

ISSN 1327-810X

CHOICE MODELLING RESEARCH REPORTS

A COMPARISON OF STATED PREFERENCE TECHNIQUES FOR ESTIMATING ENVIRONMENTAL VALUES

By M.D. Morrison, R.K. Blamey, J.W. Bennett, J.J. Louviere

Research Report No. 1 December 1996

About the authors

Mark Morrison is a PhD student in the School of Economics and Management, University College, The University of New South Wales.

Russel Blamey is a Post Doctoral Fellow in the Urban Research Program, Research School of Social Sciences, Australian National University.

Jeff Bennett is Associate Professor in the School of Economics and Management, University College, University of New South Wales.

Jordan Louviere is Professor and Chair in the Department of Marketing, University of Sydney.

Choice Modelling Research Reports are published by the School of Economics and Management, University College, The University of New South Wales, Canberra 2600 Australia.

These reports represent the provisional findings of the research project 'Using Choice Modelling to Estimate Non-Market Values'.

The project is being funded by the Land & Water Resources Research and Development Corporation and Environment Australia under their joint program on the conservation and management of remnant vegetation. Support for the project is also being provided by the Queensland Department of Primary Industries, the Queensland Department of Natural Resources and the New South Wales National Parks and Wildlife Service.

The views and interpretations expressed in these reports are those of the author(s) and should not be attributed to the organisations associated with the project.

Because these reports present the results of work in progress, they should not be reproduced in part or in whole without the written authorisation of the Research Project Leader, Associate Professor Jeff Bennett.

Any comments will be gratefully received and should be directed to Associate Professor Jeff Bennett, School of Economics and Management, University College, The University of New South Wales, Canberra ACT Australia 2600. Telephone (06) 268 8833. Email: j.bennett@adfa.oz.au Fax: (06) 268 8450

Table of Contents

ABSTRACT	i
1 INTRODUCTION	1
2 OVERVIEW OF STATED PREFERENCE TECHNIQUES	2
2.1 Contingent Valuation	2
2.2 Contingent Rating	4
2.3 Contingent Ranking	6
2.4 Paired comparison	7
2.5 Choice Modelling	9
2.6 Summary	10
3 APPLYING STATED PREFERENCE TECHNIQUES IN THE NON-MARKET VALUATION CONTEXT	11
3.1 Biases associated with using SP techniques to estimate non-market values	12
3.2 Theoretical validity	17
3.3 Predictive validity	17
3.4 Reliability	19
3.5 Summary	20
4 PRACTICAL CONSIDERATIONS IN THE USE OF STATED PREFERENCE TECHNIQUES FOR ESTIMATING NON-MARKET VALUES	20
5 CONCLUSION	21
BIBLIOGRAPHY	23

ABSTRACT

The use of stated preference (SP) techniques for estimating environmental values has increased substantially in recent years. However, criticism about the most widespread SP technique used for valuing environmental resources, the contingent valuation method (CVM), suggests that there is a need to not only refine the CVM, but to develop alternative SP techniques. In this paper the CVM is compared with four other SP techniques: contingent rating, contingent ranking, paired comparison and choice modelling. The techniques are compared in terms of their methodologies and the validity and reliability of the results they produce. The appropriateness of using each of the SP techniques in different environmental valuation applications is also discussed. It is concluded that while the CVM is prone to bias and has some practical limitations, when applied appropriately it can be used to produce theoretically valid results. Three of the other techniques—contingent rating, contingent ranking and paired comparison—are found to have weak theoretical bases and do not produce economically valid valuation estimates. The final SP technique examined, choice modelling, appears to have considerable potential for providing useful and valid estimates of environmental values.

1 Introduction

The environmental impacts of resource use decisions have generated increasing community concern over the past twenty years. This has been the result of numerous factors including a growing awareness of environmental issues on the part of the general public and an increasing scarcity of environmental assets. Policy makers considering alternative uses for resources now increasingly look to include environmental values in their weighing up of the consequences of various resource use options. So the task of allocating instream flows between irrigation uses and wetland conservation is now seen to involve the weighing up of the benefits of irrigation against the benefits of wetland conservation. Similarly, the decision maker contemplating the imposition of tree clearing limitations is faced with an assessment of the benefits of biodiversity protection in stands of remnant woodlands relative to the value of agricultural production that could be achieved if those woodlands were to be cleared.

An important obstacle facing the decision maker in these and many other resource use choices is a lack of information about the value the community places on the environmental impacts. Value information relating to the resource uses such as irrigation and cattle grazing can be readily obtained from markets. However, environmental outcomes from resource use choices are rarely valued in conventional markets. Without information on the value of environmental outcomes there is a danger that resource use decisions will be inappropriate. Ignoring these values would result in a misallocation of resources away from environmental uses whilst treating the environment as having an infinite value would ensure that too many resources were allocated to preserving the environment.

Economists have sought to develop techniques which go beyond conventional markets to estimate the value of environmental outcomes. These techniques are basically of two generic types: those involving the use of revealed preferences and those involving the use of stated preferences. Revealed preference (RP) techniques such as the travel cost method and the hedonic pricing technique rely on observations of people's behaviour in markets that are somehow related to the environmental outcomes under investigation. For example, the value of noise pollution caused by an airport can be estimated through observations of actions in the market for residential property surrounding the airport. Similarly, the value of recreation at a natural area can be inferred from the costs incurred during travel to and from the area.

However, the revealed preference techniques are limited in their usefulness because they are retrospective and hence unable to value changes that have not been experienced. In addition, they can only be used where some relationship exists between the environmental outcomes of interest and an established, well functioning market. They cannot be used to estimate 'non-use' values.

Stated preference (SP) techniques have been developed to overcome the limitations of revealed preference techniques. SP techniques are characterised by the uses of surveys in which respondents' preferences for various environmental outcomes are identified. The use of surveys means that SP techniques are very flexible. A prospective stance can be adopted and outcomes which are in no way related to marketed goods can be investigated. SP techniques have frequently been used to value environmental goods (Mitchell & Carson 1989, Wilks 1990, NSW EPA 1995), and to evaluate new marketable products (Wittink & Cattin 1989) and transport options (Hensher 1991).

The appealing features of SP techniques are, however, countered by some drawbacks. The SP technique most widely used for estimating the value of environmental goods is the contingent valuation method (CVM). In Australia, decision makers are cautious about using this method, partly as a result of the criticism it received following the Resource Assessment Commission's application of the CVM to estimate the value of environmental damage if mining was to be allowed in the Kakadu Conservation Zone (Bennett 1996). The controversy surrounding the use of CVM in Australia was paralleled by the debate in the United States of America following the use of the method to estimate the value of the damage caused to the natural environment following the grounding of the Exxon Valdez (Diamond & Hausman 1994, Hanemann 1994).

So while there is a strong demand from policy makers for information relating to the value placed on environmental outcomes by the community, economists have had only limited success in providing valid valuation estimates. This gap, between the demand and supply of valuation estimates, can be addressed through the refinement of existing methods such as the CVM, but also by the development of alternative SP methods for the estimation of environmental values. Alternatives to the CVM include contingent rating, contingent ranking, paired comparison and choice modelling (CM)¹.

This paper is the first in a series which will periodically report on the progress of a research project designed to develop and test the CM alternative. The purpose of this paper is to provide a comparison of these techniques when applied to the valuation of environmental goods, both in terms of the methodology involved and the validity and reliability of the results. In Section 2, each of the five main SP techniques are outlined and in Section 3 the validity and reliability of applications of the five SP techniques is considered. Section 4 contains a discussion of the appropriateness of using each SP technique for different environmental valuation applications.

2 Overview of Stated Preference Techniques

2.1 Contingent Valuation

The CVM involves asking a sample of respondents whether (or how much) they are willing to pay to prevent or obtain a particular environmental outcome. The use of surveys to estimate the value of environmental goods was first suggested by Ciriacy-Wantrup (1947). Since then the number of CVM studies has grown exponentially, with Carson, Wright, Carson, Alberini & Flores (1994) producing a bibliography of CVM applications and studies containing 1674 entries. In Australia, the growth has been less dramatic, with the ENVALUE database listing 26 studies in its 1995 version (NSW EPA 1995).

CVM applications provide estimates of the aggregate value of changes in the quantity or quality of environmental goods involving usually only one or two resource use options. For most environmental applications, responses are used to estimate compensating or equivalent surplus (Mitchell & Carson 1989).

In practice most contingent valuation studies contain several well defined elements (see Portney 1994). The first is a description of the status quo and the environmental changes that will result from a proposed management or policy option. The description should be 'fair and accurate' and should include all information that is of relevance for the respondent.

The second element is a mechanism for eliciting the willingness to pay of respondents to achieve the environmental improvement or prevent the decline in environmental quality under the proposed option. There are various mechanisms or formats available for eliciting willingness to pay. The main formats are listed in Table 1. The dichotomous choice format (Bishop & Heberlein 1979) is generally considered to be the state of the art elicitation format. Various modifications to this format have also been used, such as the double-bounded dichotomous choice (Hanemann 1985) and polychotomous choice (Ready, Whitehead & Blomquist 1995; Bennett, Blamey & Morrison forthcoming). In all formats of the CVM, a payment method such as increased income taxes, levies on water and house rates, or voluntary donations is usually specified.

¹ In the marketing literature contingent rating and ranking, paired comparison and CM are collectively known as 'conjoint analysis'. Green & Srinivasan (1978) define conjoint analysis as 'any decompositional method that estimates the *structure* of a consumers's preferences given his/her overall evaluations of a set of alternatives that are prespecified in terms of levels of different attributes'.

Table 1: Main forms of the contingent valuation method

Form	Main features
Open-ended question	Respondents are directly asked their maximum willingness to pay
Iterative bidding/bidding game	Respondents are asked whether they are willing to pay a certain amount, and then extra or reduced increments, depending on whether they respond positively or negatively to the initial bid, until their maximum willingness to pay is reached
Payment card	Respondents are given a card that contains a number of payment bids and choose one
Dichotomous/discrete choice	Often referred to as the 'referenda model', respondents are asked whether they support a change in environmental quality given a specified additional payment
Double-bounded dichotomous choice	Similar to dichotomous choice except that if respondents support the payment they are then asked whether they would pay a slightly higher amount. If they don't support the payment they are asked if they are willing to pay a slightly lower amount
Polychotomous choice	Similar to dichotomous choice except respondents are able to indicate the extent to which they are in favour of or oppose the good

CVM questionnaires also contain questions to determine respondents' socioeconomic characteristics, as well as other information about their environmental attitudes and anything else that may affect willingness to pay. These data are used to demonstrate whether theoretical relationships between willingness to pay and independent variables such as income hold.

The final part of a CVM study is the statistical analysis. Responses are usually regressed against the socio-economic and attitudinal characteristics of respondents, as well as price in the case of discrete choice formats. The estimated equation is used to provide aggregate estimates of mean or median Hicksian surplus.

