# The use of economic analysis for water quality improvement investments

Rodriguez, Diego Juan

# Chapter 4:
## Valuation Methods

### 4.1. Introduction

Chapter 3 provided a conceptual framework for the cost-benefit analysis (CBA) of water quality improvement investments. CBA of proposed investments in an inherently uncertain enterprise because it involves the future, which we cannot predict. Costs and project performance can be different from our expectations. The economy in which the project/program is embedded may change, and the tastes, incomes and preferences of the population affected by the project may changes as well in ways that are hard to predict. When the proposed project involves environmental public goods, such as improved water quality, another widely recognized source of uncertainty is the behavior of the natural system involved. Completing this familiar list is the fact that many of the benefits of water quality investments cannot be found in markets so techniques in economic valuation have been designed to capture these types of benefits.

The goal of this chapter is to review the existing nonmarket benefit estimation techniques and introduce one in particular, the contingent valuation method (CVM)[44], which would be applied in a case study later in this document. We begin this review by a brief illustration of the principal categories of benefits and those derived from water quality improvement investments. The review on the benefit estimation techniques will focus on the major problems that plague the estimation of benefits from ambient water quality improvement (as distinct from potable water provision) by the traditional approaches to benefit estimation, and on the alternative approaches that might solve, or at least avoid, these problems. Then the traditional, revealed-preference methods are presented, with a focus on their applicability, special problems associated with them, and the extent to which these methodologies have been applied in developing countries. Readers interested in a more detailed analysis of the theory and practice of nonmarket benefit estimation is directed to more exhaustive reviews (Freeman 2003 and 1993; Champ et al. 2003; Haab and McConnell 2002; Herriges and Kling 1999). Following, we turn to the stated-preference techniques, assessing the state of the art, and again focusing on applications in developing countries. The estimation of health benefits is discussed separately in the next section. Two special topics are then explored: the combination of revealed- and stated-preference methodologies, and the pos-

44 Given the fact that this method is the one applied in this case study, we will devote greater time and detail to the discussion of this technique.

sibility of so-called "benefit transfer." These are considered in terms of their general potential as valuation techniques and their specific promise for water quality benefit estimation in developing countries. A brief conclusion suggests the major lesson of the review.

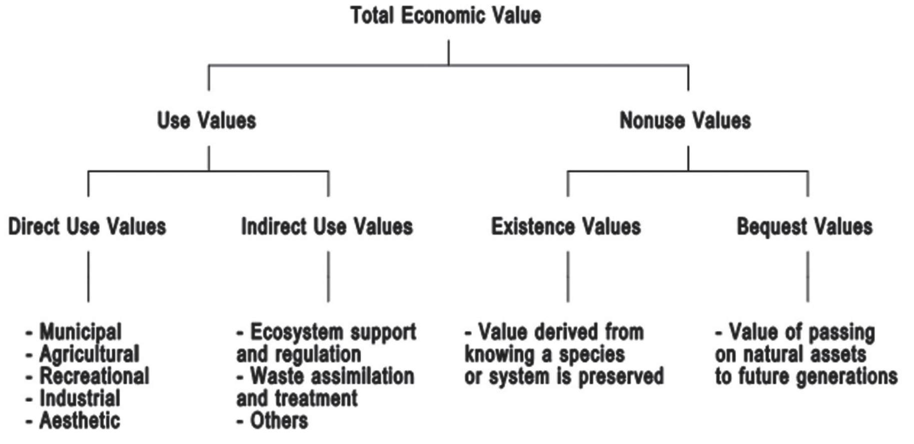
## 4.2. Principal Categories of Benefits

The benefit framework used throughout this chapter consists of the traditional division of values into use and nonuse values. Figure 4.1 illustrates this taxonomy[45].

