I T L S

WORKING PAPER
ITLS-WP-05-04

Efficient Designs for Alternative Specific Choice Experiments

By

Michiel CJ Bliemer* & John M Rose

Adjunct Lecturer at ITLS, The University of Sydney and
A/Professor Transportation Modeling
Delft University of Technology
Faculty of Civil Engineering & Geosciences
Transport & Planning Section
Delft The Netherlands

March 2005

ISSN 1440-3501

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-05-04
TITLE:	**Efficient Designs for Alternative Specific Choice Experiments**

ABSTRACT:

In the past, research on the construction of efficient designs for stated choice experiments has been limited to unlabeled experiments with generic parameter estimates. In this paper, by deriving the asymptotic (co)variance matrix for the alternative-specific MNL model, the authors are able to generate efficient alternative-specific experiments. The authors show that D-error assuming prior parameter values equal to zero is unable to explain statistical efficiency in orthogonal designs and that wide attribute levels are likely to yield more reliable parameter estimates than using narrow attribute levels. The authors also show that the D-optimality criterion may yield inefficient parameter estimates for some design attributes given that trade-offs are made between the efficiencies of different parameter estimates.

.

KEY WORDS: ***Stated Choice, alternative specific and D-Efficiency***

AUTHORS: Michiel CJ Bliemer & John M Rose

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

DATE: March, 2005

1. Introduction

The growing evidence on the ability of stated choice (SC) experiments to represent decisions made in real markets (Burke et al. 1992; Carson et al. 1994) has made them a popular data paradigm in the elicitation of behavioral responses of individuals, households and organizations over diverse choice situations and contexts. An acknowledged limitation of SC experiments is that in order to produce asymptotically efficient parameter estimates, it is necessary that choice data from a number of respondents be pooled (Huber and Zwerina 1996), unless the number of person-specific observations captured is very large. A typical SC experiment might involve respondents being asked to undertake a number of choice tasks involving the choice from amongst a number of labeled or unlabeled alternatives defined on a number of attribute dimensions, each in turn described by pre-specified levels drawn from some underlying experimental design. The number of choice tasks undertaken will be up to the total number of choice sets drawn from the experimental design. Consequently, an archetypal SC experiment might require choice data collected from 200 respondents, each of whom were observed to have made eight choices each, thus producing a total of 1600 choice observations.

The necessity to pool data has lead several authors to seek ways to reduce the number of choice observations necessary for reliable analysis of choice data (e.g., Huber and Zwerina 1996; Sándor and Wedel 2001; Carlsson and Martinsson 2002; Kanninen 2002). Primarily, these research efforts have attempted to produce more statistically efficient experimental designs that for a given level of accuracy, allow for either a reduction in the number of choice set profiles shown to individual respondents or alternatively, a reduction in the number of respondents required to complete the experiment. Such designs have been widely studied within the literature. For example, Bunch, Louviere and Anderson (1994) studied statistically efficient main effects designs whilst Anderson and Wiley (1992) and Laziri and Anderson (1994) introduce methods to generate statistically efficient cross-effect designs.

More recently, Huber and Zwerina (1996), Sándor and Wedel (2001) and Kanninen (2002), showed that the use of logit models to analyze discrete choice data requires that *a priori* information be known about the parameter estimates in order to derive greater statistical efficiency in the generation of SC experimental designs (Kanninen demonstrates, however, how the efficiency of a designs may be updated during the course of the experiment). Information on the parameter estimates may be used to calculate the expected utilities for each of the alternatives present within the design, which in turn may be used to calculate the likely choice probabilities via the now familiar logit formula. Given knowledge of the attribute levels, expected parameter estimate values and choice probabilities, it becomes a straightforward exercise to calculate the asymptotic variance-covariance matrix. By manipulating the attribute levels of the alternatives, for fixed parameter values, the analyst is able to minimize the elements within the variance-covariance matrix, which in the case of the diagonals means lower standard errors and hence greater reliability in the estimates at a fixed sample size.

A common assumption within the literature dealing with the design of optimal SC experiments has been that the parameter estimates for each attribute are generic across the alternatives within the experiment. Carlsson and Martinsson (2002) remain the sole exception, examining the case of optimal SC designs in which the utility functions of the alternatives present within the experiment differ only by a constant term. As we demonstrate, however, the designs generated by Carlson and Martinsson (2002) are not strictly optimal given a misspecification of the log-likelihood function they used in the calculation of the asymptotic variance-covariance matrix for the case of alternative-specific choice models.

The assumption of generic parameter estimates arises as a direct result of the way the log-likelihood function for the MNL model 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 is limited to that of the MNL assuming generic parameter estimates. The assumption 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 variance-covariance matrix of discrete choice models, without which, attempts to minimize the elements of the asymptotic variance-covariance matrix cannot be guaranteed.

In this paper, we derive the log-likelihood function for the alternative specific MNL model and contrast this to that derived by McFadden. We then use the alternative specific derivation of the MNL log-likelihood function 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. Derivation of the Generic and Alternative Specific MNL models

In this section, we outline the derivation of the MNL model for both the generic and alternative specific case. Both the alternative-specific and generic cases follow the work of McFadden (1974) on random utility theory (RUT) and are 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 n and j refer to choice task n = 1, 2, …, N, and alternative j = 1, 2, …, J. RUT posits that the utility possessed by an individual for alternative j present in choice set n may be expressed as:

(1) $$U_{jn} = V_{jn}(x_{jn} \mid b_j) + e_{jn},$$

where U_{jn} is the overall utility associated with alternative j in choice set n, V_{jn} is the component of utility observed by the analyst for alternative j in choice set n, and x_{jn} is a vector of attribute levels for different alternative-specific attributes k = 1, 2, …, K_j of alternative j. Further, b_j denotes the vector of K_j parameters for each of the alternative-specific attributes, and e_{jn} represents the component of utility that is not observed by the analyst.

