-0.11	0.01	-0.25	**2.13**	0.10
λ^{pt}	-0.52	-0.16	-0.29	-0.49	0.45	0.25	1.27	0.31	0.98	0.10	**0.67**

AVC matrix for orthogonal design (least efficient)

	β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
β_0^{cart}	**224.3**	9.09	31.52	-36.1	55.65	-0.28	-4.03	0.75	4.16	17.74	-1.70
β_1^{car}	9.09	**1.48**	1.35	1.39	4.02	0.01	-0.59	-0.01	1.20	2.85	0.13
β_2^{car}	31.52	1.35	**7.49**	-4.57	4.80	-0.14	1.95	0.13	-0.63	2.95	-0.43
β_3^{cart}	-36.1	1.39	-4.57	**13.51**	-2.40	0.10	-0.55	-0.20	2.25	2.61	0.74
β_0^{bus}	55.65	4.02	4.80	-2.40	**422.6**	-4.18	-31.4	2.42	25.95	4.26	2.30
β_1^{bus}	-0.28	0.01	-0.14	0.10	-4.18	**0.24**	0.67	0.16	0.71	0.06	0.26
β_2^{bus}	-4.03	-0.59	1.95	-0.55	-31.4	0.67	**19.60**	0.92	3.29	-0.07	1.71
β_1^{train}	0.75	-0.01	0.13	-0.20	2.42	0.16	0.92	**0.28**	0.48	0.01	0.33
β_2^{train}	4.16	1.20	-0.63	2.25	25.95	0.71	3.29	0.48	**15.74**	2.13	1.92
λ^{car}	17.74	2.85	2.95	2.61	4.26	0.06	-0.07	0.01	2.13	**5.99**	0.52
λ^{pt}	-1.70	0.13	-0.43	0.74	2.30	0.26	1.71	0.33	1.92	0.52	**0.75**

As expected, the efficient design (without the orthogonality restriction) produces the lowest D-error value (0.1421), while the two orthogonal designs produce higher D-error values (0.2983 and 1.0477 for the most and least efficient orthogonal designs, respectively), see Table 2. The D-error of the most efficient orthogonal design is 2.1 times greater than the D-error value of the efficient design, while D-error of the least efficient orthogonal design is even 7.4 times that of the efficient design. This suggests that on average, the asymptotic standard errors of the parameter estimates using the orthogonal designs will be $\sqrt{2.1} \approx 1.4$ to $\sqrt{7.4} \approx 2.7$ times larger than the average asymptotic standard errors of the efficient design. This is confirmed by examining the AVC matrices of the designs shown in Table 3. Clearly, the efficient design is able to provide more reliable parameter estimates than any orthogonal design (given that the prior parameter values are correct), which is a conclusion that holds in general.

Looking at individual parameters, in the efficient design the parameter that is the most difficult to estimate (having the highest theoretical minimum sample size to be statistically significant in estimation) is β_3^{cart}, needing a minimum sample size of

$T^* = 5.9$ (that is, $-1.3/\sqrt{2.61/5.9} \approx -1.96$, see also Table 4). In the most efficient orthogonal design parameter β_2^{bus} is most difficult to estimate, requiring a minimum sample size of $T^* = 13.1$ (as $-1.5/\sqrt{7.69/13.1} \approx -1.96$), and in the least efficient orthogonal design this is parameter β_2^{car} with a minimum sample size of $T^* = 35.5$ (as $-0.9/\sqrt{7.49/35.5} \approx -1.96$). Again, the efficient design is (much) more efficient than the orthogonal designs, this time in terms of sample size requirements for individual parameter estimates.

5.3 Effect of prior parameter misspecification

Efficient designs are constructed under the assumption that the prior parameter values are the true parameter values, which implies that any misspecification of these priors can lead to losses in the efficiency of the design. For each design, the effect of misspecifying these priors can be tested by computing the D-errors for that design for different values of the priors. As an example, we analyze the effect of misspecifying the scale parameter of the public transport branch, λ^{pt} (which was assumed to be 0.6), on the efficiency of the efficient design. In Figure 2 the solid line indicates the D-error for the efficient design for different values of λ^{pt}, ranging from 0.4 to 0.8. Clearly, the efficient design has the lowest D-error in case $\lambda^{\text{pt}} = 0.6$, as the efficient design was optimized for this parameter value. However, any deviation from this value leads to higher D-error (if $\lambda^{\text{pt}} = 0.4$, the D-error is four times higher). Hence, the design is rather sensitive to misspecification of this scale parameter.

In order to create a more robust design that is less sensitive to prior parameter misspecification, one could consider a Bayesian efficient design, in which the expected D-error is minimized over a probability distribution of prior parameter values instead of having the minimum D-error for only a single prior parameter value. Table 4 presents a Bayesian efficient design in which the scale parameter λ^{pt} is assumed to be uniformly distribution on the range [0.4, 0.8], which has also been generated using Ngene 0.7. Generating Bayesian efficient designs is much more computationally intensive than generating efficient designs given fixed priors, as the expected D-error can only be computed through simulated integration (see Bliemer *et al.*, 2006), using similar techniques as in mixed logit models (see e.g., Train, 2003). The D-error for this Bayesian efficient design is plotted in Figure 2 (see dashed line) for different values of scale parameter λ^{pt}. Compared to the efficient design, the D-error for this Bayesian efficient design is slightly higher in the area around 0.6, but outside this area the D-error does not increase as rapidly as the D-error for the efficient design does (e.g., if $\lambda^{\text{pt}} = 0.4$, the D-error for the Bayesian design is much smaller). This indicates that the Bayesian efficient design is indeed more robust to prior parameter misspecification.

