Ryan, M. and S. Wordsworth (2000), 'Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments', *Scottish Journal of Political Economy*, 47(5), 504–24.

Ryan, M., A. Bate, C. Eastmond and A. Ludbrook (2001), 'Using discrete choice experiments to elicit preferences', *Quality in Health Care*, **10**, 155–60.

Ryan, M., E. McIntosh, T. Dean and P. Old (2000), 'Trade-offs between location and waiting times in the provision of health care: the case of elective surgery on the Isle of Wight', *Journal of Public Health Medicine*, **22**, 202–10.

San Miguel, F., M. Ryan and M. Amaya-Amaya (2005), '"Irrational" stated preferences: a quantitative and qualitative investigation', *Health Economics*, **14**(3), 307–22.

San Miguel, F., M. Ryan and E. McIntosh (2000), 'Demonstrating the use of conjoint analysis in economic evaluations: an application to menorrhagia', *Applied Economics*, **32**, 822–33.

Sculpher, M., S. Bryan, Frt P. de Winter, P. H. Pasyne and M. Emberton (2004). Patient preferences for the management of non-metastatic prostate cancer: a discrete choice experiment. BMJ, **328**: 382–4.

Swait, J. (2001), 'A non-compensatory choice model incorporating attribute cut-offs', *Transportation Research*, **35**, 903–28.

Towse, A., C. Prichard and N. Devlin (eds) (2002), *Cost-effectiveness Thresholds: Economic and Ethical Issues*, London: King's Fund and Office of Health Economics.

Ubach, C., A. Scott, F. French, M. Awramenko and G. Needham (2003), 'What do hospital consultants value about their jobs? A discrete choice experiment', *British Medical Journal*, **326**, 1432–42.

Vale, L., L. Wyness, K. McCormack, L. McKenzie, M. Brazzelli and S.C. Stearns (2002), *Systematic Review of the Effectiveness and Cost-effectiveness of Metal on Metal Hip Resurfacing Arthroplasty for Treatment of Hip Disease*, London: National Institute for Clinical Excellence.

van der Pol, M. and J. Cairns (1998), 'Establishing patient preferences for blood transfusion support: an application of conjoint analysis', *Journal of Health Services Research and Policy*, 3(2), 70–76.

Viney, R., E. Lancsar and J. Louviere (2002), 'Discrete choice experiments to measure preferences for health and health care', *Expert Review Pharmacoeconomics Outcomes Research*, **2**, 319–26.

Watson, V., M. Ryan and E. Watson (2003), 'Testing the validity of discrete choice experiments: evidence from a chlamydia screening study', paper presented at the Scottish Economic Society Conference, Glasgow University.

39 Design of choice experiments in health economics

Leonie Burgess, Deborah J. Street, Rosalie Viney and Jordan Louviere

1. Introduction

In many areas of applied economics, economists use observations of actual choices, or revealed preference (RP) data, to model behaviour. Individuals are assumed to make utility maximizing choices, and utility functions are estimated by analysing observed choices. A need for stated preference (SP) data arises when there are limited or no RP data available, because the good or service is new, not provided in a market context, or there is insufficient variability in choice attributes to obtain reliable estimates of their effects. Such situations often arise in the health and health care sectors. Discrete choice experiments (DCEs) are an SP method of increasing interest in health economics because they allow analysis of preferences for complex, multi-attribute goods like health care. DCEs were developed in marketing and transport research and are applied elsewhere (for example, environmental economics and telecommunications) (Louviere et al., 2000). Since the first application of DCEs in health (Propper, 1990) there has been rapid growth in their use (Viney et al., 2002; Ryan and Gerard, 2003).

DCEs ask individuals to state preferences in surveys designed to simulate market choices. Respondents evaluate a series of choice sets that each have m options, and choose their preferred option in each choice set. Options in each choice set are described by attributes (which may be quantitative or qualitative) that are varied over a plausible and policy-relevant range, generally expressed as a set of discrete levels. Responses are discrete observations (one option in the choice set is chosen, or 'yes/no' if only one option is in each choice set), and discrete choice methods are used to estimate preferences from the choices. By varying the attribute levels of options across choice sets, one can generate data to estimate the impact of attributes on utility. Inclusion of respondents' personal characteristics allows estimation of these effects as well.

As numbers of attributes and levels in experiments increase, numbers of options and numbers of possible ways of combining them to form choice sets grow large. Typically, it is infeasible to present respondents with all possible choice sets: hence, experimental design principles are used to select samples from the set of all the possible choice sets to allow efficient estimation of attribute main effects and interactions.

