I T L S

WORKING PAPER
ITLS-WP-05-07

Constructing Efficient Choice Experiments

By

*John M Rose & Michiel CJ Bliemer**

*Assistant Professor Transportation Modeling
Delft University of Technology
Faculty of Civil Engineering and Geosciences
Transport and Planning Section
Delft, The Netherlands

Adjunct Professor
Institute of Transport & Logistics Studies

* corresponding author

ISSN 1832-570X

April 2005

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-05-07

TITLE: **Constructing Efficient Choice Experiments**

ABSTRACT: Research on the construction of efficient designs for stated choice (SC) experiments has been limited to either unlabeled experiments with generic parameter estimates or labeled experiments with alternative specific parameter estimates. Designs combining both generic and alternative specific parameters have not yet been addressed. In this paper, by deriving the asymptotic (co)variance matrix for the most general MNL model, the authors are able to demonstrate how efficient experiments that allow for the estimation of both types of estimates may be generated. The authors go onto show how estimation of the asymptotic (co)variance matrix may also be used to determine sample size requirements for SC experiments.

KEY WORDS: ***Stated Choice, Efficient Experimental Designs, D-Efficiency, Alternative Specific, Generic, Sample Size***

AUTHORS: John M Rose & Michiel CJ Bliemer

CONTACT: Institute of Transport and Logistics Studies (Sydney)
The Australian Key Centre in Transport and Logistics Management, (C37)
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: April 2005

1. Introduction

Stated choice (SC) data has proven useful in solving many transportation related problems. For example, SC data has been used to examine the demand for a cycle-way networks (e.g., de Dios 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 modellers to estimate models of choice behaviour, 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 design the data 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 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 situations using randomised, cyclical, Bayesian or foldover procedures (see for example, Bunch et al. 1994; Louviere and Woodworth 1983; Huber and Zwerina 1996; Kanninen, 2002; Sandor and Wedel 2002).

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., Anderson and Wiley 1992; Laziri and Anderson 1994; Bunch et al. 1994; Huber and Zwerina 1996; Sándor and Wedel 2001; Carlsson and Martinsson 2002; Kanninen 2002). With regards to SC experiments, a number of different criteria exist which may be used to both define and measure statistical efficiency. Most commonly used is the D-optimality criterion, which seeks to simultaneously minimise all the elements of the asymptotic (co)variance matrix of models to be estimated from data collected from an experimental design. Independent of the specific criterion used to define statistical efficiency, all are related to the asymptotic (co)variance matrix of the choice model to be estimated.

In order to estimate the likely asymptotic (co)variance 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). These priors are used to calculate the expected utilities as well as choice probabilities of each of the alternatives. From these choice probabilities, it becomes a straightforward exercise to calculate the asymptotic (co)variance matrix of the model to be estimated. Through manipulation of the design, the analyst is able to minimize the elements within the asymptotic (co)variance matrix, which in the case of the diagonals means lower standard errors and hence greater reliability in the estimates at a fixed sample size.

Traditionally, researchers were limited to the examination of the statistical efficiency of unlabeled SC experiments assuming generic parameter estimates. The assumption of generic parameter estimates arose as a direct result of the way the log-likelihood function for discrete choice models have been presented in the past. The literature on generation of optimal designs for SC experiments state as their basis, the seminal work by McFadden (1974) and described in detail in Ben-Akiva and Lerman (1985) and Louviere, Hensher and Swait (2000). An examination of the original derivation of the MNL model offered by McFadden (1974) reveals that this work was limited to that of the MNL assuming generic parameter estimates. Recently, Bliemer and Rose (2005) demonstrated the presence of alternative-specific parameter estimates (or the presence/absence of different attributes across alternatives) requires a different derivation of the log-likelihood function used to obtain the asymptotic (co)variance matrix of discrete choice models, without which, attempts to minimize the elements of the asymptotic (co)variance matrix cannot be guaranteed. This research however, was incomplete; being itself limited to the specific case of models estimated solely with alternative specific parameter estimates.

In this paper, we derive the log-likelihood function for the MNL model allowing for both generic and alternative specific parameter estimates. We then use derivation to demonstrate how optimal designs for alternative specific experiments may be generated, doing so for orthogonal and non-orthogonal designs. We next evaluate these designs by comparing the resulting asymptotic variance-covariance matrices, and in doing so, demonstrate how one can directly compare these results for any sample size without the use of Monte Carlo experiments. In the last section, we discuss limitations and extensions to our proposed methodology.

2. The MNL model with generic and alternative-specific parameters

In this section, we outline the derivation of the most general case of the MNL model. The model follows the seminal work of McFadden (1974) on random utility theory (RUT) which has been summarized in a number of sources (e.g., Ben-Akiva and Lerman 1985; Louviere et al. 2000; Train 2003; Hensher et al. 2005). To demonstrate RUT, consider a situation in which an individual is faced with a number of choice tasks in each of which they must make a discrete choice from a universal but finite number of alternatives. Let subscripts s and j refer to choice situation s = 1, 2, …, S, and alternative j = 1, 2, …, J. RUT posits that the utility possessed by an individual for alternative j present in choice set s may be expressed as:

(1) $$U_{js} = V_{js} + e_{js},$$

where U_{js} is the overall utility associated with alternative j in choice situation s, V_{js} is the component of utility associated with alternative j that is observed by the analyst in choice situation s, and e_{js} represents the component of utility that is not observed by the analyst.

RUT assumes that individuals attach parameter weights to each of the attributes associated with the alternatives specified within an experiment. For a given attribute, a parameter weight may be the same for any two alternatives (i.e., generic) or different across alternatives (i.e., alternative-specific). Let there be K^* attributes which have generic parameter weights, and K_j attributes with alternative specific parameters. Assuming a linear additive utility function, the observed component of utility may be expressed as:

$$V_{js} = \sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}, \quad \forall j = 1,\ldots,J, \forall s = 1,\ldots,S. \tag{2}$$

The generic and alternative-specific parameters are denoted by b_k^* and b_{jk}, respectively, with their associated attribute levels x_{jks}^* and x_{jks} for each choice situation *s*. Under the assumption that the unobserved component of utility, e_{js}, are independently and identically extreme value type I distributed, we are able to derive the multinomial logit model in which P_{js} is the probability of choosing alternative *j* in choice situation *s*:

$$P_{js} = \frac{\exp(V_{js})}{\sum_{i=1}^{J} \exp(V_{is})}, \quad \forall j = 1,\ldots,J, \forall s = 1,\ldots,S. \tag{3}$$