The subscript j in b_j allows for the estimation of alterative-specific parameter estimates across the j utility specifications, such that the number of estimated parameters is equal to $S_j K_j$. In the case of an unlabeled or generic experiment, for each attribute k, a single parameter is estimated independent of the number of alternatives resulting in the estimation of only K parameter estimates across all j, where K is the common number of attributes for all alternatives. Assuming linear additive utility functions, the observed components of the alternative-specific and generic cases are given in Equations (2a) and (2b) respectively.

(2a) $$V_{jn}(x_{jn} \mid b_j) = \sum_{k=1}^{K_j} x_{jnk} b_{jk},$$ (alternative-specific attributes)

(2b) $$V_{jn}(x_{jn} \mid b) = \sum_{k=1}^{K} x_{jnk} b_k.$$ (generic attributes)

Under the assumption that the unobserved component of utility, e_{jn}, are independently and identically extreme value type I distributed, we are able to derive the multinomial logit model in which P_{in} is the probability of choosing alternative i in choice set n:

(3a) $$P_{in}(x_n \mid b) = \frac{\exp\left(V_{in}(x_{in} \mid b_i)\right)}{\sum_{j=1}^{J} \exp\left(V_{jn}(x_{jn} \mid b_j)\right)},$$ (alternative-specific attributes)

(3b) $$P_{in}(x_n \mid b) = \frac{\exp\left(V_{in}(x_{in} \mid b)\right)}{\sum_{j=1}^{J} \exp\left(V_{jn}(x_{jn} \mid b)\right)}.$$ (generic attributes)

The log-likelihood as a function of the parameters is given by

(4) $$L(b \mid x, y) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{jn} \log P_{jn}(x_n | b),$$

where the vector y describes the outcomes of all choice tasks, that is, y_{jn} is one if alternative j is chosen in choice task n and is zero otherwise. The asymptotic variance-covariance matrix can be derived from the second derivative of the log-likelihood function. For the alternative-specific case this leads to the following (see Appendix A):

(5a) $$\frac{\partial^2 L}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = \begin{cases} -\sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n|b)\left(1-P_{j_2n}(x_n|b)\right), & \text{if } j_1 = j_2, \\ \sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n|b)P_{j_2n}(x_n|b), & \text{if } j_1 \neq j_2. \end{cases}$$

This differs to that of the second derivative in the case of generic parameters, which is given by (see e.g., Kanninen 2002)

(5b) $$\frac{\partial^2 L}{\partial b_{k_1}\partial b_{k_2}} = -\sum_{n=1}^{N}\left(\sum_{j=1}^{J} x_{jk_1n}P_{jn}(x_n|b)\left(x_{jk_2n} - \sum_{i=1}^{J} x_{ik_2n}P_{in}(x_n|b)\right)\right).$$

Note that these second derivatives do not depend on the outcomes *y*. Suppose there are *M* respondents each completing the same *N* choice tasks. Then the second derivates are merely multiplied by *M*.

The maximum likelihood (ML) estimates of b 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 $\hat{b}$, that is:

(6) $$\hat{b} = \arg\max_{b} L(b|x,y).$$

McFadden (1974) has shown for the generic case that the ML estimates $\hat{b}$ are asymptotically normally distributed with mean b and variance-covariance matrix Ω which is equal to the negative inverse of the Fisher information matrix. The Fisher information matrix I is defined as the expected values of the second derivative of the log-likelihood function, that is:

(7) $$I(b \mid x) = M \cdot \frac{\partial^2 L}{\partial b \partial b'}.$$

Hence, the asymptotic variance-covariance matrix can be computed as

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

It can be shown that the same holds for the alternative-specific case. Clearly, the (co)variances become smaller with larger sample sizes, that is, with an increasing number of respondents *M*.

The resulting asymptotic variance-covariance matrix for the alternative-specific case using equations (5a), (7), and (8), will be a matrix of $\sum_j K_j$ rows and columns in which each row/column represents a separate alternative-specific parameter estimate. The asymptotic variance-covariance matrix derived for the generic case using Equation (5b) will possess only *K* rows and columns.

This last point is important in light of claims made by Carlsson and Martinsson (2002). In their paper, Carlson and Martinsson claim to construct alternative-specific designs (albeit, only through the addition of an alternative specific constant) using algorithms designed to locate optimal designs for unlabeled or generic choice experiments. Using the procedures outlined here, it can be shown that the D-efficient design that they derive has a near singular variance-covariance matrix producing an extremely large D-efficiency value. Indeed, examination of the design reveals that the X3 attribute for alternatives one and two are perfectly negatively correlated.

3. Measuring Statistical Efficiency in SC Experimental Designs

A statistically efficient design is a design that minimizes the elements of the asymptotic variance-covariance 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 can be computed by taking the determinant of the asymptotic variance-covariance matrix and applying a scaling factor $1/\sum_j K_j$ in order to take the number of parameters into account:

(9) $$\text{D-error} = \left(\det \Omega\right)^{1/\sum_j K_j} = -\frac{1}{M}\left(\det\left(\frac{\partial L^2}{\partial b \partial b'}\right)\right)^{-1/\sum_j K_j},$$

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 due to the fact that the covariance matrix is positive definite as 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.