Hanemann (1984) and Hanemann & Kanninen (1996) show that the behavioural basis of the discrete choice format CVM is random utility theory (RUT). RUT assumes that the probability of an individual choosing a good from an array of goods is dependent on the utility of the good relative to the utility of other goods. In other words, an individual q will choose alternative i over alternative j if and only if $U_{iq} > U_{jq}$ ($i \neq j \in A$), where A is the choice set faced by the individual. In RUT the utility of a good is considered to depend on observable components, including a vector of attributes (x) and individual characteristics (s), as well as unobservable components (e). Unobservable components are treated as if they are random, and are assumed to follow some distribution function. The utility of good i can be represented as, $U_{iq} = V(s_i, x_{iq}) + e_{iq}$ where V is an indirect utility function. The probability of choosing alternative i can be written as:

$$P(i|j \in A) = P[(V_{iq} + e_{iq}) > (V_{jq} + e_{jq})]$$

The probability of someone choosing i instead of j is equal to the probability that their deterministic utility (V) plus their random utility (e) for i is greater than for j . By rearranging the terms in the above equation it can be seen that the probability that an individual drawn randomly from a sampled population will choose a given alternative is equal to the probability that the difference between the random utility of alternative j and i is less than the difference between the deterministic utility of alternative i and j :

$$P(i|j \in A) = P[(V_{iq} - V_{jq}) > (e_{jq} - e_{iq})]$$

2.2 Contingent Rating

The two main differences between CVM and contingent rating (and other SP techniques) are the number of resource use alternatives that are evaluated by respondents and the method of evaluation. In CVM applications there are usually only one or two resource use alternatives evaluated by respondents and, under the dichotomous choice format, respondents indicate whether they support or oppose the alternative. In contingent rating applications a series of resource use alternatives are evaluated by respondents, one at a time, through the use of a ratings scale. Respondents are not asked to compare the different alternatives, but rather are asked to rate each separately. For example, the following resource use alternative may be presented to a respondent:

Table 2: Example from a contingent rating survey

Wetland management survey									
Please circle one of the numbers below to show your preferences for the following alternative:									
Water quality	fair								
Number of waterbirds	50,000								
Area of wetland	60,000 ha								
Household cost	\$40								
1	2	3	4	5	6	7	8	9	10
<i>Weakly Preferred</i>					<i>Strongly Preferred</i>				

The use of contingent rating as a stated preference technique originated in the marketing area. Wittink & Cattin (1989) report that there were about 400 commercial conjoint studies undertaken each year in the USA during the early 1980s, and that 49% of these involved ratings scales. There have also been several applications involving the valuation of environmental goods, including recreational fishing (Roe, Boyle & Teisl forthcoming) and preservation of the bandicoot (Jakobsson, Kennedy & Elliott 1995).

The theoretical basis of contingent rating is information integration theory (IIT) (Anderson 1982, Lynch 1985, Louviere 1988a). Under IIT respondents are assumed to evaluate separately each piece of information about an option presented to them and assign a value to each piece of information. The information is then integrated by respondents to produce an overall evaluation which is transformed into a rating. As well as assuming that people are able to transform their evaluations into ratings, IIT theory usually assumes that the errors people make in deriving ratings are normally distributed. If these assumptions hold, ordinary least squares can be used to diagnose and test utility functions.

As shown in Table 3, the information that respondents 'integrate' involves a set of attributes (or resource use characteristics) that have multiple levels. The choice of attributes and attribute levels is based on factors such as which attributes are relevant to respondents' decisions, attribute levels actually existing at the site, policy relevant attributes and levels and the need for parsimony. For example, in the evaluation of a wetland, an analyst may select the following four attributes and attribute levels from which to develop resource use alternatives. Notice that in this example two attributes have two levels and two have three levels.

Table 3: Attributes and levels for a wetland

Attributes	Levels
Water quality	Good, fair, poor
Number of waterbirds	100,000 waterbirds, 50,000 waterbirds
Area	60,000 ha, 20,000 ha
Cost	\$70, \$40, \$10

These attributes can be combined to form $2^2 \times 3^2$ or 36 different resource use options. The total number of resource use options or alternatives which can be formed from a set of attributes is known as a complete factorial. Complete factorials have two significant properties. In a complete factorial each attribute level occurs in an equal number of choice alternatives, and occurs with other attribute levels an equal number of times. The attributes are said to be *orthogonal* or independent, which means that multicollinearity between attributes is eliminated. Secondly, because all possible combinations of attributes and attribute levels are included within a complete factorial, it is possible to estimate all two way and higher level *interactions* in addition to *main effects*² when attempting to model respondents' behaviour using regression techniques.

In many studies it is not possible to evaluate all resource use alternatives within a complete factorial because the number of possible resource use alternatives increases exponentially with the number of attributes and attribute levels. Given that it is desirable to have multiple evaluations of each resource use alternative to allow tests of variance, a method is needed to reduce the number of choice alternatives. In practice, most applications systematically sample from complete factorials by using fractional factorials. A fractional factorial consists of a 'fraction' of the resource use alternatives within a complete factorial. It is possible to maintain the property of orthogonality within a fractional factorial, although information about some or all interactions will be lost. However, in some cases it is possible to select a fractional factorial which excludes information about potentially unimportant interactions. Basic main effects and main effects designs with interactions are available in Box & Hunter (1961), Addelman (1962a, 1962b) and Dey (1985).

After the resource use alternatives have been generated using complete or fractional factorials they are evaluated by respondents using ratings scales. Respondents' ratings are then regressed against the attributes using ordinary least squares (OLS)³. Where an appropriate experimental design has been used the existence of significant interaction effects can be demonstrated by the significance of the interaction coefficient. If there are no interaction effects, the marginal rate of substitution between each attribute and price provides a measure of the value or 'part-worth' of each attribute. Part-worths are calculated by dividing the estimated coefficients of each of the attributes by the coefficient of price. A per respondent estimate of the value of an aggregate change in environmental quality is found by adding

² A main effect refers to the contribution of change in one of the attributes (eg X_1 , X_2 or X_3) on the dependent variable, which in this case is ratings. A two way interaction effect occurs when the magnitude of a main effect changes at different levels of a second attribute (and is represented by $X_i X_j$); a three way interaction occurs when the magnitude of a two way interaction changes at different levels of a third attribute and so on.

³ Ratings have, however, only ordinal and not cardinal interval significance and therefore should not be analysed by OLS which implicitly treats ratings as cardinal intervals. The cardinality assumption in OLS implies that a unit increase in the ratings scale represents a similar increase in utility regardless of the level of the ratings scale. It also implies that if a respondent gives bundle Z^0 a rating of 8 and bundle Z^1 a rating of 4, that the respondent is indifferent between two bundles of Z^1 and a bundle of Z^0 (Mackenzie 1992). Secondly, OLS multiple regression analysis yields biased and inefficient estimates of coefficients as ratings are discrete and not continuous. Several recent studies have used contingent rating but have analysed the data using ordered logit or probit (eg Gan & Luzar 1993; Mackenzie 1992, 1993; Jakobsson et al 1995). These studies are a hybrid contingent rating-ranking study. While they elicit ratings, the ratings are treated ordinally which is identical to how the data from a contingent ranking study is treated.

part-worths. However, these are only *unconditional* or relative estimates of willingness to pay. Contingent rating estimates are not conditional on respondents agreeing to purchase a good. Hence while valid estimates of the marginal rate of substitution between attributes may be derived from respondent's ratings of resource use alternatives, estimates of value derived from the marginal rate of substitution will be biased if respondents would not actually choose to purchase the good⁴.

2.3 Contingent Ranking

The characteristic feature of contingent ranking is the ranking by respondents of three or more resource use alternatives from most to least preferred, as shown in Table 4. There have been a number of applications of the contingent ranking method that have attempted to estimate the value of environmental goods, including electric cars (Beggs, Cardell & Hausman 1981), improved air quality (Rae 1983), improved water quality (Smith & Desvousges 1986), hazardous waste risk reduction (Smith, Desvousges & Freeman 1985), reduced diesel odours from motor vehicles (Lareau & Rae 1987) and recreational hunting (Mackenzie 1993).

Table 4: Example from a contingent ranking survey

Wetland management survey			
Please rank the three alternatives below, from most to least preferred by placing the numbers 1, 2 and 3 in the boxes below:			
	Alternative 1	Alternative 2	Alternative 3
Water quality	fair	good	poor
Number of waterbirds	50,000	100,000	50,000
Area of wetland	60,000 ha	60,000 ha	20,000 ha
Household cost	\$40	\$70	\$10
	<input type="text"/>	<input type="text"/>	<input type="text"/>

As with contingent rating, contingent ranking allows the estimation of part-worths as well as the aggregate value of environmental goods. It also, however, shares the weakness of contingent rating that respondents are not able to express opposition to payment for the environmental good, except through providing a low ranking. Hence the valuation estimates derived in contingent ranking studies are also unconditional.

The use of contingent ranking for valuing environmental goods was stimulated by the seminal study by Beggs et al (1981) who showed that contingent ranking could be based on RUT (Thurstone 1927, McFadden 1974), and suggested statistical methods for analysing data that corresponded to RUT. Prior to the study by Beggs et al (1981) researchers had no direct way of statistically analysing the results of a ranking exercise. Because of the dearth of statistical techniques, researchers either converted rankings to a ratings scale or asked respondents to place alternatives which they had initially ranked onto a ratings scale, and then analysed the data using OLS multiple regression (eg Whitmore & Cavadias 1974)⁵. Hence many of the early contingent ranking studies were conceptually very similar to contingent rating studies, and in a statistical sense, virtually identical. Beggs et al (1981) showed, theoretically, that the ranking of alternatives is equivalent to the probability of choosing the good with the highest utility from a choice set A, multiplied by the probability of choosing the good with the next highest utility out of the remaining goods and so on. This is shown in the following conditional

⁴ Using labels to denote the lowest rating as 'oppose payment' is unlikely to make the results conditional as respondents use ratings scales in various ways and may use several of the lowest ranks to indicate that they oppose payment.

⁵ In marketing applications before the availability of ordered logit models algorithms such as MANANOVA were used to evaluate ranked data (eg Green & Rao 1971, Hargreaves, Claxton & Siller 1976). A weakness of MANANOVA is that there is no theory of errors, hence there are no tests of significance (Louviere 1988b).

probability distribution, where m represents a resource use alternative, J is the set of resource use alternatives, h represents rank, and H is the total number of rankings which is equivalent to the lowest ranking:

$$\Pr(U_1 > U_2 > \dots > U_H \text{ for } H \leq J) = \prod_{h=1}^H [e^{V_h} / \sum_{m=1}^H e^{V_m}]$$

Based on their theoretical exposition, Beggs et al (1981) proposed the use of the ordered logit model to explain rankings in terms of attributes⁶.

Despite the work of Beggs et al (1981), some questions remain about the ability of contingent ranking to satisfy the assumptions of RUT. Respondents may use strategies when ranking alternatives that are contrary to RUT, especially when rankings are complex (Chapman & Staelin 1982, Smith & Desvousges 1986). For example, they may rank the alternative with the highest and lowest utility and randomly rank those in the middle. Or they may just rank the top one or two and randomly assign the remaining alternatives. Alternatively respondents may use a bottom-up ranking methodology⁷. Violations of RUT have been indicated by several studies which have found that the sampling variance is much greater with lower ranked alternatives (eg Chapman & Staelin 1982; Hausman & Ruud 1987; Ben-Akiva, Morikawa & Shiroshi 1992). This violates RUT because greater variance at lower ranks affects the probability of choice. Consider the following probability expression reported earlier:

$$P(i|j \in A) = P[(V_{iq} - V_{jq}) > (e_{jq} - e_{iq})]$$

As the variance increases at lower ranks, $e_{jq} - e_{iq}$ increases thereby affecting the probability of choice. Hausman & Ruud (1987) suggested introducing a scale parameter, τ , that would vary across rankings and would allow for differences in cognitive burden. However, this model is not consistent with the random utility model, as it implies a different set of preferences for each choice set within a ranking activity (Hanemann & Kanninen 1996). Moreover, Hausman & Ruud (1987) found that the scale parameter did not increase monotonically with the number of items ranked, while Ben-Akiva et al (1992) found that re-scaling did not produce completely consistent rankings.