3. Derivation of the asymptotic covariance matrix

Most commonly used to determine the parameters (b^*, b) in the MNL model (2)-(3) is a method known as maximum likelihood estimation. Consider a single respondent facing *S* choice situations. The log-likelihood as a function of the parameters is given by

$$\begin{aligned} L(b^*, b) &= \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \log P_{js} \\ &= \sum_{s=1}^{S} \left[\sum_{j=1}^{J} y_{js} \left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks} \right) - \log \left(\sum_{i=1}^{J} \exp \left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{jks} \right) \right) \right] \end{aligned} \tag{4}$$

where the vector *y* describes the outcomes of all choice tasks, that is, y_{js} is one if alternative *j* is chosen in choice task *s* and is zero otherwise. The asymptotic (co)variance matrix can be derived from the second derivative of the log-likelihood function. Allowing for both alternative-specific and generic parameters, this leads to the following (see Appendix A):

$$\frac{\partial^2 L(b^*, b)}{\partial b_{k_1}^* \partial b_{k_2}^*} = -\sum_{s=1}^{S} \sum_{j=1}^{J} x_{jk_1s}^* P_{js} \left(x_{jk_2s}^* - \sum_{i=1}^{J} P_{is} x_{ik_2s}^* \right), \quad \forall k_1, k_2 = 1,\ldots,K^*, \tag{5a}$$

$$\frac{\partial^2 L(b^*, b)}{\partial b_{j_1k_1} \partial b_{k_2}^*} = -\sum_{s=1}^{S} x_{j_1k_1s} P_{j_1s} \left(x_{j_1k_2s}^* - \sum_{i=1}^{J} x_{ik_2s}^* P_{is} \right), \quad \forall j_1 = 1,\ldots,J, k_1 = 1,\ldots,K_{j_1}, k_2 = 1,\ldots,K^*, \tag{5b}$$

(5c) $$\frac{\partial^2 L(b^*, b)}{\partial b_{j_1 k_1} \partial b_{j_2 k_2}} = \begin{cases} \sum_{s=1}^{S} x_{j_1 k_1 s} x_{j_2 k_2 s} P_{j_1 s} P_{j_2 s}, & \text{if } j_1 \neq j_2; \\ -\sum_{s=1}^{S} x_{j_1 k_1 s} x_{j_2 k_2 s} P_{j_1 s} \left(1 - P_{j_2 s}\right), & \text{if } j_1 = j_2. \end{cases} \qquad \forall j_i = 1, \ldots, J, k_i = 1, \ldots, K_{j_i}.$$

Note that these second derivatives do not depend on the outcomes *y*. It is also worth noting that assuming *M* respondents each complete the same *S* choice situations, then equations (5a,b,c) will be simply multiplied by *M*.

The maximum likelihood (ML) parameter estimates (both generic and alternative specific) can be found by maximizing the log-likelihood function, or alternatively, setting the first derivatives (the score vector) equal to zero (it can be shown that the log-likelihood function is concave). Call these ML estimates $(\hat{b}^*, \hat{b}),$ then

(6) $$(\hat{b}^*, \hat{b}) = \arg \max_{(b^*, b)} L(b^*, b).$$

Suppose that the true parameter values are $(\bar{b}^*, \bar{b}).$ McFadden (1974) has shown for the case with only generic parameters, the ML estimates $\hat{b}^*$ are asymptotically normally distributed with mean $\bar{b}^*$ and (co)variance matrix Ω, which is equal to the negative inverse of the Fisher information matrix. It can be shown that the same holds for the more general specification of the MNL model allowing for generic and alternative specific parameter estimates. The Fisher information matrix I is defined as the expected values of the second derivative of the log-likelihood function, hence with *M* respondents

(7) $$I(\hat{b}^*, \hat{b}) = M \cdot \frac{\partial^2 L(\bar{b}^*, \bar{b})}{\partial b \partial b'}.$$

Therefore, the asymptotic (co)variance matrix may be computed as

(8) $$\Omega = -\left[I(\hat{b}^*, \hat{b}) \right]^{-1} = -\frac{1}{M} \left[\frac{\partial^2 L(\bar{b}^*, \bar{b})}{\partial b \partial b'} \right]^{-1}.$$

This symmetric asymptotic (co)variance matrix will be of dimension corresponding to the total number of parameters, $\bar{K},$ where $\bar{K} = \sum K^* + \sum_j K_j.$ Clearly, the (co)variances become smaller with larger sample sizes, that is, with an increasing number of respondents *M*. Summarising,

(9) $$(\hat{b}^*, \hat{b}) \rightarrow N\left((\bar{b}^*, \bar{b}), -\frac{1}{M} \left[\frac{\partial^2 L(\bar{b}^*, \bar{b})}{\partial b \partial b'} \right]^{-1} \right).$$

4. Measuring Statistical Efficiency in SC Experimental Designs

A statistically efficient design is a design that minimizes the elements of the asymptotic (co)variance matrix, resulting in more reliable parameter estimates for a fixed number of choice observations. In order to be able to compare the statistical efficiency of SC experimental designs, a number of alternative approaches have been proposed within the literature (see e.g., Bunch et al. 1994). The most commonly used measure within the literature is that of D-error. The D-error of a design may be computed by taking the determinant of the asymptotic (co)variance matrix and applying a scaling factor $1/\bar{K}$ in order to take the number of parameters into account:

$$\text{(10)} \qquad \text{D-error} = \left(\det \Omega\right)^{1/\bar{K}} = -\frac{1}{M}\left(\det\left(\frac{\partial L^2(\bar{b}^*,\bar{b})}{\partial b \partial b'}\right)\right)^{1/\bar{K}},$$

where usually only one complete design for a single respondent is taken into account, that is, M = 1. The determinant will always yield a positive value as the asymptotic (co)variance matrix is positive definite given that the log-likelihood function is concave. If the D-error is low, meaning that the (co)variances of the parameter estimates are low, then the statistical efficiency is high.