Two popular approaches exist for computing the D-error. The first approach assumes there is no information on the true parameter values; that is, the prior parameters for all b are zero. This leads to the so-called D_z-error measure. In contrast, if prior information is available, then these priors for b can be used to compute the D-error, yielding the so-called D_p-error measure. Hence, the D_z-error can be computed as:

(10) $$D_z\text{-error} = \left(\det I(0 \mid x)\right)^{-1/\sum_j K_j},$$

while the D_p-error assuming knowledge of prior parameter estimates $\tilde{b}$ can be computed as

(11) $$D_p\text{-error} = \left(\det I(\tilde{b} \mid x)\right)^{-1/\sum_j K_j}.$$

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 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, Tobias and Garrett (1994), Huber and Zwerina (1996), Sándor and Wedel (2001), Kanninen (2002), Carlsson and Martinsson (2002), and Burgess and Street (2005) amongst other sources.

4. Generating D-efficient Alternative Specific Stated Choice Designs

Using Equations (5a), (7), and (8), we generate a number of alternative-specific SC designs for a choice experiment involving two alternatives, one with three attributes and the second with two attributes. Fixing the prior parameter estimates, we generate designs with wide and narrow attribute level ranges and construct D_p-efficient designs assuming both orthogonality and non-orthogonality in the construction of these designs. 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). For purpose of comparison, we also generate the least D_p-efficient orthogonal designs for the wide and narrow cases. The assumed parameter estimates and attribute levels 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 alternative-specific SC design with this number of attributes and attribute levels.

Table 1: Prior parameters and design attribute levels

Alternative	Attribute	Parameter	Wide levels	Narrow levels
A	Constant	-0.5	-	-
A	x_{11}	0.8	1, 3, 5	2, 3, 4
A	x_{12}	0.4	3, 5, 7	4, 5, 6
A	x_{13}	0.3	3, 6	4, 5
B	x_{21}	0.9	1, 3, 5	2, 3, 4
B	x_{22}	0.5	3, 5, 7	4, 5, 6

A total of six alternative-specific SC designs were generated and are shown in Tables 2 and 3. Designs 1 through 3 represent designs generated using the wider attribute level ranges, designs 4 through 6 represent the designs with narrow attribute levels. Designs

1, 2, 4, and 5 are orthogonal whilst designs 3 and 6 are non-orthogonal. Designs 1 and 4 are the most D_p-efficient (balanced) orthogonal designs we were able to construct using wide and narrow attribute levels, whilst designs 2 and 5 represent the worst possible D_p-efficient (balanced) orthogonal designs given the parameter priors assumed. The reason for including these worst possible D_p-efficient designs is because many researchers tend to consider only one orthogonal design, which can have a high D_p-efficiency or a low D_p-efficiency (which is generally not computed). The generation of the two designs will allow for an examination of the impact upon the reliability in parameter estimates obtained from two different orthogonal designs of the same dimensions. Designs 3 and 6 have even a lower D_p-error than the designs 1 and 4 as we do not impose orthogonality on these designs.

Table 2: Wide 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)**						
	Alt A			Alt B				Alt A			Alt B				Alt A			Alt B			
Cset #	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B
1	1	3	6	3	3	0.29	0.71	3	5	3	1	5	0.80	0.20	5	5	3	3	5	0.77	0.23
2	1	7	3	3	3	0.45	0.55	1	3	3	5	3	0.03	0.97	3	3	6	5	3	0.25	0.75
3	5	5	6	5	5	0.57	0.43	5	3	6	3	3	0.91	0.09	3	3	3	1	7	0.40	0.60
4	3	3	6	3	3	0.67	0.33	3	3	6	5	7	0.04	0.96	3	5	3	3	7	0.20	0.80
5	1	7	6	1	7	0.62	0.38	1	7	6	1	3	0.92	0.08	1	5	3	1	3	0.69	0.31
6	5	5	3	1	5	0.95	0.05	5	7	3	5	5	0.55	0.45	3	7	6	5	3	0.62	0.38
7	5	5	6	1	5	0.98	0.02	5	5	3	3	5	0.77	0.23	1	7	3	3	3	0.45	0.55
8	3	7	3	3	3	0.80	0.20	3	7	3	3	5	0.60	0.40	5	3	6	3	5	0.79	0.21
9	5	5	3	5	5	0.35	0.65	5	5	6	1	7	0.95	0.05	5	7	6	5	7	0.52	0.48
10	1	3	3	5	7	0.00	1.00	1	7	6	5	7	0.04	0.96	1	7	6	1	7	0.62	0.38
11	3	7	6	5	7	0.18	0.82	1	3	3	1	7	0.12	0.88	5	5	3	5	5	0.35	0.65
12	3	3	3	1	7	0.40	0.60	3	5	6	3	3	0.82	0.18	1	3	6	1	5	0.48	0.52
	$D_p = 0.2694, D_z = 0.1514$							$D_p = 0.4501, D_z = 0.1514$							$D_p = 0.2233, D_z = 0.1906$						