DCEs share some common features with conjoint methods commonly used in marketing, and DCEs in health have sometimes been described as 'conjoint analysis' (Ryan and Hughes, 1997; Ratcliffe and Buxton, 1999). Both methods describe goods/services in terms of underlying attributes, use experimental designs to develop sets of descriptions for preference elicitation, and use statistical models to estimate the effects of each attribute on preferences (Louviere, 2001a). However, in DCEs respondents are asked to choose one option from those presented, rather than ranking or rating all options, and the analysis uses the random utility model (RUM) as a behavioural theory. Thus, DCEs are consistent with economic theory (Hanley and Mourato, 2001), and can be designed to simulate real market situations (Ryan and Farrar, 2000), particularly if the options presented in choice sets include 'do not choose' when not choosing is feasible (Louviere, 2001a).

Economic analysis based on DCEs differs from most empirical research in economics because the data are derived from experiments designed by the researchers to answer specific questions, rather than from administrative data or statistical collections. DCE researchers decide what data are to be collected, and design the data collection instruments, including deciding the combinations of options to present to respondents. This requires the researcher to consider in advance of data collection what models and forms of utility function are to be estimated, and, particularly, any interactions among attributes. As data collection is costly, one should design experiments to maximize the information that can be obtained, ensuring all relevant attributes are included with an appropriate range of levels. However, DCEs can be complex cognitive tasks, so it is also important to consider how complexity might impact on respondent choices, and in turn on model estimates. This chapter focuses on identifying issues relevant to the design and development of highly efficient DCEs. The chapter provides a guide to the principles of efficient experimental design, and directs readers to key sources for more information.

2. Understanding choice behaviour in designing choice experiments

Conceptually, individuals' choices are based on an underlying choice process, which is assumed to be utility maximization. Potentially unknown factors can affect consumer choices so some factors relevant to choices will be unobserved or unobservable, and individuals' choice processes and preferences are likely to differ. One cannot measure and include all relevant factors in any choice experiment, which has two important implications: 1) the data collected limit the models that can be estimated; and 2) which options are presented (and how) and which responses are obtained, and how they are measured, can affect response variability, which impacts on the quality of inferences. The goal of data collection and modelling is to minimize the unexplained variability in the observed choices by including as many factors as possible that systematically affect choices and by minimizing random variability (noise) in choices.

Factors not explicitly measured or included in models contribute to random error variance, and there is error in the measurement of responses in choice experiments (like all measurement instruments). There is also inherent variability in individuals' responses to survey questions, which can increase if tasks are difficult or surveys do not provide complete information about factors relevant to decisions. In designing DCEs one must consider how all aspects of data collection affect response variability and how to ensure that DCEs provide the maximum information at each stage.

Conceptualizing the choice process involves understanding individuals' decision-making contexts, options likely to be available, how options are presented, and factors that are likely to drive choices. The policy context must be considered, as well as how respondents are likely to interpret choice sets that they evaluate. Accurate conceptualization of choice processes involves literature reviews, qualitative research, and

(iterative) pilot studies. Design and framing of DCE tasks should consider whether choices are once-off or repeated, the importance of the outcome of the choice (is it a life-or-death decision, for instance), and how familiar individuals are with the decision-making context. In many DCE contexts like transport mode choice, choices are familiar. However in health care, decisions may be made under stress, options may be less familiar, and there may be serious consequences. Decisions like choice of health insurance may be familiar to all individuals, but other health decisions, such as choices about medical interventions, may require provision of detailed information about the choice context and the attributes of options.

One must also consider whether combinations of attributes and levels presented in choice sets are not feasible or credible, in which case adjustments must be made to the design and presentation. As previously noted, because parameters of choice models cannot be estimated independently of random error variances, factors that increase error variability like task complexity or unrealistic attribute options may lead to more variability and possibly biased parameter estimates; see Louviere et al. (2002) and (2003). There is often confusion about the idea of feasibility in DCEs. What researchers or policy-makers know to be 'feasible' is irrelevant in DCEs. Concerns about feasiblity should arise only if respondents regard particular combinations of attributes as nonsensical, resulting in credibility issues. Feasibility issues can be addressed by framing an appropriate explanation in the instructions for what might seem unlikely. Eliminating 'infeasible' options/sets can result in loss of efficiency and orthogonality, which can cause serious bias, particularly in small designs. Thus, efforts should be made prior to fieldwork to determine whether feasibility is a real issue for respondents.

3. Discrete choice experiments, random utility theory and choice models

A DCE consists of a set of N choice sets, each with m options. Respondents evaluate each choice set and choose one option (or possibly 'none of these'). Each option in a choice set is described by k attributes, where the qth attribute has l_q levels, $q = 1, \ldots, k$. The attributes may be generic (common across options) or option-specific (belonging only to a particular option). The options may be labelled, to allow utilities specific to a product or brand (for example, medical intervention, surgical intervention), or they may be unlabelled (intervention A, intervention B).

Analysis of DCEs is based on Lancaster's theory of value, which assumes that utility is derived from the underlying characteristics or attributes of goods/services (Lancaster, 1966), and on the Random Utility Model (RUM) (McFadden, 1973; Manski, 1977). Utility is not directly observable but can be estimated from observed choices.