A practical problem exists, however, in that rarely will the true parameters, $(\bar{b}^*,\bar{b})$, be known *a priori*. Therefore, in computing the D-error it is necessary to assume a set of prior parameter values, $(\tilde{b}^*,\tilde{b})$, which it is hoped will be close to the true underlying population parameter values. This necessity has resulted in two popular approaches for computing the D-error of experimental choice designs. The first approach assumes that there exists no information on the true parameter values (including sign). Assuming no *a priori* information exists, it is common to set the prior parameters values to zero. This leads to what the literature has termed the D_z-error measure. In contrast, if prior information is available, then these priors can be used to compute the D-error, yielding what is known as the D_p-error measure. Assuming a single respondent, the D_z-error can be computed as:

$$\text{(11)} \qquad D_z\text{-error} = \left(\det\left(\frac{\partial L^2(0,0)}{\partial b \partial b'}\right)\right)^{-1/\bar{K}},$$

while the D_p-error assuming knowledge of prior parameter estimates $(\tilde{b}^*,\tilde{b})$ may be computed as

$$\text{(12)} \qquad D_p\text{-error} = \left(\det\left(\frac{\partial L^2(\tilde{b}^*,\tilde{b})}{\partial b \partial b'}\right)\right)^{-1/\bar{K}}.$$

For designs of the same dimensions (i.e., number of choice sets, alternatives, attributes and attribute levels), the design(s) with the lowest D-error is (are) termed the D-optimal design(s). Given the large number of possible attribute level combinations for a design of fixed dimensions, it will be unlikely that for all but the smallest of designs, that the D-error measure will be calculable for all possible design permutations. Unless one can examine all design permutations keeping the design dimensions constant, it will therefore be impossible to demonstrate that a design has the lowest possible D-error, and hence, it will often be more appropriate to discuss D-efficient designs rather than D-optimal designs.

Manipulation of the attribute levels of the alternatives within a design will result in different D-error values (D_z or D_p), assuming fixed prior parameter estimates. Over a number of iterations, it may be possible to locate designs with lower D-error values. Methods of manipulating the attribute levels so as to generate and locate D-efficient designs are discussed in detail in Kuhfeld et al. (1994), Huber and Zwerina (1996), Sándor and Wedel (2001), Kanninen (2002), Carlsson and Martinsson (2002), and Burgess and Street (2005) amongst other sources.

5. Generating D-efficient Stated Choice Designs

Using Equations (5a,b,c) and (11) or (12) in order to compute the statistical efficiency of a design, it is possible to assess SC experiments of any dimension allowing for both generic and alternative specific parameter estimates. In this section, we generate a number of choice experiments involving two labeled alternatives, both with four attributes; two of which are generic. Fixing the prior parameter estimates, we construct a number of different D_p-efficient designs, including both orthogonality and non-orthogonality as a criterion in their construction. In all cases, we have assumed attribute level balance, though such an assumption is not necessary to locate either D_p-efficient orthogonal or non-orthogonal designs (indeed, it is possible that such an assumption will result in less than efficient designs). Given the predominance in the use of orthogonal designs in SC studies, we also generate the worst D_p-efficient orthogonal design. Whilst it is improbable that one would specifically set out to generate and use such a design in practice, the fact that most orthogonal are randomly generated suggests that such a design could indeed actually be used in reality. Therefore, the generation of such a design allows for an examination of the worst case scenario.

The parameter estimates and attribute levels used in the construction of the SC designs are shown in Table 1. The designs were generated with 12 choice sets each, which is the minimum amount of choice sets for a balanced orthogonal SC design with this number of attributes and attribute levels.

Table 1: Prior parameters and design attribute levels

Alternative	Attribute	Parameter	Parameter Prior	Attribute levels
A	x_{11}	G_1	0.4	2, 4, 6
A	x_{12}	G_2	0.3	1, 3, 5
A	x_{13}	b_{13}	0.3	2.5, 3, 3.5
A	x_{14}	b_{14a}	0.6	4, 6, 8
B	Constant	$ß_{20}$	-1.2	-
B	x_{21}	G_1	0.4	2, 4, 6
B	x_{22}	G_2	0.3	1, 3, 5
B	x_{23}	b_{23}	0.4	2.5, 4, 5.5
B	x_{24}	b_{14a}	0.7	4, 6, 8

Three SC designs were generated and are shown in Table 2. Design 1 represents the most D_p-efficient (balanced) orthogonal design located whilst design 2 represents the worst D_p-(in)efficient (balanced) orthogonal design. As stated previously, the reason for including the worst D_p-efficient design is because many researchers tend to consider only one orthogonal design, which may have either a high or a low D_p-efficiency (which is generally not computed). Design 3 is the most efficient non-orthogonal design that we were able to construct. All designs were generated using algorithms programmed in Matlab, which used a heuristic to generate a large number of orthogonal and non-orthogonal designs in a smart way and determine which of these was the most D_p-(in)efficient. For the non-orthogonal designs, the algorithm employed a simple swapping procedure similar to that discussed in Huber and Zwerina (1996) and Sándor and Wedel (2001).

Table 2: SC experimental designs

		Design 1: Best D_p-error Orthogonal Design					**Design 2: Worst D_p-error Orthogonal Design**					**Design 3: Best D_p-error (Non-Orthogonal Design)**				
Cset #	**Alternative**	x_{11} x_{21}	x_{12} x_{22}	x_{13} x_{23}	x_{14} x_{24}	**Prob**	x_{11} x_{21}	x_{12} x_{22}	x_{13} x_{23}	x_{14} x_{24}	**Prob**	x_{11} x_{21}	x_{12} x_{22}	x_{13} x_{23}	x_{14} x_{24}	**Prob**
1	1	4	5	3.5	8	0.56	4	1	3.5	6	0.84	4	5	3	6	0.43
	2	6	1	4	8	0.44	2	3	4	4	0.16	6	1	5.5	6	0.57
2	1	4	3	3	4	0.33	2	1	3.5	8	0.05	4	1	2.5	4	0.39
	2	6	5	2.5	4	0.67	6	5	4	8	0.95	2	5	4	4	0.61
3	1	2	5	3.5	6	0.89	2	3	3	4	0.18	2	5	3.5	4	0.32
	2	2	3	4	4	0.11	6	1	5.5	4	0.82	6	1	5.5	4	0.68
4	1	6	1	3.5	8	0.81	2	3	3	6	0.87	6	5	3	4	0.67
	2	4	5	2.5	6	0.19	2	3	2.5	4	0.13	2	1	2.5	8	0.33
5	1	4	1	2.5	6	0.19	4	3	3	6	0.21	2	5	2.5	8	0.49
	2	2	3	4	8	0.81	2	3	5.5	8	0.79	4	1	4	8	0.51
6	1	6	5	2.5	4	0.76	6	1	2.5	4	0.08	4	1	3.5	8	0.28
	2	4	1	2.5	6	0.24	4	5	5.5	6	0.92	6	3	2.5	8	0.72
7	1	6	1	3.5	4	0.28	6	5	3.5	8	0.93	2	3	3	6	0.62
	2	4	1	5.5	6	0.72	4	1	5.5	6	0.07	4	5	2.5	4	0.38
8	1	6	5	2.5	8	0.76	6	1	2.5	8	0.91	6	3	2.5	8	0.63
	2	4	5	5.5	6	0.24	4	1	2.5	6	0.09	4	5	5.5	6	0.37
9	1	2	3	3	4	0.00	4	5	2.5	8	0.82	4	3	3	6	0.15
	2	6	5	5.5	8	1.00	6	5	4	4	0.18	6	3	2.5	8	0.85
10	1	2	3	3	6	0.29	6	5	3.5	4	0.56	6	1	3.5	8	0.81
	2	2	3	2.5	8	0.71	4	5	2.5	6	0.44	4	3	4	6	0.19
11	1	4	3	3	6	0.82	4	3	3	4	0.09	6	1	3.5	4	0.32
	2	2	3	5.5	4	0.18	6	1	2.5	8	0.91	2	3	5.5	6	0.68
12	1	2	1	2.5	8	0.68	2	5	2.5	6	0.26	2	3	2.5	6	0.63
	2	6	1	4	4	0.32	2	3	4	8	0.74	2	5	4	4	0.37
		D_p = 0.31470 D_z = 0.19031					D_p = 0.45368 D_z = 0.19031					D_p = 0.24836 D_z =0.20930				

