Monetary Valuation of Environmental Goods: Alternatives to Contingent Valuation

Baarsma, B.E.

Citation for published version (APA):
Chapter 4
Alternatives to Contingent Valuation

"Faced with situations demanding resolution, critique remains a luxury if it does not point clearly to alternatives." (Lowe et al., 1993, p. 101)

In the previous chapter we saw that the contingent valuation method (CVM) is very popular among valuation researchers. However, we also saw that the CVM has many serious problems. This popularity of the CVM is not caused solely by its scientific superiority, but also by the strong and effective pr-offensive of the proponents. Be that as it may, critique alone does not get us any further if we want to solve environmental problems. Taking the above mentioned quote to heart, we will now review some other valuation methods that do not necessarily replace the CVM but complement it. These methods deal with part of the problems of the CVM. However, sometimes the alternative methods suffer from the same drawbacks that the CVM and other valuation methods suffer from (see the discussion on neoclassical assumptions about individual behaviour in chapter 2, section 2.1.2). Moreover, these methods have problems of their own.

4.1 Introduction

In this chapter, three alternative methods will be discussed, namely the conjoint measurement method, the welfare evaluation method and the well-being evaluation method. Before moving on to the discussion of these alternative methods, two issues that apply to all three alternatives are addressed. The first issue is about the measurement of non-use values, the second one is about the direct way of questioning in CVM surveys.
**Stated preference methods and non-use values**

It is important to notice that the conjoint measurement method, the welfare evaluation method and the well-being evaluation method are full alternatives to the CVM, in the sense that all three methods are capable of measuring the total value of a good, and not just the use part of this value. Actually, in this respect all stated preference methods could serve as full alternatives to the CVM, because stated preference methods use hypothetical markets where all kinds of goods can be surveyed, and thus also goods that individuals do not have a user-relation with. This point was already made in section 3.3 in the previous chapter.

**Strategic answers and cognitive stress**

CVM asks people directly for a monetary value for an environmental good. Because of this direct way of questioning at least two problems arise, namely strategic answers and cognitive stress. The three alternative methods discussed in this chapter outperform the CVM on these two points.

The first problem is perhaps one of the most serious criticisms of the CVM. In a CVM setting very few incentives exist for people to express their true value. They are not punished in any way if they do not express their true value but a value that is lower or higher, in an attempt to influence the provision or the price of a good. In other words, posing this kind of direct questions entails in all probability a severe strategic bias. CVM researchers claim that, by carefully constructing the survey, the potential for strategic bias can be minimized. But this potential for strategic bias is still very large when compared to other, more 'veiled' questioning methods like the three alternatives discussed in this chapter.

Secondly, it is extremely difficult for respondents to come up with an absolute monetary value. This is particularly true for the decisions for which CVM techniques are most needed, which concern novel and abstract judgements and which depend on a lot of information. Although the CVM assumes that respondents have clear and pre-defined ideas about the worth and desirability of possible objects and events, it appears that most people do not have well-structured preferences (cf. chapter 2, section 2.1.2.f). In that case, so much cognitive constriction and perseverance may occur that thought processes are disrupted and thinking becomes simplistic (Harris et al., 1989). The CVM no longer gives reliable and valid results, since coming up with the true values takes too much effort. Therefore, it is of the utmost importance that people's information-processing and decision-making processes are not overburdened and that the questions are framed as easy as possible. The direct CVM question is probably not satisfactory in this regard. By asking indirect questions (as with conjoint
measurement, welfare evaluation and well-being evaluation) the cognitive stress will probably decline.

4.2 Conjoint Measurement

Conjoint measurement (CM) has been widely used in consumer market research and in transportation studies since the 1970s. For instance, when introducing a new product the producer is interested in the relative importance of attributes of that product, like colour, weight, technical specifications, package, price and so on, and not just in the valuation of the product as a whole. Green and Srinivasan (1978, p. 104) define conjoint measurement as:

"any decompositional method that estimates the structure of a consumer's preferences [...] given his or her overall evaluation of a set of alternatives that are prespecified in terms of levels of different attributes."

The typical CM question presents each respondent with a number of commodity descriptions or situations (called vignettes), that differ according to the attributes described, and survey respondents are then asked to rank and rate the desirability of each vignette. The inclusion of price as one of the attributes allows for the derivation of implicit prices for each of the other attributes. Unlike the CVM, CM does not directly ask for a willingness to pay, but requires that respondents rank possible outcomes from most preferred to least preferred, while several attributes of the good are varied. This results in a relative value, in the sense that the expressed value depends upon the other alternatives that have to be ranked.

At least one important behavioural arguments exists in support of the decomposition of preferences by using CM (Louviere, 1996). This argument lies in the theory of individual behaviour, especially Lancaster's (1966) work, which assumes that a consumer's utility for a good can be decomposed into utilities for separate attributes or benefits provided by that good.

CM departs from the random utility maximization model, which admits for the fact that, from a researcher's point of view, consumers do not always seem to choose what they prefer, and that some choices vary over choice occasions (McFadden, 1974). Suppose that an individual has a vector of measured attributes $s$ (like age, income and so on), and that he or she faces $J$ vignettes, indexed $j = 1,...,J$ which are described by the vector of attributes $x_j$. The researcher can only explain part of the individual's
behaviour, namely the deterministic part $V(s,x)$. The other part is stochastic and includes the variations in choice that the researcher cannot explain ($\varepsilon(s,x)$). The utility model is then defined as follows:

$$ U = V(s,x) + \varepsilon(s,x) $$

where $\varepsilon(s,x)$ is a random term, the distribution of which depends on $s$ and $x$. The explainable component of choice, $V(s,x)$, is specified by the researcher. Subsequently, the probability is modelled that an individual will choose a particular vignette $j$ over the other vignettes in the choice set. The analytical model to estimate these probabilities for a particular rank order is a rank-ordered logit model (Beggs et al., 1981). This model provides that set of parameter weights on the attributes which maximizes the likelihood of realizing that rank ordering. Provided that one of the attributes is a money measure, parameter weights can then be used to calculate the WTP for an attribute, as will be explained below. Because CM is based on the random utility maximization model, it is possible to come up with the exact welfare measures (viz. the compensating variation or the equivalent variation), just as the CVM does.