Table 4: Bayesian efficient design for NL model (Bayesian D-error = 0.1908)

s	TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{cart}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{bus}$ (A$)	TT_s^{train} (min.)	$FARE_s^{train}$ (A$)
1	10	1	4	20	4	50	3	40	3
2	10	3	4	20	2	60	2	50	3
3	20	1	2	20	4	60	2	50	3
4	20	3	2	30	3	40	3	30	4
5	10	3	2	40	3	40	2	30	2
6	10	3	2	20	2	50	2	40	3
7	20	2	3	40	3	40	1	30	4
8	30	2	3	40	2	50	3	50	2
9	30	1	3	30	2	50	3	40	4
10	20	1	4	30	4	60	1	50	2
11	30	2	3	30	4	60	1	40	4
12	30	2	4	40	3	40	1	30	2

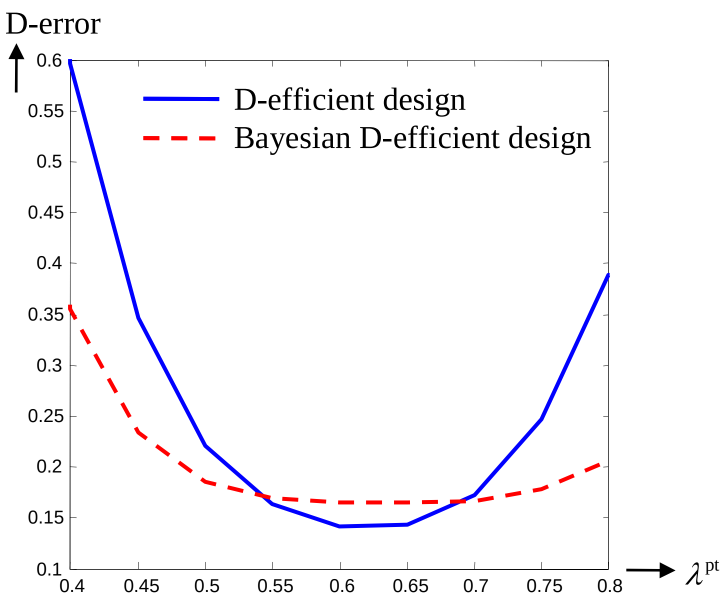

Figure 2: Effect of misspecifying the scale parameter for public transport

5.4 Effect of model type misspecification

As this paper deals with differences in the error variances across subsets of alternatives, we would like to investigate the implications of creating a design that ignores these differences (as in the MNL model) while the true model has these differences (as in the NL model), and the other way around. For this purpose, the efficient design in Table 1, generated based on the NL model in Section 5.1, is used to compute the D-error assuming an MNL model specification. Furthermore, we will generate an efficient design for the corresponding MNL model, and use this design to compute the D-error assuming the NL model specification.

The MNL prior parameter values are the same as in the NL model in Section 5.1, except that the parameters in each of the branches are multiplied by the scale parameter of that branch in order to keep the comparison as consistent as possible. An efficient design for the MNL model is presented in Table 5. The D-error of this MNL efficient design is 0.0717, however, the D-error of this MNL design assuming NL is 0.1697. Recall that the NL efficient design has a D-error of 0.1421, which means that the MNL design is not efficient if the actual model is an NL model rather than an MNL model. Reversely, the D-error of the NL efficient design (see Table 1) assuming MNL is 0.0756, which is slightly higher than the D-error for the MNL efficient design. Hence, in this example the NL design is efficient for the NL model and looses only a little of bit efficiency when applied to an MNL model, but the MNL design looses almost 20 percent efficiency when applied to an NL model. These results are summarized in Table 6 (under model M1), and demonstrate the importance of designing an experiment specifically for the model to be estimated.

Table 5: Efficient design for MNL model (D-error = 0.0717)

s	TT_s^{cart} (min.)	RC_s^{cart} (A$)	$TOLL_s^{cart}$ (A$)	TT_s^{carnt} (min.)	RC_s^{carnt} (A$)	TT_s^{bus} (min.)	$FARE_s^{bus}$ (A$)	TT_s^{train} (min.)	$FARE_s^{train}$ (A$)
1	30	2	3	40	3	40	1	30	3
2	20	2	4	30	2	60	1	50	2
3	10	3	4	40	3	40	3	40	3
4	20	1	2	20	4	50	3	50	2
5	30	1	2	30	2	60	3	50	4
6	10	1	4	20	4	50	2	30	4
7	10	3	2	40	4	40	1	40	3
8	10	3	2	40	3	50	2	30	3
9	20	1	4	30	3	60	1	50	2
10	30	2	3	20	2	50	2	30	4
11	30	2	3	20	4	40	3	40	2
12	20	3	3	30	2	60	2	40	4

Table 6: D-errors of efficient designs under different model (mis)specifications

			Assumed model for estimation	
			MNL	NL
M1	Assumed model for design	MNL	0.0717	0.1697 (+19%)
		NL	0.0756 (+5%)	0.1421
M2	Assumed model for design	MNL	0.0543	0.2702 (+32%)
		NL	0.0622 (+15%)	0.2045
M3	Assumed model for design	MNL	0.1476	0.3109 (+5%)
		NL	0.1550 (+5%)	0.2966

The effect of model misspecification is very case specific. In Table 6 two other models have been used to compare MNL and NL models and designs.[8] In model M2 two branches are considered with two alternatives each. Each alternative has two attributes and within each branch all parameters are generic, leading to four parameters in total (excluding two scale parameters). Efficient designs are generated with 12 choice situations. In Model M3 there are three branches of which the first branch is a degenerate branch with a single alternative, and the other two branches have two

