is that the non-market values for unfamiliar, complex phenomena (such as biodiversity) are likely to depend crucially on how this valuation is carried out, and are likely to vary over contexts.

11.3 Empirical methods for valuing the environment

In this section, we give a brief overview of the main methods available to economists wishing to estimate the non-market, economic value of changes in environmental quality. Valuation methods are usually divided into two approaches: stated and revealed preference methods. Stated preference (SP) methods seek to measure individuals' values for environmental quality directly, by asking them to state their preferences for the environment. In contingent valuation surveys, for example, this might consist of asking people for either their maximum willingness to pay (WTP) for an increase in environmental quality or their minimum willingness to accept compensation (WTA) to forgo such an increase. Because the level of environmental quality is usually chosen by a third party (such as the government requiring a certain level of air quality to be achieved) and not by the individual being questioned, these measures correspond to either the Hicksian equivalent or the compensating surplus (rather than equivalent or compensating variation, where quality/quantity is chosen by the individual). Another SP method, growing in popularity amongst environmental economists, is known as choice modelling, and we also describe this in Section 11.3.2.

Revealed preference methods seek to recover estimates of individuals' WTP for environmental quality by observing their behaviour in related markets. In the hedonic pricing model, for example, the related market for urban air quality is the housing market, and economists seek to infer individuals' valuation of air quality improvements by considering their behaviour in the house market. In the travel cost model, the analyst tries to infer the value people place on changes in quality at outdoor recreational sites through their expenditure on travel to the site, and choices over alternative sites. We explain all of these in more detail in what follows.

STATED PREFERENCE APPROACHES

11.3.1 The contingent valuation method

The contingent valuation method (CVM) for the valuation of environmental goods was first used by Davis (1963) in a study of hunters in Maine. However, it was not until mid-1970s that the method's development began in earnest (Randall *et al.*, 1974; Brookshire *et al.*, 1976). Since then, the method has become the most widely used (and perhaps most controversial) of all environmental valuation techniques. Comprehensive accounts of the method may be found in Bateman and Willis (1999) and Bateman *et al.* (2002). In what follows, we first run through the stages of a CVM, then review some major current research issues (Carson *et al.*, 2001). Finally, evidence on the validity of CVM is

considered. Since a very large literature now exists on CVM, we focus on a selection of issues only.

Most CVM exercises can be split into five stages: (1) setting up the hypothetical market, (2) obtaining bids, (3) estimating mean WTP and/or WTA, (4) aggregating the data and (5) carrying out validity checks.

Stage 1: The hypothetical market

The first step is to set up a hypothetical market for the environmental service flow in question. For example, take a policy to restore old civic buildings in a city centre. Respondents might be told that the local government could engage in such restoration activities, describe what these would consist of and their effects and explain that the operation could only go ahead if extra funds are generated. This sets up a reason for payment for services (the aesthetic quality of the built environment in this example). How funds will be raised also needs to be described: the bid vehicle must be decided upon – for example, property taxes, local income taxes or trust fund payments. In this example, the bid vehicle could be higher property taxes. The survey instrument (questionnaire) should also describe whether all consumers will pay if the change goes ahead, and how the decision on whether to proceed with the project will be made. The questionnaire should be pre-tested before the main survey occurs: this is often done using small focus groups assembled to discuss their reactions to a questionnaire prior to a pilot study – indeed, such groups are key to successful study design. The information given to respondents about all aspects of the hypothetical market, together with such information as is provided on the good being valued, constitutes the 'framing' of the good.

Stage 2: Obtaining bids

Once the survey instrument is set up, the survey is administered. This can be done by either face-to-face interviewing, telephone interviewing, or mail. Telephone interviews are probably the least-preferred method since conveying information about the good may be difficult over the telephone, partly owing to a limited attention time span. Mail surveys are frequently used, but suffer from potential non-response bias and often from low response rates. Interviews offer the most scope for detailed questions and answers, but are relatively costly. Typically, a CVM survey will ask some general questions about environmental attitudes; test for knowledge of the good in question and provide information on the hypothetical scenario; collect WTP/WTA information; ask for socio-economic data on the respondent and pose some 'de-briefing' questions such as how hard the respondent; found the exercise.

