Using Choice Experiments for Non-Market Valuation

Francisco Alpízar, Fredrik Carlsson[1]
and Peter Martinsson

ABSTRACT

This paper provides the latest research developments in the method of choice experiments applied to valuation of non-market goods. Choice experiments, along with the, by now, well-known contingent valuation method, are very important tools for valuing non-market goods and the results are used in both cost-benefit analyses and litigations related to damage assessments. The paper should provide the reader with both the means to carry out a choice experiment and to conduct a detailed critical analysis of its performance in order to give informed advice about the results. A discussion of the underlying economic model of choice experiments is incorporated, as well as a presentation of econometric models consistent with economic theory. Furthermore, a detailed discussion on the development of a choice experiment is provided, which in particular focuses on the design of the experiment and tests of validity. Finally, a discussion on different ways to calculate welfare effects is presented.

1. INTRODUCTION

THE METHODS OF VALUATION of non-marketed goods have become crucial when determining the costs and benefits of public projects. Non-market valuation exercises have been conducted in many different areas, ranging from health and environmental applications to transport and public infrastructure projects. In the case of a good that is not traded in a market, an economic value of that good obviously cannot be directly obtained from the market. Markets fail to exist for some goods either because these goods simply do not exist yet, or because they are public goods, for which exclusion is not possible. Nevertheless, if one wants to compare different programs by using cost-benefit analysis, the change in the quality or quantity of the non-market goods should be expressed in monetary terms. Another crucial application of valuation techniques is the determination of damages associated with a certain event. Under the Comprehensive Environmental Response, Compensation and Liability Act of 1980 in the US, and after the events that followed the Exxon Valdez oil spill in 1989, the methods of valuation have become a central part of litigation for environmental and health related damages in the United States and in several other countries.

Over the years, the research on valuation of non-market goods has developed into two branches: revealed preference methods and stated preference methods. The first branch, the revealed preference method, infers the value of a non-market good by studying actual (revealed) behaviour on a closely related market. The two most-well-known revealed preference methods are the hedonic pricing method and the travel cost method (see Braden and Kolstad, 1991). In general, the revealed preference approach has the advantage of being based on actual choices made by individuals. However, there are also a number of drawbacks; most notably that the valuation is conditioned on current and previous levels of the non-market good and the impossibility of measuring non-use values, i.e. the value of the non-market good not related to usage such as existence value, altruistic value and bequest value. Thus, research in the area of valuation of non-market goods has therefore seen an increased interest in the stated preference method, during the last 20 years.

Stated preference methods assesses the value of non-market goods by using individuals' stated behaviour in a hypothetical setting. The method includes a number of different approaches such as conjoint analysis, contingent valuation method (CVM) and choice experiments. In most applications, CVM has been the most commonly used approach. In particular, closed-ended CVM surveys have been used, in which respondents are asked whether or not they would be willing to pay a certain amount of money for realizing the level of the non-market good described or, more precisely, the change in the level of the good (see Bateman and Willis, 1999 for a review). The idea of CVM was first suggested by Ciriacy-Wantrup (1947), and the first study ever done was in 1961 by Davis (1963). Since then, CVM surveys have become one of the most commonly used methods for valuation of non-market goods, although its use has been questioned (see e.g. Diamond and Hausman 1994; Hanemann, 1994, for critical assessments). At the same time as CVM was developed, other types of stated preference techniques, such as choice experiments, evolved in both marketing and transport economics (see Louviere, 1993; Polak and Jones 1993, for overviews).

In a choice experiment, individuals are given a hypothetical setting and asked to choose their preferred alternative among several alternatives in a choice set, and they are usually asked to perform a sequence of such choices. Each alternative is described by a number of attributes or characteristics. A monetary value is included as one of the attributes, along with other attributes of importance, when describing the profile of the alternative presented (see figure 1). Thus, when individuals make their choice, they implicitly make trade-offs between the levels of the attributes in the different alternatives presented in a choice set.

Figure 1: A choice-experiment - the basics

This is an example of a choice set containing two profiles of a given alternative (a park). Each profile is described in terms of 4 attributes, including the entrance fee. Each attribute has two or more levels. A choice experiment contains a sequence of such choice sets.

