

Using Choice Experiments for Non-Market Valuation

Francisco Alpizar
Fredrik Carlsson
Peter Martinsson^A

Working Papers in Economics no. 52
June 2001
Department of Economics
Göteborg University

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.

Keywords: Choice experiments; non-market goods; stated preference methods; valuation.

JEL classification: H41, D61, Q20

We have received valuable comments from Gardner Brown, Henrik Hammar, G.S. Haripriya, Gunnar Köhlin, David Layton, Karl-Göran Mäler, Olof Johansson-Stenman and Thomas Sterner. Financial support from the Swedish International Development Agency, the Bank of Sweden Tercentenary Foundation and the Swedish Transport and Communications Research Board is gratefully acknowledged.

^A Department of Economics, Gothenburg University, Box 640, SE-405 30 Gothenburg, Sweden. E-mail: Francisco.Alpizar@economics.gu.se, Fredrik.Carlsson@economics.gu.se, Peter.Martinsson@economics.gu.se.

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 another branch, the stated preference method, during the last 20 years.

Stated preference method 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 and Hanemann, 1994, for a critical assessment). 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 and 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.

>>> Insert Figure 1

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) and 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 choices (Hanemann, 1984 and 1999).

In many situations, 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 in a consistent manner.

In general, choice experiments applied to non-marketed goods assume a specific continuous dimension as 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). A CVM survey assumes the same specific continuous dimension since the objective is to obtain the value of a certain predefined program that includes a given continuous decision. 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.

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

$$\begin{aligned}
 & \text{Max}_{c,x} U[c_1(A_1), \dots, c_N(A_N); z] \\
 & \text{s.t.} \quad \text{i. } y = \sum_{i=1}^N p_i c_i(A_i) + z \quad (1) \\
 & \quad \quad \text{ii. } c_i c_j = 0, \forall i \neq j \\
 & \quad \quad \text{iii. } z \geq 0, c_i(A_i) \geq 0 \text{ for at least one } i
 \end{aligned}$$

where, $U[.]$ 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 ; p_i is the price of each profile; z is a composite bundle of ordinary goods with its price normalized to 1 and 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. For example, one such profile could be a visit to a national park in a rainforest, with 50 kms of marked walking tracks through the park and a visitor centre. Additionally, the choice of any profile is for a fixed, and given, amount of it, e.g. a day or a unit. There are N such profiles, where N is in principle given by all relevant profiles. However, in practice, N will be determined depending on the type of design used to construct the 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, including the given continuous dimension, for example price per day or per visit.

3. Restriction *ii* defines the number of alternatives that can be chosen. 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.¹ 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.

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:

$$z = y - p_j c_j \quad (2)$$

5. Restriction *iii* specifies that the individual will choose a non-negative quantity of the composite good and the goods being studied. 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 ($c_i > 0$ 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. $c_j = c_j^{fixed}$, $c_i = 0 \quad \forall i \neq j$, where c_j^{fixed} is the fixed continuous measure of the given profile. 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:

$$\text{if } c_i = 0, \text{ then } \frac{\partial U}{\partial A_i} = 0, \quad \forall i \neq j. \quad (3)$$

¹ In the case of perfect substitutes, it is the form of the utility function rather than restriction *ii* that ensures a single choice.

Using (2) and (3) we can write the conditional utility function, given $c_j = c_j^{fixed}$ as:

$$U_j = V_j[c_j(A_j), p_j, y, z] = V_j(A_j, y - p_j c_j). \quad (4)$$

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

$$V[A, p, y] = \max[V_1(A_1, y - p_1 c_1), \dots, V_N(A_N, y - p_N c_N)], \quad (5)$$

where the function $V[\cdot]$ captures the discrete choice, given an exogenous and fixed quantitative assumption regarding the continuous choice. Thus, it follows that the individual chooses the profile j if and only if:

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

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 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 $V_i^1[c_i(A_i^1), y - bid] > V_i^0[c_i(A_i^0), y]$, 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 if and only if:

$$V_j(A_j, y - p_j c_j, \epsilon_j) > V_i(A_i, y - p_i c_i, \epsilon_i); \forall i \neq j \quad (7)$$

In terms of probabilities, we write:

$$P\{\text{choose } j\} = P\{V_j(A_j, y - p_j c_j, \epsilon_j) > V_i(A_i, y - p_i c_i, \epsilon_i); \forall i \neq j\} \quad (8)$$

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:

$$P\{\text{choose } j\} = P\{V_j(A_j, y - p_j c_j) + \epsilon_j > V_i(A_i, y - p_i c_i) + \epsilon_i; \forall i \neq j\} \quad (9)$$

In order to specify a utility function, we need to specify the functional form for $V(\dots)$ 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 (Ben-Akiva and Lerman, 1985). Most often researchers have been inclined to use a simpler linear in the parameters utility function. Since the need for simple functional forms is linked to the estimation of welfare measures, we will postpone the discussion to section 6, where we investigate in more detail the implications of the chosen functional form on the calculation of exact welfare estimates.