At least one contingent ranking study has attempted to value multiple goods (Rae 1983). However, this is only possible using ordered logit if the goods are not close substitutes. This is because of the irrelevance of irrelevant alternatives (IIA) assumption which results from the error term distribution of the ordered logit model. This assumption is discussed more fully in Section 2.6. With the ordered logit model there is limited prospect of resolving IIA violations.

2.4 Paired comparison

In a paired comparison, respondents are presented with two choice alternatives and are asked to rate the difference between the two alternatives, usually on a five point scale⁸ (see Table 5). The paired comparison method has been used to estimate the value of several environmental goods, including

⁶ The index function of the ordered logit model is of the form $y^* = \sum \alpha_i + \beta'x$ for $i = 1 \dots j$, where y^* is a ranking, and α_i are dummy variables that represent ranking thresholds.

⁷ Chapman & Staelin (1982) comment that 'if decision makers reverse this procedure...the choice probabilities generated from such a bottom-to-top procedure will be equal to the top-to-bottom procedure only if the alternatives in each choice set are equally likely to be chosen, a rather restrictive situation'.

⁸ Some researchers define studies where respondents are asked to choose their preferred alternative out of a set of two alternatives as paired comparison (eg Lockwood 1996; Peterson, Brown, McCollum, Bell, Birjulin & Clarke 1996). We define any studies where respondents choose their most preferred alternative from a set of two or more alternatives as CM.

recreation (Sinden 1974)⁹, morbidity risk reductions (Magat, Viscusi & Huber 1988; Viscusi, Magat & Huber 1991; Krupnick & Cropper 1992; Desvousges, Johnson, Hudson, Gable & Ruby 1996), improved water quality (Griner forthcoming) environmentally friendly electricity generation (Johnson, Desvousges & Wood 1995) and use of electric cars (Segal 1995).

Table 5: Example from a paired comparison survey

Remnant vegetation management survey						
Please indicate which of the two alternatives you prefer most by circling one of the numbers below:						
	Alternative 1			Alternative 2		
Rarity of species	Fairly rare			Not rare at all		
Ease of visit	No visits allowed			Easy to visit with full facilities		
Area	100 ha			500 ha		
Household cost	\$10			\$20		
	<i>Strongly</i>					<i>Strongly</i>
	<i>Prefer</i>					<i>Prefer</i>
	<i>Alternative 1</i>	1	2	3	4	5
						<i>Alternative 2</i>

The pairs of resource use alternatives presented to respondents in paired comparison exercises have predominantly been generated using computer algorithms. The main motivation for using computer algorithms is that it saves the researcher from having to manually develop an experimental design. Some computer algorithms used for the paired comparison method, such as 'Adaptive Conjoint Analysis', claim to produce designs that are close to orthogonal. Adaptive Conjoint Analysis initially randomly selects a pair of resource use alternatives for respondents to evaluate. Based on a respondent's evaluation, the algorithm then successively seeks pairs of alternatives that are closest in utility for respondents to evaluate so that points of indifference can be found more efficiently (Johnson et al 1995). The algorithm seeks points of indifference because they provide the greatest amount of information about the value of attributes. Whilst intuitively appealing, some applications of Adaptive Conjoint Analysis have experienced problems with non-convergence (eg Johnson et al 1995), so that more recent studies have resorted to more traditional design methods (eg Desvousges et al 1996).

The data from paired comparison have been analysed using OLS (eg Magat et al 1989, Viscusi et al 1991, Krupnick & Cropper 1992) and ordered logit/probit procedures (Johnson et al 1995, Griner forthcoming). Similar to contingent rating and ranking, the paired comparison method produces estimates of the value of unit changes in attributes as well as estimates of the aggregate value of changes in environmental quality. Paired comparison also shares the limitation of contingent rating and ranking that estimates of value are unconditional as respondents are not able to oppose payment. A possible method of providing respondents with the ability to express opposition to payment is to include a constant base alternative in each paired comparison which represents the status quo (eg Segal 1995). However, respondents may also not choose to purchase a good represented by the status quo.

The behavioural basis of paired comparison is also not clearly defined. In IIT, respondents are assumed to evaluate pieces of information separately and assign each a value, and then integrate this information into an overall evaluation which is transformed into a rating. The process whereby respondents evaluate two separate bundles of information, define a difference between them and convert this into a rating is not well understood.

⁹ The study by Sinden (1974) is slightly different to the other studies as respondents were asked how much they were willing to pay to visit a more preferred alternative over a base alternative. Sinden (1974) also comments that 'data on two further measures of preference were collected...They were rank order (1 to 5) and a scaling (1 to 100)...These were not relevant as measures of intensity of preferences...interpersonal comparisons of preferences are impossible because these lack a common qualitative unit...In any case these two measures did not prove to be statistically significant...'

2.5 Choice Modelling

CM is a stated preference technique in which respondents choose their most preferred resource use option from a number of alternatives (see Table 6). CM was developed initially in the work of Louviere & Hensher (1982) and Louviere & Woodsworth (1983). It has been frequently used in the evaluation of choices involving consumer goods (eg Louviere & Woodsworth 1983), transportation (eg Hensher et al 1989, Hensher 1991), tourism (eg Morley 1994) and the selection of landfill sites (Opaluch, Swallow, Weaver, Wessells & Wichelns 1993). There have been a few applications that have valued environmental goods, including Adamowicz, Louviere & Williams (1994) who estimated water-based recreational value and Rolfe & Bennett (1996) who estimated the value of preserving international rainforests.

Choice models can be used to analyse choices between branded as well as generic alternatives. A generic choice set contains unlabelled alternatives (eg alternative 1,2 or 3), whereas branded alternatives are labelled (eg Macquarie Marshes, Gwydir Wetlands, Narran Lakes). It is therefore possible to use CM to value simultaneously several different sites, and produce site specific estimates of the value of unit changes in attributes.

Table 6: Example from a choice modelling survey

Remnant vegetation management survey			
Please indicate the alternatives you prefer most by ticking one of the boxes below:			
	Alternative 1	Alternative 2	Alternative 3
Rarity of species	Fairly rare	Not rare at all	Choose neither
Ease of visit	No visits allowed	Easy to visit	
Area	100 ha	500 ha	
Household cost	\$10	\$20	
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Choice models produce estimates of the value of changes in individual attributes as well as the value of aggregate changes in environmental quality. Therefore CM can be used to produce estimates of the value of multiple resource use alternatives. The estimates are conditional or absolute measures of value if a 'choose neither' option is included as an alternative¹⁰ (see Table 6).

As with contingent ranking, CM is based on RUT which suggests that consumers seek to maximise utility when they make choices. As noted above, RUT holds that there is a deterministic or observable component and a random or unobserved component of utility. The most common assumption made about the unobserved or error term in CM is that it is independently and identically distributed (IID), which means it has an extreme value error distribution (Ben-Akiva & Lerman 1985). The extreme value distribution implies a multinomial logit model with the following form, where V represents the systematic component of the utility of an alternative and λ is a scale parameter:

$$P_i = \exp(\lambda V_i) / \sum \exp(\lambda V_j) \text{ where } j = 1, \dots, n$$

The structure of the multinomial logit model depends on the form of the indirect utility function. An additive indirect utility function can be used to estimate main effects, or a polynomial form may be used to calculate interaction effects if an appropriate fractional factorial is selected.

The scale parameter, λ , results from the extreme value or Gumbell distribution which is the error distribution used to derive logit models (Ben-Akiva & Lerman 1985, Swait & Louviere 1993)¹¹. The

¹⁰ Including a 'choose neither' option is also useful for testing for IIA violations. Testing IIA requires at least three options, which can be satisfied by including a 'choose neither' option (Adamowicz et al 1994).

¹¹ A scale parameter is also found in probit models which have a normal distribution (Swait & Louviere 1996).

scale parameter is inversely related to variance and is equal to $\pi^2/6\mu^2$ where μ is equal to the variance of the error term. The scale parameter cannot be identified in a specific model because it is confounded with the explanatory variables, and is usually arbitrarily set to one. This is appropriate when estimating partworths as the scale parameter is cancelled when one explanatory variable is divided by another. However, the scale parameter can affect the validity of comparisons between models, as differences in variance affect the magnitude of model parameters. The parameters of two different models may therefore appear to be unequal when all that really differs is the variance of the models. While it is not possible to estimate the scale parameter in any particular model, it is possible to estimate the ratio of scale parameters in two different data sets. By identifying this ratio accurate tests of the equality of parameter estimates in different models can be conducted (Swait & Louviere 1993).

An important aspect of the multinomial logit model and all logit models is the independence of irrelevant alternatives (IIA) property. This property arises from the IID assumption and states that the probability of choosing an alternative from a subset of alternatives depends only on the alternatives included in the choice set and is independent of any other alternatives which might exist. Ben-Akiva & Lerman (1985) suggest that the IIA property is a special case of order independence. This implies that the model coefficients will be biased if the multinomial logit model is used to value goods that are close substitutes. This is because the addition of a substitute can affect the ordering of preferences. It is however possible to use more complex models such as the nested or mother logit to value goods that are close substitutes (Hensher & Johnson 1981; Ben-Akiva & Lerman 1985). McFadden (1987) reports various methods for testing for violations of the IIA assumption.

Experimental design is required to derive the choice sets that are evaluated by respondents. Both sequential and simultaneous designs can be used to derive choice sets (Bunch, Louviere & Anderson 1993). Sequential designs involve using fractional factorials to derive a parsimonious set of resource use alternatives. Further experimental design is then used to combine resource use alternatives into choice sets. In simultaneous designs resource use alternatives and choice sets are derived at the same time.

CM is an iterative process because of the experimental design process. For example, a design where two way interactions are not confounded must be selected for two way interactions to be estimated. However, if some or all of the two way interactions prove to be insignificant, a simpler and more efficient design can be selected. To test for violations of the IIA assumption a type of simultaneous design, known as the L^m design, which allows estimation of cross effects is required. Cross effects refer to the effect of the attributes of one alternative on the utility of another alternative. If such effects exist there is a violation of the IIA assumption. Again, if cross effects are not significant a more efficient design that does not allow for their estimation may be selected. By following this iterative process greater precision in parameter estimation can be achieved.

2.6 Summary

Some of the differences between the SP techniques are summarised in Table 7. The main differences are in the methodology, statistical analysis and behavioural basis of the techniques.

In terms of methodology, the most obvious difference between SP techniques is in the elicitation question. In the discrete choice CVM respondents are asked whether they support or oppose a specified payment for a number of resource use alternative; in contingent rating respondents provide a rating for various resource use alternatives; in contingent ranking respondents rank several resource use alternatives; in paired comparison respondents indicate which of two alternatives they prefer most through use of a ratings scale; and, in CM respondents choose their preferred alternative.

SP techniques differ in the type of statistical analysis required. Discrete choice CVM usually utilises binary logit while other forms of the CVM use OLS. Contingent rating uses OLS. Contingent ranking applications use either ordered logit and probit. Early applications of paired comparison used OLS, but more recent applications have used either ordered logit or probit. Most CM application use either nested or multinomial logit.

The behavioural basis of the discrete choice CVM and CM is RUT. Several researchers also contend that contingent ranking is based on RUT. While this can be demonstrated in theory, in practice, contingent ranking violates the assumptions of RUT. Contingent rating is based on IIT which, in practice, is usually violated. The behavioural basis of paired comparison is unclear.