The logit model of choice implies certain restrictions on individuals' choices and preferences. The most notable restriction is that choices must have the property of the Independence of Irrelevant Alternatives (IIA). One implication of this property is that the addition or deletion of vignettes from the choice set does not affect the ratio of the probabilities associated with any other combination of vignettes. McFadden (1974, p. 113) provides the famous auto/bus example to illustrate the IIA assumption.

"Suppose a population faces the alternatives of travel by auto and by bus, and two-thirds choose to use auto. Suppose now a second 'brand' of bus travel is introduced that is in all essential respects the same as the first. Intuitively, two-thirds of the population will still choose auto, and the remainder will split between the bus alternatives. However, if the selection probabilities satisfy IIA, only half the population will use auto when the second bus is introduced."

One approach to dealing with IIA is to redefine the choice set so that two or more very close substitutes are modelled as one alternative. Another approach would be to guarantee the independence of the alternatives (vignettes) as much as possible, by constructing an approximately orthogonal set of vignettes. In short, this implies that
the vignettes in the choice set will have to differ as much as possible. This is described in more detail below.

Anyway, in the case of stated preferences the assumption of IIA is not too restrictive compared to revealed preferences. The reason lies in the fictitious nature of the alternatives which are completely described by their stated attributes.

"Under the usual assumption that the utility function is correctly specified and that, in particular, all relevant individual-specific explanatory variables are taken into account, this implies that the utilities generated by the alternatives do not correlate for a given individual." (Van Ophem et al., 1999, p. 118)

More recently, CM has also been applied in environmental valuation studies. Indeed, CM is very suitable for the valuation of environmental goods, since these goods are pre-eminently goods with a multidimensional character. Table 4.1 below gives an overview of valuation studies using CM. As is immediately obvious from the table, CM encompasses a variety of multi-attribute preference elicitation techniques.

**Choice experiments**
Choice experiments differ from typical CM in that individuals are asked to choose what they consider to be the best vignette from a set of vignettes, instead of ranking or rating them. These choices are then repeated for different sets of vignettes.

**Conjoint rating**
In the conjoint rating approach respondents are shown a series of products or programmes which they have to rate on a specified scale, for instance, from 1 to 10. The researcher can deduce an ordering from these ratings. Sometimes, the products or programmes are presented in pairs. In such a case of rated pairs, the respondent has to indicate how much he or she prefers one product over the other by supplying a rating.

**Contingent ranking**
Respondents to a contingent ranking survey are asked to rank one series of vignettes from least preferred to most preferred. The difference between ranking and rating is that ranking occurs relative to other vignettes, while rating does not. In a rating exercise, two or even all vignettes could receive the same score, which is not possible in a ranking exercise.
**Table 4.1: An overview of valuation studies using conjoint measurement**

<table>
<thead>
<tr>
<th>Method</th>
<th>Study Details</th>
<th>Study Details</th>
<th>Study Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice experiments: respondent has to make repeated choices between vignettes</td>
<td>Adamowicz, W., P. Boxall, M. Williams and J. Louviere (1998).</td>
<td>First application of conjoint methodology to estimate non-use values, i.e. WTP for enhancing the population of a threatened species (the Caribou). Comparison of results of CVM and choice experiments.</td>
<td>Respondent has to choose from 8 sets of 3 vignettes (7 attributes); n = 355.</td>
</tr>
<tr>
<td>Conjoint rating: respondent has to supply a rating for each vignette on a specified scale</td>
<td>Johnson, F.R. and W.H. Desvouges (1996).</td>
<td>Estimating environmental, health, and employment effects of energy programmes with rated-pair surveys.</td>
<td>Respondent has to rate 26 pairs of 2 vignettes (9 attributes); n = 205.</td>
</tr>
<tr>
<td></td>
<td>Magat, W.A., W.K. Viscusi and J. Huber (1988).</td>
<td>Comparison of WTP results for morbidity risk reductions by making products (drain openers DO and bleach B) safer based on rated-pair comparison and CVM.</td>
<td>Respondent has to rate 5 pairs of 2 vignettes (2 attributes); n_c0 = 168, n_c1 = 200.</td>
</tr>
<tr>
<td>Contingent ranking: respondent has to rank a series of x vignettes from least to most preferred</td>
<td>Desvouges, W.H., V.K. Smith and M.P. McGivney (1983).</td>
<td>Comparison of alternative approaches, among which contingent ranking and CVM, for estimating recreational and related benefits of water quality improvements.</td>
<td>Respondent has to rank 4 vignettes (2 attributes); n = 213.</td>
</tr>
<tr>
<td></td>
<td>Lareau, T.J. and D.A. Rae (1989).</td>
<td>Estimation of the WTP for diesel-vehicle odour reductions.</td>
<td>Respondent has to rank 4 vignettes (3 attributes); n = 140.</td>
</tr>
<tr>
<td></td>
<td>Rae, D.A. (1983).</td>
<td>Valuation of improving the visibility at Mesa Verde (MS) and Great Smoky (GS) National Parks.</td>
<td>Respondent to MS survey has to rank 8 vignettes (13 attributes); n_MS = 205. Respondent to GS survey has to rank 8 vignettes (14 attributes); n_GS = 213.</td>
</tr>
<tr>
<td>Paired comparison: choose the most preferred vignette of several sets of two vignettes</td>
<td>Lockwood, M. (1998).</td>
<td>Value assessment of the preservation of the Mountain Pygmy-possum, and ways to integrate values that are incompatible with economic theory (e.g., noncompensatory benefits derived from the existence of the animal).</td>
<td>Respondent has to choose the best vignette from a set of 168 pairs (2 attributes); n = 95.</td>
</tr>
<tr>
<td></td>
<td>Peterson, G. L. and T.C. Brown (1998).</td>
<td>WTA estimation of 6 public goods, with emphasis on evaluation of the transivity assumption and the reliability of paired comparison responses.</td>
<td>Respondent has to choose the best vignette from a set of 155 pairs consisting of two vignettes describing private goods, public goods and sums of money); n = 130.</td>
</tr>
<tr>
<td>Combination of conjoint rating, contingent ranking and paired comparison</td>
<td>Mackenzie, J. (1993).</td>
<td>Estimation of the WTP for waterfowl hunting trips. A comparison of the informational efficiency of the following conjoint measurement methods: conjoint rating, contingent ranking and two paired comparison methods.</td>
<td>Respondent has to compare and rate four vignettes and compare the six pairs consisting of these 4 vignettes (6 attributes); n = 652.</td>
</tr>
</tbody>
</table>
Paired comparison
Using the paired comparison method, the set of vignettes is presented to the respondent in pairs, and he or she has to choose the most preferred vignette in each pair. In most cases, only two of the attributes of the paired vignettes differ. This is very similar to the dichotomous question format of the CVM. However, in contrast to the one-step elicitation procedure of the CVM, the paired comparison approach asks each subject to make a series of comparisons between vignettes. Consequently, the informational efficiency of the paired comparison method is much higher than that of the dichotomous CVM variant: the respondents' yes or no answers merely identify the upper and lower bounds on their underlying values.