Consider a choice experiment in which there are N choice sets, each consisting of m options, $T_{i_1}, T_{i_2}, \ldots, T_{i_m}$. Given these m options, we assume that if the individual chooses option T_{i_j}, the utility of choice T_{i_j}, U_{i_j}, is the maximum among the m utilities. Under the RUM, the individual's indirect utility function for a particular option is assumed to have a systematic component, V_{i_j}, and a random component, ε_{i_j}, so $U_{i_j} = V_{i_j} + \varepsilon_{i_j} = X'_{i_j}\beta + \varepsilon_{i_j}$, where X_{i_j} is a vector of variables representing observed attributes of option T_{i_j}, and β is the vector of coefficients to be estimated. The random component may be due to unobserved or unobservable attributes of the choice, unobserved taste variation or measurement error (McFadden, 1973; Ben-Akiva and Lerman, 1985).

The estimated model depends on assumptions made about the distribution of the random component and the nature of the choice being modelled. It is commonly assumed that the ε_{i_j} are independently and identically distributed with a Gumbel distribution, leading to a multinomial logit (McFadden, 1973), where the probability that an individual chooses T_{i_1} is given by:

$$P(T_{i_1} > T_{i_2}, \ldots, T_{i_m}) = P(U_{i_1} > U_{i_n}, \forall n \neq 1) = P(V_{i_1} - V_{i_n} > \varepsilon_{i_n} - \varepsilon_{i_1}) = \frac{\exp(\mu(X'_{i_1}\beta))}{\sum_j \exp(\mu(X'_j\beta))}.$$

The parameter μ is a positive scale parameter that is inversely proportional to the variance of the random component. It is not possible to estimate the model coefficients separately from the scale parameter, and it is commonly assumed that $\mu = 1$. When responses from a series of choice sets are observed for one individual there may be correlation among the ε_{i_j} for that individual, which needs to be taken into account in the analysis. For example, one can include individual fixed effects in the systematic utility components, or estimate random parameter models such as the mixed logit model.

To discuss the design of discrete choice experiments, it is useful to follow Burgess and Street (2003) and define π_{i_j} by $\ln(\pi_{i_j}) = U_{i_j}$. Then

$$\Pr(T_{i_1} > T_{i_2}, \ldots, T_{i_m}) = \frac{\pi_{i_1}}{\sum_{j=1}^{m} \pi_{i_j}}$$

for $i_j = 1, 2, \ldots, t$ treatments, where no two treatments are the same. Maximum likelihood methods are used to estimate the π_{i_j}.

More information can be obtained from a DCE by using appropriate groupings of options. DCE designs can be compared by using the generalized variance of the parameter estimates, called the *D-optimal* value of the design. The variance–covariance matrix of the parameter estimates is the inverse of the Fisher information matrix: D-optimal designs will have the maximum determinant of the information matrix which is equivalent to the minimum value of the determinant of the variance–covariance matrix. Thus a D-optimal design is the natural extension of univariate minimum variance estimators. In El-Helbawy and Bradley (1978) the information matrix is defined to be $C = B\Lambda B'$, where B is the matrix with rows comprising the contrasts[1] for the effects to be estimated (that is, main effects or main effects plus two-factor interactions), and Λ is the matrix of second derivatives of the likelihood function.

To define Λ we first define $\lambda_{i_1, i_2, \ldots, i_m} = n_{i_1, i_2, \ldots, i_m}/N$, where $n_{i_1, i_2, \ldots, i_m} = 1$ if $(T_{i_1}, T_{i_2}, \ldots, T_{i_m})$ is a choice set and is 0 otherwise. Then the entries of Λ are given by

$$\Lambda_{i_1, i_1} = \pi_{i_1} \sum_{i_2 < i_3 < \cdots < i_m}' \frac{\lambda_{i_1, i_2, \ldots, i_m} \sum_{j=2}^{m} \pi_{i_j}}{(\sum_{j=1}^{m} \pi_{i_j})^2} = \frac{(m-1)}{m^2} \sum_{i_2 < i_3 < \cdots < i_m}' \lambda_{i_1, i_2, \ldots, i_m} \text{ if all } \pi_{i_j} = 1$$

and

$$\Lambda_{i_1, i_2} = -\pi_{i_1}\pi_{i_2} \sum_{i_3 < i_4 < \cdots < i_m}' \frac{\lambda_{i_1, i_2, \ldots, i_m}}{(\sum_{j=1}^{m} \pi_{i_j})^2} = -\frac{(m-1)}{m^2} \sum_{i_3 < i_4 < \cdots < i_m}' \lambda_{i_1, i_2, \ldots, i_m} \text{ if all } \pi_{i_j} = 1$$

where the summations are over all choice sets that contain item T_{i_1} (for the diagonal elements) and are over all choice sets that contain items T_{i_1} and T_{i_2} (for the off-diagonal elements). Under the null hypothesis of no differences between the effects of the levels of each attribute (that is, all $\pi_{i_j} = 1$), Λ contains the proportions of choice sets in which pairs of treatments appear together; see Burgess and Street (2003) and Street et al. (forthcoming) for more details.