Table 7: Overview of Stated Preference Techniques

Feature	CVM	Contingent rating	Contingent ranking	Paired comparison	CM
Behavioural basis	RUT (discrete choice CVM)	IIT	RUT (unclear)	?	RUT
Elicitation method	Discrete choice	Provide rating	Rank several alternatives	Rate difference between two alternatives	Choose preferred alternative
No of alternatives evaluated	1 or 2	Many	Many	Many	Many
Substitute goods included	No	No	Yes (as long as IIA not violated)	No	Yes
Express ambivalence/preference intensity	Yes (polychotomous choice only)	Yes	No	Yes	No
Computer algorithms	n/a			Yes	No
Experimental design	Basic designs only	Yes	Yes	Option	Yes
Statistical analysis	OLS or binary logit	OLS	Ordered logit/probit	OLS or ordered logit/probit	Multinomial, nested or mother logit
Welfare measure	Absolute	Relative	Relative	Relative	Absolute
Data produced	Aggregate value	Information on relative preferences			Aggregate + attribute value

3 Applying stated preference techniques in the non-market valuation context

The previous section was primarily concerned with identifying methodological differences between the various SP techniques. In this section the appropriateness of these techniques for use in the non-market valuation context is assessed, where ‘appropriateness’ is defined in terms of validity and reliability. An estimate is considered to be valid when the mean of the distribution of the observed values is equal to the true value. In statistical jargon, it is unbiased. An estimate is reliable when it can be reproduced. Reliability reflects the variance of observed values around the true value.

While techniques such as contingent rating and ranking, paired comparison and CM have been widely used in the marketing literature, their use for the purpose of non-market valuation has only recently begun to attract the attention of environmental economists. Although empirical results are drawn upon where possible, there is a general lack of studies in this area.

In Section 3.1 the evidence of the existence of bias in SP applications is reviewed. In Sections 3.2 and 3.3 evidence about the theoretical and predictive validity of estimates made using each of the SP techniques is examined, while the reliability of estimates made using SP techniques is assessed in Section 3.4.

3.1 Biases associated with using SP techniques to estimate non-market values

Bias in SP applications can affect both the validity and reliability of estimates. Bias can cause observed values to differ from true values, and/or increase the variance of estimators. This section lists some of the common biases found in environmental applications of SP techniques, reviews the empirical and *a priori* theoretical evidence of their existence and where evidence does not exist, considers the form it might be expected to take. It is not the intention here to review the wealth of literature that addresses biases in techniques such as the CVM. Rather we focus on a comparison of the different techniques. Table 9, which is at the end of Section 3.1, summarises the biases that can be expected in each technique.

Embedding effects

A debate that received prominence after an article was published by Kahneman & Knetsch (1992) concerns the existence of what has become known as the 'embedding effect'. The embedding effect is said to occur when the estimated mean willingness to pay for a good is lower when it is valued as part of a more inclusive good, rather than on its own. In Table 8, the embedding effect would occur if good BC1, which is a component of good ABC1, is not equal to good BC2, or goods C1, C2 and C3 are not equal. The problem with this effect is that, if it occurs, the estimated value of an environmental good will be determined by how the good is framed, and there is limited consensus about what constitutes an appropriate framing. In addition to Kahneman & Knetsch (1992), several recent studies have also shown this effect (Lockwood 1992, Brown, Barro, Manfredo & Peterson 1995). Kahneman & Knetsch (1992) suggest that the embedding effect occurs because people are seeking a 'warm-glow' associated with contribution to a good cause, while Kahneman (1986) and Blamey (1996) suggest that the embedding effect may be observed in practice if respondents are expressing more generalised attitudes and values than are sought by the researcher. Randall & Hoehn (1996) however demonstrate that embedding should occur through an analysis of the effect in a demand system for marketed goods. The primary theoretical defence given for embedding is the existence of substitution effects (see Smith 1992)¹²: when people are made aware of the existence of substitutes they will reduce their willingness to pay. Hence the critical factor is finding the appropriate level of embedding or framing to be incorporated in a survey design.

¹² Kahneman & Knetsch (1992) argue that 'Although the notions of substitution and satiation may apply to some environmental goods, they do not readily extend to existence values for beautiful sites, historic landmarks, or endangered species.' Embedding in these situations can be explained by income effects (see Lockwood 1992). As the definition of the goods requiring payment broadens, respondents have less income available to spend on each component of the good. Kahneman & Knetsch (1992) counter this point by arguing that WTP is typically a small percentage of income. However, it is not necessarily a small percentage of discretionary income or the discretionary income that people might allocate in their budget to these types of expenses.

Table 8: Kahneman & Knetsch's (1992) survey design

	Subsample 1	Subsample 2	Subsample 3
Inclusiveness of the good ↑	ABC1 BC1 C1	BC2 C2	C3

Note: in each subsample respondents were asked to value the good in the top row first and then goods in lower rows

Other SP techniques are also likely to be subject to embedding effects. Importantly, however, one would expect the effect to be reduced when respondents are made aware of relevant substitutes. In the CVM this is done by reminding respondents about relevant substitutes. In contrast, with CM (and contingent ranking in certain cases) it is possible to include substitute goods within the valuation exercise, providing a more rigorous valuation context. Information about substitutes in consumption can be found prior to an SP survey through focus groups, and information about substitutes in production or supply are established by researchers with reference to the policy context.

Part-whole bias

A bias that is related to the embedding effect is part-whole bias. This occurs when respondents value a larger or smaller good than the researcher's intended good (Mitchell & Carson 1989). It is demonstrated in Table 8 by the equality of values estimated for goods ABC1, BC2 and C3 which all have a different scope. This bias occurs in CVM studies when respondents' willingness to pay is insensitive to the scope of the good that is valued ie respondents' willingness to pay does not increase monotonically with increased provision of an environmental good. In other words, estimated willingness to pay is independent of the level of the attributes of an environmental good. In addition to the reasons given by Kahneman & Knetsch (1992), Kahneman (1986) and Blamey (1996) for the existence of the embedding effect, the main reasons given for the occurrence of part-whole bias are lack of familiarity with the good (Boyle, Desvousges, Johnson, Dunford, & Hudson 1994); that the probability that a good will be provided falls as the scope of the good increases therefore reducing willingness to pay (Carson 1995); and bounded rationality in which people are not capable of discerning differences in scope when they are below certain perceptual thresholds. Boyle et al (1994) found part-whole bias when they used three different impacts that differed substantially in aggregate numbers but differed little in terms of relative impact. The majority of studies examining part-whole bias, however, have found significant scope effects (Carson 1995)¹³, although most of these studies have been for changes that are substantial. Hence it appears that while the CVM can be used to value relatively large and significant changes, it is still questionable whether it can be used to value relatively minor changes.

Whether other SP techniques are less prone to part-whole bias than the CVM is yet to be demonstrated. However, one might expect that the more explicit attention given to differences in attribute levels with these other techniques may cause results to be less prone to this form of bias.

Hypothetical bias

Hypothetical bias is sometimes cited in applications of the CVM, and is a potential problem for all SP techniques. Hypothetical bias occurs when respondents do not believe that their answers to an elicitation question will have any policy significance. If this view is taken, there is little incentive for respondents to think carefully about either the environmental good in question or payment outcomes. Response strategies may thus be subject to greater influence from internal psychological motives related to honesty, value-expression and civic duty (Blamey 1996). Hypothetical bias is likely to occur in studies which inform respondents that scenarios are hypothetical, and others such as student surveys which clearly have little input to policy. Hypothetical bias does not necessarily lead to over or under estimates of mean or median willingness to pay, although it is almost certainly associated with greater variance of parameter estimates. Alternatives to CVM that involve stating preferences for a series of policy options may be prone to hypothetical bias if the impression is given that the exercise is little

¹³ Carson (1995) reports 33 studies which have found statistically significant scope effects, 19 of which estimated preservation values.

more than an 'academic game'. However, where surveys are clearly not 'academic', respondents may view the evaluation of multiple resource use alternatives as more realistic than CVM referenda. This may act to reduce hypothetical bias.

Payment vehicle bias

Payment vehicle bias refers to any biases stemming from problems with the payment vehicle. Generally speaking, these problems take one of three forms. First, a given payment vehicle such as an increase in water rates may not apply to a portion of the sample causing these respondents to discount payment information. For example, farmers who rely solely on bore water do not pay water rates and can therefore 'free ride' in a CVM survey. Second, respondents may not believe that the specified payment vehicle would ever be implemented, or that the bid value is implausible. Finally, respondents may object on ethical grounds to the implementation of the specified payment vehicle, which, in CVM surveys, is likely to lead to protest no responses. The existence of payment vehicle bias in CVM applications has been widely researched and many studies have found evidence of payment vehicle bias (eg Daubert & Young 1981; Greenley, Walsh & Young 1982; Carlos 1991; Schechter & Kim 1991; Bennett et al forthcoming).

It has been suggested that other SP techniques do not suffer payment vehicle bias to the same extent as the CVM. Rolfe & Bennett (1996) suggest that CM may reduce the likelihood of payment vehicle bias by de-emphasising the payment mode¹⁴. However, there is little empirical evidence to support this hypothesis. Indeed for contingent ranking it is unlikely that payment vehicle bias will be less than with the CVM because in most empirical applications there have only been a few attributes used. For example, Smith & Desvousges (1986) used only two attributes (one of which was price) and Rae (1983) used only three attributes. With so few attributes the possibility of deemphasising the payment vehicle will be substantially reduced. This is reflected in evidence of payment vehicle bias in the study by Rae (1983) where 18% of respondents refused to trade off increased entry fees for improved visibility. However contingent rating, paired comparison and CM applications usually involve greater numbers of attributes.

Strategic bias

One of the first biases that researchers identified as a potential problem for the CVM was strategic bias. Since Samuelson's (1955) declaration that it was inherently difficult to obtain true valuation of public goods because of 'free-riding', many economists have believed that the use of CVM would be open to strategic bias. Strategic bias would occur, say, when somebody with a pro-conservation disposition deliberately overstates, or somebody with a pro-development disposition deliberately understates, their true bid, in order to affect the final outcome¹⁵. Strategic bias can have both weak and strong forms: weak forms occur when true bids are only partly reduced or increased; strong forms occur, for example, when a respondent would bid zero when they have a positive willingness to pay. Overall the empirical evidence suggests that only weak forms of strategic bias will occur, if at all. Several studies have found some empirical evidence of weak forms (Bohm 1984; Brubaker 1984; Bennett 1987). Other laboratory and field tests have found little empirical evidence of strategic bias (Bohm 1972; Brookshire, Ives & Schulze 1976; Marwell & Ames 1981; Milon 1989). Only one study has found strong strategic bias in the valuation of the arts in Australia (Throsby & Withers 1986). However, the result found by Throsby & Withers (1986) is possibly due to payment vehicle bias or respondents' inability to pay additional taxes¹⁶. Various theoretical reasons have been given for the lack of strategic bias found in empirical

¹⁴ It is also possible that deemphasising price is not a desirable feature of non-market valuation SP applications, since a primary concern is that respondents become fully aware of payment implications.

¹⁵ An alternative explanation of strategic behaviour is that it simply reflects the way respondents deal with uncertainty. Milon (1989) found that behaviour that appears strategic is actually related to uncertainty, and that uncertain respondents are more likely to reduce their bids or abstain rather than overstate values. While uncertainty is a valid explanation for reduced bids, it is also possible that uncertainty could cause bids to increase where respondents have a pro-environment disposition.

¹⁶ Throsby & Withers (1986) found honest revelation amongst 65% of respondents, with 35% free-riding. Of the free-riders, about one-third were strong free-riders. Throsby & Withers (1986) initially asked people how much they wanted to pay for the arts out of existing taxes and then compared this with how much they would be willing to pay in addition to current taxes. Honest respondents were defined as those who did not change their responses.

studies, including the opportunity costs of strategic behaviour (Smith 1977), and the value that respondents place on honesty (Johansen 1977, Brubaker 1984, Bennett 1987).