Table 3: Narrow designs

	Design 4: Best D_p-error Orthogonal Design							**Design 5: Worst D_p-error Orthogonal Design**							**Design 6: Best D_p-error (Non-Orthogonal Design)**						
	Alt A			Alt B				Alt A			Alt B				Alt A			Alt B			
Cset #	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	P_A	P_B
1	4	4	4	4	4	0.50	0.50	2	5	5	2	5	0.29	0.71	3	6	4	3	4	0.69	0.31
2	3	6	5	3	4	0.43	0.57	4	6	5	4	5	0.44	0.56	3	5	5	4	6	0.23	0.77
3	4	6	5	4	6	0.40	0.60	3	5	4	3	6	0.33	0.67	4	6	5	4	6	0.50	0.50
4	3	6	4	3	4	0.43	0.57	4	4	5	2	6	0.25	0.75	4	4	4	2	6	0.67	0.33
5	3	6	4	2	5	0.29	0.71	2	4	5	4	6	0.40	0.60	2	4	5	3	4	0.38	0.62
6	3	4	5	2	5	0.29	0.71	3	6	4	3	5	0.38	0.63	2	5	5	2	5	0.57	0.43
7	2	5	4	2	6	0.25	0.75	3	5	4	3	5	0.38	0.63	4	4	4	4	5	0.35	0.65
8	4	4	4	3	6	0.33	0.67	4	4	4	2	4	0.33	0.67	3	5	5	2	5	0.75	0.25
9	2	5	4	4	5	0.44	0.56	2	6	5	2	4	0.33	0.67	2	6	4	2	6	0.48	0.52
10	4	5	5	2	5	0.29	0.71	4	5	5	4	4	0.50	0.50	3	6	4	4	4	0.48	0.52
11	2	5	5	4	6	0.40	0.60	2	4	4	4	4	0.50	0.50	4	5	5	3	4	0.82	0.18
12	2	4	5	3	4	0.43	0.57	3	6	4	3	6	0.33	0.67	2	4	4	3	5	0.21	0.79
	$D_p = 0.6972, D_z = 0.5503$							$D_p = 0.7734, D_z = 0.5503$							$D_p = 0.6633, D_z = 0.5774$						

All designs were generated using algorithms programmed in Matlab, which used a heuristic to generate a large number of orthogonal and non-orthogonal designs 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).

Note that the D_z–efficiency measures for the orthogonal designs with the same attribute level ranges are identical. This will hold in general, as the following theorem states.

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

Proof: Consider the Fisher information matrix of such designs. For the D_z-efficiency measure, the Fisher information matrix, $I(0\,|\,x)$, assumes that all true parameters are equal to zero. Since $P_{jn}(x_n\,|\,0) = 1/J$ for all alternatives *j* and all choice tasks *n*, the Fisher information matrix will be a $\Sigma_j K_j \times \Sigma_j K_j$ matrix, which can be written as follows.

$$(12) \qquad I(0\,|\,x) = -\left.\frac{\partial^2 L}{\partial b_{j_1k_1}\partial b_{j_2k_2}}\right|_{b=0} = \begin{cases} \frac{1}{J}\left(1-\frac{1}{J}\right)\sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}, & \text{if } j_1 = j_2, \\ -\frac{1}{J^2}\sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}, & \text{if } j_1 \neq j_2. \end{cases}$$

Given the imposed restriction of attribute level balance, the diagonals of this matrix will be the same for any design of the same dimension. The off-diagonals however, may be different for a balanced design. Since we also assumed that the design matrix is orthogonal, all correlations between two (alternative-specific) attributes must be zero. Using the definition of correlation it therefore holds that $N\sum_n x_{j_1k_1n}x_{j_2k_2n} = \sum_n x_{j_1k_1n}\sum_n x_{j_2k_2n}$ for any combination of any two attributes. Since the design is balanced in the attribute levels, the right-hand side will be the same for each orthogonal design and hence the Fisher information matrix will also be the same for all orthogonal designs. ?

Furthermore, we conjecture that any balanced orthogonal design has minimum D_z-error, as we have not been able to find non-orthogonal designs with a lower D_z-error.

Conjecture 1 – A balanced orthogonal alternative-specific design has minimum D_z-error among all other balanced alternative-specific designs using the same attribute levels.

As such, the D_z-efficiency measure represents an inadequate method of comparison between orthogonal alternative-specific designs. Although the D_z-error may be useful in order to create efficient generic designs, the D_z-error will fail to distinguish between different design efficiency of any two alternative-specific orthogonal designs. Thus, the most efficient alternative-specific design when there is no prior information on the parameters available is an orthogonal design.

Table 4 demonstrates the asymptotic variance-covariance matrix derived for the first design, assuming a single respondent. From this, it can be seen that despite the use of an orthogonal design, the resulting covariances 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, Hensher and Swait 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 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, lower D_p-efficiency values are obtained from the two non-orthogonal designs represented in Tables 2 and 3 than for the equivalent best-case orthogonal designs.

Table 4: Asymptotic variance-covariance matrix of Design 1

	const	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}
const	16.95	-0.89	-1.12	-1.26	0.01	0.49
x_{11}	-0.89	0.46	0.15	0.14	0.33	0.16
x_{12}	-1.12	0.15	0.23	0.08	0.15	0.09
x_{13}	-1.26	0.14	0.08	0.28	0.14	0.08
x_{21}	0.01	0.33	0.15	0.14	0.48	0.18
x_{22}	0.49	0.16	0.09	0.08	0.18	0.25

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 variance-covariance matrix by *M* will produce the asymptotic variance-covariance matrix for that sample size. This will be equivalent to the asymptotic variance-covariance 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 $se_M(\hat{b}_{jk})$. Then $se_M(\hat{b}_{jk}) = se_1(\hat{b}_{jk})/\sqrt{M}$. For example, for design 1, $se_1(\hat{b}_{11}) = \sqrt{0.46}$. The asymptotic standard error with 50 respondents will therefore be $se_{50}(\hat{b}_{11}) = \sqrt{0.46}/\sqrt{50} \approx 0.096$. Using this property, we are able to plot the asymptotic standard errors for each of the designs shown in Tables 2 and 3 over sample sizes ranging from 50 to 300 respondents. These are shown in Figure 1.