Since optimal designs maximize the determinant of the information matrix, we denote the largest determinant by $\det(C_{opt})$. For any design d with information matrix C, the *D-efficiency* of d is given by $(\det(C)/\det(C_{opt}))^{1/p}$ where p is the number of parameters to be estimated.

Example 1 Suppose that there are two attributes, so $k = 2$, and that $l_1 = 2$ and $l_2 = 4$. Suppose that the levels are 0 and 1 for the first attribute and 0, 1, 2 and 3 for the second attribute. Then there are $2 \times 4 = 8$ possible options that can be described: 00, 01, 02, 03, 10, 11, 12 and 13. Suppose that the choice sets are of size $m = 3$. Then there are 56 distinct choice sets of this size: (00, 01, 02), (00, 01, 03), (00, 01, 10) and so on to (11, 12, 13). Suppose that we use (00, 01, 12) and (02, 13, 10) to be the choice experiment. Under the null hypothesis the matrices for these two choice sets are B, Λ_1 and C_1 in Table 39.1. These two choice sets are 86.6 per cent efficient. A diagonal C matrix requires at least eight choice sets; for example, (00, 01, 12), (01, 02, 13), (02, 03, 10), (03, 00, 11), (10, 11, 02), (11, 12, 03), (12, 13, 00) and (13, 10, 01) have a diagonal C matrix and are optimal for the estimation of main effects. Under the null hypothesis the matrices for these eight choice sets are B, Λ_2 and C_2 in Table 39.1.

4. Design of choice experiments

Binary response experiments

In binary response experiments respondents are shown treatment combinations one-at-a-time, and are asked whether they would choose/use each or not. If one wants to estimate the main effects of attributes independently of each other, the optimal set of treatment combinations to present to respondents are those from an orthogonal main effects plan (OMEP) (often called a resolution 3 design)[2] in which all levels of all attributes are equally replicated. If the goal of the experiment is to estimate both main effects and two-factor interactions independently of each other, the optimal set of treatment combinations to show to respondents form what is known as a fractional factorial design of resolution 5.[3] Some OMEPs and designs of resolution 5, together with definitions, may be found in Sloane (2005).

Generic forced choice experiments for main effects only

Burgess and Street (2005) provide an upper bound for $\det(C)$ for estimating main effects only. It applies for any choice set size for any number of attributes with any number of levels. All attributes can have the same number of levels, or attributes can have different numbers of levels. Recall that B is a matrix of contrasts for the effects of interest. If the

Table 39.1 Matrices for choice sets in Example 1 (k = 2, $l_1 = 2$, $l_2 = 4$)

$$B = \begin{bmatrix} \frac{-1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{-3}{2\sqrt{10}} & \frac{-1}{2\sqrt{10}} & \frac{1}{2\sqrt{10}} & \frac{3}{2\sqrt{10}} & \frac{-3}{2\sqrt{10}} & \frac{-1}{2\sqrt{10}} & \frac{1}{2\sqrt{10}} & \frac{3}{2\sqrt{10}} \\ \frac{1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{-1}{2\sqrt{10}} & \frac{3}{2\sqrt{10}} & \frac{-3}{2\sqrt{10}} & \frac{1}{2\sqrt{10}} & \frac{-1}{2\sqrt{10}} & \frac{3}{2\sqrt{10}} & \frac{-3}{2\sqrt{10}} & \frac{1}{2\sqrt{10}} \end{bmatrix}$$

$$\Lambda_1 = \frac{1}{18}\begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ -1 & 2 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & -1 & 0 & 0 & 2 \end{bmatrix} \qquad \Lambda_2 = \frac{1}{72}\begin{bmatrix} 6 & -1 & 0 & -1 & 0 & -1 & -2 & -1 \\ -1 & 6 & -1 & 0 & -1 & 0 & -1 & -2 \\ 0 & -1 & 6 & -1 & -2 & -1 & 0 & -1 \\ -1 & 0 & -1 & 6 & -1 & -2 & -1 & 0 \\ 0 & -1 & -2 & -1 & 6 & -1 & 0 & -1 \\ -1 & 0 & -1 & -2 & -1 & 6 & -1 & 0 \\ -2 & -1 & 0 & -1 & 0 & -1 & 6 & -1 \\ -1 & -2 & -1 & 0 & -1 & 0 & -1 & 6 \end{bmatrix}$$