Combination of conjoint measurement methods
Mackenzie (1993; shown in the last row of table 4.1) compared the informational efficiencies of all these conjoint measurement methods. He concludes (p. 602) that the informational efficiency is higher for both the rating and the ranking methods as opposed to the paired comparison methods. The use of rating or ranking methods leads to gains in estimation efficiency, which yield significantly narrower confidence intervals on derived WTP measures. Furthermore, the informational efficiency of the rating method is even higher than that of the ranking method. This is logical, since rating questions permit the survey respondent to express preference intensities as well as preference order.

4.2.1 Conjoint Measurement as used in this Thesis
CM as used in the research described in this thesis differs from the ones described above and shown in table 4.1 in that it combines ranking and rating tasks but does not include paired comparisons. The above mentioned Mackenzie results in terms of informational efficiency support this choice for rating and ranking methods. Our method is composed of the following three steps: constructing of the vignettes, administering the survey, and analysing the results (Baarsma, 1997a).

First of all, the vignettes have to be constructed. Vignettes are cards that describe fictitious but not unreal situations. Constructing the vignettes entails the determination of the different attributes of the good under valuation. These attributes are supposed to be variables in the utility function of the respondents. To facilitate monetary valuation, one of the attributes has to be a sum of money, or a price tag. Subsequently, the possible values of the attributes have to be filled in. The choice of the attributes and their accompanying values are based on literature, meetings with specialists on the subject, and pre-test sessions with focus groups. The choice of the values of the attributes is
also determined by the fact that the set of vignettes has to be approximately orthogonal. Orthogonality, as defined in this context, means that the explanatory variables are varied to the largest possible extent. For a strong correlation between these variables would imply that they would always explain the same part of the dependent variable, and, consequently, the dependent variable would only be partly explained.

By varying the values of the attributes, a very large number of different vignettes can be put together. However, not all of these vignettes represent plausible situations. By excluding very rare situations and logically or actually inconsistent vignettes, the ultimate choice-set of the vignettes is created. Depending on the subject of the survey and on the population involved, the researcher should be careful not to include too many vignettes (Freeman, 1993; Prince et al., 1991).

An example of a vignette in the case of noise nuisance in the surroundings of an airport is given in figure 4.1. If the aim of the study is to estimate the WTP for a reduction in noise nuisance, the monetary attribute could indicate the insulation costs for people living in the neighbourhood of the airport. On the other hand, if one wants to estimate the willingness to accept the current level of noise nuisance, the monetary attribute in the vignette could be a one-time or periodical premium, paid by the airport authorities and/or airline companies. Figure 4.1 represents this last case.

Figure 4.1: Example of a vignette

<table>
<thead>
<tr>
<th>Vignette j</th>
<th>Level of noise nuisance</th>
<th>How often do you experience noise nuisance</th>
<th>Where do you experience noise nuisance</th>
<th>When do you experience noise nuisance</th>
<th>What activities are disturbed chiefly</th>
<th>Degree of noise insulation in the house</th>
<th>Compensating payment per year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 units*</td>
<td>6 times per 24hrs</td>
<td>indoors</td>
<td>from 23.00 till 6.00</td>
<td>sleeping/resting</td>
<td>very good</td>
<td>£1,000</td>
</tr>
</tbody>
</table>

* These units differ internationally. The Dutch unit is termed KU (cf. chapter 6).

After the vignettes have been constructed, the survey can be held. In the survey as proposed in this thesis, the conjoint measurement tasks are three-fold. First, the respondent has to rank a set of, for instance, 4 vignettes. Secondly, he or she has to give each of the 4 vignettes a report mark, i.e., from 1 to 10. Finally, the respondent is asked to state whether he or she thinks that the vignettes are acceptable, i.e., would they really pay for the situation portrayed in the vignettes 1 to 4. Provided that one of the attributes is a monetary value, it is possible to deduce prices for other
attributes from the rank order and/or acceptability of the vignettes stated by the respondents. Moreover, the relative importance of the attributes can be traced.

The purpose of asking three questions instead of one, is that the respondent is forced to reconsider his or her rank order when asked to mark each vignette, and again to reconsider his or her ranks and marks when asked the acceptability question. Each time the respondent answers a question, the previous answer is viewed in a different context. This increases the reliability of the answers. Moreover, the three answers can be compared and inconsistencies between the answers can be eliminated from the sample. For instance, the ratings have to imply the same rank ordering as the ranking question. Or, to name another example, if the vignette ranked second is deemed unacceptable, it is impossible that the vignette ranked third is deemed acceptable. The acceptability question is included also in order to reduce the hypothetical setting of the survey, or in other words, to make the survey more realistic. People may ask themselves questions like "would I actually pay for this good?" or "is this situation acceptable to me?".