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, which is demonstrated by the shape of the curves represented in Figure 1.

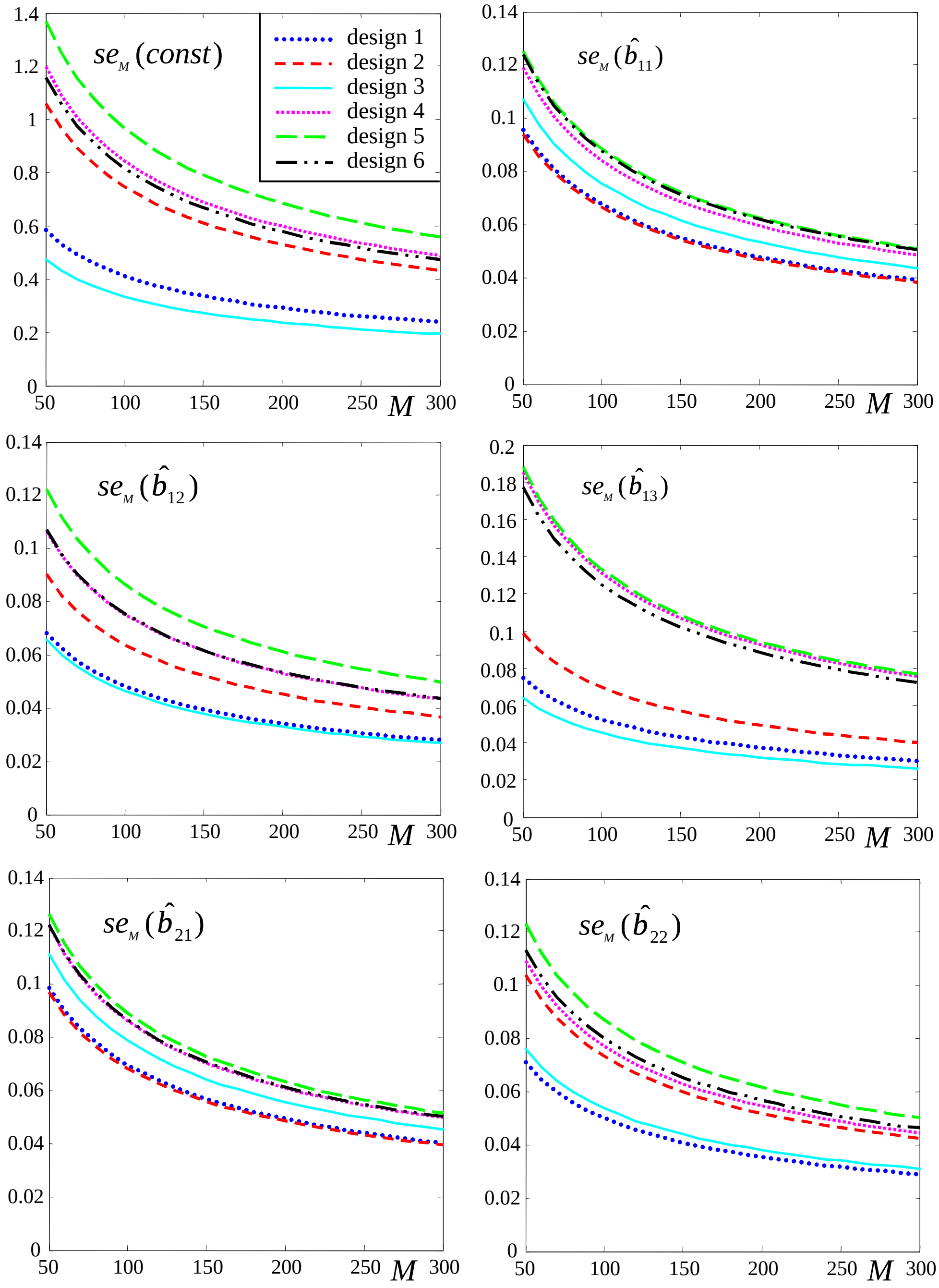

Figure 1: Expected standard errors for designs 1 through 6 for sample sizes between 50 and 300.

In general, the orthogonal and non-orthogonal D_p-efficient designs with wide attribute levels (designs 1 and 3) are competitive in producing low asymptotic standard errors compared to all other designs. The designs with narrow attribute levels (designs 4 through 6) appear to perform relatively poorly. Statistically, the wider the attribute levels used, the greater the range of values the utility functions may take for a given set of parameter values. This will allow for a greater examination of utility space, with the resulting increase in information leading to better estimates (i.e., the application area of the model is larger). This, however, ignores any behavioral influences that may result in the use of attribute level ranges which vary too far from experience in terms of existing attributes, or from expectations for currently non-existing attributes. Comparing the two orthogonal designs with wide attribute levels, we observe big differences in the asymptotic standard errors based on which orthogonal design is used. The D_p-efficient design (design 1) performs in general much better than the D_p-inefficient design (design 2).