Available facilities	Visitor center	Information office
Extension of walking tracks	2 kms	10 kms
Condition of tracks	Rustic tracks	Stoned tracks
Entrance fee	8 US$	10 US$

Which of the two options would you prefer for a one-day visit

☐ Park A ☐ Park B

The purpose of this paper is to give a detailed description of the steps involved in a choice experiment and to discuss the use of this method for valuing non-market goods. Choice experiments are becoming ever more frequently applied to the valuation of non-market goods. This method gives the value of a certain good by separately evaluating the preferences of individuals for the relevant attributes that characterize that good, and in doing so it also provides a large amount of information that can be used in determining the preferred design of the good. In fact, choice experiments originated in the fields of transport and marketing, where it was mainly used to study the trade-offs between the characteristics of transport projects and private goods, respectively. Choice experiments have a long tradition in those fields, and they have only recently been applied to non-market goods in environmental and health economics. We believe that applications of this technique will become more frequent in other areas of economics as well. Only recently has the aim of damage assessment in litigation shifted from monetary compensation to resource compensation. Therefore identification and evaluation of the different attributes of a damaged good is required in order to design the preferred restoration project (Adamowicz *et al.*, 1998b; Layton and Brown, 1998). Choice experiments are especially well-suited for this purpose, and one could expect this method to be a central part of future litigation processes involving non-market goods.

The first study to apply choice experiments to non-market valuation was Adamowicz *et al.* (1994). Since then there has been an increasing number of studies, see e.g. Adamowicz *et al.* (1998a); Boxall *et al.* (1996); Layton and Brown (2000) for applications to environment, and e.g. Ryan and Hughes (1997); Vick and Scott (1998) for applications to health. There are several reasons for the increased interest in choice experiments in addition to those mentioned above: (i) reduction of some of the potential biases of CVM, (ii) more information is elicited from each respondent compared to CVM and (iii) the possibility of testing for internal consistency.

In a choice experiment, as well as in a CVM survey, the economic model is intrinsically linked to the statistical model. The economic model is the basis of the analysis, and as such, affects the design of the survey and the analysis of the data. In this sense, we argue that the realization of a choice experiment is best viewed as an integrated and cyclical process that starts with an economic model describing the issue to analyse. This model is then continually revised as new information is received from the experimental design, the statistical model, focus groups and pilot studies, etc. In this paper, we pay special attention to the link between the microeconomic and the statistical foundations of a choice experiment, when it comes to designing the choice experiment, estimating the econometric model as well as calculating welfare measures. Furthermore, we address the issue of internal and external validity of a choice experiment, and provide a discussion of the possibility of misrepresentation of preferences by strategic responses. The literature on choice experiments has been reviewed by other authors, e.g. Adamowicz *et al.* (1998b); Hanley *et al.* (1998); Louviere *et al.* (2000). This paper contributes to providing a thorough description of each of the steps needed when performing a choice experiment on a non-market good, with special attention to the latest research results in design and estimation.

The rest of the paper is organized as follows; Section 2 discusses the underlying economic theory of choice experiments. In Section 3, econometric models are discussed and linked to the section on economic theory. Section 4 concentrates on the design of a choice experiment, given the theoretical and empirical models presented in the two previous sections. Respondent behaviour and potential biases are discussed in Section 5. Section 6 presents different techniques to apply when estimating welfare effects. Finally, Section 7 concludes this paper.

2. THE ECONOMIC MODEL

The basis for most microeconomic models of consumer behavior is the maximization of a utility function subject to a budget constraint. Choice experiments were inspired by the Lancasterian microeconomic approach (Lancaster, 1966), in which individuals derive utility from the characteristics of the goods rather than directly from the goods themselves. As a result, a change in prices can cause a discrete switch from one bundle of goods to another that will provide the most cost-efficient combination of attributes. In order to explain the underlying theory of choice experiments, we need to link the Lancasterian theory of value with models of consumer demand for discrete/continuous choices (Hanemann, 1984 and 1999).

In general, an individual's decisions can be partitioned into two parts: (i) which good to choose and (ii) how much to consume of the chosen good. Hanemann (1984) calls this a discrete/continuous choice. An example of this choice structure is the case of a tourist deciding to visit a national park. The decision can be partitioned into which park to visit, and how long to stay. In order to obtain a value of a certain park, both stages of the decision-making process are crucial to the analysis and should be modelled or assumed in a consistent manner.

In general, choice experiments applied to non-marketed goods assume a specific continuous dimension, which is a given part of the framework in which a discrete choice takes place. By referring to the example above, one could ask for a discrete choice (which type of park do you prefer to visit?) given a one-week (day, month) trip. In this case, the decision context is constructed so that it isolates the discrete choice, therefore allowing the individual to make a purely discrete choice (Hanemann, 1999). In a CVM survey the researcher also makes such an assumption since the objective is to obtain the value of a certain predefined program that includes a given continuous dimension. Finally, note that many non-marketed goods are actually public in nature, especially in the sense that the same quantity of the good is available for all agents. In such cases, each individual can only choose one of the offered alternatives, given its cost and its continuous dimension.

