I T L S

WORKING PAPER
ITLS-WP-05-13

Sample optimality in the design of stated choice experiments

By

John M Rose & Michiel CJ Bliemer*

July 2005

*Faculty of Civil Engineering and Geosciences
Delft University of Technology

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-05-13

TITLE: **Sample optimality in the design of stated choice experiments**

ABSTRACT: Stated choice (SC) experiments represent the dominant data paradigm in the study of behavioral responses of individuals, households as well as other organizations, yet little is known about the sample size requirements for models estimated from such data. Current sampling theory does not adequately address the issue and hence researchers have had to resort to simple rules of thumb or ignore the issue and collect samples of arbitrary size, hoping that the sample is sufficiently large enough to produce reliable parameter estimates. In this paper, we demonstrate how to generate efficient designs (based on D-efficiency and a newly proposed sample size S-efficiency measure) using prior parameter values to estimate multinomial logit models containing both generic and alternative-specific parameters. Sample size requirements for such designs in SC studies are investigated. Using a numerical case study, we show that using S-efficiency can substantially reduce the sample size required of SC studies.

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 **(C37)**
An Australian Key Centre
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: July 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 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 2003; 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, et al. (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 (co)variance matrix. By manipulating the attribute levels of the alternatives, for fixed parameter values, the analyst is able to minimize the elements within the (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.

A number of different efficiency criteria have been proposed within the literature. Initially, the preferred criterion was A-error, which attempts to minimize the trace of the asymptotic (co)variance matrix. Given that the trace of an asymptotic (co)variance

matrix only takes into account the variances and ignores the covariances, the literature soon turned towards the use of D-error as a measure of efficiency. The D-error measure of the asymptotic (co)variance matrix utilizes the determinant of the matrix, which summarizes all the elements of the matrix, including the covariances. The determinant of a matrix, however, is a complex calculation, involving the multiplication and subtraction of various elements of the matrix. Given the complexity calculations used in determining the determinant of the asymptotic (co)variance matrix for SC designs, it is not uncommon in attempting to minimize the D-error of a design, to minimize some elements of the asymptotic (co)variance matrix at the expense of others. This often results in an uneven treatment of the asymptotic standard errors in arriving at a final 'efficient' design. That is, for the final design solution, some parameters may be more reliable than others.

In this paper, we demonstrate an alternative efficiency criterion, which we call S-efficiency. Given that the asymptotic (co)variance matrix is directly scalable to the size of the sample contained within a data set, it is possible to determine what values the elements contained within the asymptotic (co)variance matrix will take at any sample size. Using this property, combined with the fact that the asymptotic *t*-ratios of the parameter estimates are simply equal to the parameters themselves divided by the square roots of their related variances, we are able to determine what sample size would theoretically be required for each parameter to be observed to be statistically significant. Given that the sample size requirement for statistical significance can be calculated for each parameter, it becomes possible to generate designs that minimize the sample size required for all parameters to be statistically significant, rather than attempt to globally maximize all asymptotic *t*-ratios using a criterion, which may minimize the sample required for only some of the parameters. Using the S-efficiency measure, we show how substantial gains may be obtained in the sample size requirements of SC designs.

The remainder of the paper is arranged as follows. In section 2, we outline derive the log-likelihood function for the MNL model. So as to dispel some of the misconceptions that are widely held within the literature, we go onto derive the asymptotic (co)variance matrix of the MNL model using both matrix algebra as well as the second derivatives of the log-likelihood function. We show in this section why the use of matrix algebra, which corresponds with the current state of practice, is limited to models estimated with generic parameters only. Section 3 outlines the concepts of A- and D-error and introduces the concept of S-efficiency (Sample-efficiency). Section 4 provides a numerical example of efficient designs, in which we demonstrate the potential gains that may be achieved from using the S-efficient measure when compared with designs generated using the D-error measure. The substantive implications of these comparisons are then set out followed by some conclusions and directions for ongoing research.

2. A tail of two approaches: The MNL model

It is possible to derive the asymptotic (co)variance matrix of the MNL model, necessary to optimize the statistical efficiency of SC experiments, in one of two ways. To date, the literature has tended towards using matrix algebra to derive the asymptotic (co)variance matrix of the MNL model (see for example, and Zwerina 1996, Sándor and Wedel 2001 and Kanninen 2002). An alternative approach is to use the second derivatives of the log-

likelihood function. As we show in this section, the first method will only allow for the derivation of the asymptotic (co)variance matrix for the MNL model assuming generic parameters only. Use of the second derivatives of the log-likelihood function, however, allows for both generic and alternative-specific parameters in the design. To demonstrate, we will derive asymptotic (co)variance matrix of the MNL model utilizing both methods.