It is worthwhile noting that orthogonality represents a constraint on the statistical efficiency of SC experiments and hence it will typically be possible to construct non-orthogonal designs with lower D_p-errors. This is borne out in the designs shown in Table 2. Table 2 also shows the D_z-error (i.e., assuming the priors are all equal to zero) for the three designs. Consistent with the findings of Bliemer and Rose (2005), the D_z-errors for the orthogonal designs are lower than for the best D_p-efficient non-orthogonal design and the D_z-error of the orthogonal designs are also the same[1]. This demonstrates a danger in assuming priors equal to zero, as this will generally result in the generation of an orthogonal design, which will more than likely be less efficient than a best D_p-efficient non-orthogonal design if the parameter estimates are something other than zero in reality.

Table 3 demonstrates the asymptotic (co)variance matrix derived for each design shown in Table 2, assuming a single respondent. Despite design 1 being orthogonal, Table 3 clearly demonstrates that the resulting covariances for this design are non-zero. This result demonstrates an important property of the MNL model. Whilst the design (data) employed may be orthogonal, the estimation procedure works by taking the differences in the attribute levels of the chosen and non-chosen alternatives (see Louviere et al. 2000; Lindsey 1996). Thus, whilst the design itself may be orthogonal, the differences between the chosen and non-chosen alternatives will likely be correlated, resulting in non-zero covariances from the estimated model. This result will hold for any orthogonal design when the parameter estimates from the experiment are non-zero. The enforcement of orthogonality may represent a limiting assumption and actually result in greater covariances than would be induced from a non-orthogonal design given the (much) greater number of possible combinations of attribute levels available for non-orthogonal designs in which to locate designs with lower D_p-efficiency values. As such, non-orthogonal designs may actually produce more reliable estimates than orthogonal designs when estimating MNL models. In the example above, assuming that the specified priors are correct, a better D_p-efficiency (i.e., a lower D_p-error) is obtained from the non-orthogonal design represented in Table 2 than for the best-case orthogonal design.

[1] All orthogonal designs of the same dimensions will produce the same D_z-error, see Appendix B. However, when there are generic parameters, an orthogonal design will not necessarily give the minimum D_z-error, as is believed to be the case when only alternative-specific parameters are included in the utility functions of the model (see Bliemer and Rose, 2005).

Table 3: (co)variance matrix for each design

	Design 1						
	G_1	G_2	ß$_{13}$	ß$_{14}$	ß$_{20}$	ß$_{23}$	ß$_{24}$
G_1	**0.17**	0.04	0.05	0.09	-0.29	0.12	0.10
G_2	0.04	**0.11**	0.02	0.03	-0.32	0.06	0.07
ß$_{13}$	0.05	0.02	**2.88**	0.07	7.72	0.10	0.16
ß$_{14}$	0.09	0.03	0.07	**0.25**	0.66	0.13	0.10
ß$_{20}$	-0.29	-0.32	7.72	0.66	**39.00**	-1.41	-1.02
ß$_{23}$	0.12	0.06	0.10	0.13	-1.41	**0.47**	0.13
ß$_{24}$	0.10	0.07	0.16	0.10	-1.02	0.13	**0.28**
	Design 2						
	G_1	G_2	ß$_{13}$	ß$_{14}$	ß$_{20}$	ß$_{23}$	ß$_{24}$
G_1	**0.18**	0.05	-0.26	-0.02	-1.14	-0.05	0.07
G_2	0.05	**0.22**	0.05	0.01	0.01	-0.01	0.05
ß$_{13}$	-0.26	0.05	**5.23**	0.44	20.32	-0.17	-0.25
ß$_{14}$	-0.02	0.01	0.44	**0.33**	3.33	0.02	-0.03
ß$_{20}$	-1.14	0.01	20.32	3.33	**103.70**	-2.16	-2.49
ß$_{23}$	-0.05	-0.01	-0.17	0.02	-2.16	**0.49**	-0.03
ß$_{24}$	0.07	0.05	-0.25	-0.03	-2.49	-0.03	**0.28**
	Design 3						
	G_1	G_2	ß$_{13}$	ß$_{14}$	ß$_{20}$	ß$_{23}$	ß$_{24}$
G_1	**0.14**	0.09	0.21	0.14	0.07	0.10	0.17
G_2	0.09	**0.12**	0.15	0.13	-0.30	0.11	0.18
ß$_{13}$	0.21	0.15	**2.84**	0.40	7.47	0.20	0.43
ß$_{14}$	0.14	0.13	0.40	**0.35**	1.00	0.13	0.30
ß$_{20}$	0.07	-0.30	7.47	1.00	**40.09**	-1.64	-0.83
ß$_{23}$	0.10	0.11	0.20	0.13	-1.64	**0.39**	0.24
ß$_{24}$	0.17	0.18	0.43	0.30	-0.83	0.24	**0.49**

Minimization of a single global measure (i.e., either D_p-error or D_z-error) representing all elements contained within the asymptotic (co)variance matrix explains why in this case, no single design performs best in terms of producing the lowest standard errors for all attributes considered. The D-error criterion will minimize the (co)variances of all attributes concurrently resulting in trade-offs being made between the efficiencies displayed for each of the individual parameter estimates (e.g., the best and worst orthogonal designs will produce a lower standard error for ß$_{24}$ than will the best D_p-efficient design generated, assuming a correct specification of the priors). Thus, only in the special case where there exists a design in which all elements in the asymptotic variance-covariance matrix are smaller than all for other designs, will that design produce lower asymptotic standard errors for all attributes. The existence of such a design on the efficiency frontier in design space, however, will likely be rare.