The economic model presented in this section deals *only* with such purely discrete choices. For more information on discrete/continuous choices, see Hanemann (1984). Formally, each individual solves the following maximization problem:

$$Max_{c,\delta,x} U\left[\delta_1 c_1(A_1),\ldots,\delta_N c_N(A_N); z\right]$$

$$\begin{aligned} s.t. \qquad & i.\ y = \sum_{i=1}^{N} p_i \delta_i c_i(A_i) + z \\ & ii.\quad \delta_i \delta_j = 0\,, \quad \forall i \neq j \\ & iii.\quad z \geq 0, \quad \delta_i \geq 0 \quad \text{for at least one } i \end{aligned} \tag{1}$$

where, $U[\ldots]$ is a quasiconcave utility function; $c_i(A_i)$ is alternative combination i (profile i) as a function of its generic and alternative specific attributes, the vector A_i; δ_i is a dichotomous variable equal to one if profile i is chosen and equal to zero if not; p_i is the price of each profile; z is a composite bundle of ordinary goods with its price normalized to 1and y is income. A number of properties follow from the specification of the maximization problem:

1. The c_i's are profiles defined for all the relevant alternatives and described by all the relevant attributes. Additionally, the profiles contain a fixed, and given continuous dimension, e.g. a day or a unit. For example, one such profile could be a one-day visit to a national park in a rainforest, with 50 kms of marked walking tracks through the park and a visitor center. There are N such profiles, where N is in principle given by all relevant combinations of attributes into profiles. However, in practice, N will be determined depending on the type of design used to construct these profiles, the number of attributes, and the attribute levels included in the choice experiment. Consequently, with the selection of attributes and attribute levels for a choice experiment we are already limiting or defining the utility function.

2. The price variable in the budget restriction must be related to the complete profile of the alternative. If profile i is chosen ($\delta_i = 1$), then this profile will be enjoyed by the individual for the 'duration' of the continuous dimension associated to it. In order to correctly specify the budget restriction, the price must reflect this continuous dimension. In order to keep the model simple we prefer not to introduce further notation at this stage, and hence slightly abuse the definition of the c_i's by associating the profile with its actual continuous dimension.

3. Restriction *ii* defines the number of alternatives that can be chosen in a given choice set. In general, in a choice experiment we are interested in obtaining a single choice. For example, in the case of perfect substitutes, there will be a corner solution with only one profile chosen.[2] Alternatively, the choice experiment can specify the need for a single choice. If the alternatives refer to different public goods or environmental amenities, one can specify that only one will be available. Even if the alternatives refer to private goods such as a specific treatment program, the researcher can specify that only one of them can be chosen each time.

4. In a purely discrete choice, the selection of a particular profile $c_j(A_j)$, which is provided in an exogenously fixed quantity, implies that, for a given income, the amount of ordinary goods z that can be purchased is also fixed. Combining this with the restriction that only a single profile, c_j , can be chosen results in:

$$z = y - p_j c_j \tag{2}$$

5. Restriction *iii* specifies that the individual will choose a non-negative quantity of the composite good and that an opt-out decision can take place. Alternatively, we could include a *status quo* profile with all attributes at their actual levels, and force the respondent to make a choice between the profiles. If we believe that the good is essential to the individual or that an environmental program has to be implemented, then we have to force the respondent to make a choice ($\delta_i > 0$ for at least one i).

To solve the maximization problem we follow a two-step process. First we assume a discrete choice, profile j is chosen, i.e. $\delta_j = 1,\ \ \delta_i = 0 \quad \forall i \neq j$, and c_j is enjoyed in its specified, given continuous measure. We further assume weak complementarity, i.e. the attributes of the other non-selected profiles do not affect the utility function of profile j (Mäler, 1974;

Hanemann, 1984). Formally we write:

$$\text{if} \quad \delta_i = 0, \quad \text{then} \quad \frac{\partial U}{\partial A_i} = 0 \qquad \forall i \neq j \tag{3}$$

Using (2) and (3) we can write the conditional utility function, given $\delta_j = 1$ and c_j equal to its continuous dimension as:

$$U_j = V_j\left[c_j(A_j), p_j, y, z\right] = V_j(A_j, y - p_j c_j) \tag{4}$$

In the next step we go back to the unconditional indirect utility function:

$$V[A, p, y] = \max[V_1(A_1, y - p_1 c_1), \ldots, V_N(A_N, y - p_N c_N), \tag{5}$$

where the function V[..] captures only the discrete choice, and given the exogenous and fixed quantitative assumptions for each c_i, $i = 1,2,\ldots N$. Thus, it follows that the individual chooses the profile j if and only if:

$$V_j(A_j, y - p_j c_j) > V_i(A_i, y - p_i c_i), \qquad \forall i \neq j \tag{6}$$

Equations (5) and (6) complete the economic model for purely discrete choices. These two equations are the basis for the econometric model and the estimation of welfare effects that are discussed in the following sections.