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 (MNL) model. 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 $S_m = \{A_{1m}, \dots, A_{K_m}\}$, 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

$$\begin{aligned} P\{j | S_m\} &= P\{V_j(A_{jm}, y - p_j c_j) + \varepsilon_j > V_i(A_{im}, y - p_i c_i) + \varepsilon_i; \forall i \in S_m\} = \\ &= P\{V_j(\dots) + \varepsilon_j - V_i(\dots) > \varepsilon_i; \forall i \in S_m\}. \end{aligned} \quad (10)$$

We can then express this choice probability in terms of the joint cumulative density function of the error term as:

$$P(j | S_m) = CDF_{\varepsilon | S_m}(V_j + \varepsilon_j - V_1, V_j + \varepsilon_j - V_2, \dots, V_j + \varepsilon_j - V_n). \quad (10')$$

The MNL model assumes that the random components are independently and identically distributed with an extreme value type I distribution (Gumbel). This distribution is characterized by a scale parameter μ and location parameter δ .² The scale parameter is related to the variance of the distribution such that $\text{var}_\epsilon = \pi^2/6\mu^2$. If we assume that the random components are extreme value distributed, the choice probability in (10) can be written as:

$$P(j | S_m, \beta) = \frac{\exp(\mu V_j)}{\sum_{i \in S_m} \exp(\mu V_i)}. \quad (11)$$

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 scale is doubled, the estimated parameters in the linear specification will adjust to double their previous values.³ The presence of a scale parameter raises several issues for the analysis of the estimations. First consider the variance of the error term: $\text{var}_\epsilon = \pi^2/6\mu^2$. An increase in the scale reduces the variance; therefore high fit models have larger scales. The two extreme cases are $\mu \rightarrow 0$ where, in a binary model, the choice probabilities become $1/2$, and $\mu \rightarrow \infty$ 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

² In practice, the distribution chosen is the standard Gumbel distribution with $\mu = 1$ and $\delta = 0$.

³ In a linear specification, $\beta^{estimated} = \mu \beta^{true}$, and $\beta^{estimated}$ will adjust to changes in μ . The issue of the scale parameter is not specific to multinomial models and Gumbel distributions. For the case of probit

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 homoskedasticity 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 subgroups 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:

$$U_{jq} = \beta x_{jq} + \varepsilon_{jq} = \bar{\beta} x_{jq} + \tilde{\beta}_q x_{jq} + \varepsilon_{jq}. \quad (12)$$

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

models, the scale parameter of the normal distribution is $1/\sigma$. Everything we say here about the scale parameter of the Gumbel distribution applies to nested MNL and probit models as well.

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 $f(\beta | \theta)$, 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 (11) over all values of β :

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

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. 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).

Customisation 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% higher and 15% 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

$$D - efficiency = [|\Omega|^{1/K}]^{-1}. \quad (14)$$

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).⁴ 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:

$$\Omega = (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1} = \left[\sum_{n=1}^N \sum_{j=1}^{J_m} \mathbf{z}'_{jn} P_{jn} \mathbf{z}_{jn} \right]^{-1}, \quad (15)$$

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

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.⁵ 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. Carlsson and Martinsson (2000) 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 update during the experiment depending on the results of the estimated parameters. The results from these estimations may not only require a new

⁴ The SAS code is available at <ftp://ftp.sas.com/techsup/download/technote/ts643/>.

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 (2001) 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.⁶ Furthermore, Kanninen (2001) 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

⁵ This is an important difference from the design of linear models where the covariance matrix is proportional to the information matrix, i.e. $\Omega = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$.

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 sufficiently differentiated to ensure trade-offs. Another possibility is 'yea' saying or 'nay' saying, where the respondent, for example, always opt for the most environmentally friendly alternative. Finally, lexicographic orderings may be an indication of strategic behaviour of the respondent. In practice, it is difficult to separate these cases from preferences that are genuinely lexicographic, in which case the respondents have a

⁶ The design is derived under the assumption that all attributes are quantitative variables.

ranking of the attributes, but the choice of an alternative is based solely on the level of their most important attribute. Genuine lexicographic preferences in a choice experiment are not a problem, although they provide us with little information in the analysis compared to the other respondents. However, if a respondent chooses to use a lexicographic strategy because of its simplicity, systematic errors are introduced, which may bias the results. One strategy for distinguishing between different types of lexicographic behaviour is to use debriefing questions, where respondents are asked to give reasons why they, for example, focused on only one or two of the attributes in the choice experiment. However, in a thoroughly pre-tested choice experiment using focus groups and pre tests, these problems should have been detected and corrected.

An issue related to task complexity in is the stability of preferences. In choice experiments the utility function of each individual is assumed to be stable throughout the experiment. The complexity of the exercise might cause violations of this assumption, arising from learning and fatigue effects. Johnson et al. (2000) test for stability by comparing responses to the same choice sets included both at the beginning and at the end of the experiment. They find a strong indication of instability of preferences. However, there is a potential problem of confounding effects of the sequencing of the choice sets and the stability of the preferences. An alternative approach, without the confounding effect, is applied in Carlsson and Martinsson (2001) in a choice experiment on donations to environmental projects. In their exercise, half of the respondents receive the choice sets in the order {A,B} and the other half in the order {B,A}. A test for stability is then performed by comparing the preferences obtained for the choices in subset A, when it was given in the sequence {A,B}, with the preferences obtained when the choices in subset A were given in the sequence {B,A}. This can then be formally tested in a likelihood ratio test between the pooled model of the choices in subset A and the separate groups. A similar test can be performed for subset B. By using this method Carlsson and Martinsson (2001) find only a minor problem with instability of preferences. Layton and Brown (2000) conduct a similar test of stability in a choice experiment on policies for mitigating impacts of global climate change; they did not reject the hypothesis of stable preferences. Bryan et al. (2000) compare responses in the same way, but with the objective of testing for reliability, and find that 57 percent of the respondents did not change their responses when given the same

choice set in a two-part choice experiment. Furthermore, in an identical follow-up experiment two weeks after the original experiment, 54 percent of the respondents made the same choices on at least eleven out of twelve choice situations.