The presence of M in Equation (8) provides a useful result for comparing designs over various sample sizes without having to resort to the use of Monte Carlo experimentation. Dividing each element of the asymptotic (co)variance matrix for the single respondent case by M will produce the asymptotic (co)variance matrix for that sample size. This will be equivalent to the asymptotic (co)variance matrix obtained from Monte Carlo experiments conducted over a large number of iterations, thus negating the need to conduct such experiments for problems of this type. Denote the asymptotic standard errors when the number of respondents are M by $se_{M}(\hat{b}_{k}^{*})$ and $se_{M}(\hat{b}_{jk})$ for each of the generic and alternative-specific parameters. Then it holds that

(13) $$se_{M}(\hat{b}_{k}^{*}) = se_{1}(\hat{b}_{k}^{*})/\sqrt{M}, \quad \text{and} \quad se_{M}(\hat{b}_{jk}) = se_{1}(\hat{b}_{jk})/\sqrt{M}.$$

For example, for design 1, $se_{1}(\hat{b}_{14}) = \sqrt{0.25} = 0.5.$ The asymptotic standard error with 50 respondents will therefore be $se_{50}(\hat{b}_{14}) = \sqrt{0.25}/\sqrt{50}$ » 0.07. It is worth noting that dividing the asymptotic standard errors by the square root of M as explained above will produce diminishing improvements to $se_{M}(\hat{b}_{jk})$ as M increases. As such, the MNL model will exhibit diminishing increases in reliability (as measured by lower asymptotic standard errors) as we increase the sample size

This property allows for an examination of the influences of sample size upon the statistical significance of the parameter estimates likely to be obtained from the experiment. Given that the asymptotic *t*-statistic is calculated as the ratio of the parameter estimate to the asymptotic standard error, it is possible to determine what sample size will be required in order to demonstrate statistical significance of the parameter estimates. Table 4 demonstrates the predicted asymptotic *t*-statistics for each of the designs at various sample sizes. In calculating the asymptotic *t*-statistics for the three designs, we have assumed that the parameter priors used in the construction of the designs are correct. From Table 4, the best D_p-efficient (balanced) orthogonal design would require a sample size of 123 in order to determine that all attributes are statistically significant (at the 95 percent confidence level), whilst the corresponding worst D_p-efficient (balanced) orthogonal design would require a sample size of 223 respondents (ignoring the constant term which is often dispensed with in SC experiments; see Hensher et al., 2005). The non-orthogonal D_p-efficient design would require a minimum sample size of 121 respondents.

Table 4: Sample size influences upon attribute level significance (assuming correct priors)

Design 1: Best D_p-error Orthogonal Design							
M	**G_1**	**G_2**	**$ß_{13}$**	**$ß_{14}$**	**$ß_{20}$**	**$ß_{23}$**	**$ß_{24}$**
1	0.98	0.91	0.18	1.20	-0.19	0.58	1.31
2	1.39	1.28	0.25	1.70	-0.27	0.83	1.85
3	1.70	1.57	0.31	2.08	-0.33	1.01	2.27
4	1.96	1.81	0.35	2.40	-0.38	1.17	2.62
5	2.20	2.02	0.40	2.68	-0.43	1.31	2.93
123	10.89	10.04	1.96	13.30	-2.13	6.49	14.54
Design 2: Worst D_p-error Orthogonal Design							
1	0.94	0.64	0.13	1.04	-0.12	0.57	1.32
2	1.32	0.90	0.19	1.47	-0.17	0.81	1.86
3	1.62	1.11	0.23	1.80	-0.20	0.99	2.28
4	1.87	1.28	0.26	2.08	-0.24	1.15	2.63
5	2.09	1.43	0.29	2.32	-0.26	1.28	2.94
223	13.98	9.53	1.96	15.52	-1.76	8.57	19.64
Design 3: Best D_p-error (Non-Orthogonal Design)							
1	1.08	0.88	0.18	1.01	-0.19	0.64	1.00
2	1.53	1.24	0.25	1.43	-0.27	0.90	1.41
3	1.87	1.52	0.31	1.75	-0.33	1.11	1.73
4	2.16	1.76	0.36	2.02	-0.38	1.28	2.00
5	2.42	1.96	0.40	2.26	-0.42	1.43	2.23
121	11.89	9.65	1.96	11.13	-2.08	7.02	10.97

One benefit of the methodology is that it is easy to calculate the sample sizes assuming incorrect specification of the parameter priors. Table 5 shows the sample size requirements for the three designs assuming a different set of parameter priors (given in Table 5) than those used in generating the designs. Ignoring the constant, a misspecification of the priors in this case would result in a requirement of a smaller sample size than required if the priors were correctly specified. For the new priors assumed, Design 1 would require only 69 respondents whilst Designs 2 and 3 would require 125 and 68 respondents respectively.

Note that, keeping the design constant, a misspecification of a parameter prior for any attribute will have an impact upon the asymptotic standard errors for all parameter estimates within the model. This is because for any given design, a change in any parameter value for an attribute will influence the choice probabilities within all choice sets *n* where that attribute appears. Changes in the choice probabilities will in turn feed through to the asymptotic (co)variance matrix and hence influence the resulting expected standard errors for all parameters.

Table 5: Sample size influences upon attribute level significance (assuming incorrect priors)

	Design 1: Best D_p-error Orthogonal Design						
	0.5	***0.4***	***0.4***	***0.7***	***-1.2***	***0.5***	***0.8***
M	**G_1**	**G_2**	**$ß_{13}$**	**$ß_{14}$**	**$ß_{20}$**	**$ß_{23}$**	**$ß_{24}$**
1	1.23	1.21	0.24	1.40	-0.19	0.73	1.50
2	1.74	1.71	0.33	1.98	-0.27	1.03	2.12
3	2.13	2.09	0.41	2.42	-0.33	1.27	2.60
4	2.45	2.41	0.47	2.80	-0.38	1.46	3.00
5	2.74	2.70	0.53	3.13	-0.43	1.64	3.35
69	10.20	10.03	1.96	11.62	-1.60	6.07	12.45
	Design 2: Worst D_p-error Orthogonal Design						
1	1.17	0.85	0.17	1.21	-0.12	0.72	1.50
2	1.66	1.20	0.25	1.71	-0.17	1.01	2.13
3	2.03	1.47	0.30	2.10	-0.20	1.24	2.60
4	2.34	1.70	0.35	2.43	-0.24	1.43	3.01
5	2.62	1.90	0.39	2.71	-0.26	1.60	3.36
125	13.08	9.51	1.96	13.56	-1.32	8.02	16.81
	Design 3: Best D_p-error (Non-Orthogonal Design)						
1	1.35	1.17	0.24	1.18	-0.19	0.80	1.14
2	1.91	1.65	0.34	1.67	-0.27	1.13	1.61
3	2.34	2.03	0.41	2.04	-0.33	1.38	1.97
4	2.70	2.34	0.48	2.36	-0.38	1.60	2.28
5	3.02	2.62	0.53	2.64	-0.42	1.78	2.55
68	11.14	9.65	1.96	9.73	-1.56	6.58	9.40