The MNL model was first fully derived by McFadden (1974) based on random utility theory (RUT). 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:

$$U_{js} = V_{js} + \varepsilon_{js}, \qquad \textbf{(1)}$$

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 ε_{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:

$$V_{js} = \sum_{k=1}^{K^*} \beta_k^* x_{jks}^* + \sum_{k=1}^{K_j} \beta_{jk} x_{jks}, \quad \forall j = 1,\ldots,J, \forall s = 1,\ldots,S. \qquad \textbf{(2)}$$

The generic and alternative-specific parameters are denoted by β_k^* and β_{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, ε_{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*:

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

Most commonly used to determine the parameters (β^*, β) 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:

$$L(\beta^*, \beta) = \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \log P_{js} \quad \textbf{(4)}$$

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 probability P_{js} will depend on the values of x^*_{Jks} and x_{jks} as well as upon (β^*, β).

In section 2.1, we derive the asymptotic (co)variance matrix for the MNL model using matrix algebra. We show using a numerical example, that the use of matrix algebra that using this method has implications as to the estimation of alternative-specific parameters.

2.1 Deriving asymptotic (co)variance matrix of the MNL using matrix algebra

Maximising equation (4) will yield the maximum likelihood estimator, $(\hat{\beta}^*, \hat{\beta})$, for the choice model, for given values of x^*_{Jks} and x_{jks}. McFadden (1974) showed that the distribution of $(\hat{\beta}^*, \hat{\beta})$ is asymptotically normal with a mean (β^*, β) and (co)variance

$$\Omega^{-1} = (Z'PZ)^{-1} = \left[\sum_{s=1}^{S} \sum_{J=1}^{J} Z'_{js} P_{js} Z_{js} \right]^{-1}, \quad \textbf{(5)}$$

where P is an $M \times M$ diagonal matrix of choice probabilities with elements P_{jn}, and Z is an $M \times K$ matrix with rows, such that

$$Z_{js} = x_{js} - \sum_{i=1}^{J_s} x_{is} P_{is}. \quad \textbf{(6)}$$

Taking a numerical example, it is possible to show that use of equations (5) and (6) will only hold if the analyst is prepared to assume that all parameters of the model are generic. Consider an experiment involving two alternatives, each with three attributes defined by two levels each. The utility functions for the experiment may be represented as follows.

$$U_1 = \beta_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13}$$
$$U_2 = \beta_{21} x_{21} + \beta_{22} x_{22} + \beta_{23} x_{23}.$$

Let the priors for the experiment be alternative specific, as shown in Table 11 and let β_{10} = 0.1. Table 1 also shows a design and resulting Z matrix obtained for the design, derived by using equation (1).

Table 1: Efficient 'alternative-specific' design using current design methods

Situations		X Matrix			Parameters			Probs	Z Matrix		
Situation	**alternative**	**A**	**B**	**C**	β_{j1}	β_{j2}	β_{j3}	P_{jn}	Z_A	Z_B	Z_C
1	1	-1	-1	-1	0.3	0.4	0.5	0.27	0	0	-1.46212
1	2	-1	-1	1	0.2	0.3	0.4	0.73	0	0	0.537883
2	1	-1	1	-1	0.3	0.4	0.5	0.40	-1.29131	0	0
2	2	1	1	-1	0.2	0.3	0.4	0.60	0.708687	0	0
3	1	1	1	1	0.3	0.4	0.5	0.69	0.708687	0	0
3	2	-1	1	1	0.2	0.3	0.4	0.31	-1.29131	0	0
4	1	1	-1	1	0.3	0.4	0.5	0.73	0	0	0.537883
4	2	1	-1	-1	0.2	0.3	0.4	0.27	0	0	-1.46212
5	1	-1	1	1	0.3	0.4	0.5	0.60	-0.90033	0.900332	0
5	2	1	-1	1	0.2	0.3	0.4	0.40	1.099668	-1.09967	0
6	1	-1	-1	1	0.3	0.4	0.5	0.55	0	-0.90033	0.900332
6	2	-1	1	-1	0.2	0.3	0.4	0.45	0	1.099668	-1.09967
7	1	1	-1	-1	0.3	0.4	0.5	0.20	0	-1.7163	-1.7163
7	2	1	1	1	0.2	0.3	0.4	0.80	0	0.283702	0.283702
8	1	1	1	-1	0.3	0.4	0.5	0.77	0.395632	0.395632	0
8	2	-1	-1	-1	0.2	0.3	0.4	0.23	-1.60437	-1.60437	0

The resulting (co)variance matrix obtained from $Z'PZ$ is shown as equation (7).

$$\Omega^{-1} = (Z'PZ)^{-1} = \begin{bmatrix} 0.29 & 0.03 & 0.01 \\ 0.03 & 0.34 & 0.06 \\ 0.01 & 0.06 & 0.34 \end{bmatrix} \quad \textbf{(7)}$$

Note that the (co)variance matrix represented by equation (7) is a (3×3) matrix, when in fact it should be a (7×7) matrix if each 'alternative-specific' parameter (including the alternative-specific constant) is to have its own row and column in the matrix, as it should if the model were truly alternatively specific. Whilst we have allowed for alternative-specific parameter estimates in arriving at the choice probabilities, the asymptotic (co)variance matrix that we arrive at using matrix algebra treats the parameter estimates as generic. As such, whilst it is possible to attempt to optimize the design based on (Z), doing so will not minimize the elements of the asymptotic (co)variance matrix for each 'alternative-specific' parameter to be estimated (see for example, Carlsson and Martinsson (2003) who use matrix algebra to generate alternative-specific efficient designs).