Another issue to consider in the development of the questionnaire is whether or not to include a base case scenario or an opt-out alternative. This is particularly important if the purpose of the experiment is to calculate welfare measures. If we do not allow individuals to opt for a status quo alternative, this may distort the welfare measure for non-marginal changes. This decision should, however, be guided by whether or not the current situation and/or non-participation is a relevant alternative. A non-participation decision can be econometrically analysed by e.g. a nested logit model with participants and non-participants in different branches (see e.g. Blamey et al., 2000). A simpler alternative is to model non-participation as an alternative where the levels of the attributes are set to the current attribute levels. Another issue is whether to present the alternatives in the choice sets in a generic (alternatives A, B, C) or alternative specific form (national park, protected area, beach). Blamey et al. (2000) discuss advantages of these two approaches and compare them in an empirical study. An advantage of using alternative specific labels is familiarity with the context and hence the cognitive burden is reduced. However, the risk is that the respondent may not consider trade-offs between attributes. This approach is preferred when the emphasis is on valuation of the labelled alternatives. An advantage of the generic model is that the respondent is less inclined to only consider the label and thereby focus more on the attributes. Therefore, this approach is preferred when the emphasis is on the marginal rates of substitution between attributes.

In the random utility model, unobservable effects are modelled by an error term and, in general, we assume that respondents have rational, stable, transitive and monotonic preferences. Also, we assume they do not have any problems in completing a choice experiment, and that there are no systematic errors, such as respondents getting tired or changing their preferences as they acquire experience with the experiment, i.e. learning effects. Internal tests of validity are designed to check these standard assumptions. These tests can be directly incorporated into the design of an experiment. There have been several validity tests of choice experiments in the marketing and transport literature, for example Ben-Akiva et al. (1992) and Leigh et al. (1984). The evidence

from a large proportion of studies is that choice experiments generally pass these tests of validity. However, it is not obvious that these results carry over into choice experiments done in an environmental or health economic context. The reason is that these non-market goods in many respects differ from, for example, transportation, which is a good that most respondents are familiar with. It is therefore of importance to test the validity of choice experiments in the context of valuation of general non-marketed goods. Since there are few applications of choice experiments in valuation, few tests of internal validity have been performed.

In order to test for transitive preferences, we have to construct such a test. For example, in the case of a pair-wise choice experiment we have to include three specific choice sets: (1) Alt. 1 versus Alt. 2, (2) Alt. 2 vs. Alt. 3, and (3) Alt. 1 vs. Alt. 3. For example if the respondent chooses Alt. 1 in the first choice set and Alt. 2 in the second choice set, then Alt. 1 must be chosen in the third choice if the respondent has transitive preferences. Carlsson and Martinsson (2001) conduct tests of transitivity and they do not find any strong indications of violations. Internal tests of monotonicity can also be implemented in a choice experiment and in a sense tests of monotonicity are already built-in in a choice experiment as the level of an attribute changes in an experiment. Comparing the expected sign to the actual sign and significance of the coefficient can be seen as a weak test monotonicity. Johnson et al. (2000) discuss a simple test of dominated pair, which simply tests if a respondent chooses a dominated alternative.

4.4 Sample and Sampling Strategy

The choice of survey population obviously depends on the objective of the survey. Given the survey population, a sampling strategy has to be determined. Possible strategies include a simple random sample, a stratified random sample or a choice-based sample. A simple random sample is generally a reasonable choice. One reason for choosing a more specific sampling method may be the existence of a relatively small but important sub-group which is of particular interest to the study. Another reason may be to increase the precision of the estimates for a particular sub-group. In practice the selection of sample strategy and sample size is also largely dependent on the budget available for the survey.

Louviere et al. (2000) provide a formula to calculate the minimum sample size. The size of the sample, n , is determined by the desired level of accuracy of the estimated probabilities, \hat{p} . Let p be a true proportion of the relevant population, a is the percentage of deviation between \hat{p} and p that can be accepted and α is the confidence level of the estimations such that: $\Pr(|\hat{p} - p| \leq ap) \geq \alpha$ for a given n . Given this, the minimum sample size is defined as:

$$n \geq \frac{1-p}{pa^2} \Phi^{-1}\left(\frac{1+\alpha}{2}\right). \quad (16)$$

Note that n refers to the size of the sample and not the number of observations. Since each individual makes r succession of choices in a choice experiment, the number of observations will be much larger (a sample of 500 individuals answering 8 choice sets each will result in 4000 observations). One of the advantages of choice experiments is that the amount of information extracted from a given sample size is much larger than, for example, using referendum based methods and, hence, the efficiency of the estimates is improved. The formula above is only valid for a simple random sample and with independency between the choices. For a more detailed look at this issue see e.g. Ben-Akiva and Lerman (1985). In a health economic context, the availability of potential respondents can in certain cases be limited and hence the equation above can be used to solve for a , i.e. the percentage deviation between \hat{p} and p that we must accept given the sample size used.

5. Elicitation of preferences in choice experiments

There has been an extensive discussion about the possibility of eliciting preferences for non-market goods in hypothetical surveys. While the discussion has focused on CVM (see e.g. Diamond and Hausman, 1994 and Hanemann, 1994) most of the results are valid for choice experiments as well. We believe that there are particular problems with measuring so-called non-use values in hypothetical surveys. We do not take the position that non-use values should not be measured, but rather that there are some inherent problems with measuring these values. The reason for this is that non-use values are largely motivated by "purchase of moral satisfaction" (Kahneman and Knetsch, 1992) and "warm glow" (Andreoni, 1989), and that they often involve an "important perceived

ethical dimension" (Johansson-Stenman and Svedsäter, 2001). We are not questioning these values per se; on the contrary, they may even be important shares of total value. The problem is that the cost of acquiring a "warm glow" or a satisfaction of acting ethical is much lower in a hypothetical survey situation than in an actual situation. This leaves us in a difficult position, since stated preference methods are essentially the only methods available for measuring non-use values. However, there are reasons to believe that choice experiments may be less prone to trigger this type of behaviour than CVM surveys. The reason for this is that in a choice experiment individuals have to make trade-offs between several attributes, several of which may contain non-use values.