The last step in a CM study is the analysis of the results. Due to the indirect way of questioning, the analysis is statistically more demanding than the rather simple analyses in CVM studies, where prices are elicited directly.

The analysis of the rank orderings is based on the random utility maximization model, which states that the probability of ranking a particular vignette highest is equal to the probability that this vignette yields the highest level of utility to the respondent. In general, researchers use a rank-ordered logit model to specify the indirect utility function $U$ (Beggs et al., 1981; Hausman and Ruud, 1987). This function consists of a deterministic part $V$ and a stochastic part $\varepsilon$. Respondent $i$ ($i = 1,...,N$) has a vector of measured attributes $s$, and he or she faces $J$ vignettes indexed $j = 1,..,J$, which are described by the vector of attributes $x_j$. The utility model is then defined as follows:

$$U_i = V_i + \varepsilon_i = V(s_i, x_i) + \varepsilon_i = Z_i \beta + \varepsilon_i$$

From the survey results, the researcher knows the various rank orderings of the respondents, for instance:

$$U_1 > U_2 > ... > U_J$$
The stochastic terms $\varepsilon_i$ are assumed to be independently and identically Weibull distributed (McFadden, 1974, p. 111). The probability of a particular rank order can then be described by (Begg et al., 1981, p. 6):

$$Pr[U_1 > U_2 > \ldots > U_J] = \prod_{j=1}^{J} \frac{\sum_{k=j}^{J} e^{v_j}}{e^{v_j}}$$

We then substitute $Z_i\beta$ for $V_i$, where $Z_i$ is a combination of $s_i$ and $x_i$. The parameters $\beta$ represent the effects of the attributes of the vignettes and the effects of the personal characteristics on utility (if at least one interaction term of $s_i$ and $x_i$ is included). Subsequently, if the index of the alternative ranked $j^{th}$ is denoted by $r_j$, the probability of observing the rank ordering for a particular individual $i$, $R_i = (r_1, \ldots, r_j, \ldots, r_J)$, is (Begg et al., 1981, p. 7; Hausman and Ruud, 1987, p. 86):

$$Pr[R_i] = Pr[U_{r_1} > U_{r_2} > \ldots > U_{r_J}] = \prod_{j=1}^{J-1} \frac{\sum_{k=j}^{J} e^{Z_{ir_j}\beta}}{e^{Z_{ir_j}\beta}}$$

The parameters $\beta$ are estimated using a maximum likelihood procedure. For a sample of $N$ individuals the log-likelihood function is:

$$L(\beta) = \sum_{i=1}^{N} \log Pr[R_i] = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=j}^{J} Z_{ir_j}\beta - \sum_{i=1}^{N} \sum_{j=1}^{J-1} \log \sum_{k=j}^{J} e^{Z_{ir_k}\beta}$$

To analyse the report marks we apply an ordered logit model (e.g., Maddala, 1992). The answers to the acceptability question are analysed using a logit model, as the variable acceptability can only take two values (0 if the vignette is not acceptable, and 1 if the vignette is acceptable).

The CM method can estimate welfare measures like the compensating or equivalent variation (Desvouges et al., 1983). For instance, the compensating variation corresponds to that change in income that will keep the level of utility constant while changing the environmental quality. Imagine the example presented above in figure 4.1 of the noise nuisance in the surroundings of an airport. Suppose the deterministic part of the utility function is specified as follows, with $Y$ as income and $C$ as monetary
compensation, with an interaction term of a personal characteristic $X$ and noise nuisance $N$, and with $P$ variables $S_p$ describing various other personal characteristics:

$$V = \beta_1 N + \beta_2 C + \beta_3 \left( \frac{C}{Y} \right) + \beta_4 N \cdot X + \sum_{p=1}^{P} \alpha_p S_p$$

Now imagine an increase in the noise nuisance ($\Delta N > 0$). The question is then: how much should someone receive as compensation ($C > 0$) in order to stay at the same level of utility ($\Delta V = 0$)? Departing from the utility function specified and estimated above, it is now possible to state the WTA for an increase in $N$.

$$\Delta C = - \frac{\beta_1 + \frac{\beta_4 X}{\beta_3 Y}}{\beta_3} \Delta N$$

### 4.2.2 Advantages of Conjoint Measurement over the CVM

The indirect way of questioning in the CM approach has many advantages over the CVM approach. These advantages relate to strategic behaviour, protest voters, cognitive stress, embedding effects, information requirements and overestimation.

The potential for strategic behaviour declines by asking respondents to express relative rankings or ratings instead of absolute values, because respondents do not focus their attention solely on the monetary attribute of the valuation question but also concentrate on the trade-off between different attributes of the good under valuation. In short, the monetary valuation component is more veiled in CM surveys.

Moreover, because of the veiled monetary valuation component, the number of protest voters will be lower in a CM survey. Protest voters are typically respondents who are not willing to pay anything since they oppose to the concept of having to pay for an environmental, and thus often public, good. By naming the monetary value as one of many attributes of the good, protest voters are probably less aroused.

Thirdly, by asking indirect questions the cognitive stress will decline, since it is easier (less stressful) to respond in terms of relative values than in terms of absolute values. Ranking methods are based on the conjecture that a ranking procedure is easier for people to handle conceptually when faced with the situation of putting a money value
on a non-market good, relative to the procedure involved in contingent valuation (Freeman, 1993). The reason is that individuals are more likely to be capable of ordering hypothetical combinations of environmental amenities and fees than to be able to directly express their WTP for any specific change in these amenities. An example in support of this conjecture is provided by Green and Srinivasan (1978), who found that ranked data are likely to be more reliable. Furthermore, Lareau and Rae (1989) and Hausman and Ruud (1987) showed that respondents can provide consistent and stable rankings.