Note that the economic model underlying a closed-ended CVM study can be seen as a special case of the model above, where there are only two profiles. One profile is the 'before the project' description of the good, and the other is the 'after the project' description of the same good. Thus a certain respondent will say yes to a bid if $V_i^1\left[c_i(A_i^1), y - bid\right] > V_i^0\left[c_i(A_i^0), y\right]$, where A_i^t entirely describes the good, including its continuous dimension.

Until now we have presented and discussed a deterministic model of consumer behaviour. The next step is to make such a model operational. There are two main issues involved; one is the assumption regarding the functional form of the utility function and the other is to introduce a component into the utility function to capture unobservable behaviour. In principle, these issues are linked, since the form of the utility function determines the relation between the probability distribution of the disturbances and the probability distribution of the indirect utility function.

3. The econometric model

Stated behaviour surveys sometimes reveal preference structures that may seem inconsistent with the deterministic model. It is assumed that these inconsistencies stem from observational deficiencies arising from unobservable components such as characteristics of the individual or non-included attributes of the alternatives in the experiment, measurement error and/or heterogeneity of preferences (Hanemann and Kanninen, 1999). In order to allow for these effects, the Random Utility approach (McFadden, 1974) is used to link the deterministic model with a statistical model of human behaviour. A random disturbance with a specified probability distribution, ε, is introduced into the model, and an individual will choose profile j (i.e. $\delta_j = 1$) if and only if:

$$V_j(A_j, y - p_j c_j, \varepsilon_j) > V_i(A_i, y - p_i c_i, \varepsilon_i) \; ; \quad \forall i \neq j \tag{7}$$

In terms of probabilities, we write:

$$P\{\delta_j = 1\} = P\{V_j(A_j, y - p_j c_j, \varepsilon_j) > V_i(A_i, y - p_i c_i, \varepsilon_i); \forall i \neq j\} \tag{8}$$

The exact specification of the econometric model depends on how the random elements, ε, enter the conditional indirect utility function and the distributional assumption. Let us divide the task into two parts: (i) specification of the utility function, and (ii) specification of the probability distribution of the error term.

3.1 Specification of the utility function

The most common assumption is that the error term enters the utility function as an additive term. This assumption, although restrictive, greatly simplifies the computation of the results and the estimation of welfare measures. In section 3.2 we present a random parameter model, which is an example of a model with the stochastic component entering the utility function via the slope coefficients, i.e. non-additively (Hanemann, 1999).

Under an additive formulation the probability of choosing alternative j can be written as:

$$P\{\delta_j = 1\} = P\{V_j(A_j, y - p_j c_j) + \varepsilon_j > V_i(A_i, y - p_i c_i) + \varepsilon_i; \forall i \neq j\} \tag{9}$$

In order to specify a utility function, we need to specify the functional form for $V(\ldots)$ and to select the relevant attributes (A_i) that determine the utility derived from each alternative. These attributes should then be included in the choice experiment.

When choosing the functional form, there is a trade-off between the benefits of assuming a less restrictive formulation and the complications that arise from doing so. This is especially relevant for the way income enters the utility function. A simpler functional form (e.g. linear in income) makes estimation of the parameters and calculation of welfare effects easier, but the estimates are based on restrictive assumptions. One crucial assumption concerns how income enters the utility function. Usually a constant marginal utility of income has been assumed (Herriges and Kling, 1999), not because this seems like the most reasonable assumption, but mainly because of difficulties with estimating welfare measures without this assumption. We will postpone the discussion about how income enters the utility function to section 6, where we investigate in more detail the implications of the chosen functional form on the calculation of exact welfare estimates. Note, that a linear in parameters utility function does not rule out non-linear effects on utility, for example through a quadratic utility function. However, as discussed by Layton (2001) such an approach is not likely to be suitable when the choice experiments includes both small and large changes in attributes (utility). Other approaches could then be considered such as the Box-Cox or the inverse hyperbolic sine transformation (Layton, 2001).