[8] In order to reduce the length of the paper, the complete model specification and the generated designs will not be presented in this paper, however, they can be found on http://www.itls.usyd.edu.au/about_itls/ staff/johnr.asp.

alternatives each. Again, each alternative has two attributes and now all parameters in the model are alternative-specific, leading to 12 parameters in total (excluding three scale parameters, of which the first one is fixed to one as it is a degenerate branch). Again, efficient designs are generated, this time with 16 choice situations. As can be observed from Table 6, M2 is rather sensitive to misspecification of the model. If a design assuming an MNL model is generated while the actual model is an NL model, then more than 30 percent of efficiency is lost over generating a design for the NL model. M3 is much less sensitive to misspecification, where only five percent efficiency is lost by choosing the wrong model to generate the design for. It is clear that model misspecification can lead to (smaller or larger) efficiency losses.

5.5 Effect of nesting misspecification

Similar to misspecification of priors and model type, any misspecification in the nesting may lead to loss in efficiency of the design. To illustrate, we use a case study based on a data set collected using a SC experiment for studying mode choice in Australia. This data set and the SC experiment is described in detail in Hensher *et al*. (2005). The alternatives consist of car on toll road, car on non-toll road, bus, train, busway, and light rail. Attributes for the car include fuel (with levels A$1, A$2, and A$3 for car on tolled road, and A$3, A4, and A$5 for car on non-tolled road), toll (A$1, A$1.5, and A$2), and travel time (10, 12, and 15 minutes for car on tolled road and 15, 20, and 25 minutes for car on non-tolled road), while for the public transport models they include fare (A$1, A$3, and A$5), travel time (10, 15, and 20 minutes), and frequency (every 5, 15, or 25 minutes). The underlying experimental design used in the study is an orthogonal design with 81 choice situations, blocked in sets of three.

Different nesting structures have been used to estimate different NL models. The parameter estimates for each NL model are presented in Table 7 (MNL results have been added as well), where each nest is indicated by a different shade of gray, and the corresponding estimated scale parameters are indicated at the bottom of the table. In order to allow for different nesting structures, each parameter has been treated as alternative-specific. According to the adjusted rho-squared, NL model 5 seems to fit the data best. This conforms with the earlier observation that the most appropriate nesting structure is not always obvious and need not follow an underlying decision tree. In particular, NL model 5 does not put the tolled and non-tolled car alternatives in the same branch, as some may expect.

Table 7: Parameter estimates for different nesting structures in Sydney mode choice study

	Estimated models					
parameters	MNL	NL1	NL2	NL3	NL4	NL5
CT_CON	-1.043	-1.340	-1.042	-1.335	-1.325	--
CT_FUEL	-0.103	-0.107	-0.095	-0.100	-0.102	-0.138
CT_TOLL	-0.176	-0.227	-0.268	-0.234	-0.236	-0.201
CT_TIME	-0.014	-0.013	-0.009	-0.011	-0.012	-0.034
CN_CON	-0.067	--	--	--	--	0.413
CN_FUEL	-0.101	-0.140	-0.107	-0.149	-0.146	-0.105
CN_TIME	-0.035	-0.049	-0.045	-0.048	-0.048	-0.037
BU_CON	-0.102	-0.101	0.153	-0.170	-0.119	0.507
BU_FARE	-0.169	-0.249	-0.315	-0.250	-0.221	-0.187
BU_TIME	-0.009	-0.023	-0.029	-0.010	-0.014	-0.014
BU_FREQ	-0.044	-0.062	-0.063	-0.050	-0.052	-0.048
TR_CON	0.706	--	0.662	0.906	0.732	1.140
TR_FARE	-0.208	-0.266	-0.358	-0.277	-0.275	-0.206
TR_TIME	-0.080	-0.071	-0.093	-0.100	-0.093	-0.081
TR_FREQ	-0.023	-0.023	-0.033	-0.029	-0.027	-0.022
BW_CON	0.403	0.396	-0.002	--	--	0.862
BW_FARE	-0.264	-0.311	-0.352	-0.287	-0.318	-0.279
BW_TIME	-0.056	-0.065	-0.057	-0.047	-0.052	-0.054
BW_FREQ	-0.007	-0.010	-0.017	-0.004	-0.004	-0.008
LR_CON	--	--	--	--	--	--
LR_FARE	-0.252	-0.300	-0.384	-0.325	-0.251	-0.242
LR_TIME	-0.037	-0.045	-0.057	-0.050	-0.036	-0.016
LR_FREQ	-0.006	-0.012	-0.020	-0.008	-0.006	-0.002
SCALE1	1.000 *	0.638	0.365	0.617	0.600	1.000 *
SCALE2	--	0.623	0.259	0.815	0.707	0.676
SCALE3	--	0.773	--	0.749	1.000 *	--
Adj. ρ^2	0.246	0.246	0.215	0.246	0.266	0.286

* Fixed, not estimated.

Suppose that we would now like to create an efficient design for a similar mode choice study in which NL model 5 is taken as the appropriate model to be estimated. Then the parameter estimates of NL model 5 can be used as prior parameter values and an efficient design can be determined as outlined in this paper. Now suppose that the efficient design was not based on NL model 5, but on one of the other models with a different nesting structure. How much efficiency would be lost when such a design would be used to estimate NL model 5?