2.2 Deriving asymptotic (co)variance matrix using the second derivatives of the log-likelihood function

The asymptotic (co)variance matrix of the MNL model may also be derived by taking the second derivatives of the log-likelihood function. Allowing for both alternative-specific and generic parameters, this leads to the following:

$$\frac{\partial^2 L(\beta^*,\beta)}{\partial\beta^*_{k_1}\partial\beta^*_{k_2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J} x^*_{jk_1s}P_{js}\left(x^*_{jk_2s} - \sum_{i=1}^{J} P_{is}x^*_{ik_2s}\right), \quad \forall k_1,k_2 = 1,\ldots,K^*, \tag{8a}$$

$$\frac{\partial^2 L(\beta^*,\beta)}{\partial\beta_{j_1k_1}\partial\beta^*_{k_2}} = -\sum_{s=1}^{S} x_{j_1k_1s}P_{j_1s}\left(x^*_{j_1k_2s} - \sum_{i=1}^{J} x^*_{ik_2s}P_{is}\right), \quad \forall j_1 = 1,\ldots,J, k_1 = 1,\ldots,K_{j_1}, k_2 = 1,\ldots,K^*, \tag{8b}$$

$$\frac{\partial^2 L(\beta^*,\beta)}{\partial\beta_{j_1k_1}\partial\beta_{j_2k_2}} = \begin{cases} \sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}P_{j_1s}P_{j_2s}, & \text{if } j_1 \neq j_2; \\ -\sum_{s=1}^{S} x_{j_1k_1s}x_{j_2k_2s}P_{j_1s}\left(1-P_{j_2s}\right), & \text{if } j_1 = j_2. \end{cases} \quad \forall j_i = 1,\ldots,J, k_i = 1,\ldots,K_{j_i}. \tag{8c}$$

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 (8a,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 $(\hat{\beta}^*,\hat{\beta})$, then

$$(\hat{\beta}^*,\hat{\beta}) = \arg\max_{(\beta^*,\beta)} L(\beta^*,\beta). \tag{9}$$

Suppose that the true parameter values are $(\bar{\beta}^*,\bar{\beta})$. The ML estimates $\hat{\beta}^*$ are asymptotically normally distributed with mean $\bar{\beta}^*$ and (co)variance matrix, Ω, which is equal to the negative inverse of the Fisher information matrix (McFadden 1974). 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

$$I(\hat{\beta}^*,\hat{\beta}) = M \cdot \frac{\partial^2 L(\bar{\beta}^*,\bar{\beta})}{\partial\beta\partial\beta'}. \tag{10}$$

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

$$\Omega = -\left[I(\hat{\beta}^*,\hat{\beta})\right]^{-1} = -\frac{1}{M}\left[\frac{\partial^2 L(\bar{\beta}^*,\bar{\beta})}{\partial\beta\partial\beta'}\right]^{-1}. \tag{11}$$

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,

$$(\hat{\beta}^*, \hat{\beta}) \rightarrow N\left((\overline{\beta}^*, \overline{\beta}), -\frac{1}{M}\left[\frac{\partial^2 L(\overline{\beta}^*, \overline{\beta})}{\partial\beta\partial\beta'}\right]^{-1}\right). \quad \textbf{(12)}$$

3. Measuring Statistical Efficiency in SC Experimental Designs: Statistical Efficient 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 two most commonly used measures found within the literature are those of A-error and D-error.

$$\text{A-error} = \left(\text{trace}\,\Omega\right)^{(\sum K + \sum_j K_j)} = -\frac{1}{M}\left(\text{trace}\left(\frac{\partial L^2}{\partial\beta\partial\beta'}\right)\right)^{(\sum K + \sum_j K_j)} \quad \textbf{(13)}$$

$$\text{D-error} = \left(\det\Omega\right)^{1/(\sum K + \sum_j K_j)} = -\frac{1}{M}\left(\det\left(\frac{\partial L^2}{\partial\beta\partial\beta'}\right)\right)^{-1/(\sum K + \sum_j K_j)}. \quad \textbf{(14)}$$

where it is usual to assume $M = 1$[1].