Regarding the selection of attributes it is important to be aware that the collected data come from a specific design based on a priori assumptions regarding estimable interaction effects between attributes. Once the experiment has been conducted we are restricted to testing for only those effects that were considered in the design. This shows the importance of focus groups and pilot studies when constructing the experiment.

3.2 Specification of the probability distribution of the error term

The most common model used in applied work has been the Multinomial Logit model (MNL). This model relies on restrictive assumptions, and its popularity rests on its simplicity of estimation. We begin by introducing the MNL model and discussing its limitations, and then we introduce less restrictive models. Suppose that the choice experiment consists of M choice sets, where each choice set, S_m, consists of K_m alternatives, such that $S_m = \{A_{1m},\ldots,A_{Km}\}$, where A_i is a vector of attributes. We can then write the choice probability for alternative j from a choice set S_m as

$$P\{\delta_j = 1 \mid S_m\} = P\{V_j(A_{jm}, y - p_j c_j) + \varepsilon_j > V_i(A_{im}, y - p_i c_i) + \varepsilon_i; \forall i \in S_m; \forall i \neq j\} = P\{V_j(\ldots) + \varepsilon_j - V_i(\ldots) > \varepsilon_i; \forall i \in S_m; \forall i \neq j\} \quad (10)$$

The MNL model assumes that the random components are independently and identically distributed with an extreme value type I distribution (Gumbel). The variance of an extreme value distribution is $\mathrm{var}_\varepsilon = \pi^2/6\mu^2$, where μ is a scale parameter and δ is a location parameter.[3] If we assume that the random components are extreme value distributed with mean zero and variance $\pi^2/6$, the choice probability in equation (10) can be written as:

$$P(\delta_j = 1 \mid S_m, \beta) = \frac{\exp(\mu V_j)}{\sum_{i \in S_m} \exp(\mu V_i)} \quad (11)$$

In principle, the size of the scale parameter is irrelevant when it comes to the choice probability of a certain alternative (Ben-Akiva and Lerman, 1985), but by looking at equation (11) it is clear that the true parameters are confounded with the scale parameter. Moreover, it is not possible to identify this parameter from the data. For example, if the inverse of the scale parameter is doubled, the estimated parameters in the linear specification will adjust to double their previous values.[4] The presence of a scale parameter raises several issues for the analysis of the estimates. First consider the variance of the error term: $\mathrm{var}_\varepsilon = \pi^2/6\mu^2$. An increase in the scale reduces the variance; therefore high fit models have larger scales. The two extreme cases are $\mu \to 0$ where, in a binary model, the choice probabilities become ½, and $\mu \to \infty$ where the model becomes completely deterministic (Ben-Akiva and Lerman, 1985). Second, the impact of the scale parameter on the estimated coefficients imposes restrictions on their interpretation. All parameters within an estimated model have the same scale and therefore it is valid to compare their signs and relative sizes. On the other hand, it is not possible to directly compare parameters from different models as the scale parameter and the true parameters are confounded. Nevertheless, it is possible to compare estimated parameters from two different data sets, or to combine data sets (for example stated and revealed preference data). Swait and Louviere (1993) show how to estimate the ratio of scale parameters for two different data sets. This procedure can then be used to compare different models or to pool data from different sources (see e.g. Adamowicz *et al.*, 1994; Ben-Akiva and Morikawa, 1990).

There are two problems with the MNL specification: (i) the alternatives are independent and (ii) there is a limitation in modelling variation in taste among respondents. The first problem arises because of the IID assumption (constant variance), which results in the independence of irrelevant alternatives (IIA) property. This property states that the ratio of choice probabilities between two alternatives in a choice set is unaffected by changes in that choice set. If this assumption is violated the MNL should not be used. One type of model that relaxes the homoscedasticity assumption of the MNL model is the nested MNL model. In this model the alternatives are placed in subgroups, and the variance is allowed to differ between the sub-groups but it is assumed to be the same within each group. An alternative specification is to assume that error terms are independently, but non-identically, distributed type I extreme value, with scale parameter μ_i (Bhat, 1995). This would allow for different cross elasticities among all pairs of alternatives, i.e. relaxing the IIA restriction. Furthermore, we could also model heterogeneity in the covariance among nested alternatives (Bhat, 1997).