$$C_1 = B\Lambda_1 B' = \begin{bmatrix} \frac{1}{9} & \frac{1}{18\sqrt{5}} & \frac{1}{36} & \frac{-1}{36\sqrt{5}} \\ \frac{1}{18\sqrt{5}} & \frac{1}{9} & \frac{-1}{9\sqrt{5}} & 0 \\ \frac{1}{36} & \frac{-1}{9\sqrt{5}} & \frac{1}{9} & \frac{1}{18\sqrt{5}} \\ \frac{-1}{36\sqrt{5}} & 0 & \frac{1}{18\sqrt{5}} & \frac{1}{9} \end{bmatrix} \qquad C_2 = B\Lambda_2 B' = \begin{bmatrix} \frac{1}{9} & 0 & 0 & 0 \\ 0 & \frac{1}{9} & 0 & 0 \\ 0 & 0 & \frac{1}{9} & 0 \\ 0 & 0 & 0 & \frac{1}{9} \end{bmatrix}$$

main effect of attribute q, which has l_q levels, is of interest then B will contain $l_q - 1$ rows that correspond to $l_q - 1$ independent contrasts (one for each degree of freedom) associated with that attribute. Any set of $l_q - 1$ independent contrasts results in the same information matrix and hence the same variance–covariance matrix of the parameter estimates. For each attribute one finds an appropriate set of contrasts and these are used as the rows of the matrix B to calculate the information matrix $C = B\Lambda B'$.

Once the C matrix for a DCE is calculated, the statistical efficiency of the design can be calculated. The *D-efficiency* is given by $[\det(C) / \det(C_{opt})]^{1/p}$, where p is the number of parameters to be estimated in the model. For designs that estimate main effects only, $p = \Sigma_i(l_i - 1)$. The maximum possible value of the determinant of C is

$$\det(C_{opt}) = \prod_{q=1}^{k}\left(\frac{2S_q}{m^2(l_q - 1)\prod_{i=1, i\neq q}^{k} l_i}\right)^{l_q - 1}$$

where

$$S_q = \begin{cases} (m^2 - 1)/4 & l_q = 2, m \text{ odd}, \\ m^2/4 & l_q = 2, m \text{ even}, \\ (m^2 - (l_q x^2 + 2xy + y))/2 & 2 < l_q < m, \\ m(m-1)/2 & l_q \geq m \end{cases}$$

and positive integers x and y satisfy the equation $m = l_q x + y$ for $0 \leq y < l_q$. S_q is the maximum number of differences in the levels of attribute q in each choice set. Hence, an optimal design is one where the maximum number of level differences is attained for each attribute.

To construct designs that are optimal or near-optimal, an OMEP is required. The OMEP is used to represent the treatments in the first option of each choice set. Systematic level changes are made so that there are as many pairs of options as possible with different levels for each attribute in each choice set. For attributes with two levels this gives the foldover pairs advocated in Louviere, Hensher and Swait (2000), for example. These systematic changes are equivalent to adding generators, using modular arithmetic, to the OMEP to create the rest of the options in the choice sets. In most situations these generators are not unique and different generators can give rise to equally good designs; see Street et al. (forthcoming) for instance.

Example 2 Suppose that there are $k = 4$ attributes and that $l_1 = l_2 = 2$ and $l_3 = l_4 = 4$. There is an OMEP with 16 treatment combinations available; each level of the 2-level attributes is replicated 8 times and each level of the 4-level attributes is replicated 4 times. The 16 treatment combinations in the OMEP are used as the first option in the 16 choice sets. The other options in each choice set are obtained by adding a generator. For instance, use the generator 1111 (or 1112 or 1131 or 1133) to get choice sets of size $m = 2$. The addition is done modulo 2 in the first two positions (so $0 + 0 \equiv 1 + 1 \equiv 0$ and $0 + 1 \equiv 1 + 0 \equiv 1$) and is done modulo 4 in the third and fourth positions (so $1 + 3 \equiv 2 + 2 \equiv 0$ and $2 + 3 \equiv 1$, for example). Thus the first choice set is (0000, 0000 + 1111) which is (0000, 1111), for

Table 39.2 Main effects only: two attributes with two levels, two attributes with four levels

m	generators	# sets	Eff	Choice sets
2	1111	16	96%	(0000,1111), (0011,1122), (0122,1033), (0133,1000), (0012,1123), (0003,1110), (0130,1001), (0121,1032), (1023,0130), (1032,0103), (1101,0012), (1110,0021), (1031,0102), (1020,0131), (1113,0020), (1102,0013)
3	1111, 1122	16	100%	(0000,1111,1122), (0011,1122,1133), (0122,1033,1000), (0133,1000,1011), (0012,1123,1130), (0003,1110,1121), (0130,1001,1012), (0121,1032,1003), (1023,0130,0101), (1032,0103,0110), (1101,0012,0023), (1110,0021,0032), (1031,0102,0113), (1020,0131,0102), (1113,0020,0031), (1102,0013,0020)
4	0111, 1022, 1133	16	100%	(0000,0111,1022,1133), (0011,0122,1033,1100), (0122,0033,1100,1011), (0133,0000,1111,1022), (0012,0123,1030,1101), (0003,0110,1021,1132), (0130,0001,1112,1023), (0121,0032,1103,1010), (1023,1130,0001,0112), (1032,1103,0010,0121), (1101,1012,0123,0030), (1110,1021,0132,0003), (1031,1102,0013,0120), (1020,1131,0002,0113), (1113,1020,0131,0002), (1102,1013,0120,0031)

instance. The 16 choice sets that result are given in Table 39.2. Examples with $m = 3$ and $m = 4$ are also given in the table. See Burgess and Street (2005) or Street et al. (forthcoming) for more on choosing generators.