The A-error is computed by taking the trace of the asymptotic (co)variance matrix, whilst the D-error is calculated by taking the determinant, with both scaled to take into account the number of parameters to be estimated. The trace of a matrix is calculated as the sum of the diagonals of that matrix. As such, minimizing the trace of the asymptotic (co)variance matrix will minimize the variances (standard errors) of the associated parameter estimates, without consideration being given to the covariances. Given that the trace is calculated as the sum of the diagonal elements, if one of these elements is large in magnitude, then that element will tend to dominate the calculation. For this reason, the A-error measure has fallen out of favor. The D-error computation is a little more complicated as the determinant of a matrix is calculated as a series of multiplications and subtractions over all the elements of the matrix (see for example, Kanninen 2002). As such, the determinant (and by implication, the D-error measure) summarizes all the elements of the matrix in a single 'global' value. Thus, whilst attempts to minimize the D-error measure, on average, minimize all the elements within the matrix, it is possible that in doing so, some elements (variances and/or covariances) may in fact become larger. Despite this property, the D-error measure has become the most common measure of statistical efficiency within the literature.

In the past, the literature has considered two different approaches for computing the D-error of a design, both related to the priors assumed in generating the design. The first approach results in what has been termed the D_z-error measure and is used when the

[1] The assumption of a single respondent is not inconsistent with the MNL model which assumes that all respondents act behaviourally in a similar fashion.

choice analyst has no prior information on the true parameter values (including sign). In such cases as when there exists absolutely no information on the likely parameter estimates, the approach adopted in the past has been to assume that all the β are simultaneously equal to zero. . Hence, the D_z-error can be computed as:

$$D_z\text{-error} = \left(\det I(0 \mid x)\right)^{-1/(\sum K + \sum_j K_j)}. \quad \textbf{(15)}$$

In contrast, if information about the parameter estimates is available in advance (whether from pilot studies, other research, etc.), then this information can be used as the priors for β used to compute the D-error. In such instances, the D-error measure has been termed D_p-error. The D_p-error assuming knowledge of prior parameter estimates, $\tilde{\beta}$, can be computed as

$$D_p\text{-error} = \left(\det I(\tilde{\beta} \mid x)\right)^{-1/(\sum K + \sum_j K_j)}. \quad \textbf{(16)}$$

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.

The presence of M in equations (11) and (12) allows for an alternative optimization strategy directly linked to the sample size. 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 $se_M(\hat{\beta}_k^*)$ and $se_M(\hat{\beta}_{jk})$ for each of the generic and alternative-specific parameters. Then it holds that:

$$se_M(\hat{\beta}_k^*) = se_1(\hat{\beta}_k^*)/\sqrt{M}, \quad \text{and} \quad se_M(\hat{\beta}_{jk}) = se_1(\hat{\beta}_{jk})/\sqrt{M}. \quad \textbf{(17)}$$

Equation (15) 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 (equation (18) for the case of alternative-specific parameters), it is possible to determine what sample size will be required in order to demonstrate statistical significance for each of the parameter estimates.

$$t_{jk} = \frac{\hat{\beta}_{jk}}{\sqrt{\frac{\left(se_M(\beta_{jk})\right)^2}{M}}}. \quad \textbf{(18)}$$

Rearranging equation (18) yields:

$$M = \frac{t_{jk}^2\left(se_M(\beta_{jk})\right)}{\beta_{jk}^2}. \quad \textbf{(19)}$$

A similar equation holds for the generic parameters case. We can view the sample size requirement stated in equation (19) as a theoretical lower bound for finding a statistically significant parameter estimate for that parameter. Different parameters may have different lower bounds. Parameters with high lower bounds will be more difficult to estimate than parameters with low lower bounds. In case we would like to find the minimum theoretical sample size for which all parameters are statistically significant, then we would probably prefer to change the design in such a way that the parameters that are difficult to estimate obtain more information from the design in order to decrease its standard error. In other words, we may prefer to have all parameter estimates in the same range with their asymptotic *t*-values such that all parameters get equal attention in the design. We term a design that minimizes the sample size needed for all parameters to be statistically significant an S-efficient design.

Manipulation of the attribute levels of the alternatives within a design will result in different D-error (D_z or D_p) and S-efficiency values, assuming fixed prior parameter estimates. Over a number of iterations, it may be possible to locate designs with lower D-error and S-efficiency values. Methods of manipulating the attribute levels so as to generate and locate D-efficient and S-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.

One important point of clarification is necessary before we move on. The objective of deriving statistically efficient SC experiments is to minimize the asymptotic standard errors, and hence maximise the asymptotic *t*-ratios, of models estimated from data collected using the efficient design. Whilst this point seems obvious, it is a point that has sometimes been missed by the literature. It is extremely important that one should always use the attribute level values that are to be used in estimating the final choice model (i.e., the values that will be used in the data) rather than codes that are to be used exclusively in generating the design. To demonstrate why, consider the experiment discussed in section 2.1., the utility specifications of which are reproduced below.

$$U_1 = 0.1 + 0.3x_{11} + 0.4x_{12} + 0.5x_{13}$$
$$U_2 = 0.2x_{21} + 0.3x_{22} + 0.4x_{23}.$$

Taking the first choice situation in Table 1, and using the same coding structure, the choice probabilities for alternatives 1 and 2 respectively are 0.27 and 0.73. However, assuming that in the final data set we let -1 equal five and 1 equal 10, then the choice probabilities for the first choice situation will become 0.08 and 0.92. Clearly, the

resulting (co)variance matrix used in optimizing the design will be different to that obtained when estimating models based on the data. Given that we wish to minimize the asymptotic standard errors for models estimated from colleted data and not from the actual design itself, it stands to reason that what is of primary importance is the efficiency of the data. Whilst this may seem logical, it is a point that has sometimes been missed within the literature. For example, Huber and Zwerina (1996) use design codes to establish the choice probabilities for each of the alternatives of their design. Using these choice probabilities, they then go onto generate an efficient design after effects coding the attributes. This design will likely be inefficient when it comes to data analysis as the choice probabilities should have been derived using the effects coded variables.