To answer this question, we have generated an efficient design for each of the six models (one MNL model and five NL models), and used each of these efficient designs to compute the asymptotic standard errors assuming that the parameters in NL model 5 represent the 'true' parameters. The results are presented in Figure 3, in which the increases in standard errors, compared to using an efficient design for NL model 5, are shown in percentages. The efficient design for NL model 5 is represented as the dashed zero percent line. Each line corresponds to a different design being used (i.e., being efficient for different nesting structures). The travel time parameter for the tolled car alternative and the scale parameter for branch 2 tend to produce significantly higher standard errors (an increase of 23 percent to 35 percent and 25 percent to 43 percent respectively). Hence, if the model to be estimated is NL model 5, then using an efficient design based on another nesting will likely result in a significant loss of efficiency. In this case, the scale parameter in particular will become harder to estimate at a given sample size. Ignoring the efficient design for NL model 5, the efficient design for the MNL model looses the least efficiency when estimating NL model 5, and for some parameters the asymptotic standard errors even marginally improve. However, this design is still much less efficient than the design optimized for NL model 5.

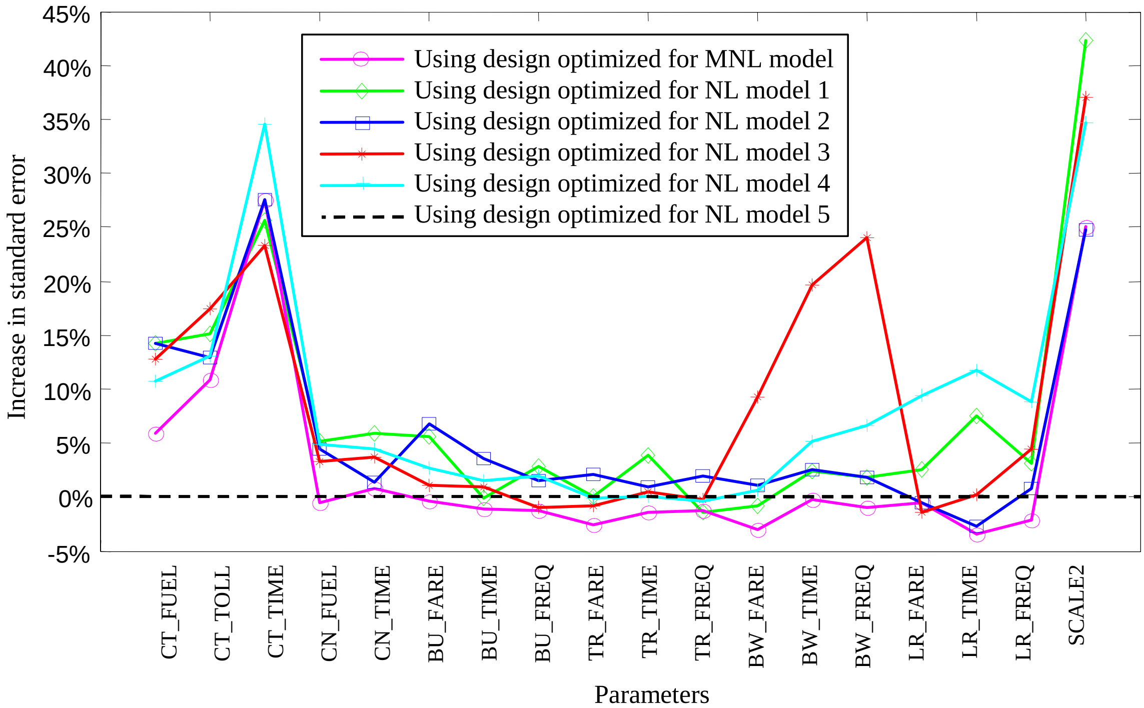

Figure 3: Increases in standard errors for estimating model NL5 when using designs optimized for other models

In addition, we have also analyzed the asymptotic standard errors in case the efficient design for NL model 5 was used to estimate the other models, and compared the results to the asymptotic standard errors corresponding to the optimized designs. The design optimized for NL model 5 becomes much less efficient (e.g., the standard error for the travel time parameter for the tolled car alternative increases 45 to 75 percent) when it is used to estimate a model with a different nesting structure. An interesting direction for further research would be to try to optimize a design over a range of nesting structures (or even model types), which has some similarities with Bayesian efficient designs that optimize a design over a range of prior values.

6. Discussion and conclusions

Violations of the IID assumption underlying the MNL model have led many to more advanced econometric models. In particular, the NL model has proven popular amongst transport researchers. The NL model is particularly appealing as it provides a closed form solution (unlike probit or random parameter logit models for example), whilst also allowing for a significant relaxation of the IID assumption. It is therefore likely that the NL model will remain a popular choice of model well into the future.

Commonly associated with the use of the NL model, is SC data. Whilst many advances have been made in the econometric modeling of such data, the generation of SC experimental designs appears to have largely lagged behind. Even to this day, the most common type of design used in transportation research appears to be orthogonal designs which have been heavily promoted for several decades. The most recent advances in the design of SC experiments, efficient SC experiments, appear to be limited largely to the generation of designs assuming the estimation of an MNL model.

Efficient SC designs have largely been limited to unlabeled experimental applications. This has had particular implications for designing efficient SC experiments for NL models. As Bliemer and Rose (2005) and Rose and Bliemer (2005) showed, however, by taking the second derivatives of the log-likelihood function, it is possible to correctly derive the AVC matrix for the MNL model allowing for alternative-specific parameter estimates.