A fourth advantage of CM over the CVM is the fact that a CM question format may provoke less embedding effects. Embedding effects occur when people cannot or will not differentiate between different scales of the good under valuation. Consequently, the valuation for good A is the same as that for good B which includes good A. One of the causes for embedding effects is the fact that the sequence in which (close) substitutes are valued matters because the marginal utility of a good diminishes (cf. chapter 3, section 3.4.1). Since respondents to CM surveys are explicitly reminded of alternative situations which could be treated as substitutes, the potential for embedding effects declines compared to the CVM setting, where only one situation is valued.

A fifth advantage relates to the fact that CM relies less than the CVM on the accuracy and completeness of the characteristics and features used to describe the situation (Boxall et al., 1996). In a CVM survey, respondents are required to answer a question which involves paying for the improved good or service. A problem with this CVM approach is its reliance on the accuracy of the information, and that any errors in the information discovered after the fact (e.g., if the alternative turns out to be technically infeasible) cannot be changed. Moreover, it is unlikely that the one alternative selected for the CVM study is the lowest-costs alternative (Mathews et al., 1995). The CM approach, however, relies on the representation of a choice situation (rather than the specific change in the good or service), using an array of attributes. Therefore, rather than being questioned in detail about a single alternative, as in CVM analysis, subjects are questioned about a sample of alternatives drawn from the universe of possible alternatives of that type. For example, in the famous Exxon Valdez survey (cf. chapter 3, section 3.3) the respondents only heard about an escort ship programme that would prevent future similar oil spills. However, this programme may not have been realistic to these respondents. Nevertheless, in the CVM survey the respondents did not get the opportunity to consider any other restoration or prevention options.
Alternatives to Contingent Valuation

A sixth advantage of CM over the CVM is the fact that the indirect questions probably entail less overestimation. Probably, since the evidence on this point is not unanimous. Boxall et al. (1996) find that CM leads to less overestimation, while Magat et al. (1988) find the opposite and Desvouges et al. (1983) find no difference whatsoever between CVM and CM results.

Boxall et al. (1996) compared the WTP results from a CVM and a CM (choice experiments) study. The study focuses on moose-hunting improvements in 15 different areas in the US. They found the CVM estimate to be 20 times higher than the CM estimate, and they also found the CM estimate to be similar to the welfare measures derived from parallel revealed preference analyses of the actual trips the same hunters took in 1992. The 20 times higher estimate of the CVM is rather strange, given the fact that the respondents (hunters with a licence) were all familiar with the good (moose hunting), and that the two methods were administered to the same sample of individuals. The researchers propose the following explanation: the CVM respondents may have ignored substitution possibilities among various hunting sites, since the CVM question only deals with one site. On the other hand, respondents to the CM survey are forced to think in terms of substitute sites, because they are asked to choose between various sites.

Magat et al. (1988) also compare the WTP results from a CVM and a CM (conjoint rating) study. Their research involves the valuation of morbidity risk reduction by making bleach and liquid drain openers safer. Their average CVM valuation is 58% lower than the average CM valuation. According to the researchers, the CM format yields higher estimates because the non-iterative CVM approach may create incentives for respondents to state values below their true valuations. For instance, the CVM respondents did not get an opportunity to revise their bids, while in reality they are more accustomed to bargaining (that is: bidding and revising these bids) to purchase a given good at a lower price. It is interesting to note that both CM and the CVM produce large values of morbidity risk reduction in comparison to estimates derived from revealed preference methods (hedonic wage).

Desvouges et al. (1983) studied the valuations for water quality in the Monongahela river in the US. They compare option prices obtained from a CVM and a CM (contingent ranking) study, and found that the CVM and CM results move in the same direction across individuals, where the CM results do not significantly differ from the CVM results. However, the authors warn that the similarity of the two results may have been caused by the fact that the same survey asked for both CVM and CM estimates.
Concluding remarks
Apart from the advantages of CM listed above, CM also has some disadvantages. However, since CM has not been used very often in environmental valuation studies, the criticism is somewhat weak when compared to the criticism of the CVM.

One disadvantage, which was already mentioned, is the fact that the analysis of indirect CM responses is more laborious than the analysis of direct CVM responses. Moreover, the information obtained from indirect questions concerns ordinal measures of the WTP, unlike the cardinal measures obtained by the CVM. Thus, a CM study requires a larger sample than does a CVM study (Hanley and Spash, 1993). Finally, the construction of the vignettes is more time-consuming and more complicated than is the specification of the rather simple CVM question, with only a monetary value and one other attribute. Also, the fact that a set of vignettes should be approximately orthogonal, complicates the construction of the survey.

Furthermore, CM is expected to suffer from some of the same drawbacks that the CVM suffers from. After all, CM, like the CVM, elicits stated preferences under hypothetical conditions. It does not employ any actual behaviour or market transactions. Therefore, researchers should be aware of all kinds of biases that typically occur in stated preference surveys, like sampling bias, compliance bias, information bias and payment vehicle bias.

Some of the researchers who use CM are very optimistic about the possibilities of CM as an alternative to the CVM (i.e., Adamowicz et al., 1998; Mathews et al, 1995). On the other hand, there are other CM researchers who are sceptical about the possibilities of replacing the CVM by CM (i.e., Roe et al., 1996). This thesis aims at widening the methodological scope, where it is shown that other methods than CVM are interesting alternatives.

4.3 Welfare Evaluation Method

Another stated preference method that can be used to put a monetary value on environmental goods is the welfare evaluation method. This is one of the methods applied in the research that underlies this thesis (viz. the Schiphol study), but in this particular Schiphol study the effects are not plausible and/or not significant. Therefore, they are only limitedly included in chapter 6 (appendix A). However, since the welfare evaluation method does constitute an alternative to the CVM, the method is nevertheless reviewed here.
The welfare evaluation method was introduced by Van Praag (1971) and was further developed by a group of researchers at the University of Leyden in the Netherlands, including Van Praag and Kapteyn (1973). The empirical base of the welfare evaluation method is the income evaluation question, which relates the valuation of income to six verbal labels. From this attitude question, an interpersonally comparable individual welfare or utility concept can be constructed. It is implicitly assumed that individual income can be used as the most important criterion for welfare, and that the valuation of income is an index for utility. The income evaluation question runs as follows:

Which monthly after-tax household income would you, in your circumstances, consider to be very bad? Bad? Insufficient? Sufficient? Good? Very good?