Individuals can be asked to state their WTP/WTA in a number of ways (in what follows, we focus on WTP alone for simplicity, and since that is what most studies estimate in practice):

- As a payment card. A range of values is presented on a card, and the respondent is asked to pick that which most closely matches their WTP. Payment ladders can also be used. Data from such modes can be treated either as continuous information on WTP or, more correctly, as interval-type data.

- As an open-ended question. Individuals are asked for their maximum WTP with no value being suggested to them. Respondents have often found it relatively difficult to answer such questions, especially where they have no prior experience of trading with the commodity in question.
- As a single valued referendum: A single payment is suggested, to which respondents either agree or disagree (yes/no reply). Such responses are often known as single-bounded dichotomous choice (SBDC) responses. Double-bounded dichotomous choice (DBDC) models present those respondents who say 'no' to the first amount with a lower amount and those respondents who say 'yes' to the first amount with a higher amount (see, for example, Carson *et al.*, 1994). Other variants exist.

Stage 3: Estimating WTP

For open-ended responses, calculating mean or median WTP is simple, although researchers must take care to separate out protest responses first: these are zero values for WTP given for reasons other than a zero value being placed on the environmental good in question. These might occur because an individual objects on moral grounds to paying for the environmental good, or finds the hypothetical scenario non-credible. Mean WTP is the relevant value for use in cost-benefit analysis, although authors often focus on median WTP since it is less impacted by extreme values, and since it is meaningful from a political consensus viewpoint. Confidence intervals for WTP should also be reported, and bid functions are usually estimated to investigate the determinants of variations in WTP. For dichotomous choice designs, the researcher must estimate WTP, since all the respondent reveals is whether she is willing to pay a given amount, not her maximum. Several approaches are available to do this, the most popular being Hanemann's 'utility difference approach', which we now explain*. An alternative, due to Cameron (1988) is also touched on. Full treatments of these issues raised here can be found in Hanemann and Kanninen (1999) and in Bateman *et al.* (2002).

Let us focus on a SBDC design, and assume that an individual derives utility from an environmental good q, and market-valued goods X. Write the indirect utility function for this person as $V(q, y)$, where y is income. Let us assume that the researcher cannot observe all of the aspects of $V(.)$: for example, we may not be able to measure her preferences well, or there may be other aspects of life which she derives utility from! This idea is known as the *random utility model*, and underlies the DC version of CVM, as well as choice modelling and multiple-site travel cost models. We capture the unobservable aspects of utility in an error term, e: it is assumptions about the distribution of this error term and about the functional form of $v(.)$ which will give rise to different models of WTP. For the individual (Sue), imagine she is offered the option that environmental quality will rise from q_0 to q_1, where q_1 is 'better'. The researcher conducts a CVM study

*The presentation here follows Hanemann and Kanninen (1999) closely.

the form of the bid function and the distribution of e. For instance, Bateman *et al.* cite two simple possible bid function models:

$$C = a + e$$

and

$$C = a_0 + by + e$$

Which method is best? There is no clear-cut answer here, but for a discussion see Bateman *et al.* (op. cit.).

Once the discrete response data has been used to estimate a probilistic choice model, such as (11.6) above, then we want to calculate welfare measures, typically mean and median WTP. This may require certain restrictions to be placed on the estimation, depending on whether we assume that WTP could equally well be negative as positive, or if we want to impose some limits on WTP (e.g. that it cannot be greater than income). Again, see the extended discussion in Hanemann and Kanninen. If WTP can only take positive values for the increase in q, then using the general expression given in (11.5), mean WTP C^* could be calculated as

$$C^* = \int_0^\infty [1 - G_c(A)]\mathrm{d}A$$

and median WTP C^{**} is given by

$$1 - G_c(A) = 0.5$$

that is, the value of A that there is a 50–50 chance a randomly selected person would agree to pay. The exact formulae for calculating the numerical value of mean and median WTP will depend on what functional form was chosen for the utility function, what distribution was chosen for the random part of utility and what theoretical restrictions, if any, are placed on willingness to pay.