The second problem arises when there is taste variation among respondents due to observed and/or unobserved heterogeneity. Observed heterogeneity can be incorporated into the systematic part of the model by allowing for interaction between socio-economic characteristics and attributes of the alternatives or constant terms. However, the MNL model can also be generalized to a so-called mixed MNL model in order to further account for unobserved heterogeneity. In order to illustrate this type of model, let us write the utility function of alternative j for individual q as:

$$U_{jq} = \beta x_{jq} + \varepsilon_{jq} = \bar{\beta} x_{jq} + \tilde{\beta} x_{jq} + \varepsilon_{jq} \,. \tag{12}$$

Thus, each individual's coefficient vector β is the sum of the population mean $\bar{\beta}$ and individual deviation $\tilde{\beta}_q$. The stochastic part of utility, $\tilde{\beta}_q x_{jq} + \varepsilon_{jq}$, is correlated among alternatives, which means that the model does not exhibit the IIA property. If the error terms are IID standard normal we have a random parameter multinomial probit model. If instead the error terms are IID type I extreme value, we have a random parameter logit model.

Let tastes, β, vary in the population with a distribution with density $f(\beta \mid \theta)$, where θ is a vector of the true parameters of the taste distribution. The unconditional probability of alternative j for individual q can then be expressed as the integral of the conditional probability in equation (11) over all values of β:

$$P_q(\delta_j = 1 \mid \theta) = \int P_q(j \mid \beta) f(\beta \mid \theta) d\beta = \int \frac{\exp(\mu \beta x_{jq})}{\sum_{i=1}^{K_m} \exp(\mu \beta x_{iq})} f(\beta \mid \theta) d\beta. \tag{13}$$

In general the integrals in equation (13) cannot be evaluated analytically, and we have to rely on simulation methods for the probabilities (see e.g. Brownstone and Train, 1999).

When estimating these types of models we have to assume a distribution for each of the random coefficients. It may seem natural to assume a normal distribution. However, for many of the attributes it may be reasonable to expect that all respondents have the same sign for their coefficients. In this case it may be more sensible to assume a log-normal distribution. For example, if we assume that the price coefficient is log-normally distributed, we ensure that all individuals have a non-positive price coefficient.

In most choice experiments, respondents make repeated choices, and we assume that the preferences are stable over the experiment. Consequently, the utility coefficients are allowed to vary among individuals but they are constant among the choice situations for each individual (Revelt and Train, 1998; Train, 1998). It is also possible to let the coefficients for the individual vary over time, in this case among the choice situations in the survey. This type of specification would be valid if we suspect fatigue or learning effects in the survey.

McFadden and Train (2000) show that under some mild regularity conditions any discrete choice model derived from random utility maximization has choice probabilities that can be approximated by a mixed MNL model. This is an interesting result because mixed MNL models can then be used to approximate difficult parametric random utility models, such as the multinomial probit model, by taking the distributions underlying these models as the parameter distributions.

4. Design of a choice experiment

There are four steps involved in the design of a choice experiment: (i) definition of attributes, attribute levels and customisation, (ii) experimental design, (iii) experimental context and questionnaire development and (iv) choice of sample and sampling strategy. These four steps should be seen as an integrated process with feedback. The development of the final design involves repeatedly conducting the steps described here, and incorporating new information as it comes along. In this section, we focus on the experimental design and the context of the experiment, and only briefly discuss the other issues.

4.1 Definition of attributes and levels

The first step in the development of a choice experiment is to conduct a series of focus group studies aimed at selecting the relevant attributes. The focus studies could be in terms of verbal protocols, group discussion and actual surveys, see for example Layton and Brown (1998) for a discussion of how to use focus groups for pretesting the question format and attributes. A starting point involves studying the attributes and attribute levels used in previous studies and their importance in the choice decisions. Additionally, the selection of attributes should be guided by the attributes that are expected to affect respondents' choices, as well as those attributes that are policy relevant. This information forms the base for which attributes and relevant attribute levels to include in the first round of focus group studies.

The task in a focus group is to determine the number of attributes and attribute levels, and the actual values of the attributes. As a first step, the focus group studies should provide information about credible minimum and maximum attribute levels. Additionally, it is important to identify any possible interaction effect between the attributes. If we want to calculate welfare measures, it is necessary to include a monetary attribute such as a price or a cost. In such a case, the focus group studies will indicate the best way to present a monetary attribute. Credibility plays a crucial role and the researcher must ensure that the attributes selected and their levels can be combined in a credible manner. Hence, proper restrictions may have to be imposed (see e.g. Layton and Brown, 1998).