I would consider about \( \$ \ldots \) per month to be very bad
I would consider about \( \$ \ldots \) per month to be bad
I would consider about \( \$ \ldots \) per month to be insufficient
I would consider about \( \$ \ldots \) per month to be sufficient
I would consider about \( \$ \ldots \) per month to be good
I would consider about \( \$ \ldots \) per month to be very good

The six answers of individual \( i (i = 1, \ldots, N) \) to this question are denoted by \( c_k \), where \( k \) stands for the verbal labels "very bad" to "very good" \((k = 1, \ldots, 6)\). A basic assumption of the welfare evaluation method is that respondents associate more or less the same feeling of welfare to such a verbal label. The implication of this assumption is that interpersonal comparisons are possible.

The responses to the income evaluation question vary between individuals. It follows that there is no uniform opinion on what is a "very good" income, a "good" income, etcetera. This does not indicate, however, that the verbal labels represent different things to different people. The log-mean and the variance of these six answers are defined as the welfare parameters \( \mu \) and \( \sigma^2 \) (dropping the individual-specific index for the sake of simplicity).

\[
\mu = \frac{1}{6} \sum_{k=1}^{6} \ln c_k \quad \text{and} \quad \sigma^2 = \frac{1}{2} \sum_{k=1}^{6} \left[ \ln c_k - \mu \right]^2
\]
The number of levels \( k \) is mostly fixed at six since, in practice, having six levels works rather well, in the sense that people are willing and able to answer (Van Praag, 1993). In the first publication in 1971, when investigating the welfare evaluation method on Belgian data, 10 levels were used, and in 1973 when investigating the welfare evaluation method in The Netherlands, 9 levels were used. Recently, five or four levels have been used as well.

The next step is to translate these verbal labels into an interval scale. It is assumed that the numerical distance between two subsequent levels is equal, viz. \( ^{2}/_{12} \). This is called the equal interval assumption (Van Praag, 1971 and 1994). The numerical values for each c-level are presented in Table 4.2.

**Table 4.2: Translating verbal labels in numerical values**

<table>
<thead>
<tr>
<th>Verbal label</th>
<th>Numerical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁ = very bad</td>
<td>( ^{1}/_{12} )</td>
</tr>
<tr>
<td>c₂ = bad</td>
<td>( ^{3}/_{12} )</td>
</tr>
<tr>
<td>c₃ = insufficient</td>
<td>( ^{5}/_{12} )</td>
</tr>
<tr>
<td>c₄ = sufficient</td>
<td>( ^{7}/_{12} )</td>
</tr>
<tr>
<td>c₅ = good</td>
<td>( ^{9}/_{12} )</td>
</tr>
<tr>
<td>c₆ = very good</td>
<td>( ^{11}/_{12} )</td>
</tr>
</tbody>
</table>

In general: numerical value related to \( c_r = (2x-1)/12 \)

The assumptions made here, i.e., the equal interval assumption and the assumption that the verbal labels have the same meaning to all respondents, are not unreasonable. The latter assumption is actually the cornerstone of a language community, that is, language should mean or is assumed to mean the same to each member of that community. The equal interval assumption is also used on a large scale in our community, for instance when translating school results. Moreover, experiments conducted by Van Praag and his colleagues (described in Van Praag, 1994), verify the two assumptions. Besides these facts and experiments, if we want to apply the utility function approach for intrapersonal and interpersonal welfare comparisons, we will have to accept some unproven assumptions as a matter of working convention.

The numerical translation enables us to specify a utility function \( U(c_r) = U(c_r, s) \), where \( s \) stands for an individual parameter vector \( S \), like the current income position of the individual \( (y_c) \) and family size \( (fs) \). An example of a specification that resulted in very
good and stable outcomes for most Western European countries (Van Praag and Plug, 1995), is:

$$\mu = \beta_0 + \beta_1 \ln fs + \beta_2 \ln y_c$$

The parameter $\beta_2$ represents the fact that people adjust their norms about what they consider to be a good or bad income when their own situation changes. More specifically, if the current household income increases, the income level that is considered good, will also rise. This is referred to as preference drift (Van Praag, 1971).

The specification of $U$ proposed by the researchers who worked with the welfare evaluation method, is the lognormal specification. Also, these researchers propose to look at relative income differences rather than absolute differences. Relative differences are studied most easily by looking at the logarithm of the answers. For instance, $\mu$ is the mean of the logarithmic answers $[\ln(c_1), ..., \ln(c_6)]$, $\sigma$ is the standard deviation of the logarithmic answers about their mean $\mu$, and the variables $y_c$ and $fs$ are introduced in the model as $\ln y_c$ and $\ln fs$. Van Herwaarden and Kapteyn (1981) found that the lognormal and the logarithmic specifications fitted best among 12 different possibilities. The resulting welfare or utility level $U$, corresponding to income position $y$, is the following utility function, the so-called individual welfare function of income:

$$U(y, y_c, fs) = N\left(\frac{\ln y - \mu}{\sigma} ; 0, 1 \right) = N\left(\frac{\ln y - \beta_0 - \beta_1 \ln fs - \beta_2 \ln y_c}{\sigma} ; 0, 1 \right)$$

By substituting $y_c$ for $y$, equivalence scales can be constructed. Such an equivalence scale indicates by how much the household income should rise, if the family size increases, keeping utility constant. For instance, if the reference situation is denoted by $fs^1$ and $y^1$ and the after-change situation is denoted by $fs^2$ and $y^2$, the corresponding compensating change in income can then be derived from:

$$U((1 - \beta_2)\ln y_c^1 - \beta_0 - \beta_1 \ln fs^1) = U((1 - \beta_2)\ln y_c^2 - \beta_0 - \beta_1 \ln fs^2)$$

$$\ln y_c^2 - \ln y_c^1 = \frac{\beta_1}{(1 - \beta_2)} (\ln fs^2 - \ln fs^1)$$

Note that $\beta_1(1-\beta_2)$ represents the shadow price elasticity of extra household members.
If a plausible cardinalisation of utility is applied, such that welfare differences between levels are equalised (welfare evaluation of income $y = \%$ below that income in the social reference group), a cardinal welfare measure is found that is useful for normative intra- and interpersonal welfare comparisons (Van Praag, 1993, pp. 383-384).