Use values follow from the use of an environmental good or service in a consumptive or productive process. Direct use values of water can be apportioned into five different categories: municipal, agricultural, recreational, industrial, and aesthetic. They can include both goods (environmental commodities), such as drinking water or shellfish, and services (environmental functions), such as water-based transportation or recreation. The key distinction between direct and indirect uses is that in the case of direct uses, the environmental components or environmental functions themselves enter into the consumption or production activity whereas in indirect uses, the environmental function affects the consumption or production activity only through an intermediate good or service. For example, a coral reef that protects a beach from storm surges is an example of an indirect use because the protection affects the consumption of beach-front recreation only because it helps to preserve the beach. In addition to storm and sea protection, indirect use values include groundwater recharge, flood control, sediment and nutrient retention, and the support of ecosystems.

45 This taxonomy is a general scheme. In applied cases, such as the one to be depicted in this study, the challenges are related to identifying the potential benefits that will be generated by the investments. In particular, the question arises on the objective of the project and its main outcome. In the case of water quality improvement investments, we need to raise the question of the goal that the project is trying to achieve and how this goal will achieve certain objectives and benefits. We can attempt to improve water quality; the question is on what are we trying to achieve with this improvements: is it the option of recreational activities; is it improvement in aesthetic conditions; environmental quality; recovery of certain aquatic species?

**Figure 4.1: Benefit Framework Taxonomy**




Nonuse value is the willingness to pay for the preservation or improvement of natural resources without any intent of in-situ use (Haab and McConnell 2002). These values were first described by Kutrilla (1967) and they show the satisfaction that individuals derive from the existence of environmental assets *per se* (intrinsic value), for the pleasure of others (altruism) or for future generations (bequest value) (Plottu and Plottu 2007). Although a third form of value—option or quasi-option values—has traditionally been included in this taxonomy, these values are omitted from our discussion because they are no longer viewed as a separate category of value. Barbier (1994) states that:

> *A special category of value is option value, which arises because an individual may be uncertain about his or her future demand for a resource and/or its availability as a wetland in the future. There is a general consensus in the economics literature that option values are not a separate form of value but represent a difference between ex ante and ex post valuation. If an individual is uncertain about the future value of a wetland but believes it may be high or that current exploitation and conversion may be irreversible, then there may be quasi-option value derived from delaying the development activities. Quasi-option value is simply the expected value of the information derived from delaying exploitation and conversion of the wetland today. Again, there is consensus that quasi-option value is not a separate component of benefit but involves the analyst properly accounting for the implications of gaining additional information.* (Barbier, 1994, p. 156)

Many authors have employed the term "total economic value" (TEV) to describe the sum of these different values. This description can be misleading because the different values are not necessarily additive. Some uses of a resource may preclude other uses or values, so the values can only be additive to the extent that they are consistent with any one management strategy for the resource. A closely related problem is that there is often an imperfect fit or overlap between the valuation methodologies and the benefit categories, so that even if the values are additive, our estimates of these values often are not. Thus, attempts to generate TEV by summing estimates of individual values obtained using several methodologies may encounter problems with inconsistency and nonadditivity, either because the resource cannot physically generate all of the different values purported to be measured or because of overlaps or inconsistencies in the measurement of these values. Conversely, estimates of TEV are not in general decomposable into underlying categories. A survey that asks for WTP to preserve a resource generates a TEV for that resource, but this TEV cannot be distributed among individual values (Cummings and Harrison, 1995).

## 4.3. Overview of Valuation Techniques

The traditional methodologies are often categorized in a taxonomy similar to the one provided by Freeman (1993), which differentiates the methodologies on the basis of whether they are based on observed behavior or on responses to hypothetical questions, and on the basis of whether monetary values are generated directly or must be inferred, "through some indirect technique based on a model of individual behavior and choice." (Freeman, 1993, p. 23)[46].

The methodologies relying on observations of behavior are often referred to as revealed-preference methodologies, while those relying on answers to hypothetical questions are often referred to as stated-preference methodologies. The interaction of these two characteristics generates four categories in which the different methodologies may be grouped, as shown in Table 4.1.