Another issue involves incentives for truthfully revealing preferences in hypothetical surveys. Carson et al. (1999) argue that given a consequential survey a binary discrete choice is incentive compatible for the cases of (i) a new public goods with coercive payments, (ii) the choice between two public goods and (iii) a change in an existing private or quasi-public good. A consequential survey is defined as one that is perceived by the respondent as something that may potentially influence agency decisions, as well as one where the respondent cares about the outcome. The problem arises when the individual faces not one but a sequence of binary choices. Let us assume we are dealing with a public good, i.e. everybody will enjoy the same quantity and composition of the good after the government has decided its provisions. The respondents could then perceive the sequence of binary choices as a voting agenda, and, if they expect one of their less preferred outcomes to be chosen, they would have an incentive to misrepresent their true preferences. The same type of problem arises with multinomial choices. If only one alternative is to be chosen, the multinomial choice is reduced to a binary choice between the two alternatives that the respondent believes are most likely to be chosen, even if these two alternatives are not the most preferred ones. The problem with these incentives is that the preference profile constructed from the survey is not a reflection of the true preferences, but rather a reflection of strategic behaviour. The choice experiment would then be flawed and any welfare estimate would not be reliable. This issue clearly demands attention from researchers, although we believe that the importance of these results should not be overemphasized.

It is in general more difficult to behave strategically in a choice experiment, when compared to a CVM survey. In a CVM survey the respondent "only" has to consider a

single change in a project involving a certain payment. A typical choice experiment consists of two to four alternatives, where each alternative is described by at least three or four attributes. The selection of all attributes is done under the premise that they are relevant determinants of choice behaviours of individuals and the levels are set such that they imply meaningful changes in utility. Furthermore, there is, generally, no clearly identifiable agenda in a sequence of choices, where almost all levels of attributes change from one choice set to another. Thus, it is more difficult for a respondent to behave strategically in a choice experiment. First they need to create an expectation regarding the values of each of the alternatives in the choice set. Based on this expectation they need to calculate the decision weights for each pair-wise decision. Of particular importance is the fact that most choice experiments, as well as CVM surveys, deal with situations that are not familiar to the respondent. The fact that there are no markets for some of the evaluated goods means that there is limited, if any, information about the preferences of other individuals. There are seldom any opinion polls, prices or other types of information that the respondent can use. Thus in general the respondent is in an unfamiliar situation and with limited prior information on the preferences of others.

The assumption that each respondent has perfect information regarding the preferences of other respondents is unrealistic and the question is how uncertainty affects the incentives for truthful revelation. Here we illustrate this with the model of Gutowski and Georges (1993). Each respondent has a subjective value of each of three alternatives, a_1 , a_2 and a_3 . A respondent with the preference ordering $a_1 \succ a_2 \succ a_3$, where the subjective value of the most preferred alternative, $v(a_1)$, is equal to one, and the subjective value of the least preferred alternative, $v(a_3)$, is equal to zero. The subjective value of a_2 , $v(a_2)$, is uniformly distributed between zero and one. Any particular respondent does not have perfect information regarding other respondents' preference orderings, but is assumed to form subjective beliefs regarding the chances of various scenarios. These are represented by decision weights that measure the extent to which each of the pair wise competitions are incorporated into a respondent's choice among admissible strategies. There are three possible pair-wise competitions, and consequently three decision weights, w_{12} , w_{13} , and w_{23} , where $w_{12} + w_{13} + w_{23} = 1$. The decision weight w_{ij} is the weight associated with the competition between a_i and a_j ,

and it reflects the expected probability from the respondents perspective, that the outcome of the survey is defined only by the competition between a_i and a_j . We assume that the value of a strategy is the weighted average of the possible outcomes of that strategy. Finally, we assume that the respondent is only interested in the survey if the response is critical in determining the alternative. Let us now analyze the incentives for a respondent with the preference ordering $a_1 \succ a_2 \succ a_3$. Gutowski and Georges (1993) show that the only admissible strategies are to choose a_1 or a_2 , i.e. it can never be optimal to choose the least preferred alternative. Consequently, with three alternatives the respondent has to make a choice between the most preferred or the second most preferred alternative. Setting the value distributions of choosing a_1 and a_2 equal, we can find the critical value of $v(a_2)$ at which the respondent is indifferent between a strategic and a non-strategic behaviour, defined as

$$v(\bar{a}_2) = \frac{2w_{12} + w_{13}}{2w_{12} + w_{23}}. \quad (17)$$

If the true subjective value of outcome a_2 is larger than the critical value then the respondent acts strategically and chooses alternative a_2 , although alternative a_1 is the most preferred alternative. A number of interesting conclusions can be drawn from this expression: (i) A respondent will always choose truthfully if $w_{13} \geq w_{23}$, since $v(a_2) < 1$. This means that if the perceived competition between a_1 and a_3 is larger than that between a_2 and a_3 , the respondent will choose truthfully. Furthermore, this implies that in the case of equal decision weights the respondent will always choose truthfully. The latter case would perhaps be likely when the respondent does not have much information regarding other individuals' preferences, and therefore puts equal decision weights on all pair-wise competitions. (ii) The probability of acting truthfully is decreasing in $v(a_2)$. This means that if the utility of the two alternatives is close, then there is a higher probability of strategic behaviour. (iii) A respondent will in general only choose strategically if w_{23} is considerably larger than both w_{12} and w_{13} . Three straightforward and important conclusions can be drawn from above. First, introducing imperfect information does not ensure that the degree of strategic behaviour is reduced.