In this paper, we extend the work of Bliemer and Rose by deriving the AVC matrix for the NL model and show that it is possible to generate efficient SC designs allowing for differences in the error components across subsets of alternatives. Through use of a numerical example, we demonstrate that orthogonal designs may not be very efficient, hence generating an efficient design given a certain model specification is clearly preferred. Furthermore, failure to allow for violations of the IID assumption in generating efficient SC experiments may result in the generation of sub-optimal designs. In particular, we show that assuming an MNL model form when in fact an NL model form is correct, may lead to losses in efficiency in the design, leading to larger standard errors for the parameter estimates and larger sample size requirements for a given design. Results from the three model exercises may suggest that one is perhaps be better off by determining a design for the NL model instead of the MNL model, as the losses in efficiency are smaller when it comes to model misspecification. However, this result may not hold in general, and needs to be investigated in more depth. We further show that not only model misspecification leads to efficiency losses, also misspecifications of priors can have large impacts. By generating a Bayesian efficient design, these impacts are largely reduced, as shown in the case study.

In this paper, we also show for that orthogonality as a design property may also result in the generation of sub-optimal designs, both in terms of the overall efficiency of the design, as well as in terms of theoretical sample size requirements. Our findings here continue a trend within the literature which suggests a move away from orthogonal designs for SC studies, towards designs which relate to the econometric models being fitted to such data.

An area of further research would be to test different assumptions on the model specification (model type and model parameters) and create different SC experiments for a real SC study and compare the estimation outcomes. Furthermore, generating designs for a range of model types and nesting structures seems another interesting direction of research, which would increase the robustness of the design under model and nesting misspecification.

References

Anderson, D.A. and Wiley, J.B. (1992) Efficient Choice Set Designs for Estimating Cross Effect Models, *Marketing Letters*, 3 (October), 357-370.

Ben-Akiva, M. and Lerman, S.R. (1985) Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press.

Bhat, C.R. and Castelar, S. (2002) A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research B,* 36 (7) 593-616.

Bliemer, M.C.J and Rose, J.M. (2005) Efficient Designs for Alternative Specific Choice Experiments, Working Paper: ITLS-WP-05-05, Institute of Transport and Logistics Studies, University of Sydney, February.

Bliemer, M.C.J., and Rose, J.M. (2006) Designing Stated Choice Experiments: State-of-the-art, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan, August.

Bliemer, M.C.J., and Rose, J.M. (2008) Construction of Experimental Designs for Mixed Logit Models Allowing for Correlation Across Choice Observations. *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Washington DC, USA.

Bliemer, M.C.J., Rose, J.M. and Hess, S. (2006) Approximation of Bayesian Efficiency in Experimental Choice Designs, Submitted for publication to *Journal of Marketing Research*.

Brownstone, D. and Small, K.A. (2005) Valuing time and reliability: assessing the evidence from road pricing demonstrations, *Transportation Research B,* 39(4), 279-293.

Bunch, D.S., Louviere, J.J., and Anderson, D. (1994) A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models, Working Paper, Graduate School of Management, University of California, Davis.

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

Carrasco, J.A. and Ortúzar, J. de D. (2002) Review and assessment of the nested logit model, *Transport Reviews*, 22(2), 197-218.

Cherchi, E. and Ortúzar, J. de Dios (2002) Mixed RP/SP models incorporating interaction effects, *Transportation*, 29 (4) 371-395.

Cook, R.D., and Nachtsheim, C.J. (1980) A comparison of algorithms for constructing exact *D*-optimal designs. *Technometrics*, 22, 315-324.

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

Garrod, G.D., Scarpa, R. and Willis, K.G. (2002) Estimating the Benefits of Traffic Calming on Through Routes: A Choice Experiment Approach, *Journal of Transport Economics and Policy*, 36(2), 211-232.

Hensher, D.A. (2001a) The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications, *Transportation*, 28(2), 101-118.

Hensher, D.A. (2001b) Measurement of the Valuation of Travel Time Savings, *Journal of Transport Economics and Policy*, 35(1), 71-98.

Hensher, D.A. and Greene, W.H. (2002) Specification and estimation of the nested logit model: alternative normalisations, *Transportation Research B*, 36, 1-17.

Hensher, D.A. and King, J. (2001) Parking demand and responsiveness to supply, pricing and location in the Sydney central business district, *Transport Research A*, 35(3), 177-196.

Hensher, D.A., Rose, J.M. and Greene, W.H. (2005) *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge.

Hess, S., Daly, A., Rohr, C. and Hyman, G. (2006), On the development of time period and mode choice models for use in large scale modeling forecasting systems, *submitted to Transportation Research A*, January 2006.

Hess, S. and Polak, J.W. (2006a), Airport, airline and access mode choice in the San Francisco Bay area, *Papers in Regional Science*, 85(4), pp. 543-567.

Hess, S. and Polak, J.W. (2006b), Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area, *Transportation Research E*, 42, pp. 63-81.

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

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

Kessels, R., P. Goos, and M. Vandebroek (2006) A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research*, 43(3), 409-419.

Koppelman, F.S. and Wen C.H. (1998a) Alternative nested logit models: structure, properties and estimation, *Transportation Research Part B*, 32(5), 289-298.

Koppelman, F.S. and Wen C.H. (1998b) Different nested logit models: which are you using? *Transportation Research Record*, 1645, 1-7.

Kuhfeld, W.F., Tobias, R.D., and Garratt, M. (1994) Efficient Experimental Design with Marketing Research Applications, *Journal of Marketing Research*, 21 (November), 545-557.

Lazari, A.G. and Anderson, D.A. (1994) Designs of Discrete Choice Experiments for Estimating Both Attribute and Availability Cross Effects, *Journal of Marketing Research*, 31 (3), 375-383.

Louviere, J.J. and Hensher, D.A. and Swait, J.D. (2000) Stated Choice Methods: Analysis and Application, Cambridge University Press, Cambridge.