**Table 4.1. Benefit Estimation Methods**

| | Observed Behavior | Hypothetical |
|---|---|---|
| **Direct Path to Values** | Market (1)<br>Simulated markets | Open-ended contingent valuation (3)<br>Bidding game contingent valuation |
| **Indirect Path to Values** | Hedonic wage analysis (2)<br>Hedonic property analysis<br>Travel cost<br>Averting behavior<br>Voting in actual referenda | Referendum contingent valuation (4)<br>Conjoint analysis<br>Contingent choice (between two vectors)<br>Contingent ranking or rating (of three or more vectors) |

*Source: Adapted from Freeman (1993, p. 24).*

46 This four-way schematic is an elaboration of the older "direct" vs. "indirect" division, where the former involved direct questioning of a sample of people and the latter meant that the method sought indirect evidence of willingness to pay via data from related markets.

The first category, *direct observed*, can only be used to value market benefits because these methods require observations of behavior that can be directly translated into willingness-to-pay measures. Although the direct observed methodologies (particularly the approaches utilizing dose–response functions) are often used to value environmental benefits, they can only provide a partial measure of these benefits because they are incapable of measuring the nonmarket aspects of improvements in environmental quality.

The second category, *indirect observed*, infers values for a nonmarket good from observed behavior in some market by recognizing the link between this behavior and the nonmarket good to be valued. While there are many ways to categorize the methods falling under this label, this study concentrates on three examples that seem most relevant: hedonic analysis, travel cost, and averting behavior.

The third category, *direct hypothetical*, is composed of surveys that elicit maximum WTP by asking direct questions about WTP. This box contains only techniques that allow more or less direct inference of WTP. These techniques may include open-ended questions ("What is the maximum amount that you would be willing to pay for this benefit?") or questions or bidding games composed of a series of questions ("Would you pay this amount for this benefit?"). The methods in this category are referred to as contingent valuation because the value elicited is contingent upon the details of the hypothetical market created to elicit the value, i.e., how the benefit is to be achieved, how the WTP is to be paid, and so on.

The last category, *indirect hypothetical*, consists of methods that infer WTP through a series of exercises in which respondents are asked to choose between, rank, or rate sets of hypothetical options, essentially the various forms of what is often called "conjoint analysis" (CA). Overall, these methodologies apply unevenly to the major routes to environmental benefits (or damage), as shown in Table 4.2.

Table 4.2. Environmental Impacts and Valuation Methods

| Impact | Valuation Methods |
|---|---|
| **Productivity** | Market<br>Averting behavior |
| **Health** | Averting behavior<br>Hedonic wage and property analysis<br>Dose–response coupled with hedonic human<br>Capital or contingent valuation-based values<br>"Pure" contingent valuation<br>Conjoint analysis |
| **Amenity/recreation** | Hedonic market<br>Travel cost<br>Contingent valuation<br>Conjoint analysis |
| **Existence and bequest values** | Contingent valuation<br>Conjoint analysis |

*Source: Adapted from Winpenny (1995 p. 43, Table 3.2).*

It hardly need be said that there is an extensive literature on the theory of the economic valuation of nonmarket goods and services. Freeman (1993) and Haab and McConnell (2002) provide some of the most complete sources.

## 4.4. Indirect Observed or Revealed-Preference Methods

As indicated in Table 4.1, the indirect observed methods for benefit estimation include hedonic analyses, travel costs, and measures of averting behavior.

### *4.4.1. Hedonic Market*

Hedonic market valuation is used to estimate direct and indirect use values, typically the effect of changes in environmental quality on housing values (most hedonic applications are on this) or of risk of morbidity or mortality on wages. The hedonic methodology is based on the premise that a good (or service) can be defined as a bundle of characteristics or attributes that together determine the price of the good via the demand for and supply of the characteristics in the good market. For example, the price of a residence can be viewed as a function of its size, age, physical condition, proximity to schools or places of work, the quality of the environment in which it is located, and so on. Houses are treated simply as a bundle of attributed (Haab and McConnell 2002). Similarly, the wage paid for a job can be expressed as a function of, among other things, the risk of injury associated with the job. These functions are called "hedonic price functions." Once estimated, these functions can be used to generate implicit marginal prices for each of the individual characteristics of the good[47]. The earliest application of hedonics to the housing markets was done by Ridker and Henning (1967) to determine the impact of air pollution in property values by performing a multiple regression of property values on air pollution, housing characteristics, accessibility, and neighborhood attributes using cross-section data for the St. Louis (Missouri) metropolitan area (Harrison and Rubinfeld 1978). More formally, the hedonic price function is described as (Markandya et al. 2002):