Taking the above into account, it is not necessarily clear what is meant by those who call for the generation of tables of statistically optimal designs (for example, Viney et al. 2005) similar to those produced for orthogonal designs (for example, Hahn and Shapiro 1966; Lazari and Anderson 1994). Whilst it may be possible to do so for designs for data that will adhere to strict coding guidelines (such as for data that will use effects or dummy coding only), there may exist an infinite number of possible applications using different continuous attributes with different attribute levels, making it impossible to produce such tables for all but a small subset of problems. It is therefore more likely that practitioners will have to generate statistically efficient designs as required.

4. A Numerical comparison of Efficiency Measures

In order to illustrate the theory of efficient designs using both D-error and S-efficiency, we will consider the following discrete choice problem. Suppose there are two alternatives, each having several generic and alternative-specific attributes. Assume that all attributes have three attribute levels. For this experiment, the first two attributes are to be treated as generic across both alternatives. The third attribute of both alternatives will be assigned an alternative specific parameter, whilst the forth will be effects coded and also allow for alternative specific parameters. For both alternatives, we will also allow for the generation of an interaction effect between the second and third attributes of the design. The second alternative will also be assigned an alternative-specific constant. The two utility functions are therefore:

$$U_1 = \beta_1^* x_{11} + \beta_2^* x_{12} + \beta_{13} x_{13} + \beta_{14}^1 x_{14}^1 + \beta_{14}^2 x_{14}^2 + \beta_{15} x_{12} x_{13}$$
$$U_2 = \beta_{20} + \beta_1^* x_{21} + \beta_2^* x_{22} + \beta_{23} x_{23} + \beta_{24}^1 x_{24}^1 + \beta_{24}^2 x_{24}^2 + \beta_{25} x_{22} x_{23}.$$

In total, there are 11 parameters to estimate, whilst there are eight attributes that change attribute levels (i.e., $x_{11}, x_{12}, x_{13}, x_{14}, x_{21}, x_{22}, x_{23}$ and x_{24}. The values of x_{14}^1 and x_{14}^2 will depend on changes of x_{14}. Likewise, x_{24}^1 and x_{24}^2 will depend on changes of x_{24}.). The constant, β_{21}, has a fixed attribute level of one. Within the SC experiment, the eight attributes can take on different levels over the different choice situations shown to respondents. Assume that each respondent observes twelve choice situations and let the attributes take on the following levels: $L_{11}^* = L_{21}^* = L_{22} = \{2, 4, 6\}$,

$L_{12}^{*} = L_{22}^{*} = L_{21} = \{1,3,5\}$, and $L_{12} = L_{23} = \{4,6,8\}$. Following common practice, we constrain ourselves to designs which are balanced in the attribute levels observed over choice situations (although such a constraint may result in the generation of a sub-optimal design).

In generating an efficient design, it is necessary to assume a set of prior parameter estimates. For the numerical example used within this paper, the priors were

$$U_1 = \quad 0.8x_{11} + 0.4x_{12} + 0.7x_{13} - 0.2x_{14}^1 + 0.5x_{14}^2 + 0.3x_{12}x_{13}$$
$$U_2 = 1 + 0.8x_{21} + 0.4x_{22} + 0.8x_{23} - 0.3x_{24}^1 + 0.4x_{24}^2 + 0.4x_{22}x_{23}.$$

We generate three different design types: (a) a D_p-efficient design, (b) an orthogonal design, and (c) an S-efficient design. For any given SC experiment, it is possible to generate multiple orthogonal designs. In line with practice, the orthogonal design we report here was randomly selected. The three designs are presented in Table 2. Although we show the interaction columns of the designs in Table 2, it is important to note that in generating the D_p-efficient and S-efficient designs, the interactions were directly computed from the relevant attributes and hence were not directly manipulated as part of the designs.

As is to be expected, the D_p-efficient design produces the lowest D_p-error measure, however, this design produces the highest D_z-error. The randomly selected orthogonal design produces the worst D_p-error measure but the lowest D_z-error. The S-optimal design produces D_z and D_p-error values in between. Table 2 also reports the minimum sample size required for each of the three designs at which all attribute related parameters would likely be statistically significant (i.e., the *t*-ratios for each attribute related parameter are greater than or equal to 1.96). In calculating these minimum sample sizes, we ignore the constant. The constant is typically ignored in SC studies, given that typically the constant is of less importance to the researcher (indeed the constant is often considered meaningless in SC experiments as it is based on the choice shares over the hypothetical situations, *S*). Further, in many SC studies, it is often the ratios of two parameter values (e.g., to derive willingness to pay) that is of primary importance. The constant is not eliminated whilst obtaining the asymptotic (co)variance matrix however (although it is ignored in obtaining the S-efficient design).