Stage 4: Aggregating the data

Aggregation refers to the process whereby the mean bid or bids are converted to a population total value figure. Decisions over aggregation revolve around three issues. First is the choice of the relevant population. This should have been decided when constructing the sampling frame from which the sample was drawn. The aim is to identify either (a) all those whose utility will be significantly affected by the action or (b) (which is the same or a smaller group) all those within a relevant political boundary who will be affected by the action. A decision must be made over the criteria to be used in deciding on who counts in (a) or (b). This group might be the local population, the regional population, the population of Scotland, or the population of the UK, or the whole of Europe. Clearly, where significant non-use values are involved,

this population of beneficiaries could be very large. One recently emerging approach to estimating the size of the benefiting population is to use distance-decay functions (Hanley *et al.*, 2003b), whereby a model is estimated which relates how far people live from an environmental resource to how much they are WTP to protect it/improve it. If such a function can be estimated, then the distance at which the expected WTP goes to some very small number, or to zero, could be calculated, and the relevant population estimated using GIS. But there are some big unresolved questions about this approach, such as the existence and transferability of distance-decay functions, their existence for some environmental goods and about how small a welfare change is 'too small' to count.

Second is moving from the sample mean to a mean for the total population. Several alternatives have been proposed. The sample mean could be multiplied by the number of households in the population, *N*. However, the sample might be a biased reflection of the relevant population; for instance, it might have higher income levels or show a lower level of educational achievement. If these variables have been included in a bid curve (see p. 334), an estimated population mean bid can be derived by inserting population values for the relevant variables in the bid curve. This number could then be multiplied by *N*.

Third is the choice of the time period over which benefits should be aggregated. This will depend on the setting within which the CVM exercise is being performed. If the present value of environmental benefit flows over time is of interest, then benefits are discounted. Where an irreversible environmental loss is involved, then the present value is calculated by taking a perpetuity (i.e. by dividing the constant real benefit flow by the discount rate). In all cases of benefit or cost flows over time where the time period is sufficiently long, society is confronted by the necessity of using current preferences to measure future preferences, as well as with the equity implications of discounting.

Stage 5: Carrying out validity checks

How good are the CVM estimates which the analyst produces? This is clearly an important question from a policy perspective, and in terms of the credibility of environmental valuation. Several 'validity checks' have emerged. These are

- scope tests
- convergent validity
- calibration factors
- protest rates and
- construct validity

Scope tests consist of examining whether WTP varies significantly with the quantity of q on offer. A simple scope test would be to test the null hypothesis that WTP $(q_2) >$ WTP (q_1), where we assume $q_2 > q_1$. For example, this could mean that WTP to protect all wetlands in one region of France was greater than WTP to protect a single

wetland. Scope tests arose as a validity criterion because of a worry that the failure of WTP to show scope sensitivity would imply a poor description of the environmental change in question had been provided, or that people's WTP amounts were largely symbolic donations which could not be interpreted as compensating surplus/equivalent surplus – although sometimes a CVM survey may fail a scope test due to a small sample size. Convergent validity is a test for whether WTP for a given environmental quality change estimated using CVM is significantly different from WTP for the same change using some other technique; for instance, comparing CVM and travel cost estimates for a day's fishing. This assumes that CVM and, in this instance, travel costs measure the same underlying value – which may not be true when non-use values are concerned.

Calibration factors address a fundamental weakness of CVM: that the values stated are hypothetical commitments, not real ones. A calibration factor is calculated by comparing a WTP value obtained from a CVM survey with a comparable real commitment – obtained, typically, through experimental economics methods (Fox *et al.*, 1998), or occasionally by means of a comparison with actual voting behaviour (Schlapfer *et al.*, 2004). If WTP (CVM) is bigger than WTP (real), then doubt is cast on the CV estimate. Harrison and Ruström's (2005) review shows 34 of 39 tests revealing hypothetical bias, ranging from 2 to 2600 per cent. A review is also provided by Murphy *et al.* (2005), who find a mean calibration factor of 1.35 (i.e. stated values exceed actual monetary values by 35%), although they note that for public goods, this hypothetical bias increases on average. These results reinforce the argument that people tend to overstate their actual WTP when confronted with hypothetical questions. Conversely, in a study of 616 comparisons of contingent valuation results and estimates derived from actual markets via revealed preference methods, Carson *et al.* (1996) found that CVM estimates were on average *lower* than revealed preference estimates.