$$P_h = f_h(S_{h1},\ldots,S_{hj}; N_{h1}, \ldots N_{hk}; E_{h1},\ldots E_{hm}) \quad (4.1)$$

for all houses h where $P_h$ is the sales price of the house, $f_h$ relates to house price to property, neighborhood and environmental characteristics, $S_h$ are physical characteristics of the house and property, $N_h$ are neighborhood characteristics; and $E_h$ are environmental characteristics.

Equation 4.1 is the estimated using a multivariate regression technique. Three important considerations must be taken into account for the estimation of hedonic analysis: a) functional form, b) model specification, and c) the treatment of spatial dependencies (for detail discussion on these issues see Neupane and Gustavson 2008 and Palmquist 2005).

This method has been widely used in air pollution but much less to determine the effects of water pollution on housing prices. This is primarily due to the difficulties of disentangling the water quality effect from the effects of other variables that are part of the 'location on the water front' effect such as improved view, access to water recreation, and some potential disamenities such as congestion from boat traffic (Haab and McConnell 2002).

Under certain assumptions, the implicit marginal prices can be deemed to represent marginal willingness-to-pay (WTP) for the characteristic. Unfortunately, some of these assumptions are problematic. For one thing, the market is assumed to be in equilibrium. Taking the example of a housing market, this assumption requires full information, no moving or transactions costs, and instantaneous price adjustments to changes in supply and demand. Although no existing housing market can be expected to meet these stringent requirements, a reasonably well-functioning market will not, in general, be unduly burdened by its imperfections[48]. However, where the market is relatively inactive,

47 If the hedonic price function has been estimated for an entire market, then the marginal implicit price for any characteristic is simply the partial derivative of the hedonic price function with respect to that characteristic.

48 It should be noted that although the errors introduced by these imperfections are generally random, non-random errors will be introduced by the continuous movement, or the expectation of movement, of market

controlled, or distorted, the application of hedonic analysis becomes much more problematic. Second, market actors must be aware of and appreciate the effects of the environmental change to be valued. If there is little information on environmental risks or the risks are not fully appreciated by individual actors in the market, hedonic analysis will not generate a valid estimate of the WTP to avoid these risks. A third problematic assumption is that a sufficiently wide variety of the good is required to approximate a continuous spectrum of choices among characteristics, allowing consumers to choose their optimum bundles without being constrained by the unavailability of some bundle. (That is, they are not forced into a corner solution.) Concerns over this assumption have been partially addressed with discrete choice or random utility models[49].

If these assumptions are accepted and if the change in the characteristic of interest is to be measured ex post, where the accumulated data allow calculation of before and after marginal implicit prices, then the difference in aggregate marginal WTP provides a measure of the value of the environmental change. However, the measure derived from the hedonic function will generally overstate the "true" value of the change[50]. On the other hand, if the analysis is to occur ex ante, or does not have access to before and after data, and the change is nonmarginal (as most environmental changes that result from the implementation of a policy or project would be), then a second, more problematic, stage of the analysis is required in order to derive a marginal WTP or inverse demand function. These functions are needed to map nonmarginal changes in the characteristic to changes in total WTP. The problem is identifying (in the technical, econometric sense) these functions from the observations of marginal implicit prices and quantities. Since the hedonic price function is typically nonlinear[51], the marginal implicit price for any characteristic is not parametric to the individual, but is a function of the individual's choice of bundle. If all individuals had identical incomes and utility, then this would not be a problem because all individuals would have the same inverse demand curve, which would simply be the marginal implicit price schedule, i.e., the locus of equilibrium for marginal implicit prices and quantities. However, if individuals have different inverse demand functions,

---

forces in any one direction, resulting in biased marginal implicit prices (see Freeman, 1993, especially p. 383).