Customization is an issue in the selection of attributes and their levels. It is an attempt to make the choice alternatives more realistic by relating them to actual levels. If possible an alternative with the attribute levels describing today's situation should be included which would then relate the other alternatives to the current situation. An alternative is to directly relate some of the attributes to the actual level. For example, the levels for visibility could be set 15 per cent higher and 15 per cent lower than today's level (Bradley, 1988).

The focus group sessions should shed some light on the best way to introduce and explain the task of making a succession of choices from a series of choice sets. As Layton and

Brown (1998) explain, choosing repeatedly is not necessarily a behavior that could be regarded as obvious for all goods. When it comes to recreation, for example, it is clear that choosing a site in a choice set does not preclude choosing another site given different circumstances. However, in the case of public goods, such repeated choices might require further justification in the experiment.

A general problem with applying a choice experiment to an environmental good or to an improvement in health status is that respondents are not necessarily familiar with the attributes presented. Furthermore, the complexity of a choice experiment in terms of the number of choice sets and/or the number of attributes in each choice set may affect the quality of the responses; this will be discussed in Section 4.3. Basically, there is a trade-off between the complexity of the choice experiment and the quality of the responses. The complexity of a choice experiment can be investigated by using verbal protocols, i.e. by asking the individual to read the survey out loud and/or to think aloud when responding; this approach has been used in CVM surveys (e.g. Schkade and Payne, 1993), thereby identifying sections that attract the readers' attention and testing the understanding of the experiment.

4.2 Experimental design

Experimental design is concerned with how to create the choice sets in an efficient way, i.e. how to combine attribute levels into profiles of alternatives and profiles into choice sets. The standard approach in marketing, transport and health economics has been to use so-called orthogonal designs, where the variations of the attributes of the alternatives are uncorrelated in all choice sets. Recently, there has been a development of optimal experimental designs for choice experiments based on multinomial logit models. These optimal design techniques are important tools in the development of a choice experiment, but there are other more practical aspects to consider. We briefly introduce optimal design techniques for choice experiments and conclude by discussing some of the limitations of statistical optimality in empirical applications.

A design is developed in two steps: (i) obtaining the optimal combinations of attributes and attribute levels to be included in the experiment and (ii) combining those profiles into choice sets. A starting point is a full factorial design, which is a design that contains all possible combinations of the attribute levels that characterize the different alternatives. A full factorial design is, in general, very large and not tractable in a choice experiment. Therefore we need to choose a subset of all possible combinations, while following some criteria for optimality and then construct the choice sets. In choice experiments, design techniques used for linear models have been popular. Orthogonality in particular has often been used as the principle part of an efficient design. More recently researchers in marketing have developed design techniques based on the *D*-optimal criteria for non-linear models in a choice experiment context. *D*-optimality is related to the covariance matrix of the *K*-parameters, defined as

$$D - efficiency = [|\Omega|^{1/K}]^{-1} \tag{14}$$

Huber and Zwerina (1996) identify four principles for an efficient design of a choice experiment based on a non-linear model: (i) orthogonality, (ii) level balance, (iii) minimal overlap and (iv) utility balance. Level balance requires that the levels of each attribute occur with equal frequency in the design. A design has minimal overlap when an attribute level does not repeat itself in a choice set. Finally, utility balance requires that the utility of each alternative within a choice set is equal. The last property is important since the larger the difference in utility between the alternatives the less information is extracted from that specific choice set. At the same time, this principle is difficult to satisfy since it requires prior knowledge about the true distribution of the parameters. The theory of optimal design for choice experiments is related to optimal design of the bid vector in a CVM survey. The optimal design in a CVM survey depends on the assumption regarding the distribution of WTP (see e.g. Duffield and Patterson, 1991; Kanninen, 1993).

Several design strategies explore some or all of the requirements for an efficient design of a choice experiment. Kuhfeld *et al.* (1994) use a computerized search algorithm to minimize the D-error in order to construct an efficient, but not necessarily orthogonal, linear design. However, these designs do not rely on any prior information about the utility parameters and hence do not satisfy utility balance. Zwerina *et al.* (1996) adapt the search algorithm of Kuhfeld *et al.* (1994) to the four principles for efficient choice designs as described in Huber and Zwerina (1996).[5] In order to illustrate their design approach it is necessary to return to the MNL model. McFadden (1974) showed that the maximum likelihood estimator for the conditional logit model is consistent and asymptotically normally distributed with the mean equal to β and a covariance matrix given by:

$$\Omega = (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1} = [\sum_{n=1}^{N}\sum_{j=1}^{J_n} \mathbf{z}'_{jn} P_{jn} \mathbf{z}_{jn}]^{-1} \tag{15}$$

$$\text{where} \qquad \mathbf{z}_{jn} = \mathbf{x}_{jn} - \sum_{i=1}^{J_n} \mathbf{x}_{in} P_{in}$$