Similar to payment vehicle bias, other SP techniques may not suffer from strategic bias to the same extent as the CVM. This may result if de-emphasising payment reduces the propensity of respondents to act strategically (Rolfe & Bennett 1996). Another possibility is that the cognitive requirements to act strategically may be greater with other SP techniques. However, there is also no empirical evidence to support this hypothesis.

Starting point bias

Starting point bias occurs in CVM studies when the amount at which the initial bid is set affects the bid distribution. It has been suggested that this occurs because the starting bid may suggest to the individual the approximate range of 'appropriate' bids, or because a good is poorly defined or not distinctly perceived by respondents, or that respondents are not willing to go through an iterative bidding process because they value time highly or become bored (Brookshire, d'Arge, Schulze & Thayer 1981; Cummings, Brookshire & Schulze 1986). A number of studies have analysed whether starting point bias exists when the iterative bidding format is used and the results from these studies have been mixed. Studies by Brookshire et al (1981), Thayer (1981) and Green & Tunstall (1991), using different starting points, found no statistically significant evidence of starting point bias. However two of these studies had small sample sizes and the third did not substantially change starting points. Other studies by Rowe, d'Arge & Brookshire (1980), Boyle, Bishop & Welsch (1985) and Desvousges, Smith & Fisher (1987) did find evidence of starting point bias. Overall it appears that starting point bias can be a significant problem in iterative bidding formats. In recent studies researchers have also found that starting point bias occurs in discrete choice CVM formats (Holmes & Kramer 1995, Herriges & Shogren 1996).

A form of starting point bias may also occur in other SP techniques if the magnitude of attribute levels affects the estimates of part-worths. For example, this may occur if three levels for costs are \$20, \$50 and \$80 instead of \$10, \$60, \$110. While there is no available evidence about the effect of *increasing* the magnitude of attribute levels on parameter estimates, Meyer & Eagle (1982), in a study of shopping centre choice using student respondents, show that *reduced* variability in the levels of 'major' attributes can affect the importance of minor attributes. They suggest that this occurs because insufficient variation in major attributes cause respondents to focus on less important attributes when choosing between alternatives. Meyer & Eagle's (1982) study used both choice based and ratings based elicitation formats.

Information bias

Information bias occurs when respondents are sensitive to the quantity or quality of information provided in scenario descriptions that are 'true' and 'accurate'. For the CVM, most researchers have examined whether including background information about costs, existing payments or causes of the problem affects estimates (eg Samples, Dixon & Gowen 1986; Boyle 1989), or alternatively whether informational cues designed to stimulate altruism affects estimates (Ajzen, Brown & Rosenthal 1996). These studies have found that information can affect mean estimates, or at least their precision. However, this is to be expected from the CVM, as well as other SP techniques, which are *contingent* on the information supplied. Defining the appropriate amount of information to present to respondents is part of all SP surveys and depends on the purpose of the survey and the requirements of respondents. The effect of specific information can be ascertained through focus groups and removed or modified if it is not appropriate or biased.

Metric bias

While strategic bias may be occurring it is possible, and even quite likely, that respondents who reduced their bids either were unwilling to pay extra taxes or could not afford to do so. In a recent telephone survey and focus groups held to develop a CVM survey for a public good in South Australia, it was found that many believed that they were already paying sufficient taxes or that public good provision was simply the government's responsibility (Bennett et al forthcoming). This would cause the strategic bias that Throsby & Withers (1986) found, yet it is probably closer to being a form of payment vehicle bias or due to people's inability to pay additional taxes.

Metric bias occurs in contingent rating and paired comparison applications because of the use of ratings scales. This bias relates to the difficulty of cardinal measurement of utility and the problems of interpersonal comparison of cardinal measurement of utility first exposed by Robbins (1937). Respondents may use ratings scales in different ways: one respondent may use from 6 to 10 on the scale and another may use 1 to 10 on the scale, while they both have similar preferences¹⁷. The use of rating scales also rests on the assumption of a monotonic transformation of utility into ratings. It is not clear that this assumption will hold if respondents rank preferred options very highly and other options towards the bottom of the scale, or if the ratings scale is insufficiently small to allow respondents to differentiate between alternatives. Metric bias may bias parameter estimates as well as lead to increased variance.

Non-response bias

Non-response bias arises when there is a low response rate which means that a survey sample may not be representative of the population. Some researchers have argued that certain SP techniques will be less prone to non-response bias because the questionnaires are easier to answer. For example, contingent rating surveys may be easier for respondents to answer as providing ratings allows respondents to express ambivalence (Mackenzie 1993). Mackenzie (1993) reports some evidence of lower protest rates in a survey where respondents were initially asked to rate various alternatives¹⁸. Only 1.4% of respondents refused to answer the rating questions¹⁹.

Rolfe & Bennett (1996) contend that respondents find CM exercises easier than some other SP exercises because people are familiar with choices in market transactions and that, cognitively, it is simpler. Choosing a preferred alternative *a priori* may be less demanding than rating or ranking the same number of alternatives.

¹⁷ Some researchers have tried to overcome this problem by labelling ratings scales (eg Griner forthcoming) or by adjusting each respondent's ratings to a common mean (Green & Srinivasan 1978).

¹⁸ Ready et al (1995) found slightly higher response rates in the polychotomous choice CVM format, which allows respondents to express ambivalence, than was found in the discrete choice format.

¹⁹ However there may be other negative effects on response rates. Mackenzie (1993) also reported that 6.2% of respondents provided identical ratings, reflecting either confusion or laziness, and 2.1% of respondents misread the question. Hence non-useable responses are close to 10% which is similar to what is found in CVM surveys.

Table 9: Evidence of bias in applications of stated preference techniques

Bias	CVM	Contingent rating	Contingent ranking	Paired comparison	CM
Embedding effects	Yes	? (Yes)	? (Yes)	? (Yes)	Allows explicit inclusion of substitutes
Hypothetical bias	Yes	?	?	?	?
Metric bias	No	Yes	No	Yes	No
Non-response bias	Yes	? (Ability to express ambivalence)	?	?	? (Respondents may find choices easier)
Part-whole bias	Yes	? (Techniques more suited to evaluating changes of a minor nature)			
Payment vehicle bias	Yes	? (Multiple attributes de-emphasises payment)	Yes	? (Multiple attributes de-emphasises payment)	? (Multiple attributes de-emphasises payment)
Starting point bias	Yes	?	?	?	?
Strategic bias	Yes (weak forms)	?	?	?	?

Note: ? indicates absence of empirical evidence; information in brackets is theoretical evidence

3.2 Theoretical validity

The biases referred to in the previous section can affect the results of SP applications in several ways. One way in which the effect of bias can be assessed is through an examination of theoretical validity. This involves examining whether *a priori* theoretical relationships between the dependent variable and certain independent variables hold (Mitchell & Carson 1989). Theoretical validity indicates that systematic forms of bias may be limited, as systematic forms of bias are likely to confound theoretical relationships. The independent variables usually included in models are income, age, education, sex and certain attitudinal variables.

For the CVM it is standard practice to include socio-economic and attitudinal variables in valuation functions. Many CVM studies have shown that *a priori* theoretical relationships hold, including Walsh, Loomis & Gillman (1984), Lant & Roberts (1990), Kaoru (1993) and Kosz (1996). The main variable examined has typically been income as it was believed that income should have a strong and positive effect on willingness to pay. However, Carson & Flores (1995) have shown theoretically that income sensitivity should be low. Other socio-economic variables that are often included in valuation equations are level of education, age and sex. Yet these variables can affect estimates in different ways and there is less certainty about the sign and magnitude of their coefficients.

Very few environmental applications of SP techniques other than the CVM have examined theoretical validity. Three exceptions in applications of contingent ranking are Rae (1983), Smith & Desvousges (1986) and Lareau & Rae (1987), with only the latter two studies finding significant *a priori* relationships. We could only find two paired comparison studies that have examined theoretical validity (Krupnick & Cropper 1992, Desvousges et al 1996), and no contingent rating or CM studies. Some non-environmental studies using CM have, however, examined theoretical validity, including Louviere, Fox & Moore (1993), Morley (1994) and Hensher (1995). In each of these studies most or all of the *a priori* relationships were found to be significant.

3.3 Predictive validity

Another way of assessing the validity of SP results is through their predictive power (Mitchell & Carson 1989). This is demonstrated when the choices or estimates predicted using SP techniques are equal to choices or estimates revealed in actual market situations. Predictive validity also indicates that systematic forms of bias may be limited.

Where SP techniques attempt to estimate non-market values, testing predictive validity is a difficult task because there are usually no equivalent markets. For the CVM many researchers have sought to validate estimates by comparing them with estimates generated using other non-market valuation techniques, particularly the hedonic price and travel cost methods. The results from these comparisons indicate some convergence between estimates made using CVM and other non-market techniques. For example, Carson, Flores, Martin & Wright (1995) found in an analysis of 83 studies where there were 616 comparisons of CVM and revealed preference (RP) estimates, the sample mean CVM/RP ratio was 0.89 with a 95% confidence interval of 0.81-0.96. However, a weakness of this validity test is that the travel cost and hedonic price methods are also prone to substantial variation²⁰ and are limited to the estimation of values which can be estimated concurrently by the two techniques.

Some researchers have, however, sought to compare estimates from the CVM with market data. Diamond & Hausman (1994) report the estimates of five CVM studies which compared hypothetical and actual statements of willingness to pay. They comment that these studies 'often find large and significant differences'. Hanemann (1994) however countered that Diamond & Hausman (1995) only examined five of the ten studies which made this comparison and the remaining studies found no significant difference. Moreover, Hanemann (1994) contended that in two of the studies Diamond & Hausman (1994) cite the difference is not statistically significant, while the remaining three studies did not use 'state of the art' CVM techniques. Hanemann's (1994) observations suggest that discrete choice CVM estimates may accurately predict willingness to pay, however, it is difficult at this stage to draw firm conclusions, especially in relation to non-use values.

Simple tests of validity for contingent rating studies have been conducted by comparing predicted and actual market choices. These tests have produced mixed results. Leigh, McKay & Summers (1984) compared actual choices with predictions from an exercise where respondents rated the difference between two different calculators. They found that only 35% of choices were predicted correctly. Bateson, Reibstein & Boulding (1987) report the results of three other studies by Davidson (1973), Robinson (1980) and Montgomery & Wittink (1980). Montgomery & Wittink (1980) examined job choice at an individual level, and found that contingent rating predicted 63% of actual choices correctly. The other two studies examined group data for transportation choice and these studies found a close correlation between predicted and actual group shares. However, group data can disguise non-convergence between predicted and actual data at an individual level.

Few tests on the predictive validity of contingent ranking or paired comparison exist. Smith & Desvousges (1986) compared the estimates made using contingent ranking with estimates made using the CVM in the valuation of improved recreational water quality. They found a close correspondence between the estimates made using the two techniques, although it was not tested whether the difference between estimates was statistically significant. However, as noted above, comparing the results of two non-market valuation techniques is not a strong test of predictive validity.