Despite producing a higher D_p-error than the D_p-efficient design, the S-efficient design would theoretically require a substantially smaller sample size in order for all design related parameters to be found to be statistically significant. Indeed, for the numerical example explored within this paper, the S-efficient design would theoretically require a sample of 108 respondents compared with a minimum sample size of 137 respondents for the D_p-efficient design (representing a reduction of 21.17 percent in sample size) before all parameters were likely to be statistically significant. The orthogonal design found for this example performs extremely badly in terms of sample size requirements, requiring a sample size of 188,469 before all parameters would be expected to be statistically significant.

Table 2: SC experimental designs

Design 1: D_p-efficient design

S	x_{11}	x_{12}	x_{13}	x_{14}^1	x_{14}^2	$x_{12}x_{13}$	x_{21}	x_{22}	x_{23}	x_{24}^1	x_{24}^2	$x_{22}x_{23}$	**P1**	**P2**
1	2	1	8	1	0	8	2	1	6	-1	-1	6	0.43	0.57
2	6	1	4	1	0	4	6	3	2	1	0	6	0.15	0.85
3	6	5	4	-1	-1	20	2	5	4	0	1	20	0.29	0.71
4	6	3	8	0	1	24	6	3	6	1	0	18	0.65	0.35
5	4	3	4	0	1	12	2	5	2	-1	-1	10	0.77	0.23
6	4	5	6	1	0	30	4	3	6	1	0	18	0.75	0.25
7	2	3	6	-1	-1	18	4	1	6	0	1	6	0.48	0.52
8	4	1	8	-1	-1	8	2	5	2	-1	-1	10	0.77	0.23
9	6	5	8	0	1	40	6	3	4	0	1	12	1.00	0.00
10	4	5	6	0	1	30	6	5	4	-1	-1	20	0.50	0.50
11	2	1	6	1	0	6	4	1	2	0	1	2	0.60	0.40
12	2	3	4	-1	-1	12	4	1	4	1	0	4	0.45	0.55

D_Z-error = 0.2249 D_p-error = 0.3494 Min M = 137

Design 2: Orthogonal Design

S	x_{11}	x_{12}	x_{13}	x_{14}^1	x_{14}^2	$x_{12}x_{13}$	x_{21}	x_{22}	x_{23}	x_{24}^1	x_{24}^2	$x_{22}x_{23}$	**P1**	**P2**
1	4	5	4	-1	-1	20	6	1	4	-1	-1	4	0.94	0.06
2	4	3	6	1	0	18	6	5	2	1	0	10	0.67	0.33
3	2	1	8	-1	-1	8	6	1	4	1	0	4	0.27	0.73
4	4	3	6	0	1	18	2	3	6	1	0	18	0.27	0.73
5	6	1	4	-1	-1	4	4	5	2	0	1	10	0.04	0.96
6	6	1	4	1	0	4	4	1	6	0	1	6	0.04	0.96
7	4	1	8	0	1	8	2	3	4	-1	-1	12	0.60	0.40
8	2	3	6	0	1	18	2	3	2	-1	-1	6	0.99	0.01
9	2	3	6	1	0	18	6	5	6	-1	-1	30	0.00	1.00
10	6	5	8	-1	-1	40	4	5	6	0	1	30	0.67	0.33
11	6	5	8	1	0	40	4	1	2	0	1	2	1.00	0.00
12	2	5	4	0	1	20	2	3	4	1	0	12	0.80	0.20

D_Z-error = 0.2085 D_p-error = 1.857 Min M = 188,469

Design 3: S-optimal Design

S	x_{11}	x_{12}	x_{13}	x_{14}^1	x_{14}^2	$x_{12}x_{13}$	x_{21}	x_{22}	x_{23}	x_{24}^1	x_{24}^2	$x_{22}x_{23}$	**P1**	**P2**
1	2	1	8	1	0	8	2	1	6	-1	-1	6	0.43	0.57
2	6	1	4	1	0	4	4	3	2	1	0	6	0.48	0.52
3	6	5	4	0	1	20	2	5	4	0	1	20	0.48	0.52
4	6	3	8	0	1	24	6	3	6	1	0	18	0.65	0.35
5	4	3	4	0	1	12	2	5	2	-1	-1	10	0.77	0.23
6	4	5	6	1	0	30	4	3	6	1	0	18	0.75	0.25
7	2	3	6	-1	-1	18	4	1	6	0	1	6	0.48	0.52
8	2	1	8	-1	-1	8	2	3	2	-1	-1	6	0.88	0.12
9	6	5	8	0	1	40	6	5	4	0	1	20	1.00	0.00
10	4	5	6	-1	-1	30	6	5	4	-1	-1	20	0.31	0.69
11	4	1	6	1	0	6	6	1	2	0	1	2	0.60	0.40
12	2	3	4	-1	-1	12	4	1	4	1	0	4	0.45	0.55