Generic choice experiments for main effects plus two-factor interactions

If all attributes have two levels, the optimal design consists of all choice sets in which the number of attributes that differ between any pair of profiles in the choice set is $(k + 1)/2$, if k is odd, or $k/2$ or $k/2 + 1$ if k is even (Burgess and Street, 2003). Furthermore, the maximum possible determinant of C for any choice set size has been determined and is given by

$$\det(C_{opt}) = \begin{cases} \left(\dfrac{(m-1)(k+2)}{m(k+1)2^k}\right)^{k+k(k-1)/2} & k \text{ even} \\ \left(\dfrac{(m-1)(k+1)}{mk2^k}\right)^{k+k(k-1)/2} & k \text{ odd.} \end{cases}$$

The *D-efficiency* of any design is given by $[\det(C) / \det(C_{opt})]^{1/p}$, where p is given by $p = k + k(k-1)/2$ when all the attributes are binary. In general $p = \Sigma_i (l_i - 1) + \Sigma\Sigma_{i<j} (l_i - 1)(l_j - 1)$.

Table 39.3 Main effects and two factor interactions: four attributes with two levels

m	Generators	# sets	Eff	Choice sets
2	1110, 1011, 0111	24	94%	(0000,1110), (0001,1111), (0010,1100), (0011,1101), (0100,1010), (0101,1011), (0110,1000), (0111,1001), (0000,1011), (0001,1010), (0010,1001), (0011,1000), (0100,1111), (0101,1110), (0110,1101), (0111,1100), (0000,0111), (0001,0110), (0010,0101), (0011,0100), (1000,1111), (1001,1110), (1010,1101), (1011,1100)
3	1100, 0110 and 1100, 0111	32	97%	(0000,1100,0110), (0001,1101,0111), (0010,1110,0100), (0011,1111,0101), (0100,1000,0010), (0101,1001,0011), (0110,1010,0000), (0111,1011,0001), (1000,0100,1110), (1001,0101,1111), (1010,0110,1100), (1011,0111,1101), (1100,0000,1010), (1101,0001,1011), (1110,0010,1000), (1111,0011,1001), (0000,1100,0111), (0001,1101,0110), (0010,1110,0101), (0011,1111,0100), (0100,1000,0011), (0101,1001,0010), (0110,1010,0001), (0111,1011,0000), (1000,0100,1111), (1001,0101,1110), (1010,0110,1101), (1011,0111,1100), (1100,0000,1011), (1101,0001,1010), (1110,0010,1001), (1111,0011,1000)
4	1100, 0110, 1011	16	99%	(0000,1100,0110,1011), (0001,1101,0111,1010), (0010,1110,0100,1001), (0011,1111,0101,1000), (0100,1000,0010,1111), (0101,1001,0011,1110), (0110,1010,0000,1101), (0111,1011,0001,1100), (1000,0100,1110,0011), (1001,0101,1111,0010), (1010,0110,1100,0001), (1011,0111,1101,0000), (1100,0000,1010,0111), (1101,0001,1011,0110), (1110,0010,1000,0101), (1111,0011,1001,0100)

Example 3 Suppose there are $k = 4$ binary attributes. Then the only resolution 5 design consists of all 16 treatment combinations. Table 39.3 shows some good designs for choice sets of size 2, 3 and 4.

When attributes can have any number of levels, and choice sets can be of any size, an explicit expression for $\det(C)$ in terms of the differences between the levels of each attribute in the choice sets is provided by Burgess and Street (2005). No general constructions are known, but Burgess and Street (2005) give optimal designs for some specific values of k and m. A method similar to that given for main effects only can be used to construct choice sets and the $\det(C)$ values compared to choose the best design.

Choice experiments with constant options

In this section we consider two types of choice experiments with constant options, the 'none of these' option and a common base option (for example, the status quo). Street and Burgess (2004) consider both types, and show that the designs that are optimal for the generic forced choice setting also are optimal when a 'none of these' option is included in each choice set, but designs with a 'none of these' option are not as efficient at estimating main effects as a forced choice design of the same size. Including a 'none of these' option may be more realistic and certainly means that interaction effects can be estimated, although inefficiently and not necessarily independently: see Street and Burgess (2004).

Sometimes each treatment combination needs to be compared to a common base option. In this situation the optimal design is one in which all the treatment combinations, including the common base option, form an OMEP. This can always be accomplished by a suitable renaming of the attribute levels.