Several studies have conducted tests of the validity of CM. These studies appear primarily in the marketing and transportation literature and have compared predicted and actual market choices. Louviere & Woodworth (1983) report on two studies on pet food and expenditure at shopping centres where the correlation between predicted and actual market choices were 0.83 and 0.96 respectively. Carson, Louviere, Anderson, Arabie, Bunch, Hensher, Johnson, Kuhfeld, Steinberg, Swait, Timmermans & Wiley (1995) report three other CM studies which found a close correspondence

²⁰ Graves, Murdoch, Thayer & Waldman (1988), in a study using hedonic pricing to estimate the costs of air pollution, comment that 'our findings cast doubt on the results of studies that have utilised hedonic-based marginal prices to evaluate the validity of other nonmarket methods'. While Smith, Desvousges & Fisher (1986), in a comparison of the CVM and the travel cost method, found that 'the contingent valuation estimates appear to be sensitive to the question format used, with the ratio of the largest estimated mean for a question type to the smallest about six to one. *Even more striking, however, is the range of estimated consumer surplus increments across the travel costs models. It more than encompasses the contingent valuation estimates...*'

between predicted and actual choices, although one study by Hensher & Battelino (1993) found a divergence between predicted utilities associated with traffic management at one point in time with actual utilities some time later.

While correlations between predicted choices, made using any SP technique, and actual market choices may be quite high, there may be differences in scale so that predicted choices consistently over or under predict actual choices. For example, Louviere & Woodworth (1983) suggested that the difference between predicted and actual market shares in the two examples cited above was large enough for the CM data to be rescaled. In other words, the predicted choices under CM were sufficiently different to observed choices in real markets that an adjustment to the CM data set, so that it would match the observed choices, was warranted. This difference may arise because of the higher variance in hypothetical markets which affects parameter estimates (see Section 2.5). If the difference between predicted and actual choices is due to differences in the scale parameter, it is possible to adjust the CM data by amalgamating CM and revealed preference or actual market data sets and rescaling the CM data. Rescaling has been undertaken in transport applications (eg Ben-Akiva & Morikawa 1990; Hensher & Bradley 1993; Ben-Akiva, Bradley, Morikawa, Benjamin, Novak, Oppewal & Rao 1994; Swait, Louviere & Williams 1994) and environmental applications (Adamowicz et al 1994; Adamowicz, Swait, Boxall, Louviere & Williams 1996; Swait & Adamowicz 1996). The capacity to re-scale data provides a powerful test of the validity of CM.

The extent to which the above findings for market goods transfer to the non-market environmental valuation context remains to be seen, and will ultimately depend on any differences in the incentive structures that respondents face in these contexts.

3.4 Reliability

In addition to validity, SP researchers often consider the reliability of their results. Reliability refers to the consistency or reproducibility of results over time. Whereas validity measures indicate whether the mean of the observed values is equal to the true value, reliability indicates whether it is possible to consistently estimate the true value. It is therefore an indicator of variance or efficiency. Reliability studies are not generally capable of detecting systematic biases in parameter estimates, therefore they should be considered of secondary importance when compared to tests of validity.

The primary method of examining reliability has been to use test-retests. The similarity of the results can be examined by testing the significance of the correlation between the estimates, or the similarity of the functional forms between the two models, with the latter test being the most rigorous.

The CVM has generally been shown to be quite reliable. Tests of stability over time have been conducted by Kealy, Montgomery & Dovidio (1990), Loomis (1990), Reiling, Boyle, Phillips & Anderson (1990) and Stevens, Moore & Glass (1994). Kealy et al (1990) examined stability with private and public goods over a two week period and found no statistical difference in willingness to pay using a log-likelihood ratio test. Loomis (1990), however, used a much longer time period of 9 months and found significant differences in willingness to pay. Loomis (1990) examined willingness to pay for two different improvements in environmental quality at Mono Lake in California using both open-ended and dichotomous choice elicitation formats. While the test-retest correlations were significantly different from zero, indicating a degree of reliability, significant differences were found using likelihood ratio tests for one of the improvements under each elicitation format. This difference, however, could be explained by changes in preferences over this time period. Reiling et al (1990) compared the willingness to pay of one sets of respondents to control black flies (a type of mosquito) at the peak of the black fly season with the willingness to pay of a second set of respondents at the end of the season (3 months later). Using a pairwise t-test they concluded that there was no statistically significant difference between the observed mean values. Stevens et al (1994) compared respondent's willingness to pay to preserve the bald eagle in 1989 and 1992 using both panel data and independent samples from each year. Using various statistical tests including a pairwise t-test and a Chow test, they concluded that there was no statistical difference between the observed mean values in the two time periods.

For contingent rating the results from studies of marketed goods have been mixed. Bateson et al (1987) summarised the results from nine studies which assessed reliability of model coefficients over periods ranging from one day to two months. Three of the more robust studies (McCullough & Best 1979, Cattin & Weinberger 1980, Segal 1982) had quite high correlation coefficients ranging from 0.63-0.88. However the study by Leigh, McKay & Summers (1981) found a correlation coefficient of only 0.49, and a later study by the same authors found an adjusted correlation coefficient of 0.48 (Leigh et al 1984).

No studies that have tested reliability have been identified for contingent ranking, paired comparison or CM.

3.5 Summary

The validity and reliability of valuation estimates made using SP techniques may be affected by the existence of a number of types of bias. Empirical evidence suggests that the CVM is prone to a number of forms of bias, including hypothetical, part-whole, payment vehicle, starting point and strategic bias although it does appear to be a reliable technique. The remaining SP techniques may be less prone to certain biases that affect the CVM, although this is yet to be demonstrated empirically. This does not mean that all of the other remaining SP techniques are more valid than the CVM, as other biases affect some of these techniques. For example contingent rating and paired comparison applications are affected by metric bias. Of the alternatives to CVM, CM appears to be the least prone to bias.

4 Practical considerations in the use of stated preference techniques for estimating non-market values

As well as validity and reliability, there are other criteria that need to be considered when selecting an SP techniques for a particular environmental valuation application. These include the complexity of the environmental issue, the number of valuation estimates required, time, cost and the need for benefit transfer²¹.

Complexity

Environmental issues have differing levels of complexity. Some issues require considerable background information about many attributes, and between causes and effects. For some issues it may not be possible or practical to define a parsimonious and representative number of attributes, as is required for CM. For particularly complex valuation exercises the CVM may be the most appropriate valuation technique, as it is possible to present large amounts of information to respondents about a single option.

²¹ Benefit transfer is the extrapolation of estimates to sites which are different from where the valuation estimate was originally made.

Number of management options under consideration

In some cases decision makers are interested in ascertaining the values associated with a single management option. In others, they require information on multiple options. If multiple valuation estimates are required, an SP technique other than CVM is likely to be most appropriate.

Time

The time required to apply each of these techniques parallels costs. Contingent valuation and contingent rating require the least time. The other techniques can take longer because of the iterative nature of the research process. However, depending on the level of accuracy required and the nature of a given application it may be possible to find quickly an adequate experimental design.

Cost

Cost is based on several factors including the complexity of the issue, accuracy required and the skills necessary to use a particular technique. The cost of all techniques increases as greater accuracy is required, as this usually involves larger sample sizes and more thorough development and testing of survey instruments.

The other main costs are in skills related to experimental design and statistical modelling and interpretation. CVM does not require the use of experimental design and the statistical techniques are relatively straightforward. Contingent rating, requires the use of simple experimental designs and straightforward statistical techniques. The experimental design and statistical techniques required for contingent ranking and paired comparison are more sophisticated than CVM or contingent rating. These techniques therefore require greater skills and result in higher costs. CM requires a similar level of skills for experimental design as contingent ranking and paired comparison, but the statistical analysis is more sophisticated, which means that it is likely to require the highest skill level of all of the SP techniques.

Benefit Transfer

Contingent valuation is the least suitable of the SP techniques for benefit transfer as only one or two valuation estimates are typically produced and considerable interpolation is required to adjust for differences between sites where studies were undertaken and sites for which values are to be inferred. The other SP techniques produce valuation estimates that can be readily adjusted for differences in key site characteristics, therefore they are particularly suited to benefit transfer.

CM has an additional advantage for benefit transfer. By simultaneously valuing several substitute goods CM is able to estimate the 'generic' value of attributes. These estimates are likely to be less site specific and more suited to benefit transfer. CM is also able to control for brand effects, such as the sentiment associated with the protection of well known environmental areas such as Kakadu.

5 Conclusion

Contingent valuation is the SP technique most widely used to estimate the value of environmental goods. Because it is prone to bias it has been widely criticised, and there is a rationale for refining and developing other SP techniques. The CVM also can only be used to estimate the value of one or two resource use options. Therefore where multiple estimates are required or where the results are needed for benefit transfer it is not suitable. Nonetheless the CVM does have a strong behavioural basis, it is relatively cheap and quick to implement. The CVM also provides absolute estimates of the value of environmental goods.

Contingent rating is unlikely to be offer any improvement over the CVM. While it can be used to estimate the value of multiple resource use options, and may be less prone to certain biases experienced

by the CVM, it has some distinct disadvantages. It has a weak theoretical basis as most contingent rating applications violate the assumptions of information integration theory. It suffers from estimation bias as OLS produces inefficient and biased estimates when applied to discrete data. Metric bias occurs because of problems with combining ratings across individuals. Moreover, the estimates of value derived using contingent rating are only relative because respondents are not able to express opposition to payment.

Contingent ranking is also problematic. Its behavioural basis is random utility theory, however, in practice, it violates the theory's assumptions. Contingent ranking applications are likely to be prone to biases found in CVM applications such as payment vehicle bias. Similar to contingent rating, the estimates of value are only relative because respondents are not able to express opposition to payment.

Paired comparison shares most of the advantages and disadvantages of contingent rating. It can be used to estimate the value of multiple resource use options and may be less prone to certain biases experienced by the CVM. However, the behavioural basis of paired comparison has not clearly been defined and it suffers from metric bias. The estimates of value are also relative.

CM appears to be the most promising SP technique. It can be used to estimate the value of multiple resource use options. It appears that CM applications may be less prone to some of the biases experienced in applications of the CVM and other SP techniques. It has a strong behavioural basis in random utility theory. It also has the advantage of being able to include explicitly substitute goods within valuation exercises. While the experimental design involved in CM exercises may mean that it is more time consuming and costly than CVM applications, its validity and suitability for benefit transfer may mean that it will be the most cost-effective option in cases where there is little *a priori* justification for an isolated valuation using the CVM.

Bibliography

Adamowicz, W., Louviere, J. & Williams, M. (1994). 'Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities'. *Journal of Environmental Economics and Management*, 26: 271-292.

Adamowicz, W.L., Swait, J., Boxall, P., Louviere, J. & Williams, M. (1996). 'Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation'. *Journal of Environmental Economics and Management*, Forthcoming.

Addelman, S. (1962a). 'Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments'. *Technometrics*, 4(1): 21-46.

Addelman S. (1962b). 'Symmetrical and Asymmetrical Fractional Factorial Plans'. *Technometrics*, 4(1): 47-57.

Ajzen, I., Brown, T.C. & Rosenthal, L.H. (1996). 'Information Bias in Contingent Valuation: Effects of Personal Relevance, Quality of Information, and Motivational Orientation'. *Journal of Environmental Economics and Management*, 30: 43-57.

Anderson, N.H. (1982). *Methods of Information Integration Theory*. Academic Press, New York.

Bateson, J.E.G., Reibstein, D. & Boulding, W. (1987). 'Conjoint Analysis Reliability and Validity: A Framework for Future Research'. *Review of Marketing*, 451-481.

Beggs, S., Cardell, S. & Hausman, J. 'Assessing the Potential Demand for Electric Cars'. *Journal of Econometrics*, 16: 1-19.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. & Rao, V. (1994). 'Combining Revealed and Stated Preference Data'. *Marketing Letters: Special Issue on the Duke Invitation Conference on Consumer Decision-Making and Choice Behavior*, 5(4)

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

Ben-Akiva, M. & Morikawa, T. (1990). 'Estimation of Switching Models from Revealed Preferences and Stated Intentions'. *Transportation Research A*, 24A(6): 485-495.