D_Z-error = 0.2505 D_p-error =0.3575 Min M = 217

Table 3 shows the asymptotic (co)variance matrices of the three designs. As would be expected, the D_p-efficient design generally produces lower asymptotic standard errors than the two other designs. Examination of the asymptotic (co)variance matrices reveals, however, that despite being more efficient overall, the D_p-efficient design will not necessarily produce lower asymptotic standard errors for all parameters. As a consequence of this, whilst overall the asymptotic *t*-ratios of the D_p-efficient design will generally be higher than those of the S-optimal (and orthogonal) design, the use of a 'global' measure of efficiency allows for the possibility that some elements of the matrix being measured will be sub-optimal in terms of efficiency. For the numerical example shown here, this has significant implications for sample size requirements of the design. In particular, we observe the asymptotic standard error for β_{14}^1 to be 1.42 for the D_p-efficient design but only 1.12 for the S-efficient design, and given the priors assumed, it is this attribute which determines the theoretical minimum sample size required for the design.

Table 3: Asymptotic (co)variance matrices of SC experimental designs

Design 1: D_p-efficient design											
	β_1^*	β_2^*	β_{13}	β_{14}^1	β_{14}^2	β_{15}	β_{20}	β_{23}	β_{24}^1	β_{24}^2	β_{25}
β_1^*	**0.83**	0.56	0.46	-0.12	0.48	0.20	-0.11	0.59	-0.20	0.28	0.31
β_2^*	0.56	**2.62**	1.18	-0.77	0.82	-0.26	-1.17	1.63	-0.01	1.00	-0.27
β_{13}	0.46	1.18	**0.90**	-0.41	0.38	-0.02	0.79	1.00	-0.22	0.40	0.00
β_{14}^1	-0.12	-0.77	-0.41	**1.42**	-1.20	0.14	0.50	-0.36	0.14	-0.14	0.09
β_{14}^2	0.48	0.82	0.38	-1.20	**2.29**	-0.07	-1.35	0.33	-0.08	-0.25	0.08
β_{15}	0.20	-0.26	-0.02	0.14	-0.07	**0.15**	0.27	-0.01	-0.06	-0.04	0.19
β_{20}	-0.11	-1.17	0.79	0.50	-1.35	0.27	**9.73**	-0.46	-0.28	-0.67	0.16
β_{23}	0.59	1.63	1.00	-0.36	0.33	-0.01	-0.46	**1.50**	-0.23	0.70	0.00
β_{24}^1	-0.20	-0.01	-0.22	0.14	-0.08	-0.06	-0.28	-0.23	**1.06**	-0.49	-0.10
β_{24}^2	0.28	1.00	0.40	-0.14	-0.25	-0.04	-0.67	0.70	-0.49	**1.61**	-0.04
β_{25}	0.31	-0.27	0.00	0.09	0.08	0.19	0.16	0.00	-0.10	-0.04	**0.26**
Design 2: Orthogonal Design											
	β_1^*	β_2^*	β_{13}	β_{14}^1	β_{14}^2	β_{15}	β_{20}	β_{23}	β_{24}^1	β_{24}^2	β_{25}
β_1^*	**2771.71**	1717.23	1696.61	1361.45	-3317.58	-6.78	-768.11	1866.09	-3496.94	1452.41	297.34
β_2^*	1717.23	**1104.65**	1058.51	886.78	-2102.60	-15.24	-555.89	1188.67	-2174.53	914.29	167.27
β_{13}	1696.61	1058.51	**1041.07**	840.59	-2036.19	-5.88	-479.77	1148.24	-2140.83	887.71	179.47
β_{14}^1	1361.45	886.78	840.59	**719.38**	-1684.49	-15.27	-459.93	949.90	-1727.01	733.21	127.63
β_{14}^2	-3317.58	-2102.60	-2036.19	-1684.49	**4038.44**	21.81	1019.05	-2270.15	4199.50	-1770.35	-334.46
β_{15}	-6.78	-15.24	-5.88	-15.27	21.81	**3.11**	24.59	-13.38	11.12	-8.53	4.04
β_{20}	-768.11	-555.89	-479.77	-459.93	1019.05	24.59	**407.95**	-583.98	986.36	-438.10	-47.25
β_{23}	1866.09	1188.67	1148.24	949.90	-2270.15	-13.38	-583.98	**1283.37**	-2360.59	988.31	186.61

β_{24}^1	-3496.94	-2174.53	-2140.83	-1727.01	4199.50	11.12	986.36	-2360.59	**4417.23**	-1842.47	-370.97
β_{24}^2	1452.41	914.29	887.71	733.21	-1770.35	-8.53	-438.10	988.31	-1842.47	**788.52**	147.44
β_{25}	297.34	167.27	179.47	127.63	-334.46	4.04	-47.25	186.61	-370.97	147.44	**39.31**