Street and Burgess (2004) give an expression for det(C) for this situation. Assume that all levels of all attributes are equally replicated and let $p = \sum_{q=1}^{k}(l_q - 1)$ (total degrees of freedom for main effects), $L = \prod_{q=1}^{k} l_q$ and assume that the OMEP has t treatments. Then

$$\det(C) = \frac{1+p}{(4(t-1))^p}\left(\frac{t}{L}\right)^p$$

Thus we can also evaluate the efficiency of the DCEs with a common base relative to generic forced choice designs as well as observing that the smallest OMEP is the most efficient if a common base must be used.

5. Unresolved issues in choice experiments

Results exist for the design of optimal choice experiments when only some attributes are presented to respondents in each choice set and all attributes are binary (Grasshoff et al., 2003), and for the general case (Burgess and Street, 2005). However, it is unclear how respondents deal with unpresented attributes, as noted by Bradlow et al. (2004) and Islam et al. (2004).

There appear to be no results on the design of optimal choice experiments when certain combinations of attribute levels cannot appear together, nor if one wants to avoid having choice sets in which one option dominates all others on all attribute levels.

Sandor and Wedel (2001) discuss Bayesian design of choice experiments. Kanninen (2002) constructs optimal designs when attributes are assumed to be continuous. More work is needed on both of these problems.

This chapter has focused on models in the MNL family and there appear to be no results on the design of optimal choice experiments for any model more complicated than this. Much recent work in economics, transport and marketing has focused on new classes of discrete choice models that relax various aspects of the assumptions that underlie MNL models. For example, mixed logit models (for example, McFadden and Train, 2000) allow one to capture forms of preference heterogeneity, while retaining IID Extreme Value Type I errors; and variants of heteroscedastic error models (for example, Swait and Adamowicz, 2001) allow one to capture and parameterize non-constant error variances. Unfortunately, at this time we do not know whether designs developed for MNL models can be used to estimate such other models. That is, there are identification issues associated with these more complex models, and there has been no work that maps the model properties and constraints into design specifications.

Those who wish to apply more complex choice models should note that a number of serious questions have been raised recently about them. For example, Louviere (2001b, 2004a, 2004b), Louviere and Islam (2004), Louviere et al. (2004), Train and Weeks (2004)

and Sonnier et al. (2004) have raised concerns about likely failure of real choice data to satisfy constant variance assumptions. If the errors associated with choices do not exhibit constant variances, there are serious confounding issues that can lead to incorrect and biased estimation of model parameters and associated variances, with evidence now revealing that estimates of willingness-to-pay not only can be highly unrealistic, but also can differ by large orders of magnitude. More recent work on what they call a 'Scale Decomposition Model' (for example, Islam et al., 2004) demonstrates that the errors do not have constant variance, but can be parameterized by specifying them as a function of certain design factors and/or individual characteristics. While promising, it is unclear at this time how to construct designs that are consistent with these models, or how to construct designs that will maximize the statistical efficiency of the resulting model estimates. Thus, more research on these issues would be welcome as theoretical and methodological advances are outpacing advances in design research.

Notes

1. For example for a 2-level attribute a contrast is the difference between the sum of the π_i with the high and low levels of the attribute. Polynomial contrasts are defined for attributes with more than two levels.
2. In a resolution 3 design any combination of levels from any pair of attributes appears together in the same number of treatment combinations in the stated choice experiment.
3. In a resolution 5 design any combination of levels from any four attributes appears together in the same number of treatment combinations in the stated choice experiment.

References

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

Bradlow, E.T., Y. Hu and T.H. Ho (2004), 'A learning-based model for imputing missing levels in partial conjoint profiles', *Journal of Marketing Research*, **41**, 369–81.

Burgess, L. and D. Street (2003), 'Optimal designs for 2^k choice experiments', *Communications in Statistics – Theory & Methods*, **32**, 2185–206.

Burgess, L. and D.J. Street (2005), 'Optimal designs for choice experiments with asymmetric attributes', *Journal of Statistical Planning and Inference*, **134**, 288–301.

El-Helbawy, A.T. and R.A. Bradley (1978), 'Treatment contrasts in paired comparisons: large-sample results, applications and some optimal designs', *Journal of the American Statistical Association*, **73**, 831–9.

Grasshoff, U., H. Grossmann, H. Holling and R. Schwabe (2003), 'Optimal paired comparison designs for first order interactions', *Statistics: A Journal of Theoretical and Applied Statistics*, **37**, 373–86.

Hanley, N. and S. Mourato (2001), 'Choice modelling approaches: a superior option for environmental valuation?', *Journal of Economic Surveys*, **15**, 435–62.

Islam, T., J. Louviere and P. Burke (2004), 'Modeling the effects of including/excluding attributes in choice experiments on systematic and random components', unpublished working paper, School of Marketing, Faculty of Business, University of Technology, Sydney (December).