It may well be the case that respondents form such expectations so that they act strategically even if they would not have done so with perfect information. Second, using a generic (no labels) presentation of the alternatives instead of an alternative specific (labels) form probably reduces the risk of strategic behaviour, since it increases the complexity of forming expectations regarding other respondents' preferences. Third, it is generally advisable to explicitly introduce uncertainty into the choice experiment. This can be done by saying that there is uncertainty regarding individuals' preferences for the alternatives and the attributes. We believe that this strategy should be used in general with choice experiments. It is important to convince the respondent of the importance of he/she carefully answering the questionnaire, and that his/her choice can affect the outcome. Fourth, it is not clear whether differences in utility between alternatives in a choice set should be small or large. If the utility difference is small, then it is more difficult for the respondents to form expectations regarding how other respondents will choose, thereby making it more difficult to act strategically. At the same time, if the alternatives are close in utility the cost of acting strategically and being wrong is not that high compared to choosing sincerely, thereby increasing the probability of choosing strategically.

The empirical counterpart to the above discussion is tests of external validity, i.e. comparisons of actual and hypothetical behaviour. In transport economics, validity tests are either comparative studies with both hypothetical choice/ranking data and revealed preference data (e.g. Benjamin and Sen, 1982), or comparisons of predicted market shares from hypothetical choice/ranking studies with observed market shares (e.g. Wardman, 1988). The evidence from a large proportion of studies is that choice experiments generally pass external tests of validity. However, as we have discussed it is not obvious that these results carry over to hypothetical experiments on non-market goods. Carson et al. (1996) perform a meta-analysis, comparing results of CVM studies with revealed preference studies, and they find that the CVM estimates are slightly lower than their revealed preferences counterparts. However, several other experimental tests of the validity of CVM show that individuals overstate their WTP in hypothetical settings (see e.g. Cummings et al., 1995; Frykblom, 1997). We are only aware of three external validity tests for environmental goods. Carlsson and Martinsson (2001) conduct a classroom experiment consisting of both a hypothetical and an actual choice

experiment, and they cannot reject the hypothesis of external validity. Johansson-Stenman and Svedsäter (2001) conduct a similar experiment as Carlsson and Martinsson, but allows for between-subjects tests. They find a significant difference between actual and hypothetical behaviour, arguing that the difference in results is due to their between-subjects test. Cameron et al. (1999) compare six different hypothetical choice formats with actual purchase behavior. They assume an underlying indirect utility function, which allows the data from the choice formats to be used independently or pooled with heteroskedasticity across the formats. They cannot reject the hypothesis of the same indirect utility function across the question formats: actual behavior, closed-ended CV (phone survey), closed-ended CV (mail survey) and a pair-wise choice experiment.

6. Welfare Effects

The main purpose of a choice experiment is to estimate the welfare effects of changes in the attributes. In order to obtain these, researchers have generally assumed a simple functional form of the utility function by imposing a constant marginal utility of income. We focus on purely discrete choices; this means that in some cases welfare measures have to be interpreted with care in some cases. For example in the case of a site choice experiment, the welfare measures are per trip or per week, depending on what has been defined in the survey.

Let us assume the following utility function:

$$u = h(A) + \gamma(Q, z)z + \varepsilon, \quad (18)$$

where the function $h(A)$ captures the effect of the different attributes on utility, Q is a vector of personal characteristics and z is a composite bundle. This is a flexible specification of the marginal utility of income as it may vary by both the level of income and the personal characteristics of the individual. However, let us begin with the common case of constant marginal utility of income and independence of personal characteristics. For such a utility function, the ordinary and compensated demand functions coincide. Given this functional form and the assumption of weak complementarity, we can write the conditional indirect utility function for the purely discrete choice as:

$$V_j(A_j, p_j, y, \varepsilon) = h_j(A_j) + \bar{\gamma}(y - p_j c_j) + \varepsilon. \quad (19)$$

Furthermore, we can write the probability that alternative j is preferred as:

$$\begin{aligned} P\{j\} &= P\{h_j(A_j) + \bar{\mathbf{g}}(y - p_j c_j) + \mathbf{e}_j > h_i(A_i) + \bar{\mathbf{g}}(y - p_i c_i) + \mathbf{e}_i; \forall i \neq j\} = \\ &P\{h_j(A_j) - \bar{\mathbf{g}}p_j c_j + \mathbf{e}_j > h_i(A_i) - \bar{\mathbf{g}}p_i c_i + \mathbf{e}_i; \forall i \neq j\}. \end{aligned} \quad (20)$$

Equation (20) shows that income does not affect the probability of choosing a certain alternative under the current assumptions and hence the welfare measures will have no income effects. Thus, we can express the unconditional indirect utility function as:

$$v(A, p, y, s) = \bar{\gamma}y + \max[h_1(A_1) - p_1 c_1 + \varepsilon_1, \dots, h_N(A_N) - p_N c_N + \varepsilon_N]. \quad (21)$$

The Compensating Variation (CV) is obtained by solving the equality: $V(A^0, p^0, y) = V(A^1, p^1, y - CV)$. Using the functional form in equation (21), we have:

$$\begin{aligned} \bar{\gamma}y + \max[h_1(A_1^0) - p_1^0 c_1 + \varepsilon_1, \dots, h_N(A_N^0) - p_N^0 c_N + \varepsilon_N] = \\ \bar{\mathbf{g}}(y - CV) + \max[h_1(A_1^1) - p_1^1 c_1 + \mathbf{e}_1, \dots, h_n(A_N^1) - p_N^1 c_N + \mathbf{e}_N]. \end{aligned} \quad (22)$$

We can solve for CV and this results in:

$$\begin{aligned} CV = \frac{1}{\bar{\mathbf{g}}} \{ \max[h_1(A_1^1) - p_1^1 c_1 + \mathbf{e}_1, \dots, h_n(A_N^1) - p_N^1 c_N + \mathbf{e}_N] - \\ \max[h_1(A_1^0) - p_1^0 c_1 + \varepsilon_1, \dots, h_N(A_N^0) - p_N^0 c_N + \varepsilon_N] \}. \end{aligned} \quad (23)$$