6. Conclusion and Discussion

In this paper, we have extended the proof offered by McFadden (1974) for the generic (or unlabeled) MNL model and the alternative-specific case specified by Bliemer and Rose (2005) to the more general case allowing of both alternative-specific and generic parameter estimates. In doing so, we have been able to demonstrate the appropriate asymptotic (co)variance matrix for the most general model specification, thus allowing for the first time, the correct generation of efficient designs for any form of SC experiment. Beyond the ability to generate efficient designs for alternative-specific SC experiments, a number of additional aspects contained within this paper are worth emphasizing.

First, we demonstrate that for an experiment of given dimensions, it may be possible to generate a number of different orthogonal designs, each with differing levels of efficiency as measured after model estimation (assuming that the estimated parameters are non-zero). Within this paper, we have demonstrated that the D_z-efficiency measure often employed within the literature on the generation of efficient generic (or unlabeled) SC experiments, provides a meaningless basis of comparison amongst orthogonal designs.

Second, for any given sample size, one may examine the likely standard errors and asymptotic *t*-statistics of a design to be estimated using the MNL model directly from the asymptotic (co)variance matrix. This means that for this class of models, one does not have to rely on Monte Carlo simulations to determine the expected standard errors for various sample sizes for different designs as has been done by some researchers in the past (e.g., Sándor and Wedel 2001). The ability to use the asymptotic (co)variance matrix to estimate the standard errors directly extends to being able to examine likely biases in the expected *t*-statistics given misspecification of the parameter priors. This can be done relatively quickly, allowing for an assessment of the implications of misspecification of the priors even before an experiment has been implemented.

The ability to derive efficient alternative specific designs introduces a number of possible interesting research directions. First, the limitation of being only able to estimate efficient designs for generic SC experiments has meant that the literature has not addressed the issue of efficient designs assuming differences in scale across alternatives. An interesting research direction therefore would be to extend the designing of SC experiments beyond the MNL model to models that allow for scale differences such as the nested logit model (Sándor and Wedel (2002) have examined efficient design generation for the mixed logit model). Second, the designs generated here do not assume the presence of a no-choice base alternative. Although only a simple extension, the effect of having a no-choice alternative needs to be examined for alternative-specific designs, as has occurred with the unlabeled SC case (see Carlsson and Martinsson 2002).

We would also promote research into wider aspects of constructing efficient experimental designs. Of particular interest is the construction of efficient designs for experiments in which the attribute levels are pivoted from the revealed levels obtained from respondents prior to the commencement of a SC experiment (see for example, Greene et al. 2005). Of issue for such designs is that not only are the prior parameter estimates needed to generate efficient designs not known with any certainty, but so are the attribute levels for each respondent. Urgent research examining the use of internet or CAPI technology with in-built design optimization routines is required for such experiments.

A further research issue involves the investigation of what constitutes the best source for determining the priors used in generating optimal designs. Should the analyst conduct a pilot study, and if so, what represents a sufficient sample size to obtain the priors? Alternatively, should the analyst rely upon managers and other practitioners beliefs and how best should such beliefs be captured?

Finally, we propose further research be conducted into various possible measures for defining the efficiency of designs. Although for this paper, we have relied upon D-error as our measure of, numerous other possible measures exist. One such possible measure not yet considered by the literature is that of using some form of weighting procedure to indicate which elements within the asymptotic (co)variance matrix should receive priority in terms of minimisation. Such a measure would be of interest, if for example, one were mainly interested in estimating the willingness to pay for a specific attribute. In such a case, it would be conceivable that the researcher could believe that it is more important to produce lower standard errors for both this and the cost attribute within the design whilst other attributes are of less importance to the study. In such a case, the reliance on a global measure to determine the efficiency of the overall asymptotic (co)variance matrix will be inadequate.

Appendix A

1st derivative of the loglikelihood function (Jacobian/score vector)

Derivative with respect to generic parameter $b^*_{k_1}$:

$$\frac{\partial L(b\,|x,y)}{\partial b^*_{k_1}} = \sum_{s=1}^{S}\left[\sum_{j=1}^{J} y_{js}\left(\frac{\partial}{\partial b^*_{k_1}}\sum_{k=1}^{K^*} b^*_k x^*_{jks}\right) - \frac{\partial}{\partial b^*_{k_1}}\log\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{iks} + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)\right)\right]$$

$$= \sum_{s=1}^{S}\left[\sum_{j=1}^{J} y_{js} x^*_{jk_1s} - \frac{\sum_{j=1}^{J}\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{jks} + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right) x^*_{jk_1s}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{iks} + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right]$$

$$= \sum_{s=1}^{S}\left[\sum_{j=1}^{J} y_{js} x^*_{jk_1s} - \sum_{j=1}^{J} P_{js} x^*_{jk_1s}\right]$$

$$= \sum_{s=1}^{S}\sum_{j=1}^{J}\left(y_{js} - P_{js}\right) x^*_{jk_1s}.$$

Derivative with respect to alternative-specific parameter $b_{j_1k_1}$:

$$\frac{\partial L(b\,|x,y)}{\partial b_{j_1k_1}} = \sum_{s=1}^{S}\left[y_{j_1s}\left(\frac{\partial}{\partial b_{j_1k_1}}\sum_{k=1}^{K_j} b_{jk} x_{jks}\right) - \frac{\partial}{\partial b_{j_1k_1}}\log\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{iks} + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)\right)\right]$$

$$= \sum_{s=1}^{S}\left[y_{j_1s} x_{j_1k_1s} - \frac{\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{j_1ks} + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right) x_{j_1k_1s}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b^*_k x^*_{iks} + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right]$$

$$= \sum_{s=1}^{S}\left[y_{js} x_{j_1k_1s} - P_{j_1s} x_{j_1k_1s}\right]$$

$$= \sum_{s=1}^{S}\left(y_{js} - P_{j_1s}\right) x_{j_1k_1s}.$$

2nd derivative of the loglikelihood function (Hessian/information matrix)