Design 3: S-optimal Design

	β_1^*	β_2^*	β_{13}	β_{14}^1	β_{14}^2	β_{15}	β_{20}	β_{23}	β_{24}^1	β_{24}^2	β_{25}
β_1^*	**1.08**	0.60	0.65	-0.10	-0.39	0.25	1.36	0.67	-0.14	0.32	0.34
β_2^*	0.60	**2.34**	0.87	-0.24	0.33	-0.14	-1.12	1.38	-0.03	1.08	-0.15
β_{13}	0.65	0.87	**0.85**	-0.25	-0.12	0.07	1.58	0.84	-0.26	0.29	0.09
β_{14}^1	-0.10	-0.24	-0.25	**1.12**	-0.63	0.05	-0.01	-0.16	0.26	-0.18	0.00
β_{14}^2	-0.39	0.33	-0.12	-0.63	**2.40**	0.02	-2.05	0.13	0.07	0.51	0.10
β_{15}	0.25	-0.14	0.07	0.05	0.02	**0.15**	0.43	0.07	-0.01	0.00	0.19
β_{20}	1.36	-1.12	1.58	-0.01	-2.05	0.43	**11.74**	-0.06	-0.75	-1.17	0.45
β_{23}	0.67	1.38	0.84	-0.16	0.13	0.07	-0.06	**1.29**	-0.18	0.62	0.08
β_{24}^1	-0.14	-0.03	-0.26	0.26	0.07	-0.01	-0.75	-0.18	**0.95**	-0.34	-0.02
β_{24}^2	0.32	1.08	0.29	-0.18	0.51	0.00	-1.17	0.62	-0.34	**1.51**	0.03
β_{25}	0.34	-0.15	0.09	0.00	0.10	0.19	0.45	0.08	-0.02	0.03	**0.26**

Tables 2 and 3 assume that the priors have been correctly specified. In many instances, the analyst will not have a clear picture as to what the true priors are. In such cases, the analyst may have to make a best guess as to the priors 'true' value. By fixing the design and changing the priors, the analyst is able to recalculate the asymptotic (co)variance matrix of the design (without the need to resort to Monte Carlo simulation). Doing so will allow the analyst to examine the robustness of SC designs to misspecifications of the priors. Table 4 demonstrates the new predicted sample size requirements assuming a misspecification of the β_{14}^1 prior over a range of values. For all values assumed, the S-efficient design outperforms the Dp-efficient and orthogonal designs in terms of the theoretical minimum sample size required for statistical significance of all parameters. Misspecification of other priors can be examined in the same manner.

Table 4: Asymptotic t-ratios by designs

β_{14}^1	Design 1	Design 2	Design 3
-0.1	550	253,313	434
-0.2*	137	188,469	108
-0.3	65	207,678	57
-0.4	67	228,767	57
-0.5	69	251,888	58

Note that, keeping the design constant, a misspecification of prior for any attribute will have an impact upon the asymptotic standard errors (and hence asymptotic *t*-ratios) 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.

6. Conclusion and Discussion

This paper introduces a new form of design efficiency criterion which we have titled S-efficiency. We have demonstrated through use of a numerical example that S-efficient designs may yield significant improvements in the sample size requirements necessary for SC studies. Whilst D-efficient designs will in general provide greater reliability to a greater number of parameters, it is unlikely to do so in a way that will minimize the sample sizes required for statistical significance to be observed. We have also shown that given the often random selection of orthogonal designs, that anything is possible in terms of the outcomes of SC experiments to which such designs are applied. We would therefore strongly recommend against their use, unless the analyst has no knowledge of the priors to assume in generation more efficient experimental designs.

In writing this paper, we have also attempted to demystify some of the more prevalent misconceptions that exist within the literature. The literature in its current form rely on the use of matrix algebra to derive the asymptotic (co)variance matrices used in generating statistically efficient SC designs. In doing, the literature has limited itself to the generation of designs allowing for generic parameters only. Unfortunately, this limitation is not immediately obvious and several researchers have tended to inappropriately employ matrix algebra to derive efficient designs for SC studies involving the estimation of alternative-specific parameter estimates. By deriving the asymptotic (co)variance matrix by taking the second derivatives of the log-likelihood function, it becomes possible to correctly obtain the asymptotic (co)variance matrix for models requiring the estimation of both alternative-specific and generic parameter estimates.

In using the methods we outline within this paper, we show that for any given sample size, one may determine 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 for example, Carlsson and Martinsson 2003).

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?

References

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

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.

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.

Hahn, G.J. and S.S. Shapiro (1966) A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments, Report no. 66-C-165, General Electric Research and Development Center, New York.

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.

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

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.

Viney, R. Savage, E. and Louviere, J. Empirical investigation of experimental design properties of discrete choice experiments in health 2005; 14, 349-362.