Ben-Akiva, M., Morikawa, T. & Shiroishi, F. (1991). 'Analysis of the Reliability of Preference Ranking Data', *Journal of Business Research*, 24: 149-164.

Bennett, J.W. (1987). 'Strategic Behaviour: Some Experimental Evidence'. *Journal of Public Economics*, 32: 355-368.

Bennett, J.W. (1996). 'The Contingent Valuation Method: A Post-Kakadu Assessment'. *Agenda*.

Bennett, J.W., Blamey, R. & Morrison, M.D. (forthcoming). *Contingent Valuation of Wetlands in the Upper-South-East of South Australia*. Report prepared for the Australian Bureau of Agricultural Economics.

Bishop, R.C. & Heberlein, T.A. (1979). 'Measuring Values of Extramarket goods: Are indirect Measures Biased?'. *American Journal of Agricultural Economics*, 61: 926-930.

Blamey, R.K. (1996). 'Decisiveness, Attitude Expression and Symbolic Responses in Contingent Valuation Surveys'. *Journal of Economic Behaviour and Organisation*, Forthcoming.

- Bohm, P. (1972). 'Estimating Demand for Public Goods: An Experiment'. *European Economic Review*, 3: 111-130.
- Bohm, P. (1984). 'Revealing Demand for an Actual Public Good'. *Journal of Public Economics*, 24: 135-151.
- Box, G.E.P. & Hunter, J.S. (1961). 'The 2^{k-p} Fractional Factorial Designs Part 1'. *Technometrics*, 3(3): 311-351.
- Boyle, K.J. (1989). 'Commodity Specification and the Framing of Contingent-Valuation Questions'. *Land Economics*, 65(1): 57-63.
- Boyle, K.J., Bishop, R.C. & Welsh, M.P. (1985). 'Starting Point Bias in Contingent Valuation Bidding Games'. *Land Economics*, 61(2): 188-193.
- Boyle, K.J., Desvousges, W.H., Johnson, F.R., Dunford, R.W. & Hudson, S.P. (1994). 'An Investigation of Part-Whole Biases in Contingent Valuation Studies'. *Journal of Environmental Economics and Management*, 27: 64-83.
- Brookshire, D.S., Ives, B.C. & Schulze, W.D. (1976). 'The Valuation of Aesthetic Preferences'. *Journal of Environmental Economics and Management*, 7: 325-346.
- Brookshire, D.S., d'Arge, R.C., Schulze, W.D. & Thayer, M.A. (1981). 'Experiments in Valuing Public Goods', Smith, V.K. (ed), *Advances in Applied Microeconomics*. JAI Press Inc, Connecticut, in Cummings, R.G., Brookshire, D.S. & Schulze, W.D. (1986). *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman & Allanheld, Totowa.
- Brown, T.C., Barro, S.C., Manfreda, M.J. & Peterson, G.L. (1995). 'Does Better Information About the Good Avoid the Embedding Effect?' *Journal of Environmental Management*, 44: 1-10.
- Brubaker, E.R. (1984). 'Demand Disclosures and Conditions on Exclusion: An Experiment'. *The Economic Journal*, 94: 536-553.
- Bunch, D.S., Louviere, J.J. & Anderson, D. (1993). *A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes*. Working Paper.
- Carlos, C. (1991). 'What is Town Water Worth?'. *Australian Journal of Soil and Water Conservation*, 4(3): 32-36.
- Carson, R.T. (1995 draft). *Contingent Valuation Surveys and Tests of Insensitivity to Scope*. University of California, San Diego.
- Carson, R.T., Flores, N.E., Martin, K.M. & Wright, J.L. (1995 draft). *Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods*. University of California, San Diego.
- Carson, R.T., Louviere, J.J., Arabie, P., Bunch, D.S., Hensher, D.A., Johnson, R.M., Kuhfeld, W.F., Steinberg, D., Swait, J., Timmermans, H. & Wiley, J.B. (1994). 'Experimental Analysis of Choice'. *Marketing Letters*, 5(4): 351-368.
- Carson, R.T., Wright, J., Carson, N., Alberini, A. & Flores, N. (1995). *A Bibliography of Contingent Valuation Studies and Papers*. Natural Resource Damage Assessment, Inc.
- Cattin, P. & Weinberger, M.G. (1980). 'Some Validity and Reliability Issues in the Measurement of Attribute Utilities', in Olson, J.C. (ed). *Advances in Consumer Research*, Vol.7. Ann Arbor, Michigan.
- Chapman, R.G. & Staelin, R. (1982). 'Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model'. *Journal of Marketing Research*, 19: 288-301.

- Ciriacy-Wantrup, S.V. (1947). 'Capital Returns from Soil-Conservation Practices'. *Journal of Farm Economics*, 29: 1188-1190.
- Cummings, R.G., Brookshire, D.S. & Schulze, W.D. (1986). *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman & Allanheld, Totowa.
- Currim, I.S., Weinberg, C.B. & Wittink, D.R. (1981). 'The Design of Subscription Programs for a Performing Arts Series'. *Journal of Consumer Research*, 8: 67-75.
- Daubert, J.T. & Young, R.A. (1981). 'Recreational Demands for Maintaining Instream Flows: a Contingent Valuation Approach'. *American Journal of Agricultural Economics*, 63: 667-676.
- Davidson, J.D. (1973). 'Forecasting Traffic on STOL'. *Operations Research Quarterly*, 22: 561-569.
- Desvousges, W.H., Smith, V.K. & Fisher, A. (1987). 'Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River'. *Journal of Environmental Economics and Management*, 14: 248-267.
- Desvousges, W.H., Johnson, F.R., Hudson, S.P., Gable, A.R. & Ruby, M.C. (1996). *Using Conjoint Analysis and Health-State Classifications to Estimate the Value of Health Effects of Air Pollution*. Draft Report prepared for Environment Canada. Triangle Research Institute.
- Dey, A. (1985). *Orthogonal Fractional Factorial Designs*. Wiley, New Delhi.
- Diamond, P.A. & Hausman, J.A. (1994). 'Contingent Valuation: Is Some Number Better than No Number?'. *Journal of Economic Perspectives*, 8(4): 45-64.
- Gan, C. & Luzar, E.J. (1993). 'A Conjoint Analysis of Waterfowl Hunting in Louisiana'. *Journal of Agricultural and Applied Economics*, 25(2): 36-45.
- Goodman, A.C. (1989). 'Identifying Willingness-to-pay for Heterogeneous Goods with Factorial Survey Methods'. *Journal of Environmental Economics and Management*, 16: 58-79.
- Graves, P., Murdoch, J.C., Thayer, M.A. & Waldman, D. (1988). 'The Robustness of Hedonic Price Estimation: Urban Air Quality'. *Land Economics*, 64(3): 220-233.
- Green, P.E. & Rao, V.R. (1971). 'Conjoint Measurement for Quantifying Judgmental Data'. *Journal of Marketing Research*, 8: 355-363.
- Green, P.E. & Srinivasan, V. (1978). 'Conjoint Analysis in Consumer Research: Issues and Outlook'. *Journal of Consumer Research*, 5: 103-123.
- Green, C.H. & Tunstall, S.M. (1991). 'The Evaluation of River Water Quality Improvements by the Contingent Valuation Method'. *Applied Economics*, 23: 1135-1146.
- Greenley, D.A., Walsh, R.G. & Young, R.A. (1982). 'Option Value: Empirical Evidence for a Case Study of Recreation and Water Quality'. *Quarterly Journal of Economics*, 100(1): 657-673.
- Griner, B.P. (forthcoming). *A Conjoint Analysis of Water Quality Improvements*. <http://www.griner@info.pitt.edu>
- Hanemann (1984). 'Discrete/Continuous Models of Consumer Demand'. *Econometrica*, 52: 541-562.
- Hanemann (1985). 'Some issues in continuous and discrete-response contingent valuation studies'. *Northeast Journal of Agricultural Economics*, 5-13.
- Hanemann, W.M. (1994). 'Valuing the Environment Through Contingent Valuation'. *Journal of Economic Perspectives*, 8(4): 19-43.

- Hanemann, W.M. & Kanninen, B. (1996). *The Statistical Analysis of Discrete-Response CV Data*. Working Paper No.798, Department of Agricultural and Resource Economics, University of California at Berkeley.
- Hargreaves, G., Claxton, J.D. & Siller, F.H. (1976). 'New Product Evaluation: Electric Vehicles for Commercial Applications'. *Journal of Marketing*, 74-77.
- Hausman, J.A. & Ruud, P. (1987). 'Specifying and Testing Econometric Models for Rank-Ordered Data'. *Journal of Econometrics*, 34: 83-10.
- Hensher, D.A. (1991). 'The Use of Discrete Choice Models in the Determination of Community Choices in Public Issue Areas Impacting on Business Decision Making'. *Journal of Business Research*, 23: 299-309.
- Hensher, D.A. (1995). *Value of Travel Time Savings in Personal and Commercial Automobile Travel*. Report No ITS-WP-95-4. Institute of Transport Studies, University of Sydney.
- Hensher, D.A., Barnard, P., Milthorpe, F. & Smith, N. (1989). 'Urban Tollways and the Valuation of Travel Time Savings'. *The Economic Record*, 66: 146-156.
- Hensher, D.A. & Battelino, H.C. (1993). 'The Use of Discrete Choice Models in the Determination of Community Preferences Towards Sub-arterial Traffic Management Devices'. *Proceeding of the 7th World Conference on Transportation Research*, Lyon France.
- Hensher, D. & Bradley, M. (1993). 'Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models'. *Marketing Letters*, 4(2): 139-151.
- Hensher, D.A. & Johnson, L.W. (1981). *Applied Discrete Choice Modelling*. John Wiley & Sons, New York.
- Herriges, J.A. & Shogren, J.F. (1996). 'Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning'. *Journal of Environmental Economics and Management*, 30: 112-131.
- Holmes, T.P. & Kramer, R.A. (1995). 'An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation. *Journal of Environmental Economics and Management*, 29: 121-132.
- Jakobsson, K.M., Kennedy, J.O.S. & Elliott, M. (1995). *Survey Method of Valuing the Conservation of Endangered Species*. La Trobe University, Agricultural Economics Discussion Paper 26/95.
- Johansen, L. (1977). 'The Theory of Public Goods: Misplaced Evidence?'. *Journal of Public Economics*, 7: 147-152.
- Johnson, F.R., Desvousges, W.H. & Wood, L.L. (1995). *Conjoint Analysis of Individual and Aggregate Environmental preferences*. Triangle Economic Research Technical Working Paper No T-9502.
- Kahneman, D. (1986). 'Comments by Professor Daniel Kahneman', in Cummings, R.G., Brookshire, D.S. & Schulze, W.D. (1986). *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman & Allanheld, Totowa.
- Kahneman, D. & Knetsch, J.L. (1992). 'Valuing Public Goods: The Purchase of Moral Satisfaction'. *Journal of Environmental Economics and Management*, 22: 57-70.
- Kaoru, Y. (1993). 'Differentiating Use and Nonuse Values for Coastal Pond Water Quality Improvements'. *Environmental and Resource Economics*, 3: 487-494.