The behavioral implications of attribute level range (beyond the assumed priors), although ignored above, is an important consideration in the construction of SC experiments. We note, however, that the literature examining the behavioral implications for using wider or narrow attribute level ranges appears somewhat fragmented. Verlegh et al. (2002) found that consumers judging the importance of an attribute tend to consider the range of levels for this attribute, such that an increase in range results in an increase in self-rated importance. Ohler et al. (2000) concluded that variations in attribute ranges affect the (i) complexity of model functional forms; (ii) model fits; (ii) (possibly) the power to detect non-additivity; and (iv) between-subject response variability. In their study, range differences had little to no effect upon (v) logistic regression model parameters; (vi) within subject response variability, or (vii) error variance. Mellers and Cooke (1994) on the other hand found that the effect of a given difference in attribute range was greater when presented in a narrow range than a wide range.

Minimization of a single global measure (i.e., either D_p-error or D_z-error) representing all elements contained within the asymptotic variance-covariance matrix explains why in his 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. 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 results shown in Figure 1 assume that the prior parameter estimates were correctly specified over the sample. An interesting question is what impact a misspecification of the priors has upon the reliability of a given design. Keeping the generated design constant (based on some prior parameter estimates), we are able to change the true parameter estimates over a range of values and observe the effects upon the asymptotic standard errors obtained from the resulting asymptotic variance-covariance matrices. Figure 2 shows the impact upon the standard errors of each parameter (disregarding the constant) given a range of true parameter values for the first attribute of alternative one (b_{11}), assuming a sample size of one, as would occur for generated design 3. The prior parameter value that was used to create the design was $\tilde{b}_{11} = 0.8$, while we assume that the true parameter value lies between 0 and 2.

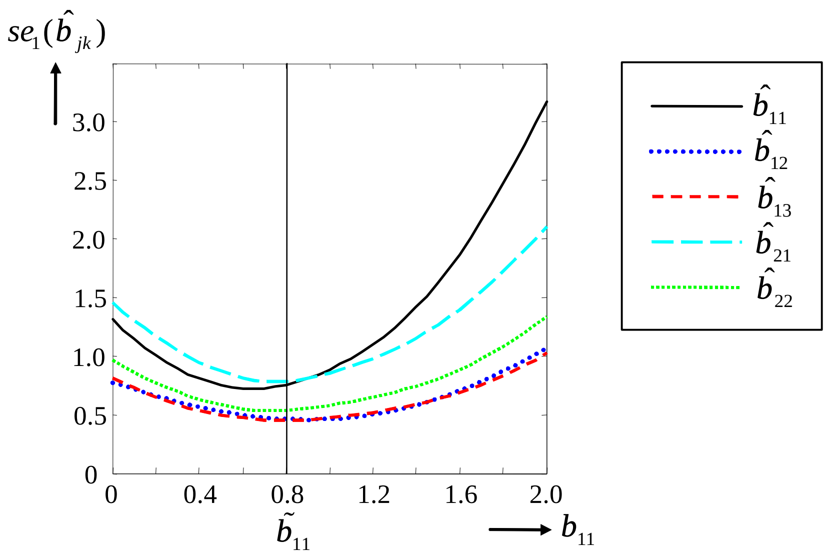

Figure 2: Expected standard errors given changes in the true value assumed by b_{11}

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 variance-covariance matrix and hence influence the resulting expected standard errors for all parameters. It is interesting to note that in Figure 2, lower standard errors will likely be obtained if the true parameter value is between 0.6 and 0.7 rather than the prior specified for that attribute (i.e., 0.8) used in generating the design. Again, this is as a result of attempting to minimize a single measure representing the overall asymptotic variance-covariance matrix rather than a measure that minimizes each individual element contained within the matrix (which may only be possible given the unlikely presence of a corner solution in efficiency space).

5. Conclusion and Discussion

In this paper, we have extended the proof offered by McFadden (1974) for the generic (or unlabeled) MNL model to the more general case of the alternative-specific model. In doing so, we have been able to demonstrate the appropriate asymptotic variance-covariance matrix for the alternative-specific model, thus allowing for the first time, the correct generation of efficient designs for alternative-specific SC experiments. 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, 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 shown 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 when used to distinguish between alternative specific experiments. Using the properties of the MNL model, one can examine the influence non-zero priors will likely have upon the statistical efficiency of generated orthogonal designs, as well as the ramifications upon the results of the experiment of mis-specifying these priors. Such information will allow for selection of which orthogonal design should then be used. The implication however, is that the D_z-efficiency measure should not be used to select between alternative-specific orthogonal SC designs.

Second, ignoring the behavioral implications of attribute level range (beyond the assumed priors), wider attribute level ranges are preferred in designing D-efficient (orthogonal and non-orthogonal) experiments. The downside of using narrow attribute ranges is twofold. First, they tend towards larger standard errors in the asymptotic efficiencies of the parameter estimates. Secondly, the application area of the model is smaller, resulting in less coverage in utility space. Attribute level range appears to have particular relevance with regards to the design efficiencies achieved for alternative-specific SC orthogonal designs, which tend to yield greater variation in efficiencies when wider attribute ranges are used. The behavioral implications associated with choice of attribute level range should not be ignored in real studies.

Third, for any given sample size, one may examine the likely standard errors of a design to be estimated using the MNL model directly from the asymptotic variance-covariance 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 variance-covariance matrix to estimate the standard errors directly extends to being able to examine likely biases in the expected standard errors 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 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).

Although not specific to the generation of alternative-specific SC experiments, 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.

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 variance-covariance matrix should receive priority in terms of minimization. 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 variance-covariance matrix will be inadequate.