Louviere, J.J. and Woodworth, G. (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, *Frontiers of Econometrics,* Zarembka, P. (ed.), Academic Press, New York, 105-142.

Ortúzar, J. de Dios, Iacobelli, A. and Valeze, C. (2000) Estimating demand for a cycle-way network, *Transport Research A*, 34(5), 353-373.

Polydoropoulou, A. and Ben-Akiva, M. (2001) Combined Revealed and Stated Preference Nested Logit Access and Mode Choice Model for Mass Transit technologies. *Transportation Research Record*, 1771, 38-45.

Rose, J.M. and Bliemer, M.C.J (2005a) Constructing Efficient Choice Experiments with Alternative Specific and Generic Parameter Estimates, Working Paper: ITLS-WP-05-07, Institute of Transport and Logistics Studies, University of Sydney, April.

Rose, J.M. and Bliemer, M.C.J. (2006) Designing Efficient Data for Stated Choice: Accounting for Socio-demographic and contextual effects in designing stated choice experiments, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan, August.

Rose, J.M. and Bliemer, M.C.J. (2007) Stated Preference Experimental Design Strategies. In Hensher, D.A. and K. Button, (eds.) *Transport Modelling*, Second Edition, Handbooks in Transport, Vol. 1, Elsevier Science, Oxford, Chapter 8

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

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

Sándor, Z. and Wedel, M. (2005) Heterogeneous Conjoint Choice Designs, *Journal of Marketing Research*, 42, 210-218.

Shiflan, Y. and Bard-Eden, R. (2001) Modeling Response to Parking Policy, *Transport Research Record*, **1765**, 27-34.

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

Van der Waerden, P., Timmermans, H. and Borgers, A. (2002) PAMELA: Parking Analysis Model for Predicting Effects in Local Areas. *Transport Research Record*, 1781, 10-18.

Yao, E. and Takayuki, M. (2005) A study of an integrated intercity travel demand model, *Transportation Research B*, 39(4), 367–381.

Appendix A: Algorithms for constructing efficient experimental designs

In order to locate the most efficient design, one could first determine the full factorial design and next evaluate each different combination of *S* choice situations taken from this full factorial. The combination with the lowest efficiency error is the optimal design. However, this procedure is often not practical as a result of the extremely high number of possible design combinations that may need to be evaluated. For the example case study problem, the full factorial has a total of $3^9 = 19,683$ choice situations. Selecting a subset of 12 choice situations from the full factorial yields 3.37×10^{51} possible different designs, all of which require evaluation. In practical terms, it may be impossible to evaluate all these designs, hence analysts often turn to smart algorithms to search a subset of the total possible number of designs to locate as efficient a design as possible.

Generally, the algorithms used are described as either row based or column based algorithms. In a *row based algorithm* choice situations are selected from a predefined canditure set of choice situations (either a full factorial or a fractional factorial) in each iteration. *Column based algorithms* create a design by selecting attribute levels over all choice situations for each attribute. Row based algorithms can easily remove bad choice situations from the canditure set at the beginning (e.g., by removing designs which don't match a particular analyst defined constraint), however in such designs it is often difficult to maintain attribute level balance (i.e., where each attribute level appears an equal number of times within each column). Column based algorithms generally have no difficulty in maintaining attribute level balance, and in general offer more flexibility and can deal with larger designs. Nevertheless, when the analyst wishes to impose some form of constraint upon the design, row based algorithms may be more suitable.

Historically, a row based algorithm known as the *Modified Federov* algorithm (Cook and Nachtsheim, 1980) was used. In this algorithm, first a candidature set is determined (either the full factorial (for small problems), or a fractional factorial (for larger problems)) and then, a design is created by selecting choice situations from the candidature set at random. Once located, the efficiency error (e.g., D-error) of the design is computed. A new design is generated from the canditure set, and the efficiency measure computed for this new design. Designs with better efficiency measures are stored, and the process repeated a number of times until all possible designs are searched or the process is terminated by either the analyst or some form of stopping criteria.

RSC (Relabeling, Swapping & Cycling) algorithms (Huber and Zwerina, 1996; Sándor and Wedel, 2001) are also used within the literature and are both row and column based algorithms. In each iteration, different columns for each attribute are created, which together form a design. This design is evaluated and if it has a lower efficiency error than the current best design, then it is stored. The columns are not created randomly, but – as the name suggests – are generated in a structured way using relabeling, swapping, and cycling techniques. With relabeling, two or more attribute levels within a column are exchanged with one another. For example, if the attribute levels 1 and 3 are relabeled, then a column containing the levels (1,2,1,3,2,3) will become (3,2,3,1,2,1). Swapping involves two or more attribute levels switching place. For example, if the attribute levels in the first and fourth choice situation are swapped, then (1,2,1,3,2,3) would become (3,2,1,1,2,3). Cycling is row based, replacing all attribute levels in each choice situation at the same time by replacing the first attribute level with the second level, the second level with the third, etc. Typically, the algorithm is set up to first relabel for a number of iterations, before moving onto swapping and finally cycling. In some cases, only subsets of the RSC algorithm are used (e.g., RS or RC).

If an efficient *orthogonal* design is required, as generated in the case study, it is possible to construct a single orthogonal design, and from this design relabel the attribute levels (as described above in the RSC algorithm) within each column to create a number of different orthogonal designs based on the initial design. Even though all these designs will be orthogonal, their efficiencies (e.g., according to the D-error) may be different. The orthogonal design with the lowest efficiency we can find we call an efficient orthogonal design.