49 Random utility models are extensively used in travel cost analysis and are discussed in more detail in the section dealing with that methodology. For a discussion of hedonic random utility models, see Freeman (1993), pp. 412–415, and Palmquist (1991), pp. 116–119.

50 See Vaughan (1988) for a concise explanation of why this is so.

51 A linear hedonic price function implies that consumers can "arbitrage" among goods by varying characteristics. "For example, if individuals are indifferent between owning two two-door cars and one four-door car, other things being equal, they can create equivalents of four-door cars by repackaging smaller units. If both sizes exist on the market, the larger size must sell at twice the price of the smaller one, and the hedonic price of a car will be a linear function of the number of doors." (Freeman, 1993, pp. 128–129; see also Rosen, 1974, pp. 37–38)

then each marginal implicit price observation represents, in general, a point on only one person's inverse demand function and conveys no information about anyone else's inverse demand function. As a result, distinguishing between the marginal implicit price schedule and the consumers' marginal bid functions requires additional data from other markets or the imposition of structure on the consumers' utility functions.

Although researchers are still trying to come to grips with exactly what is needed to validly perform this second stage of the analysis, the most reliable approach seems to be to use additional data from what is essentially a segmented or entirely separate market[52]. However, the practical problems and complications encountered in attempts to apply hedonic analysis in developing country contexts render even first-stage estimation tenuous at best. The satisfaction of additional data requirements for a second stage is generally well beyond possibility[53]. For instance, application to developing countries will only make the three "problematic" assumptions described earlier even less likely to be valid. Developing country markets are less likely to be free of substantial imperfections or rapid change; individuals are less likely to be fully informed about environmental conditions and their effects; and housing or job choices are less likely to be sufficiently broad to allow consumers to make choices unconstrained by the lack of alternatives. Most important, even where markets are well functioning and individuals well informed, data collection may be much more difficult in developing countries. A major difficulty in applying this technique to developing countries is the relative scarcity of monetary land and property transactions in many cultures. To the extent that they occur, they may not be recorded in a fashion required for economic analysis (Bojö et al., 1992).

Although these are certainly not all (or even most) of the practical problems associated with hedonic analysis[54], they are the ones most relevant to developing country applications and should adequately convey the difficulties of applying the methodology in that context. Last but by no means least, hedonic models can provide only a partial measure of a change in environmental quality because they capture only one aspect of the effect of a change in environmental quality; i.e., hedonic property models only measure WTP for the attributes of quality that are "captured" by choice of residence, while ignoring "willingness to pay for improvements in

52 See Palmquist (1991), pp. 95–102, and Freeman (1993), pp. 387–391 for overviews of the literature addressing this problem.

53 Vaughan (1988) makes a similar assertion.

54 Other problems include determining the appropriate measure of environmental quality (see Freeman, 1993, pp. 377–379, and Palmquist, 1991, pp. 92–93) and choosing a functional form (see Freeman, 1993, pp. 379–381; Palmquist, 1991, pp. 87–89; and Vaughan, 1987 and 1988, pp. 1–4).

environmental amenities at other points in the urban area—for example, in the work place, shopping areas, or parks and recreational areas." (Freeman, 1993, p. 416)

Given these practical problems, it is little wonder that few hedonic studies have actually been carried out in developing countries (see Birol et al. for recent examples 2008). Similarly, there are few examples of hedonic wage studies, for many of the same reasons[55]. Nonetheless, the partial derivative of a first-stage hedonic price function can be used to produce an approximate benefit number for neighborhood or local effects of environmental change, such as those introduced by the construction of a sewer: "[S]imple versions of the technique may be useful (indeed have been) in establishing the effects on property values of improvements in neighborhood amenities such as water supply, rubbish collection, street lighting, etc. providing there is data on property values before and after the changes, and so give rough estimates of benefits." (Pearce et al., 1994, p. 147)

However, as pointed out in Vaughan (1987), a "correct" application of first-stage estimation can generate estimates that are two to three times as large as the "true" benefits. As a result, even the reduced effort needed to generate first-stage estimates seems unlikely to be merited, given the quality of the results[56].