This covariance matrix, which is the main component in the *D*-optimal criteria, depends on the true parameters in the utility function, since the choice probabilities, P_{in} , depend on these parameters. Consequently, an optimal design of a choice experiment depends, as in the case of the optimal design of bid values in a CV survey, on the value of the true parameters of the utility function. Adapting the approach of Zwerina *et al.* (1996) consequently requires prior information about the parameters.[6] Carlsson and Martinsson (2003) discuss strategies for obtaining this information, which includes results from other studies, expert judgments, pilot studies and sequential designs strategies. Kanninen (1993) discusses a sequential design approach for closed-ended CVM surveys and she finds that this approach improves the efficiency of the design. A similar strategy can be used in designing choice experiments. The response data from the pilot studies and the actual choice experiment can be used to estimate the value of the parameters. The design can then be updated during the experiment depending on the results of the estimated parameters. The results from these estimations may not only require a new design, but changes in the attribute levels as well. There are other simpler design strategies which do not directly require information about the parameters. However, in all cases, some information about the shape of the utility function is needed in order to make sure that the individuals will make trade-offs between attributes. The only choice experiment in environmental valuation that has adopted a *D*-optimal design strategy is Carlsson and Martinsson (2001). In a health economic application by Johnson *et al.* (2000), a design partly based on *D*-optimal criteria is applied.

Kanninen (2001) presents a more general approach to optimal design than Zwerina *et al.* (1996). In her design, the selection of the number of attribute levels is also a part of the optimal design problem. Kanninen (2002) shows that in a *D*-optimal design each attribute should only have two levels, even in the case of a multinomial choice experiment, and that the levels should be set at the two extreme points of the distribution of each attribute. Furthermore, Kanninen (2002) shows that for a given number of attributes and alternatives, the *D*-optimal design results in certain response probabilities. This means that updating the optimal design is simpler than updating the design presented in Zwerina *et al.* (1996). In order to achieve the desired response probabilities the observed response probabilities from previous applications have to be calculated, and a balancing attribute is then included. This type of updating was adopted by Steffens *et al.* (2000) in a choice experiment on bird watching. They found that the updating improved the efficiency of the estimates.

There are several problems with these more advanced design strategies due to their complexity, and it is not clear whether the advantages of being more statistically efficient outweigh the problems. The first problem is obtaining information about the parameter values. Although some information about the coefficients is required for other design strategies as well, more elaborate designs based on utility balance are more sensitive to the quality of information used, and incorrect information on the parameters may bias the final estimates. Empirically, utility balance makes the choice harder for the respondents, since they have to choose from alternatives that are very close in terms of utility. This might result in a random choice. The second problem is that the designs presented here are based on a conditional logit model where, for example, homogeneous preferences are assumed. Violation of this assumption may bias the estimates. The third problem is the credibility of different combinations of attributes. If the correlation between attributes is ignored, the choice sets may not be credible to the respondent (Johnson *et al.*, 2000; and Layton and Brown, 1998). In this case it may be optimal to remove such combinations although it would be statistically efficient to include them.

4.3 Experimental context, test of validity and questionnaire development

In the previous section, we addressed optimal design of a choice experiment from a statistical perspective. However, in empirical applications there may be other issues to consider in order to extract the maximum amount of information from the respondents.

Task complexity is determined by factors such as the number of choice sets presented to the individual, the number of alternatives in each choice set, the number of attributes describing those alternatives and the correlation between attributes for each alternative (Swait and Adamowicz, 1996). Most authors find that task complexity affects the decisions (Adamowicz *et al.*, 1998a; Bradley, 1988). Mazotta and Opaluch (1995) and Swait and Adamowicz (1996) analyze task complexity by assuming it affects the variance term of the model. The results of both papers indicate that task complexity does in fact affect the variance, i.e. an increased complexity increases the noise associated with the choices. Task complexity can also arise when the amount of effort demanded when choosing the preferred alternative in a choice set may be so high that it exceeds the ability of the respondents to select their preferred option. The number of attributes in a choice experiment is studied by Mazotta and Opaluch (1995) and they find that including more than 4 to 5 attributes in a choice set may lead to a severe detriment to the quality of the data collected due to the task complexity.

In complex cases, respondents may simply answer carelessly or use some simplified lexicographic decision rule. This could also arise if the levels of the attributes are not suffi-