Appendix A

Generic case

Under the assumption of generic parameters, the log-likelihood function looks like the following:

$$
\begin{aligned}
L(b \mid x, y) &= \sum_{n=1}^{N} \sum_{j=1}^{J} y_{jn} \log P_{jn}(x_n \mid b) \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} y_{jn} \log \left[\frac{\exp\left(\sum_{k=1}^{K} x_{jkn} b_k\right)}{\sum_{i=1}^{J} \exp\left(\sum_{k=1}^{K} x_{ikn} b_k\right)} \right] \\
&= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} \sum_{k=1}^{K} x_{jkn} b_k - \log\left[\sum_{i=1}^{J} \exp\left(\sum_{k=1}^{K} x_{ikn} b_k \right) \right] \sum_{j=1}^{J} y_{jn} \right) \\
&= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} \sum_{k=1}^{K} x_{jkn} b_k - \log\left[\sum_{j=1}^{J} \exp\left(\sum_{k=1}^{K} x_{jkn} b_k \right) \right] \right).
\end{aligned}
$$

Note that we assume that respondents have to select exactly one alternative such that $\sum_{j=1}^{J} y_{jn} = 1,$ such that this term vanishes in the log-likelihood function above.

The first derivative of the log-likelihood function (i.e., the score function) is:

$$
\begin{aligned}
\frac{\partial L(b \mid x, y)}{\partial b_{k_1}} &= \sum_{n=1}^{N} \left(\frac{\partial}{\partial b_{k_1}} \left(\sum_{j=1}^{J} y_{jn} \sum_{k=1}^{K} x_{jkn} b_k \right) - \frac{\partial}{\partial b_{k_1}} \log\left[\sum_{j=1}^{J} \exp\left(\sum_{k=1}^{K} x_{jkn} b_k \right) \right] \right) \\
&= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} x_{jk_1 n} - \frac{\sum_{j=1}^{J} \exp\left(\sum_{k=1}^{K} x_{jkn} b_k\right) x_{jk_1 n}}{\sum_{i=1}^{J} \exp\left(\sum_{k=1}^{K} x_{ikn} b_k\right)} \right) \\
&= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{jn} x_{jk_1 n} - \sum_{j=1}^{J} P_{jn}(x_n \mid b) x_{jk_1 n} \right) \\
&= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} \left(y_{jn} - P_{jn}(x_n \mid b) \right) x_{jk_1 n} \right).
\end{aligned}
$$

The $K \times K$ matrix of second derivatives of the log-likelihood function can now be derived.

$$
\begin{aligned}
\frac{\partial^2 L(b\,|x,y)}{\partial b_{k_1}\partial b_{k_2}} &= \sum_{n=1}^{N}\left(\sum_{j=1}^{J} 0 - x_{jk_1n}\frac{\partial}{\partial b_{k_2}}\left(P_{jn}(x_n\,|b)\right)\right)\\
&= -\sum_{n=1}^{N}\sum_{j=1}^{J} x_{jk_1n}\left(\frac{\exp\left(\sum_{k=1}^{K} x_{jkn}b_k\right)x_{jk_2n}\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)\right)^2} - \right.\\
&\qquad\left.\frac{\exp\left(\sum_{k=1}^{K} x_{jkn}b_k\right)\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)x_{ik_2n}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)\right)^2}\right)\\
&= -\sum_{n=1}^{N}\sum_{j=1}^{J} x_{jk_1n}\left(x_{jk_2n}\frac{\exp\left(\sum_{k=1}^{K} x_{jkn}b_k\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)} - \frac{\exp\left(\sum_{k=1}^{K} x_{jkn}b_k\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)}\cdot\frac{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)x_{ik_2n}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K} x_{ikn}b_k\right)}\right)\\
&= -\sum_{n=1}^{N}\sum_{j=1}^{J} x_{jk_1n}\left(x_{jk_2n}P_{jn}(x_n\,|b) - P_{jn}(x_n\,|b)\sum_{i=1}^{J} x_{ik_2n}P_{in}(x_n\,|b)\right)\\
&= -\sum_{n=1}^{N}\sum_{j=1}^{J} x_{jk_1n}P_{jn}(x_n\,|b)\left(x_{jk_2n} - \sum_{i=1}^{J} x_{ik_2n}P_{in}(x_n\,|b)\right).
\end{aligned}
$$

This is the same equation as stated in, for example, Kanninen (2002). See McFadden (1974) for more details on the second derivatives for the generic case.

Alternative-specific case

For the alternative-specific case, the second derivatives are different as we will show here. The log-likelihood function can be stated as

$$
\begin{aligned}
L(b\,|x,y) &= \sum_{n=1}^{N}\sum_{j=1}^{J} y_{jn}\log P_{jn}(x_n\mid b)\\
&= \sum_{n=1}^{N}\sum_{j=1}^{J} y_{jn}\log\left(\frac{\exp\left(\sum_{k=1}^{K_j} x_{jkn}b_{jk}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{jk}\right)}\right)\\
&= \sum_{n=1}^{N}\left(\sum_{j=1}^{J} y_{jn}\sum_{k=1}^{K_j} x_{jkn}b_{jk} - \log\left[\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{jk}\right)\right]\sum_{j=1}^{J} y_{jn}\right)\\
&= \sum_{n=1}^{N}\left(\sum_{j=1}^{J} y_{jn}\sum_{k=1}^{K_j} x_{jkn}b_{jk} - \log\left[\sum_{j=1}^{J}\exp\left(\sum_{k=1}^{K_j} x_{jkn}b_{jk}\right)\right]\right).
\end{aligned}
$$