Derivatives with respect to generic parameter $b_{k_1}^*$:

$$
\begin{aligned}
\frac{\partial^2 L(b|x,y)}{\partial b_{k_1}^* \partial b_{k_2}^*} &= \frac{\partial}{\partial b_{k_2}^*} \sum_{s=1}^{S}\sum_{j=1}^{J}\left(y_{js} - P_{js}\right) x_{jk_1s}^* \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} x_{jk_1s}^* \frac{\partial}{\partial b_{k_2}^*}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right) \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} x_{jk_1s}^* \left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right) x_{jk_2s}^* \sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) - \exp\left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right)\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) x_{ik_2s}^*}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)\right)^2}\right) \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} x_{jk_1s}^* \left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right) x_{jk_2s}^*}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)} - \frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{jks}^* + \sum_{k=1}^{K_j} b_{jk} x_{jks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)} \cdot \frac{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) x_{ik_2s}^*}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right) \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} x_{jk_1s}^* \left(P_{js} x_{jk_2s}^* - P_{js} \cdot \sum_{i=1}^{J} P_{is} x_{ik_2s}^*\right) \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} x_{jk_1s}^* P_{js}\left(x_{jk_2s}^* - \sum_{i=1}^{J} P_{is} x_{ik_2s}^*\right).
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L(b|x,y)}{\partial b_{j_1k_1} \partial b_{k_2}^*} &= \frac{\partial}{\partial b_{k_2}^*} \sum_{s=1}^{S}\left(y_{js} - P_{j_1s}\right) x_{j_1k_1s} \\
&= -\sum_{s=1}^{S} x_{j_1k_1s} \frac{\partial}{\partial b_{k_2}^*}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right) \\
&= -\sum_{s=1}^{S} x_{j_1k_1s} \left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right) x_{j_1k_2s}^* \sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) - \exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right)\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) x_{ik_2s}^*}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)\right)^2}\right) \\
&= -\sum_{s=1}^{S} x_{j_1k_1s} \left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right) x_{j_1k_2s}^*}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)} - \frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k} x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)} \cdot \frac{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right) x_{ik_2s}^*}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik} x_{iks}\right)}\right) \\
&= -\sum_{s=1}^{S} x_{j_1k_1s} \left(P_{j_1s} x_{j_1k_2s}^* - P_{j_1s} \sum_{i=1}^{J} P_{is} x_{ik_2s}^*\right) \\
&= -\sum_{s=1}^{S} x_{j_1k_1s} P_{j_1s}\left(x_{j_1k_2s}^* - \sum_{i=1}^{J} x_{ik_2s}^* P_{is}\right).
\end{aligned}
$$

Derivatives with respect to alternative-specific parameter $b_{j_2k_2}$:

$$\frac{\partial^2 L(b\,|x,y)}{\partial b^*_{k_1}\partial b_{j_2k_2}} = \frac{\partial}{\partial b_{j_2k_2}}\sum_{s=1}^{S}\sum_{j=1}^{J}\left(y_{js}-P_{js}\right)x^*_{jk_1s}$$

$$= -\sum_{s=1}^{S}\sum_{j=1}^{J}x^*_{jk_1s}\frac{\partial}{\partial b_{j_2k_2}}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{jks}+\sum_{k=1}^{K_j}b_{jk}x_{jks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\right)$$

$$= -\sum_{s=1}^{S}\left[x^*_{j_2k_1s}\frac{\partial}{\partial b_{j_2k_2}}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\right)+\sum_{j\neq j_2}x^*_{jk_1s}\frac{\partial}{\partial b_{j_2k_2}}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{jks}+\sum_{k=1}^{K_j}b_{jk}x_{jks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\right)\right]$$

$$= -\sum_{s=1}^{S}\left[x^*_{j_2k_1s}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)-\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)\right)^2}\right)+\right.$$

$$\left.+\sum_{j\neq j_2}x^*_{jk_1s}\left(\frac{0-\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{jks}+\sum_{k=1}^{K_j}b_{jk}x_{jks}\right)\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)\right)^2}\right)\right]$$

$$= -\sum_{s=1}^{S}\left[x^*_{j_2k_1s}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}-\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\cdot\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\right)-\right.$$

$$\left.-\sum_{j\neq j_2}x^*_{jk_1s}\left(\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{jks}+\sum_{k=1}^{K_j}b_{jk}x_{jks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\frac{\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{j_2ks}+\sum_{k=1}^{K_j}b_{j_2k}x_{j_2ks}\right)x_{j_2k_2s}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*}b^*_k x^*_{iks}+\sum_{k=1}^{K_j}b_{ik}x_{iks}\right)}\right)\right]$$

$$= -\sum_{s=1}^{S}\left[x^*_{j_2k_1s}\left(P_{j_2s}x_{j_2k_2s}-P_{j_2s}P_{j_2s}x_{j_2k_2s}\right)-\sum_{j\neq j_2}x^*_{jk_1s}P_{js}P_{j_2s}x_{j_2k_2s}\right]$$

$$= -\sum_{s=1}^{S}\left[x^*_{j_2k_1s}P_{j_2s}x_{j_2k_2s}\left(1-P_{j_2s}\right)-P_{j_2s}x_{j_2k_2s}\sum_{j\neq j_2}x^*_{jk_1s}P_{js}\right]$$

$$= -\sum_{s=1}^{S}x_{j_2k_2s}P_{j_2s}\left(x^*_{j_2k_1s}-x^*_{j_2k_1s}P_{j_2s}-\sum_{i\neq j_2}x^*_{ik_1s}P_{is}\right)$$

$$= -\sum_{s=1}^{S}x_{j_2k_2s}P_{j_2s}\left(x^*_{j_2k_1s}-\sum_{i=1}^{J}x^*_{ik_1s}P_{is}\right).$$

$$\frac{\partial^2 L(b\,|x,y)}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = \frac{\partial}{\partial b_{j_2k_2}}\sum_{s=1}^{S}\left(y_{js}-P_{j_1s}\right)x_{j_1k_1s}$$

$$= -\sum_{s=1}^{S} x_{j_1k_1s}\frac{\partial}{\partial b_{j_2k_2}}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}\right)$$

If $j_1 \neq j_2$:

$$\frac{\partial^2 L(b\,|x,y)}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = -\sum_{s=1}^{S} x_{j_1k_1s}\left(\frac{0-\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_2ks}^* + \sum_{k=1}^{K_j} b_{j_2k}x_{j_1ks}\right)x_{j_2k_2s}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)\right)^2}\right)$$

$$= \sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}\cdot\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_2ks}^* + \sum_{k=1}^{K_j} b_{j_2k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}\right)$$

$$= \sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}P_{j_1s}P_{j_2s}.$$

If $j_1 = j_2$:

$$\frac{\partial^2 L(b\,|x,y)}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = -\sum_{s=1}^{S} x_{j_1k_1s}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)x_{j_2k_2s}\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)-\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_2ks}^* + \sum_{k=1}^{K_j} b_{j_2k}x_{j_1ks}\right)x_{j_2k_2s}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)\right)^2}\right)$$

$$= -\sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}\left(\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}x_{j_2k_2s} - \frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_1ks}^* + \sum_{k=1}^{K_j} b_{j_1k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}\cdot\frac{\exp\left(\sum_{k=1}^{K^*} b_k^* x_{j_2ks}^* + \sum_{k=1}^{K_j} b_{j_2k}x_{j_1ks}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K^*} b_k^* x_{iks}^* + \sum_{k=1}^{K_j} b_{ik}x_{iks}\right)}\right)$$

$$= -\sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}P_{j_1s}\left(1-P_{j_2s}\right).$$

Appendix B

Theorem 1 – All balanced orthogonal designs using the same attribute levels have the same D_z-error.