Kanninen, B.J. (2002), 'Optimal design for multinomial choice experiments', *Journal of Marketing Research*, **39**, 214–27.

Lancaster, K. (1966), 'A new approach to consumer theory', *Journal of Political Economy*, **74**, 132–57.

Louviere, J. (2001a), 'Choice experiments: an overview of concepts and issues', in J. Bennett and R. Blamey (eds), *The Choice Modelling Approach to Environmental Valuation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 13–36.

Louviere, J. (2001b), 'What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon', *Journal of Consumer Research*, **28**, 506–11.

Louviere, J. (2004a), 'Random utility theory-based stated preference elicitation methods: applications in health economics with special reference to combining sources of preference data', CenSoC Working Paper No. 04-001, Centre for the Study of Choice, Faculty of Business, University of Technology, Sydney.

Louviere, J. (2004b), 'Complex statistical choice models: Are the assumptions true, and if not, what are the consequences?', CenSoC Working Paper No. 04-002, Centre for the Study of Choice, Faculty of Business, University of Technology, Sydney.

Louviere, J. and T. Islam (2004), 'A comparison of importance weights/measures derived from choice-based conjoint, constant sum scales and best-worst scaling', CenSoC Working Paper No. 04-004, Centre for the Study of Choice, Faculty of Business, University of Technology, Sydney.
Louviere, J.J., D.A. Hensher and J.D. Swait (2000), *Stated Choice Methods: Analysis and Applications*, Cambridge, UK and New York, US: Cambridge University Press.
Louviere, J., D. Street and L. Burgess (2003), 'A 20+ years retrospective on choice experiments', in Y. Wind and P.E. Green (eds), *Marketing Research and Modeling: Progress and Prospects*, New York: Kluwer, pp. 201–14.
Louviere, J., D. Street, R. Carson, A. Ainslie, J.R. Deshazo, T. Cameron, D. Hensher, R. Kohn and T. Marley (2002), 'Dissecting the random component of utility', *Marketing Letters*, **13**, 175–91.
Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T. Cameron, R. Carson, J. DeShazo, D. Fiebig, W. Greene, D. Hensher and D. Waldman (2004), 'Recent progress on endogeneity in choice modeling', Workshop Report, University of Colorado 6th Invitational Choice Symposium, September.
Manski, C. (1977), 'The structure of random utility models', *Theory and Decision*, **8**, 229–54.
McFadden, D. (1973), 'Conditional logit analysis of qualitative choice behavior', in Zarembka, P. (ed.), *Frontiers of Econometrics*, New York, US: Academic Press, pp. 105–42.
McFadden, D. and K. Train (2000), 'Mixed MNL models for discrete response', *Journal of Applied Econometrics*, **15**, 447–70.
Propper, C. (1990), 'Contingent valuation of time spent on NHS waiting lists', *The Economic Journal*, **100**, 193–9.
Ratcliffe, J. and M. Buxton (1999), 'Patients' preferences regarding the process and outcomes of life-saving technology. An application of conjoint analysis to liver transplantation', *International Journal of Technology Assessment in Health Care*, **15**, 340–51.
Ryan, M. and S. Farrar (2000), 'Using conjoint analysis to elicit preferences for health care', *British Medical Journal*, **320**, 1530–3.
Ryan, M. and K. Gerard (2003), 'Using discrete choice experiments to value health care programmes: current practice and future research reflections', *Applied Health Economics and Health Policy*, **2**, 55–64.
Ryan, M. and J. Hughes (1997), 'Using conjoint analysis to assess women's preferences for miscarriage management', *Health Economics*, **6**, 261–73.
Sandor, Zsolt and Michel Wedel (2001), 'Designing conjoint choice experiments using managers prior beliefs', *Journal of Marketing Research*, **38**, 430–44.
Sloane, N.J.A. (2005), 'A library of orthogonal arrays', available at http://www.research.att.com/~njas/oadir/.
Sonnier, G., A. Ainsle and T. Otter (2004), 'The effect of parameterization on heterogeneous choice models', working paper, Anderson School of Management, UCLA.
Street, D.J. and L. Burgess (2004), 'Optimal stated preference choice experiments when all choice sets contain a specific option', *Statistical Methodology*, **1**, 37–45.
Street, D.J., L. Burgess and J.J. Louviere (forthcoming), 'Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments', *International Journal of Market Research*, **22**.
Swait, J. and W. Adamowicz (2001), 'Incorporating the effect of choice environment and complexity into random utility models', *Organizational Behavior and Human Decision Performance*, **86**, 141–67.
Train, K. and M. Weeks (2004), 'Discrete choice models in preference space and willingness-to-pay spaces', working paper, University of California, Berkeley.
Viney, R., E. Lancsar and J. Louviere (2002), 'Discrete choice experiments to measure consumer preferences for health and healthcare', Expert reviews in *Pharmacoeconomics and Outcomes Research*, **2**, 319–26.

PART VIII

MEASURING COSTS AND STATISTICAL ISSUES