This methodology can be applied to the valuation of other goods besides income position or family size. By including other variables in $S$, like environmental goods and services, the shadow price of these variables can be found. Unlike conjoint measurement, the welfare evaluation method has not yet been widely used in environmental studies. There are only two studies using welfare evaluation in an environmental good context. Van Praag (1988) and Frijters and Van Praag (1998) estimate equivalence scales for different climatic conditions in Europe and Russia, respectively.

The welfare evaluation method is an indirect method, since it does not directly ask for a WTP. The good that is being valued by the respondent is the income position. Subsequently, the researcher estimates the relation between other variables, like noise nuisance, and the evaluation of the income position. Using the estimated equation, the corresponding income change necessary to compensate for a change in the environmental effect can then be calculated. Compared to the conjoint measurement type of questions, the income evaluation question is even more veiling and, consequently, even more likely to reduce strategic behaviour. In fact, the respondent to the income evaluation question does not even know which good is under consideration, whereas the conjoint measurement respondent does know, as the good and monetary values are shown in the vignettes.

### 4.4 Well-Being Evaluation Method

The well-being evaluation method is a stated preference method, just like contingent valuation, conjoint measurement and welfare evaluation. The well-being evaluation method resembles the welfare evaluation method, except for the fact that it measures a broader concept than welfare, namely well-being or happiness. Welfare is that part of life that has to do with money, it corresponds to the notion of satisfaction with one's financial situation. Welfare is captured by a utility function $U$. Well-being or happiness, on the other hand, refers to satisfaction with life as a whole. Different scientific interpretations of happiness exist. For instance, "happiness is an overall evaluation made by the individual in accounting all his pleasant and unpleasant affective
experiences in the recent past", or, "happiness is a lasting, complete and justified satisfaction with life" (various references in: Hagenaars and Wunderink-van Veen, 1989, pp. 376-377).

One of the possible ways to empirically evaluate well-being, is through the ladder-of-life question devised by Cantril (1965). The question was later further developed into the well-being evaluation method by Plug and Van Praag (1995), Van Praag and Plug (1998), Frijters and Van Praag (1998) and Van Praag and Baarsma (1999). Since the method has only recently been applied for the first time, the debate on the pros and cons of the method has not yet started.

The methodological problem which Cantril faced in his study was essentially "that of devising some means to get an overall picture of the reality worlds in which people lived, a picture expressed in their own terms" (p. 21). Cantril found the solution to this problem and called it the 'Self-Anchoring Striving Scale'. It is called self-anchoring, as the respondent defines the worst (lowest) and the best (highest) step of the ladder himself or herself. For instance, in the first question Cantril asks the respondent to define the highest step: "if you imagine your future in the best possible light, what would your life look like then, if you are to be happy? " (p. 23). In the same way he lets the respondent define the lowest step of the ladder of life. After these introductory questions, the actual evaluation-of-life question is posed. It is worded as follows:

We will only use this particular ladder question in our research, and will refer to it as the Cantril question. However, the complete Cantril question encompasses more questions than the two introductory questions and this ladder question. In addition to these, Cantril
also asks respondents to state where they stood on the ladder five years ago, and where they expect themselves to be five years from now. Moreover, he then repeats all of the above mentioned questions in terms of the respondent’s country (i.e., “Where on the ladder would you put your country?” (p. 24)).

The answers to the Cantril question (that is, to the one question that we apply) are denoted by \( W \). Subsequently, \( W \) is related to various variables \( x \): \( W = W(x) \). It is assumed that the experienced happiness scores are interpersonally comparable.\(^5\) In other words, it is assumed that respondents positioning themselves on the same step of the ladder of life, experience the same level of well-being. This does not mean that a rating of, say, 6 given by one person, indicates the same thing as a 6 given by another person. For instance:

> “an American who gives himself a rating of 6 for the present may be projecting a standard of living for himself that will include enough money to own a boat and send his four children to private preparatory school, whereas the wife of a Havana worker who gives herself the same rating will say that her aspiration is to have enough food and clothes so that they don’t have to beg for things.” (Cantril, 1965, p. 25)

If it is assumed that well-being is interpersonally comparable by means of the Cantril question, then the question arises whether it is a cardinal measure. If it is a cardinal measure, then the jump in happiness from step 2 to 4 has to be equal to the jump from step 4 to 6. Van Praag (1994) shows that this is a reasonable assumption. Since there is no means to test this assumption, it stays as an unproven assumption. However, does it matter in the perspective of an empirical study whether the measure is ordinal or cardinal? In an empirical study, the main interest is in the trade-off between a change in the provision or quality of some good \( z \) and a change in net monthly income \( y \). Now, if we look at another ordinal specification of well-being, say, \( W^* = \delta(W(z, y)) \), where \( \delta \) is a monotonically increasing function, it is obvious that if

\[
W(z + \Delta z, y + \Delta y) = W(z, y)
\]

then

\[
W^*(z + \Delta z, y + \Delta y) = W^*(z, y)
\]

Hence, the conclusion is that it is not necessary to assume cardinality to derive trade-off ratios and we do not have to make a statement on the cardinal nature of the Cantril well-being measure either. It is only interpersonal comparability that matters.
Alternatives to Contingent Valuation

Relating $W$ to various variables

Although well-being may have a vague connotation to it, in the sense that it is not a hard piece of information, it is possible to relate well-being to objective and measurable variables, like income $y_c$, family size $fs$ and other variables included in $s$. Let $W^*$ be a positive monotonic transformation of the Cantril answers $W$.