### *4.4.2. Travel Cost Methodology*

A second indirect observed method of measurement of nonmarket benefits is the travel cost methodology (TCM). The TCM has been widely used over the course of more than forty years, to estimate the demand for and value of outdoor recreation (see Ward and Beal 2000 for a comprehensive self-contained treatment of TCM). The development of this model was evidenced in 1947 by Harold Hotelling's letter responding to a request from the National Park Service on how to measure benefits derived. In this letter, Hotelling wrote:

55 Other problems include determining the appropriate measure of environmental quality (see Freeman, 1993, pp. 377–379, and Palmquist, 1991, pp. 92–93) and choosing a functional form (see Freeman, 1993, pp. 379–381; Palmquist, 1991, pp. 87–89; and Vaughan, 1987 and 1988, pp. 1–4).

56 See Vaughan (1988) for a summary of some of the arguments for and against such efforts.

*Let concentric zones be defined around each park so that the cost of travel to the park from all points in one of these zones is approximately constant. The persons entering the park in a year, or a suitable chosen sample of them, are to be listed according to the zone from which they came. The fact that they come means that the service of the park is at least worth the cost, and this cost can probably be estimated with fair accuracy...A comparison of the cost of coming from a zone with the number of people who do come from it, together with a count of the population of the zone, enables us to plot one point for each zone on a demand curve for the service of the park. By a judicious process of fitting, it should be possible to get a good enough approximation to this demand curve to provide, through integration, a measure of consumers' surplus...* (Hotelling 1947)

Although a number of other economists were asked for suggestions on the subject matter, the origin of a major line of research associated with travel cost recreation demand models have been attributed to Hotelling. It is striking to note that initially the National Park Service ignored Hotelling's suggestions. It was not until 1959 that, through Clawson's work, attempts to estimate recreation demand were initiated. Clawson in 1959 and later Clawson and Knetsch in 1966, operationalized the idea of the TCM by estimating demand for a recreation site and measuring the consumer's surplus derived from it (Bockstael, Hanemann, and Strand 1983).

The concept of the TCM rests on the fact that unlike other goods, recreational sites are immobile and users must incur specific costs to access the locality. Therefore, the method proposed distance as an alternative for market price, expecting consumption of the site to decline as distance from the site and travel costs rose (Bockstael, Hanemann, and Strand 1983). The demand for a site should measure the willingness to pay of users for specified volumes of outdoor recreation; hence, consumer surplus from the attainment of recreational benefits could be measured. The model estimates the demand for site resources over a specific period of time, say a season or a year.

Consumers who use a site for outdoor recreation will incur different costs. The demand for a recreation site is not solely based on entrance fees[57] but rather on three main types of costs: non-time travel costs, travel time costs, and on-site time and non-time costs (Amoako-Tuffour and Martinez-Espineira 2008).

57 Usually entrance fees, if charged, are low or nominal in value. In instances where the fees are increased substantially, demand for the recreation site will diminish for the population located near the site.

People will use a recreation site to the extent that the level of utility is greater or equal to the cost involved. If a user receives a level of satisfaction greater than its cost, he/she will increase the number of visits until satisfaction and costs are equated at the margin. There have been extensive arguments on the way to treat and value time and on whether discretionary and on-site spending should be included in the model.

This method has progressed in many different ways over the course of the years. In the early stages, the TCM was often referred to as a zonal model because conceptually, and in order to be able to "value" recreational benefits, concentric zones had to be defined around a specific recreational site. Under this assumption, the demand for a "representative" individual was estimated by regressing trips per capita in each zone against average travel cost per trip (Bockstael, Hanemann, and Strand 1983). The zonal model was very basic in its requirements. The difficulty behind this model was on the definition of the distance zones. One of the salient issues was on the number of zones to define, i.e., the extent of the study zone called the "spatial limits" of the model in the literature. The longer the distance away from the site the less probable that the recreationist will travel for the sole purpose of visiting a specific site. Furthermore, consumer behavior may change the further away from the site.