If the error terms are extreme value distributed, i.e. the MNL model, the expected CV for a change in attributes is (Hanemann, 1999):

$$E(CV) = \frac{1}{\mu \bar{\gamma}} \left\{ \ln \sum_{i \in S} \exp(\mu V_{i1}) - \ln \sum_{i \in S} \exp(\mu V_{i0}) \right\}, \quad (24)$$

where μV_{i0} and μV_{i1} represent the estimated indirect utility before and after the change, $\mu \bar{\gamma}$ is the confounded estimate of the scale parameter and the marginal utility of money

and S is the choice set.⁷ With a linear utility function and only one attribute changing, the CV for a discrete choice is given by:

$$CV = \frac{1}{\gamma} \ln \left\{ \frac{e^{V_{11}}}{e^{V_{10}}} \right\} = \frac{1}{\gamma} (V^1 - V^0) = \frac{\beta_k}{\gamma} (A_k^1 - A_k^0). \quad (25)$$

By looking at the expression in equation (25) it is easily seen that for a linear utility function, the marginal rate of substitution between two attributes is simply the ratio of their coefficients, and that the marginal willingness to pay for a change in attribute is given by

$$MWTP_i = -\frac{\beta_i}{\gamma}. \quad (26)$$

For policy purposes it is of interest, and often necessary, to obtain the distribution of the welfare effects. This can be done either by bootstrapping or by the Krinsky-Robb method (Krinsky and Robb, 1986). With bootstrapping a number of new data sets are generated by resampling, with replacement, of the estimated residuals. The utility across alternatives, along with the parameter point estimates, is calculated in order to create the dependent variable. For each of these new data sets the model is re-estimated and welfare measures are calculated. The Krinsky-Robb method is based on a number of random draws from the asymptotic normal distribution of the parameter estimates and the welfare measure is then calculated for each of these draws. The Krinsky-Robb method is less computationally burdensome than bootstrapping, but its success critically depends on how closely the distribution of errors and asymptotically normal distribution coincide. For example Kling (1991) and Chen and Cosslett (1998) find that the two procedures give quite similar standard deviations.

The assumptions underlying the closed form solution of the welfare measures were (i) additive disturbances, (ii) an extreme value distribution⁸ and (iii) constant marginal utility of income. Let us relax the assumption of constant marginal utility of income and no effect of personal characteristics. The CV is in generally found by solving the

⁷ Note that this welfare measure is independent of the scale and, in practice, the scale parameter is set to equal one.

⁸ A closed form solution of the welfare measure does in fact exist for the GEV distribution in which the extreme value distribution is a member, see McFadden (1995).

equality: $V(A^0, p^0, y) = V(A^1, p^1, y - CV)$. The problem is how to obtain an estimate of CV, when income enters the utility function nonlinearly. In such a case the marginal utility of income is not constant and there is no closed-form solution to calculate the welfare effects. McFadden (1995) suggests either estimating the welfare effects by simulation or by calculating theoretical bounds on the welfare effects. Morey et al (1993) suggest an approach using a representative consumer approach, whereas Morey and Rossman (2000) impose piecewise constant marginal utility of income in the econometric model. The simulation approach is conducted in the following steps for a choice experiment consisting of K alternatives, and with T choice situations for each of the individuals. First, at iteration t , K randomly draws from a pre-specified distribution, e.g. an extreme value distribution, is performed. This results in the vector ϵ^t . Then, a numerical routine is applied to search for the CV^t , defined as:

$$V(A^0, p^0, y, s) \equiv E[v(A^0, p^0, y, \epsilon^t)] = E[v(A^1, p^1, y - CV^t, \epsilon^t)] \equiv V(A^1, p^1, y - CV). \quad (27)$$

This procedure is repeated T times. Second, for each individual, the expected CV is approximated by

$$E(CV^t) = \frac{1}{T} \sum_{t=1}^T CV^t. \quad (28)$$

If the sample of N individuals represents a random sample of the population under study then the expected CV for the population is

$$E(CV) = \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T CV^t. \quad (29)$$

The approach is easy in the case of an extreme value distribution, but more difficult for a GEV or a multivariate normal distribution. For a more detailed discussion on the simulation approach see e.g. McFadden (1995) or Morey (1999). The representative consumer approach, describe by Morey et al. (1993), uses a utility function of a representative individual. The result is that the repeated draws in McFadden's simulation approach can be skipped and a numerical routine can be directly applied in order to search for $E(CV)$. McFadden (1995) finds that this approach results in biased

estimations of CV and that the percentage of bias increases with the size of the welfare change. The benefit of McFadden's theoretical bounds approach is that it makes the computations less difficult by imposing bounds on the welfare effects from a change. The piecewise constant marginal utility approach by Morey and Rossman (2000) is easy to apply since the welfare effects can be calculated directly from the estimates. Furthermore, Morey and Rossman also present how to calculate the welfare effects when the CV results in a change from one income level with a specific constant marginal utility of income to another income level with a different constant marginal utility of income.

7. Conclusions

This paper has discussed valuation of non-market goods when using choice experiments. The advantages of choice experiments are that values for each attribute as well as marginal rate of substitution between non-monetary attributes can be obtained. Moreover, rigorous tests of internal validity can be performed. The success of a choice experiment depends on the design of the experiment which, as stressed several times in the paper, is a dynamic process involving definition of attributes, attribute levels and customisation, context of the experiment, experimental design and questionnaire development. Important tasks in future research include improving the knowledge about how respondents solve a choice experiment exercise and if preferences are consistent over the course of the experiments. Furthermore, the choice sets created by the chosen experimental design strategy have an important impact on the results. This paper describes the D-optimal approach. One of the problems with this approach is the criterion of utility balance. As we mention, it is not clear that utility balance necessarily improves the results and further studies are needed on this issue.