$$W^* = W^*(y_c, fs, s)$$

Plug and Van Praag (1995) and Plug (1997) applied the well-being evaluation method to estimate family equivalence scales. They found that the $W^*$ equation depends on a much richer set of variables than the one found by the welfare equation on the basis of the income equivalence question discussed in the previous section.$^6$

By including environmental goods and services in variable $s$, a monetary valuation of changes in these goods and services can be estimated. For instance, let us take the case of noise nuisance $N$, where the current level of $N$ and $y_c$ is denoted by the superscript 1 and the after-change level by 2. For the sake of simplicity, it is assumed that there is no interaction between $y_c$, $N$ and other variables. The change in income that keeps the level of well-being constant for a given change in the level of noise nuisance is then defined by:

$$W^*\left(\frac{y_c^1, N^1, s}{N^2, s}\right) = W^*\left(\frac{y_c^2, N^2, s}{N^1, s}\right)$$

$$\ln y_c^2 - \ln y_c^1 = -\frac{y_c^2}{1 - y_c^1} \left(\ln N^2 - \ln N^1\right)$$

A similar line of reasoning is used in chapter 6 of this thesis, when the results are discussed of a study in which the well-being evaluation method is used to estimate the monetary value of aircraft noise nuisance. The only other application I know of the well-being evaluation method to the valuation of environmental goods, is the climate change study of Frijters and Van Praag (1998) with Russian data, as already mentioned in the previous section.

The well-being evaluation method is an indirect valuation method, just like conjoint measurement and welfare evaluation. The good that is being valued by the respondent is the quality of his or her life. This evaluation is then transformed into a well-being function $W$ that is similar to the utility function $U$, but takes into account more aspects of life. Because of the indirect way of questioning, the potential for strategic behaviour is small compared to the direct method of contingent valuation. Furthermore, the
Cantril question is more veiling than are conjoint measurement and welfare evaluation, because the respondent does not have a clue that something is valued in monetary terms, nor does he or she have a clue as to what good is being valued.

4.5 Closure

Here ends the theoretical part of this thesis. Chapter 2 reviewed the neoclassical basis of monetary valuation methods and discussed the problems with the neoclassical assumptions of individual behaviour. Despite the critique, it is very important to notice that, if it is our purpose to establish monetary values or prices on environmental goods, we simply cannot get around certain elements of the neoclassical theory, because monetary valuation methods are based on part of this theory. Chapter 2 also looked at two other approaches to monetary valuation of the environment, viz. institutional and ecological economics, in order to place monetary valuation in a wider perspective.

Chapter 3 was specifically aimed at introducing and debating the pros and cons of the contingent valuation method, which is by far the most well-known monetary valuation method for environmental goods. It became obvious that this method is not trouble-free, to use an understatement. Therefore, in chapter 4, three alternative valuation methods were reviewed that can serve as complements to contingent valuation. However, we do not pretend to be complete with the introduction of these alternative methods. Besides the three alternatives included in this chapter, other appropriate alternative methods exist, like for instance allocation games and the priority evaluator technique. In order to be a little less incomplete, these two methods will be described here very briefly.

Allocation games and the closely related priority evaluator technique ask respondents to allocate a fixed budget (e.g., their own income, or 100 points) among a specified set of budget categories, including the environmental asset of interest, substitutes for it, and market goods. These methods allow little possibility for strategic bias or for embedding effects, because the total budget is fixed and respondents are explicitly required to consider the full range of possible allocations (Baarsma, 1997a). One methodological problem with these methods is that, generally, they do not require respondents to make clear declarations of their willingness to give up a specified amount of money in order to receive the amenity in question. Just because someone allocates a fixed budget among different categories, it does not follow that he or she is willing to actually pay the amount allocated to a particular category. This is the major
reason why allocation games and the priority evaluator technique were not explicitly included in this chapter.

Anyhow, here ends the theoretical part of this thesis. It is now time to empirically test the various valuation methods discussed. In chapter 5, the CVM and the conjoint measurement method are compared in the IJburg study, and in chapter 6 the well-being evaluation method and the conjoint measurement method are used in the Schiphol study.
Endnotes

1 In earlier applications of CM, mainly for marketing purposes to predict customer reaction to new products and services, other analytical models than the choice-probability based models were used (Green and Srinivasan, 1990). For instance, methods which assume that the dependent variable is ordinarily scaled (i.e., paired comparisons or rank order) include MONANOVA, and methods which assume that the dependent variable is measured on a rating scale (assuming approximately interval scale properties) include OLS regression.

2 Before the change in transportation modes, the probabilities were: \( \Pr(\text{auto}) = \frac{2}{3} \) and \( \Pr(\text{bus}) = \frac{3}{4} \). After the change, the odds ratios from the original choice set are unchanged according to IIA. \( \Pr(\text{bus1}) = \frac{1}{4} \), \( \Pr(\text{bus2}) = \frac{1}{4} \) and \( \Pr(\text{auto}) = \frac{1}{2} \). Obviously, the probability of choosing bus is still half the probability of choosing auto: \( \frac{3}{4} / \frac{2}{3} = \frac{1}{2} \) before the change, and \( \frac{1}{2} / \frac{1}{2} = \frac{1}{2} \) after the change.

3 Van Beek (1993), Van Beek et al. (1996) and Van Praag et al. (1996) also used this kind of conjoint measurement, with rank ordering, marking and acceptability questions.

4 Actually, the cardinality assumption is not necessary here, and \( \ln(y - \mu) / \sigma \) may be viewed as the relevant welfare index (Van Praag, 1993).

5 Most economists are sceptical on the measurability and interpersonal comparability of well-being. In our sister disciplines of psychology and sociology, but also in health economics, this scepticism is not shared. However, an analysis in terms of a function of well-being does not lead to anything, if we do not define an empirical analogue. Also see chapter 6 section 6.6.

6 Among other things, family size squared, an interaction term of family size and income, and a dummy for two breadwinners, sex, religion, health, and work are included.