If multiple initial orthogonal designs can be generated (i.e., different fractional factorials taken from the full factorial), relabeling of each new initial design can also take place. Note that finding an orthogonal design is not always an easy task, and one may not be able to find an orthogonal design for given design dimensions.

Rose and Bliemer (2007) provide a more detailed description of the construction of different types of stated choice experiments.

Appendix B: Deriving the AVC matrix

The determination of the statistical efficiency of a SC design requires the calculation of the expected AVC matrix of that design using equation (8), which requires taking the second derivative of the log-likelihood function in equation (9) that can be written as (substituting equations (2) and (6)):

$$L_T(\beta,\lambda)=\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{m=1}^{M}y_{mts}\sum_{j\in J_m}y_{jts|m}\log\left[\frac{\left(\sum_{i\in J_m}\exp\left(\bar{\mu}_m\sum_{k=1}^{K}\beta_k x_{iks|m}\right)\right)^{\lambda_m/\bar{\mu}_m-1}}{\sum_{n=1}^{M}\left(\sum_{i\in J_n}\exp\left(\bar{\mu}_n\sum_{k=1}^{K}\beta_k x_{iks|n}\right)\right)^{\lambda_n/\bar{\mu}_n}}\cdot\exp\left(\bar{\mu}_m\sum_{k=1}^{K}\beta_k x_{jks|m}\right)\right]. \tag{A1}$$

Furthermore, we made the normalization assumption that $\bar{\mu}_m=1$ for all $j\in J_m$, for all *m*. Given all respondents, *t*, make decisions from amongst each of the alternatives, *j*, it holds that $\sum_{j\in J_m}y_{jts|m}=1$ and $\sum_m y_{mts}=1$, such that equation (A1) can be rewritten as

$$L_T(\beta,\lambda)=\sum_{t=1}^{T}\sum_{s=1}^{S}\left[\sum_{m=1}^{M}y_{mts}\left(\sum_{j\in J_m}y_{jts|m}\sum_{k=1}^{K}\beta_k x_{jks|m}+(\lambda_m-1)\log\sum_{i\in J_m}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m}\right)\right)-\log\sum_{n=1}^{M}\left(\sum_{i\in J_n}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|n}\right)\right)^{\lambda_n}\right] \tag{A2}$$

Taking all second derivatives for all (and each combination of) parameters yields the following series of equations:

$$\begin{aligned}\frac{\partial^2 L_T(\beta,\lambda)}{\partial\beta_{k_1}\partial\beta_{k_2}}=\sum_{t=1}^{T}\sum_{s=1}^{S}\Bigg\{&\sum_{m=1}^{M}y_{mts}\left[(\lambda_m-1)\left(\sum_{i\in J_{mk_1k_2}}x_{ik_1s|m}P_{is|m}x_{ik_2s|m}-\sum_{i\in J_{mk_1}}x_{ik_1s|m}P_{is|m}\sum_{j\in J_{mk_2}}P_{js|m}x_{jk_2s|m}\right)\right]\\&-\left(\sum_{m=1}^{M}\lambda_m P_{ms}\left[\left(\lambda_m\sum_{i\in J_{mk_2}}P_{is|m}x_{ik_2s|m}-\sum_{n=1}^{M}\lambda_n P_{ns}\sum_{i\in J_{nk_2}}P_{is|n}x_{ik_2s|n}\right)\sum_{i\in J_{mk_1}}x_{ik_1s|m}P_{is|m}\right.\right.\\&\left.\left.+\sum_{i\in J_{mk_1k_2}}x_{ik_1s|m}P_{is|m}x_{ik_2s|m}-\sum_{i\in J_{mk_1}}x_{ik_1s|m}P_{is|m}\sum_{j\in J_{mk_2}}P_{js|m}x_{jk_2s|m}\right]\right)\Bigg\}\end{aligned} \tag{A3a}$$

$$\frac{\partial^2 L_T(\beta,\lambda)}{\partial\lambda_{m_1}\partial\beta_{k_2}} = \sum_{t=1}^{T}\sum_{s=1}^{S}\Bigg[\left(y_{m_1ts} - P_{m_1s}\right)\sum_{i\in J_{m_1}} P_{is|m_1} x_{iks|m_1}$$

$$- \log\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)P_{m_1s}\left(\lambda_{m_1}\sum_{i\in J_{m_1k_2}} P_{is|m_1} x_{ik_2s|m_1} - \sum_{n=1}^{M}\lambda_n P_{ns}\sum_{i\in J_{nk_2}} P_{is|n} x_{ik_2s|n}\right)\Bigg]$$

(A3b)

$$\frac{\partial^2 L_T(\beta,\lambda)}{\partial\lambda_{m_1}\partial\lambda_{m_2}} = \begin{cases} -\sum_{t=1}^{T}\sum_{s=1}^{S} P_{m_1s}\left(1-P_{m_1s}\right)\left(\log\left[\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)\right]\right)^2, & \text{if } m_1 = m_2 \\ \sum_{t=1}^{T}\sum_{s=1}^{S} P_{m_1s}P_{m_2s}\log\left[\sum_{i\in J_{m_1}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_1}\right)\right]\log\left[\sum_{i\in J_{m_2}}\exp\left(\sum_{k=1}^{K}\beta_k x_{iks|m_2}\right)\right], & \text{if } m_1 \neq m_2 \end{cases}$$

(A3c)

In these equations, considering all elemental alternatives in branch m, J_{mk} denotes the subset of elemental alternatives in which attribute k appears, and $J_{mk_1k_2}$ denotes the subset of elemental alternatives in which both attributes k_1 and k_2 appear. Unlike the second derivatives in the case of the MNL model, equations (A3a) and (A3b) depend on the model outcomes y_{mts}, that is, which branch was chosen in each choice situation (see Bliemer and Rose, 2005 and Rose and Bliemer, 2005). As we are interested in the Fisher information matrix when $T \to \infty$, we observe that $E(\sum_{t=1}^{T} y_{mts}) = TP_{ms}$, such that we will use P_{ms} to represent y_{mts} in the Fisher information matrix. The validity of this reformulation is illustrated later in this appendix.