The first derivative of the log-likelihood function then becomes

$$\frac{\partial L(b\,|x,y)}{\partial b_{j_1k_1}} = \sum_{n=1}^{N}\left(\frac{\partial}{\partial b_{j_1k_1}}\left(\sum_{j=1}^{J} y_{jn}\sum_{k=1}^{K_j} x_{jkn}b_{jk}\right) - \frac{\partial}{\partial b_{j_1k_1}}\log\left[\sum_{j=1}^{J}\exp\left(\sum_{k=1}^{K_j} x_{jkn}b_{jk}\right)\right]\right)$$
$$= \sum_{n=1}^{N}\left(y_{j_1n}x_{j_1k_1n} - \frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_k\right)x_{j_1k_1n}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_k\right)}\right)$$
$$= \sum_{n=1}^{N}\left(y_{j_1n} - P_{j_1n}(x_n\,|b\,)\right)x_{j_1k_1n}.$$

Finally, the $\sum_j K_j \times \sum_j K_j$ matrix of second derivatives can be computed by taking the derivatives of the above score function. We consider two cases: the case in which $j_1 = j_2$ and the case in which $j_1 \neq j_2$.

If $j_1 = j_2$:

$$\frac{\partial^2 L(b\,|x,y)}{\partial b_{j_1k_1}\partial b_{j_1k_2}} = \sum_{n=1}^{N}\left(0 - x_{j_1k_1n}\frac{\partial}{\partial b_{j_1k_2}}\left(P_{j_1n}(x_n\,|b\,)\right)\right)$$
$$= -\sum_{n=1}^{N} x_{j_1k_1n}\left(\frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)x_{j_2k_2n}\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)\right)^2} - \frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)\exp\left(\sum_{k=1}^{K_{j_2}} x_{j_2kn}b_{j_2k}\right)x_{j_2k_2n}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)\right)^2}\right)$$
$$= -\sum_{n=1}^{N} x_{j_1k_1n}\left(\frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)x_{j_2k_2n}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)} - \frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)}\cdot\frac{\exp\left(\sum_{k=1}^{K_{j_2}} x_{j_2kn}b_{j_2k}\right)x_{j_2k_2n}}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)}\right)$$
$$= -\sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n\,|b\,)\left(1 - P_{j_2n}(x_n\,|b\,)\right).$$

If $j_1 \neq j_2$:

$$\frac{\partial^2 L(b\,|x,y)}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = \sum_{n=1}^{N}\left(0 - x_{j_1k_1n}\frac{\partial}{\partial b_{j_2k_2}}\left(P_{j_1n}(x_n\,|b)\right)\right)$$

$$= -\sum_{n=1}^{N} x_{j_1k_1n}\left(\frac{0-\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)\exp\left(\sum_{k=1}^{K_{j_2}} x_{j_2kn}b_{j_2k}\right)x_{j_2k_2n}}{\left(\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)\right)^2}\right)$$

$$= \sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}\frac{\exp\left(\sum_{k=1}^{K_{j_1}} x_{j_1kn}b_{j_1k}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)}\cdot\frac{\exp\left(\sum_{k=1}^{K_{j_2}} x_{j_2kn}b_{j_2k}\right)}{\sum_{i=1}^{J}\exp\left(\sum_{k=1}^{K_i} x_{ikn}b_{ik}\right)}$$

$$= \sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n\,|b)P_{j_2n}(x_n\,|b).$$

Hence, the second derivatives of the log-likelihood function for the alternative-specific case can be summarizes as

$$\frac{\partial^2 L}{\partial b_{j_1k_1}\partial b_{j_2k_2}} = \begin{cases} -\sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n\,|b)\left(1-P_{j_2n}(x_n\,|b)\right), & \text{if } j_1 = j_2, \\ \sum_{n=1}^{N} x_{j_1k_1n}x_{j_2k_2n}P_{j_1n}(x_n\,|b)P_{j_2n}(x_n\,|b), & \text{if } j_1 \neq j_2. \end{cases}$$

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, Moshe and Steven R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press.

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

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

Burke, Raymond R., Bari A. Harlam, Barbara E. Kahn, and Leonard M. Lodish (1992) Comparing Dynamic Consumer Choice in Real and Computer-Simulated Environments, *Journal of Consumer Research*, 19 (June), 71–82.

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

Carson, Richard T., Jordan J. Louviere, Don Anderson, Phipps Arabie, David Bunch, David A. Hensher, Richard M. Johnson, Warren F. Kuhfeld, Dan Steinberg, Joffre D. Swait, Harry Timmermans, and James B. Wiley (1994) Experimental Analysis of Choice, *Marketing Letters*, 5 (October), 351-367.

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

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

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

Kuhfeld, Warren F., Randall D. Tobias, and Mark Garratt, (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, Jordan J., David A. Hensher, and Joffre D. Swait (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge.

McFadden, Dan (1974) Conditional Logit Analysis of Qualitative Choice Behaviour. In Zarembka, P. (ed.), *Frontiers of Econometrics*, Academic Press, New York, 105-142.

Mellers, Barbara A. and Alan D.J. Cooke, (1994) Tradeoffs Depend on Attribute Range, *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1055-1067.

Ohler, Tobias, Aihong Le, and Jordan Louviere (2000) Attribute Range Effects in Binary Response Tasks, *Marketing Letters*, 11(3), 249-260.

Sándor, Zsolt and Michel Wedel (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, 38 (4), 430-444.

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

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

Verlegh, Peter W.J, Hendrik N.J. Schifferstein and Dick R. Wittnick (2002) Range and Number-of-Levels Effects in Derived and Stated Measures of Attribute Importance, *Marketing Letters*, 31(1), 41-52.