However, it is hard to calculate calibration factors for many environmental goods since the reason why we undertake CVM is precisely because some aspect of the good defies market valuation; this is especially true for non-use values. Many experimental studies have shown that stated WTP is bigger than actual WTP. Could we therefore claim that CVM *always* produces numbers that are 'too big'? No. The current view is that the calibration factor varies according to the nature of the good and the nature of the valuation market, and does not lend itself to generalisation.

Protest rates are another indicator of the quality of a CVM survey. The protest rate is defined as the percentage of responses which are protest bids (see above): too high a protest rate ('too high' is a subjective matter, but a protest rate of over 50% would raise concerns) implies something wrong with the design of the hypothetical market; for example people did not find it believable, or found it morally objectionable. Finally, the worth of an individual CVM study can be assessed using the criterion of construct validity. This asks whether WTP varies in a manner which is consistent with theoretical expectations. Usually this question is addressed by estimating a bid function, and seeing

whether parameter signs are in accord with a priori expectations (e.g. do people with more experience or knowledge of the good pay more? Does higher income boost WTP?), and also by considering what percentage of the variation in WTP can be explained statistically. However, for many candidate independent variables it is hard to form definite a priori predictions (e.g. do we expect older people to value forest conservation more than young people? Do we expect locals to value it more than visitors?), whilst there are many different theoretically consistent assumptions one could make about the nature of the underlying utility function. The construct validity notion is therefore not as useful as it first seems.

BOX 11.2

A contingent valuation study of water quality improvements in Costa Rica

Note that the SBDC format gives much higher WTP values than the DBDC format (this is a common finding); that mean WTP is greater than median WTP; and that the distributional assumption does not seem to matter much to the DBDC results in this case. Box 9.7 discusses the differences between values in the two towns, and the use of benefits transfer techniques.

Water quality improvements have been the focus of many contingent valuation (CV) studies in the developed world, but are increasingly being studied in developing countries too. David Barton reports a CV study of water quality improvements in the towns of Jaco and Puntarenas in Costa Rica. The water resources included were coastal waters, rivers and groundwater, with two scenarios for improvement ('full' and 'partial'). A double-bounded dichotomous choice (DBDC) design was used, with monthly sewerage fees as the bid vehicle. Some 954 households in total were sampled, in 1998.

Results are presented according to how the data was analysed. Two alternative DBDC models are estimated. The first assumes that WTP responses are distributed normally, the second that they are distributed lognormally. The author also estimates a model based on the first response only, which gives a single-bounded dichotomous choice estimate of WTP. Results are shown in the following table.

Benefits of improving water quality in Costa Rica (all values in 1998 colones/household)

	Jaco		*Puntarenas*	
Estimation	*Mean WTP*	*Median WTP*	*Mean WTP*	*Median WTP*
Double-bounded: normal distribution	3089	2963	2404	2268
Double-bounded: log normal	3168	2557	2467	1918
Single-bounded	4789	3247	6617	3093

Source: Adapted from Barton (2002).

Design and research issues in CVM

We next review very briefly some important research and design issues in contingent valuation. For much more detail, consult Bateman *et al.* (2002) and Bateman and Willis (1999).