If a stated choice preference method has to be used to value a non-market good, either a closed-ended CVM survey or a choice experiment can be applied. As a rule of thumb we would recommend that practitioners apply a closed-ended CVM survey if the interest is purely in valuing a certain environmental change. In other cases, a choice experiment may be more suitable since it produces more information. In the future, however, more research is needed on both methods, and particularly on their abilities to elicit true preferences.

References

- Adamowicz, W., P. Boxall, M. Williams and J. Louviere (1998a) Stated preferences approaches to measuring passive use values. *American Journal of Agricultural Economics* 80, 64-75.
- Adamowicz, W., J. Louviere and J. Swait (1998b) Introduction to attribute-based stated choice methods. Report to NOAA Resource Valuation Branch, Damage Assessment Centre.
- Adamowicz, W., J. Louviere and M. Williams (1994) Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26, 271-292.
- Andreoni, J. (1989) Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy* 97, 1447-1458.
- Bateman, I. and K. Willis (1999) *Valuing Environmental Preferences*. Oxford University Press.
- Batsell, R. and J. Louviere. (1991) Experimental Analysis of Choice. *Marketing Letters* 2, 99-214.
- Ben-Akiva, M. and S. Lerman (1985) *Discrete Choice Analysis. Theory and Applications to Travel Demand*. MIT Press.
- Ben-Akiva, M. and T. Morikawa (1990) Estimation of travel demand models from multiple data sources. In Koshi, M. (ed.), *Transportation and Traffic Theory*, New York: Elsevier.
- Ben-Akiva, M., T. Morikawa and F. Shiroishi (1992) Analysis of the reliability of preference ranking data. *Journal of Business Research* 24, 149-164.
- Benjamin, J. and L. Sen (1982) Comparison of the predictive ability of four multiattribute approaches to attitudinal measurement. *Transportation Research Record* 890, 1-6.
- Bhat, C. (1995) A heteroskedastic extreme value model of intercity travel mode choice. *Transportation Research B* 29, 471-483.
- Bhat C. (1997) Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel. *Transportation Research B* 31, 11-21.
- Blamey, R., J. Bennett, J. Louviere, M. Morrison and J. Rolfe (2000) A test of policy labels in environmental choice modeling studies. *Ecological Economics* 32, 269-286.
- Boxall, P., W. Adamowicz, J. Swait, M. Williams and J. Louviere (1996) A comparison of stated preference methods for environmental valuation. *Ecological Economics* 18, 243-253.
- Braden, J. and C. Kolstad (1991) *Measuring the Demand for an Environmental Improvement*. North-Holland, Amsterdam.
- Bradley, M. (1988) Realism and Adaptation in Designing Hypothetical Travel Choice Concepts. *Journal of Transport Economics and Policy* 22, 121-137.
- Brownstone, D. and K. Train (1999) Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89, 109-129.
- Bryan, S., L. Gold, R. Sheldon and M. Buxton (2000) Preference measurement using conjoint methods: An empirical investigation of reliability. *Health Economics* 9, 385-395.
- Cameron, T., G. Poe, R. Ethier, and W. Schulze (1999) Alternative nonmarket value-elicitation methods: Are revealed and stated preferences the same? Working Paper Department of Economics, University of California.

- Carlsson, F. and P. Martinsson (2001) Do hypothetical and actual marginal willingness to pay differ in choice Experiments? – Application to the valuation of the environment. *Journal of Environmental Economics and Management* 41, 179-192.
- Carlsson, F. and P. Martinsson (2000) Design strategies for choice experiments in health economics. Working Paper, Department of Economics, Lund University.
- Carson, R., N. Flores, K. Martin and J. Wright (1996) Contingent valuation and revealed preference methodologies. Comparing the estimates for quasi-public goods. *Land Economics* 72, 80-99.
- Carson, R., R. Groves, and M. Machina (1999) Incentive and informational properties of preference questions. Paper presented at EAERE Ninth Annual Conference, Oslo.
- Chen, H. and S. Cosslett (1998) Environmental quality preference and benefit estimation in multinomial probit models: A simulation approach. *American Journal of Agricultural Economics* 80, 512-520.
- Ciriacy-Wantrup, S. (1947) Capital returns from soil-conservation practices. *Journal of Farm Economics* 29, 1181-1196.
- Cummings, R., G. Harrison and E. Rutstrom (1995) Home-grown values and hypothetical surveys: Is the dichotomous choice approach incentive compatible. *American Economic Review* 85, 260-266.
- Davis, R. (1963) *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*. Ph.D. dissertation, Department of Economics, Harvard University.
- Duffield, W. and D. Patterson (1991) Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. *Land Economics* 67, 225-239.
- Diamond, P. and J. Hausman (1994) Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives* 8, 45-64.
- Frykblom, P. (1997) Hypothetical question modes and real willingness to pay. *Journal of Environmental Economics and Management* 34, 275-287.
- Gutowski, W. and J. Georges (1993) Optimal Sophisticated Voting Strategies in Single Ballot Elections Involving Three Candidates. *Public Choice* 77, 225-247
- Hanemann, M. (1984) Discrete/Continuous Models of Consumer Demand. *Econometrica*. 52, 541-561.
- Hanemann, M. (1994) Valuing the environment through contingent valuation. *Journal of Economic Perspectives* 8, 19-43.
- Hanemann, M. (1999) Welfare analysis with discrete choice models. In Herriges, J. and C. Kling (eds.) *Valuing Recreation and the Environment*, Edward Elgar.
- Hanemann, M. and B. Kanninen (1999) The statistical analysis of discrete-response CV data. In Bateman, I. and K. Willies (eds.) *Valuing Environmental Preferences*, Oxford University Press.
- Hanley, N., R. Wright and W. Adamowicz (1998) Using choice experiments to value the environment. *Environmental and Resource Economics* 11, 413-428.
- Huber, J. and K. Zwerina (1996) The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 33, 307-317.
- Johansson-Stenman, O. and H. Svedsäter (2001) Choice experiments and self image: Hypothetical and actual willingness to pay. Working Paper Department of Economics, Gothenburg University.