- Kealy, M.J., Montgomery, M. & Dovidio, J.F. (1990). 'Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?'. *Journal of Environmental Economics and Management*, 19: 244-263.
- Kosz, M. (1996). 'Valuing Riverside Wetlands; the Case of the 'Donau-Auen' National Park'. *Ecological Economics*, 16: 109-127.
- Krupnick, A.J. & Cropper, M.L. (1992). 'The Effect of Information on Health Risk Valuations'. *Journal of Risk and Uncertainty*, 5: 29-48.
- Lant, C.L. & Roberts, R.S. (1990). 'Greenbelts in the Cornbelt: Riparian Wetlands, Intrinsic Values, and Market Failure'. *Environment and Planning A*, 22: 1375-1388.
- Laureau, T.J. & Rae, D.A. (1989). 'Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique'. *Southern Economics Journal*, 55(3): 728-742.
- Leigh, T.W., McKay, D.B. & Summers, J.O. (1981). 'On Alternative Experimental Methods for Conjoint Analysis', in Monroe, K.B. (ed). *Advances in Consumer Research*, Vol.8. Anne Arbor, Michigan.
- Leigh, T.W., MacKay, D.B. & Summers, J.O. (1984). 'Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison'. *Journal of Marketing Research*, 21: 456-462.
- Loomis, J.B. (1990). 'Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques'. *Journal of Environmental Economics and Management*, 18: 78-85.
- Lockwood, M. (1992). 'Embedding Effects in Contingent Valuation Studies and the Implications for Benefit-Cost Analysis', in Lockwood, M. & DeLacy, T. (eds). *Valuing Natural Areas: Applications and Problems of the Contingent Valuation Method*. Johnstone Centre of Parks, Recreation and Heritage, Charles Sturt University, Albury.
- Lockwood, M. (1996). *Integrated Value Assessment Using Paired Comparisons*. Unpublished Paper. Johnstone Centre of Parks, Recreation & Heritage, Charles Sturt University.
- Louviere (1988a). *Analyzing Decision Making: Metric Conjoint Analysis*. Sage Publications, Newbury Park.
- Louviere, J.J. (1988b). 'Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity'. *Journal of Transport Economics and Policy*, 22(1): 93-119.
- Louviere, J.J., Fox, M.F. & Moore, W.L. (1993). 'Cross-Task Validity Comparisons of Stated Preference Choice Models'. *Marketing Letters*, 4(3): 205-213.
- Louviere, J.J. & Hensher, D.A. (1982). 'On the Design and Analysis of Simulated Choice or Allocation Experiments in Travel Choice Modelling'. *Transportation Research Record*, 890: 11-17.
- Louviere, J.J. & Woodworth, G. (1983). 'Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data'. *Journal of Marketing Research*, 20: 350-367.
- Lynch, J.G. (1985). 'Uniqueness Issues in the Decompositional Modeling of Multiattribute Overall Evaluations: An Information Integration Perspective'. *Journal of Marketing Research*, 22: 1-19.
- Mackenzie, J. (1992). 'Evaluating Recreation Trip Attributes and Travel Time Via Conjoint Analysis'. *Journal of Leisure Research*, 24(2): 171-184.
- Mackenzie, J. (1993). 'A Comparison of Contingent Preference Models'. *American Journal of Agricultural Economics*, 75: 593-603.

- Magat, W.A., Viscusi, W.K. & Huber, J. (1988). 'Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation'. *Journal of Environmental Economics and Management*, 15: 395-411.
- McCullough, J. & Best, R. (1979). 'Conjoint Measurement: Temporal Stability and Structural Reliability'. *Journal of Marketing Research*, 16: 26-31.
- Marwell, G. & Ames, R.E. (1981). 'Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods, IV'. *Journal of Public Economics*, 15: 295-310.
- McFadden, D. (1974). 'Conditional Logit Analysis of Qualitative Choice Behaviour', in Zarembka, P. (ed). *Frontiers in Econometrics*. Academic Press, New York.
- McFadden, D. (1987). 'Regression-Based Specification Tests for the Multinomial Logit Model'. *Journal of Econometrics*, 34: 63-82.
- Meyer, R.J. & Eagle, T.C. (1982). 'Context Induced Parameter Instability in a Disaggregate Stochastic Model of Store Choice'. *Journal of Marketing Research*, 19: 62-71.
- Milon, J.W. (1989). 'Contingent Valuation Experiments for Strategic Behaviour'. *Journal of Environmental Economics and Management*, 17: 293-308.
- Mitchell, R.C. & Carson, R.T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington DC.
- Montgomery, D.B. & Wittink, D.R. (1980). 'The Predictive Validity of Conjoint Analysis for Alternative Aggregation Schemes', in Montgomery, D.B. & Wittink, D.R. (eds). *Proceedings of the 1979 ORSA/TIMS Conference in Marketing*. Marketing Science Institute, Massachusetts.
- Morley, C.L. (1994). 'Experimental Destination Choice Analysis'. *Annals of Tourism Research*, 21(4): 780-791.
- NSW EPA (1995). *ENVALUE: NSW EPA ENVIRONMENTAL VALUATION DATABASE*. NSW Environment Protection Authority, Chatswood.
- Opaluch, J.J., Swallow, S.K., Weaver, T., Wessells, W. & Wichelns, D. (1993). 'Evaluating Impacts from Noxious Facilities: Including Public Preferences in Current Siting Mechanisms'. *Journal of Environmental Economics and Management*, 24: 41-59.
- Peterson, G.L., Brown, T.C., McCollum, D.W., Bell, P.A. Birjulin, A.A. & Clarke, A. (1996). 'Moral Responsibility Effects in Valuation of WTA for Public and Private Goods by the Method of Paired Comparison', in Adamowicz, W.L., Boxall, P., Luckert, M.K., Phillips, W.E. & White, W.A. (eds). *Forestry, Economics and the Environment*. CAB International, Wallingford.
- Portney, P.R. (1994). 'The Contingent Valuation Debate: Why Economists Should Care'. *Journal of Economic Perspectives*, 8(4): 3-17.
- Rae, D.A. (1983). 'The Value to Visitors of Improving Visibility at Mesa Verde and Great Smoky National Parks', in Rowe, R.P. & Chestnut, L.G. (eds). *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*. Westview Press, Boulder, Colorado.
- Randall, A. & Hoehn, J.P. (1996). 'Embedding in Market Demand Systems'. *Journal of Environmental Economics and Management*, 30: 369-380.
- Ready, R.C., Whitehead, J.C. & Blomquist, G.C. (1995). 'Contingent Valuation When Respondents Are Ambivalent'. *Journal of Environmental Economics and Management*, 29: 181-196.
- Reibstein, D., Bateson, J.E.G. & Boulding, W. (1987). *Conjoint Analysis Reliability: Empirical Findings*, Report no. 87-102, Marketing Science Institute.

- Reiling, S.D., Boyle, K.J., Philips, M.L. & Anderson, M.W. (1990). 'Temporal Reliability of Contingent Values'. *Land Economics*, 66(2): 128-134.
- Robbins, L. (1937). *An Essay on the Nature and Significance of Economic Science*. Macmillan & CO Ltd, London.
- Robinson, P.J. (1980). 'Applications of Conjoint Analysis to Pricing Problems', in Montgomery, D.B. & Wittink, D.R. (eds). *Proceedings of the 1979 ORSA/TIMS Conference in Marketing*. Marketing Science Institute, Massachusetts.
- Roe, B., Boyle, K.J. & Teisl, M.F. (1996). 'Using Conjoint Analysis to Derive Estimates of Compensating Variation'. *Journal of Environmental Economics and Management*, Forthcoming.
- Rolfe, J. & Bennett, J. (1996). *Valuing International Rainforests: A Choice Modelling Approach*. Paper presented at the 40th Annual Conference of the Australian Agricultural Economics Society. Melbourne, February 11-16.
- Rowe, R.D., d'Arge, R.C. & Brookshire, D. (1980). 'An Experiment of the Economic Value of Visibility'. *Journal of Environmental Economics and Management*, 7: 1-19.
- Samples, K.C., Dixon, J.A. & Gowen, M.M. (1986). 'Information Disclosure and Endangered Species Valuation'. *Land Economics*, 62: 306-312.
- Samuelson, P.A. (1955). 'Diagrammatic Exposition of a Theory of Public Expenditure'. *Review of Economics and Statistics*, 37: 350-356.
- Schechter, M. & Kim, M. (1991). 'Valuation of Pollution Abatement Measures: Direct and Indirect Measurement'. *Journal of Urban Economics*, 30: 33-151.
- Segal, M.N. (1982). 'Reliability of Conjoint Analysis: Contrasting Data Collection Procedures'. *Journal of Marketing Research*, 19: 139-143.
- Segal, R. (1995). 'Forecasting the Market for Electric Vehicles in California Using Conjoint Analysis'. *The Energy Journal*, 16(3): 89-111.
- Sinden, J.A. (1974). 'A Utility Approach to the Valuation of Recreational and Aesthetic Experiences'. *American Journal of Agricultural Economics*, 56(1): 61-72.
- Smith, V.K. (1992). 'Arbitrary Values, Good Causes, and Premature Verdicts'. *Journal of Environmental Economics and Management*, 22: 71-89.
- Smith, V.K. & Desvousges, W.H. (1986). *Measuring Water Quality Benefits*. Kluwer-Nojhoff Publishing, Boston.
- Smith, V.K., Desvousges, W.H. and Fisher, A. (1986). 'A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits'. *American Journal of Agricultural Economics*, 280-290.
- Smith, V.K., Desvousges, W.H. & Freeman, A.M. (1985). Valuing Changes in Hazardous Waste Risks: A Contingent Valuation Analysis. Draft Interim Report prepared for the US Environmental Protection Agency in Smith, V.K. & Desvousges, W.H. (1986). *Measuring Water Quality Benefits*. Kluwer-Nojhoff Publishing, Boston.
- Smith, V.L. (1977). 'The Principle of Unanimity and Voluntary Consent in Social Choice'. *Journal of Political Economy*, 85: 1125-1139.
- Stevens, T.H., More, T.A. & Glass, R.J. (1994). 'Interpretation and Temporal Stability of CV Bids for Wildlife Experience: A Panel Study'. *Land Economics*, 70(3): 355-363.

Swait, J. & Adamowicz, W. (1996). *The Effect of Choice Environment and Task Demands on Consumer Behaviour: Discriminating Between Contribution and Confusion*. Working paper, Department of Rural Economy, Alberta.

Swait, J. & Louviere, J. (1993). 'The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models'. *Journal of Marketing Research*, 30: 305-314.

Swait, J., Louviere, J. & Williams, M. (1994). 'A Sequential Approach to Exploiting the Combined Strength of SP and RP Data: Application to Freight Shipper Choice'. *Transportation*, 21: 135-152.

Thayer, M. (1981). 'Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence'. *Journal of Environmental Economics and Management*, 8: 27-44.

Throsby, C.D. & Withers, G.A. (1986). 'Strategic Bias and Demand For Public Goods'. *Journal of Public Economics*, 31: 307-327.

Thurstone, L. (1927). 'A Law of Comparative Judgment'. *Psychological Review*, 34: 273-286.

Viscusi, W.K., Magat, W.A. & Huber, J. (1991). 'Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis'. *Journal of Environmental Economics and Management*, 21: 32-51.

Walsh, R.G., Loomis, J.B. & Gillman, R.A. (1984). 'Valuing Option, Existence, and Bequest Demands for Wilderness'. *Land Economics*, 60(1): 14-29.

Whitmore, G.A. & Cavadias, G.S. (1974). 'Experimental Determination of Community Preferences for Water Quality-Cost Alternatives'. *Decision Sciences*, 5:614-631.

Wilks, L. (1990). *A Survey of the Contingent Valuation Method*. Research Paper No2, Resource Assessment Commission.

Wittink, D.R. & Cattin, P. (1989). 'Commercial Use of Conjoint Analysis: An Update'. *Journal of Marketing*, 53: 91-96.