Following from this substitution of P_{ms} for y_{mts}, the sub-index t is no longer present within the equations, such that the summation over the respondents is simply the multiplicand of the value by *T*. The Fisher information matrix, now independent of *y*, thus simplifies to the equations (10a,b,c).

In order to check that the analytical computation of the Fisher information matrix (independent of the observed choices, *y*) using equations (10a,b,c) is correct, we compare the outcomes using a simulated approach by creating simulated choices for a (large) sample of respondents and then estimate the model parameters on this data (see e.g., Ferrini and Scarpa, *in press*; Kessels *et al.*, 2006) and take the inverse of the variance-covariance matrix.

We have simulated the observed choices for a sample of 10,000 respondents given the efficient SC design shown in Table 2. Based on this simulated sample, we estimate a nested logit model using Nlogit 4.0 and obtain the Fisher information matrix and estimated parameter estimates from the simulated sample.[9] The Fisher information matrix from this exercise is given in Table A (normalized to *T*=1). Using the parameters obtained in estimation, we next calculate the Fisher information matrix for the design using equations (10a,b,c) which we present also in Table A (assuming *T*=1). Note that

[9] The simulated data and Nlogit syntax can be found at http://www.itls.usyd.edu.au/about_itls /staff/johnr.asp.

the two are virtually equivalent. It is interesting to note that similar results are already obtained for sample sizes as low as 100 respondents.

Table A: Simulated and Analytically derived Fisher Information Matrices

Simulated Fisher information matrix (from Nlogit 4.0)

	β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
β_0^{cart}	1.66	14.82	1.74	5.17	-0.30	-15.88	-0.48	-14.24	-1.36	-10.84	17.50
β_1^{car}	14.82	1035.71	102.78	47.82	-12.66	-707.24	-24.25	-556.18	-40.85	-447.30	707.83
β_2^{car}	1.74	102.78	16.77	5.83	-1.14	-62.38	-2.50	-70.79	-6.13	-48.85	78.45
β_3^{cart}	5.17	47.82	5.83	17.75	-1.08	-58.84	-1.65	-42.74	-4.12	-36.07	57.91
β_0^{bus}	-0.30	-12.66	-1.14	-1.08	1.65	82.11	3.44	-54.98	-3.61	5.89	-9.17
β_1^{bus}	-15.88	-707.24	-62.38	-58.84	82.11	4233.35	168.06	-2765.29	-176.17	326.91	-510.24
β_2^{bus}	-0.48	-24.25	-2.50	-1.65	3.44	168.06	8.79	-119.44	-7.95	11.12	-17.16
β_1^{train}	-14.24	-556.18	-70.79	-42.74	-54.98	-2765.29	-119.44	3082.59	201.01	264.95	-433.25
β_2^{train}	-1.36	-40.85	-6.13	-4.12	-3.61	-176.17	-7.95	201.01	15.62	20.87	-34.52
λ^{car}	-10.84	-447.30	-48.85	-36.07	5.89	326.91	11.12	264.95	20.87	210.80	-334.61
λ^{pt}	17.50	707.83	78.45	57.91	-9.17	-510.24	-17.16	-433.25	-34.52	-334.61	535.40

Analytically derived Fisher information matrix (from equations (10a,b,c))

	β_0^{cart}	β_1^{car}	β_2^{car}	β_3^{cart}	β_0^{bus}	β_1^{bus}	β_2^{bus}	β_1^{train}	β_2^{train}	λ^{car}	λ^{pt}
β_0^{cart}	1.66	14.83	1.74	5.17	-0.30	-15.94	-0.48	-14.21	-1.36	-10.84	17.51
β_1^{car}	14.83	1035.88	102.77	47.86	-12.69	-709.28	-24.33	-554.67	-40.76	-447.38	707.98
β_2^{car}	1.74	102.77	16.77	5.84	-1.14	-62.55	-2.51	-70.66	-6.12	-48.88	78.46
β_3^{cart}	5.17	47.86	5.84	17.75	-1.08	-59.03	-1.66	-42.62	-4.12	-36.08	57.94
β_0^{bus}	-0.30	-12.69	-1.14	-1.08	1.65	82.36	3.45	-55.15	-3.62	5.91	-9.20
β_1^{bus}	-15.94	-709.28	-62.55	-59.03	82.36	4245.61	168.58	-2773.45	-176.66	327.87	-511.53
β_2^{bus}	-0.48	-24.33	-2.51	-1.66	3.45	168.58	8.81	-119.80	-7.97	11.16	-17.23
β_1^{train}	-14.21	-554.67	-70.66	-42.62	-55.15	-2773.45	-119.80	3087.93	201.33	264.26	-432.28
β_2^{train}	-1.36	-40.76	-6.12	-4.12	-3.62	-176.66	-7.97	201.33	15.64	20.83	-34.46
λ^{car}	-10.84	-447.38	-48.88	-36.08	5.91	327.87	11.16	264.26	20.83	210.85	-334.69
λ^{pt}	17.51	707.98	78.46	57.94	-9.20	-511.53	-17.23	-432.28	-34.46	-334.69	535.52