- Johnson, R., R. Ruby and W. Desvousges (2000) Willingness to pay for improved respiratory and cardiovascular health: A multiple-format, stated-preference approach. *Health Economics* 9, 295-317.
- Johnson, R., K. Matthews and M. Bingham (2000) Evaluating welfare-theoretic consistency in multiple-response, stated-preference survey. TER Working Paper T-0003.
- Kahneman, D. and J. Knetsch (1992) Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22, 57-70.
- Kanninen, B. (1993) Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics* 69, 138-146.
- Kanninen, B. (2001) Optimal design for multinomial choice experiments. Unpublished Working Paper.
- Kling, C. (1991) Estimating the precision of welfare measures. *Journal of Environmental Economics and Management* 21, 244-259.
- Krinsky, I. and R. Robb (1986) On approximating the statistical properties of elasticities. *Review of Economics and Statistics* 68, 715-719.
- Kuhfeld, W., R. Tobias and M. Garrat (1994) Efficient experimental design with marketing research applications. *Journal of Marketing Research* 31, 545-557.
- Lancaster, K. (1966) A New Approach to Consumer Theory. *Journal of Political Economy* 74, 132-157.
- Layton, D. and G. Brown (1998) Application of stated preference methods to a public good: Issues for discussion. Paper presented at the NOAA Workshop on the Application of Stated Preference Methods to Resource Compensation, Washington, DC, June 1-2, 1998.
- Layton, D. and G. Brown (2000) Heterogenous preferences regarding global climate change. *Review of Economics and Statistics* 82, 616-624.
- Leigh, T., D. MacKay and J. Summers (1984) Reliability and validity of conjoint analysis and self-explicated weights: A comparison. *Journal of Marketing Research* 21, 456-462.
- Louviere, J. (1993) Conjoint analysis. In Bagozzi, R. (ed.), *Advanced Methods in Marketing Research*, Cambridge: Blackwell Business.
- Louviere, J., D. Hensher and J. Swait (2000) *Stated Choice Methods. Analysis and Application*. Cambridge: Cambridge University Press.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (ed.) *Frontiers in Econometrics*, Academic Press, New York.
- McFadden, D. (1995) Computing Willingness-to-pay in Random Utility Models. Working Paper, Department of Economics, University of California, Berkeley.
- McFadden, D. and K. Train (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15. 447-470.
- Mazotta, M. and J. Opaluch (1995) Decision making when choices are complex: A test of Heiners hypothesis. *Land Economics* 71, 500-515.
- Morey, E. (1999) TWO RUMs uncloaked: Nested-logit models of site choice and nested-logit models of participation and site choice. In Herriges, J. and C. Kling (eds.) *Valuing Recreation and the Environment*, Edward Elgar.

- Morey, E. and K. Rossman (2000) The compensating variation for a change in states when there is only one alternative in each state and where utility is assumed a linear spline function of income. Working paper at the University of Colorado.
- Morey, E., R. Rowe and M. Watson (1993) A repeated nested-logit model of Atlantic Salmon Fishing. *American Journal of Agricultural Economics* 75, 578-592.
- Mäler, K-G. (1974) *Environmental Economics: A Theoretical Inquiry*. Resources for the Future.
- Polak, J. and P. Jones (1997) Using stated-preference methods to examine travelers preferences and responses. In Stopher, P. and M. Lee-Gosselin (eds.), *Understanding Travel Behavior in an Era of Change*, Oxford: Pergamon.
- Ryan, M. and J. Hughes (1997) using conjoint analysis to assess women's preferences for miscarriage management. *Health Economics* 6, 261-273.
- Revelt, D. and K. Train (1998) Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 80, 647-657.
- Schkade, D.A. and J.W. Payne (1993), Where do the numbers come from?: How people respond to contingent valuation questions. In: Hausman J.A. (ed.), *Contingent valuation: a critical assessment* Amsterdam, North-Holland.
- Steffens, K., F. Lupi, B. Kanninen and J. Hoen (2000) Implementing an optimal experimental design for binary choice experiments: An application to Bird Watching in Michigan. In S. Polasky (ed.) *Benefits and Costs of Resource Policies Affecting Public and Private Land*. Western Regional Research Publication.
- Swait, J. and W. Adamowicz (1996) The effect of choice environment and task demands on consumer behavior. Discriminating between contribution and confusion. Working Paper, Department of Rural Economy, University of Alberta.
- Swait, J. and J. Louviere. (1993) The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30, 305-314.
- Train, K. (1998) Recreation demand models with taste differences over people. *Land Economics* 74, 230-239.
- Vick, S. and A. Scott (1998) Agency in health care: Examining patients' preferences for attributes of the doctor-patient relationship. *Journal of Health Economics* 17, 587-605.
- Wardman M. (1988) A comparison of revealed and stated preference models of travel behaviour. *Journal of Transport Economics and Policy* 22, 71-91.
- Zwerina, K., J. Huber and W. Kuhfeld (1996) A general method for constructing efficient choice designs. Working Paper, Fuqua School of Business, Duke University.

Figure 1: 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.

	Park A	Park B
Available facilities	<i>Visitor center</i>	<i>Information office</i>
Extension of walking tracks	<i>2 kms</i>	<i>10 kms</i>
Condition of tracks	<i>Rustic tracks</i>	<i>Stoned tracks</i>
Entrance fee	<i>8 US\$</i>	<i>10 US\$</i>

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

? Park A ? Park B