Choice of response mode: One issue which has generated innumerable articles in academic journals is which response mode should be used and how data should be analysed. Open-ended CVM questions have been criticised for being too hard for respondents to complete, and for resulting in high-variance mean WTP distributions. Single-bounded DC became almost 'industry standard' following the 1992 National Oceanographic and Atmospheric Administration report into CVM produced by the US government, partly because it was alleged to be incentive-compatible (Hoehn and Randall, 1987), partly because it appeared more realistic (a fixed price for providing a public good), and because in the US respondents are familiar with voting on local public good issues. However, SBDC designs turned out to produce systematically high mean WTP estimates – 'yea-saying' being one explanation; required larger sample sizes because they are statistically inefficient, and produced mean WTP estimates which are very sensitive to assumptions about functional form and to truncation. Partly in response to these weaknesses, the DBDC design was pioneered by Hanemann and Carson, and became widely used in the 1990s. But concerns arose over the incentive compatibility of this design, and over the effects of the size of the first bid on responses to the second bid (i.e. whether both responses came from the same underlying distribution of WTP; McLeod and Bergland, 1999). Alternative mechanisms are thus still widely used. Payment cards which allow respondents to say how sure they are they would pay the amount asked, over a series of amounts; and payment ladders which allow people to say the most they are sure they would pay, and the least they are sure they would *not* pay – thus typically identifying a range of uncertainties given that people may be unsure of their preferences for some environmental goods – have become popular; non-parametric means of data analysis have been introduced to try to get around sensitivities to distributional assumptions.

Information provision: An early concern in CVM was the sensitivity of WTP estimates to the amount and nature of information provided to respondents (see the survey in Munro and Hanley, 1999). Especially where the analyst is dealing with unfamiliar environmental goods – such as biodiversity – information is crucial if we wish to elicit 'informed' preferences. Yet how best to do this? One new concept which addresses this question is the valuation workshop technique, as explained in MacMillan *et al.* (2003). Finally, an interesting new angle on the information story is concerned with what people know about *why* environmental problems occur: there is now some evidence to suggest that people are willing to pay more to cure environmental problems that they believe to be caused by human actions than identical problems caused by 'the forces of nature' (Bulte *et al.*, 2005).

Linking CVM with psychological insights and economic rationality: A large literature has emerged which tests the ability of CVM to produce results which are consistent with both the axioms of preference, as stated in conventional economic theory (noted in Section 11.1), and findings from behavioural psychology. Both of these angles could be described as the search for anomalies; either for CVM results which are in violation of the axioms of preference, or for those which are not in accord with the findings of behavioural psychologists. The work on preference construction, noted above, is one aspect of this field of enquiry, which also includes work on choice under risk and on embedding (Kahneman and Knetsch, 1991; Green and Tunstall, 1999).

Voluntary versus forced payments: In many contexts, the use of a voluntary payment mechanism is the most realistic choice in designing a CVM study. For example, if one thinks about an increase in the protection of an endangered bird species in the UK, then asking people their maximum WTP in terms of contributions to an environmental charity which acts to buy up and safeguard this bird's habitat is both realistic and in line with peoples' experience. However, some researchers (e.g. Carson *et al.*, 1999a) have recommended against using voluntary payment mechanisms, since they encourage free riding. One way of improving the demand-revealing properties of voluntary payment designs is the Provision Point Mechanism, whereby respondents are told that a minimum level of aggregate contribution is required for the public good to be supplied. This may be reinforced by either a proportional rebate rule (all excess contributions are returned weighted by your WTP) or an extending benefits rule, whereby additional amounts of the public good are provided above that set out, should aggregate contributions exceed the minimum. Poe *et al.* (2002) show that this type of design can greatly improve the demand-revealing potential of voluntary contribution CV studies, by reducing free riding.

11.3.2 Choice modelling

Choice modelling is a generic term for a variety of methods (sometimes referred to as conjoint analysis) which make use of the basic stated preference approach – direct elicitation of preferences – but take a somewhat different line in terms of defining what it is people are valuing, and how they are asked to reveal their preferences. Choice modelling includes contingent ranking, paired comparisons, contingent rating and choice experiments, and is described in great detail in Louviere, Henscher and Swait (2000). Here we focus on just one of these methods – choice experiments.

The choice experiment (CE) approach was initially developed by Louviere and Hensher (1982) and Louviere and Woodworth (1983). Respondents are asked to choose between alternative goods, defined in terms of their attributes. CE share's a common theoretical framework with the dichotomous contingent valuation approach described above in the random utility model; but we present it here slightly differently to enable a better link to the standard way of analysing CE data. As will be recalled, according to the RU framework, the indirect utility function for each respondent $i(U_i)$ can be decomposed into two