Proof: Consider the Fisher information matrix of designs using generic and alternative-specific attributes. For the D_z-error, the Fisher information matrix assumes that all parameters are equal to zero. Since $P_{js} = 1/J$ for all alternatives *j* and all choice situations *s*, the Fisher information matrix will become (after rearranging summations):

$$\frac{\partial^2 L(0,0)}{\partial b^*_{k_1} \partial b^*_{k_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J} x^*_{jk_1s} \frac{1}{J}\left(x^*_{jk_2s} - \sum_{i=1}^{J} \frac{1}{J} x^*_{ik_2s} \right) = -\frac{1}{J}\sum_{j=1}^{J}\left(\sum_{s=1}^{S} x^*_{jk_1s} x^*_{jk_2s} \right) + \frac{1}{J^2}\sum_{j=1}^{J}\sum_{i=1}^{J}\left(\sum_{s=1}^{S} x^*_{jk_1s} x^*_{ik_2s} \right)$$

$$\frac{\partial^2 L(0,0)}{\partial b_{j_1k_1} \partial b^*_{k_2}} = -\sum_{s=1}^{S} x_{j_1k_1s} \frac{1}{J}\left(x^*_{j_1k_2s} - \sum_{i=1}^{J} \frac{1}{J} x^*_{ik_2s} \right) = -\frac{1}{J}\left(\sum_{s=1}^{S} x_{j_1k_1s} x^*_{j_1k_2s} \right) + \frac{1}{J^2}\sum_{i=1}^{J}\left(\sum_{s=1}^{S} x_{j_1k_1s} x^*_{ik_2s} \right)$$

$$\frac{\partial^2 L(0,0)}{\partial b_{j_1k_1} \partial b_{j_2k_2}} = \begin{cases} \dfrac{1}{J^2}\left(\displaystyle\sum_{s=1}^{S} x_{j_1k_1s} x_{j_2k_2s} \right), & \text{if } j_1 \neq j_2; \\ -\dfrac{1}{J}\left(1-\dfrac{1}{J}\right)\left(\displaystyle\sum_{s=1}^{S} x_{j_1k_1s} x_{j_2k_2s} \right), & \text{if } j_1 = j_2. \end{cases}$$

Since orthogonality holds, the levels of one attribute for all choice situations are uncorrelated with the levels of any other attribute. Using the definition of correlation it holds that

$$\sum_{s=1}^{S} x_{j_1k_1s} x_{j_2k_2s} = \frac{1}{S}\sum_{s=1}^{S} x_{j_1k_1s} \sum_{s=1}^{S} x_{j_2k_2s}, \quad \text{for each combination of alternatives and attributes.}$$

Therefore, we can write the cells of Fisher information matrix as follows:

$$\frac{\partial^2 L(0,0)}{\partial b^*_{k_1} \partial b^*_{k_2}} = -\frac{1}{JS}\sum_{j=1}^{J}\left(\sum_{s=1}^{S} x^*_{jk_1s} \right)\left(\sum_{s=1}^{S} x^*_{jk_2s} \right) + \frac{1}{J^2 S}\sum_{j=1}^{J}\sum_{i=1}^{J}\left(\sum_{s=1}^{S} x^*_{jk_1s} \right)\left(\sum_{s=1}^{S} x^*_{ik_2s} \right)$$

$$\frac{\partial^2 L(0,0)}{\partial b_{j_1k_1} \partial b^*_{k_2}} = -\frac{1}{JS}\left(\sum_{s=1}^{S} x_{j_1k_1s} \right)\left(\sum_{s=1}^{S} x^*_{j_1k_2s} \right) + \frac{1}{J^2 S}\sum_{i=1}^{J}\left(\sum_{s=1}^{S} x_{j_1k_1s} \right)\left(\sum_{s=1}^{S} x^*_{ik_2s} \right)$$

$$\frac{\partial^2 L(0,0)}{\partial b_{j_1k_1} \partial b_{j_2k_2}} = \begin{cases} \dfrac{1}{J^2}\left(\displaystyle\sum_{s=1}^{S} x_{j_1k_1s} \right)\left(\displaystyle\sum_{s=1}^{S} x_{j_2k_2s} \right), & \text{if } j_1 \neq j_2; \\ -\dfrac{1}{J}\left(1-\dfrac{1}{J}\right)\left(\displaystyle\sum_{s=1}^{S} x_{j_1k_1s} \right)\left(\displaystyle\sum_{s=1}^{S} x_{j_2k_2s} \right), & \text{if } j_1 = j_2. \end{cases}$$

Since we also assume balancedness, it holds that $\sum_{s=1}^{S} x_{jks}$ is constant for each alternative *j* and attribute *k*, independent of the order of the attribute levels over the choice situations. Therefore, the cells of the Fisher information matrix are the same for each orthogonal design, hence also the (co)variance matrix and the D_z-error.

References

Anderson, Donald A. and James B. Wiley (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.

Bliemer, M.C.J and Rose, J.M. (2005) Efficient Designs for Alternative Specific Choice Experiments, Working Paper, University of Sydney, February.

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.

Burgess, Leonie and Deborah J. Street (2005) Optimal designs for choice experiments with asymmetric attributes, *Journal of Statistical Planning and Inference*, forthcoming.

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

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

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.

Greene, William H., David A. Hensher, and John M. Rose (2005) Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models, *Transportation Research B*, forthcoming.

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

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.

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

Lindsey, James K. (1996) *Parametric Statistical Inference*, London: Oxford.

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.

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

Shiflan, Y and Bard-Eden, R. (2001) Modeling Response to Parking Policy. *Transport Research Record 1765*, paper no. 01-2316, 27-34.

Train, K.E. (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.