As data became more available, economists moved to the use of micro data in estimating demand for outdoor recreation. The estimation of demand for a recreation site shifted from the aggregate approach to studies based on individual-specific visit (Bülov and Lundgren 2007; Parsons 2003; Sandstrom 1996). The change was from using zonal averages to using data on individuals' travel behavior.

The most basic version of TCM is a continuous-demand model for a single site, in which individuals maximize utility by choosing the number of visits to the site subject to monetary and time constraints. This maximization generates the individual's demand function for the site, from which a consumer's surplus can be calculated and aggregated across individuals.

Although there have been some applications of TCM in developing countries the focus of the TCM studies that value recreational sites is generally on international visitors and therefore of limited use[58]. Unlike hedonic analysis, the essential problem in applying the methodology in

58 In this instance, a measurement of the producer's surplus generated by international tourism would be more appropriate. Another application of TCM in developing countries is to fuelwood and drinking water supply where consumers of these goods must spend substantial time and effort to collect and transport them.

developing countries is not so much an absence of external sources of data as it is an absence (or at least alteration) of the relationship between environmental quality and the recreational market:

> *Applications are numerous in developed countries where motor cars enable easy access to sites, and where time has significant opportunity costs. This will not often be the case for developing countries. Recreation areas will often be close to urban areas (due to limited transportation and low incomes) and so travel costs will be very small. Valuation of non-work time is also more crucial since there will be more people who are non-producers than in developed countries. Also, visitors will often use a recreation area to seek a break from work. Access to sites may be subject to constraints, and so observed travel costs may not accurately reflect actual willingness to pay.* (Pearce et al., 1994, p. 142)

As a result, TCM may be of limited help in estimating the benefits of water treatment projects as they accrue via water-based recreational sites because so few nationals incur the kinds of travel costs needed for TCM to generate valid WTP estimates, and because these values, if estimated, are likely to be small in relation to other project benefits. In addition, the great majority of travel cost studies value a recreational site as it exists. Valuing the introduction of a new site or proposed changes to an existing site, as would be required for analysis of a water quality improvement project, requires more sophisticated versions of TCM, and these are necessarily more data intensive and assumption sensitive.

All TCM models face a number of common problems requiring simplifying assumptions. Perhaps the most fundamental of these is the assumption that "individuals perceive and respond to changes in the travel-related component of the cost of a visit in the same way they would respond to a change in the admission price." (Freeman, 1993, p. 446) This assumption is particularly important because it requires the accurate monetization of the opportunity costs of the time spent visiting and traveling to and from the site in order to generate an accurate demand schedule for it. However, choosing an appropriate method of valuing this time is by no means a straightforward process and is further complicated by severe restrictions on the types of data that can be gathered. This topic has spawned an extensive literature that includes Amoako-Tuffor and Martinez-Espineira (2008), Douglas and Johnson (2004), Smith (1997), McKean et al. (1995), McConnell (1992), Show (1992), Smith and Kaoru (1990), Bockstael (1995 and 1987), McConnell and Strand (1981), Wilman (1980), Bishop and Heberlein (1979) Because of these difficulties, most TCM studies use some fraction of the average wage rate—a solution that strikes many as somewhat arbitrary, especially given the importance of this value in affecting the outcome of the study and the results of some new research that concluded the most commonly used fractions in the literature would overestimate the opportunity cost (fraction of the hourly wage) cost of time and therefore overestimate the consumer surplus derived by the average visitor from access to the park (Amoako-Tuffour and Martinez-Espineira 2008).

In general, the problems that plague TCM are related to the theoretical and practical difficulties of modeling the recreation decision-making process. Sample selection issues illustrate the practical difficulties engendered by this process. Most travel cost data are collected either on site or through the identification of user groups on the basis of some publicly available criteria, such as hunting or fishing licenses or boat registration. Since not everyone uses a recreation site or sites, the sample is either truncated, if nonparticipants are not included in the sample, or censored, if they are[59]. Samples collected on the basis of criteria that suggest the respondent is a likely user of the resource face an additional problem because the availability of the site being studied most likely influenced the decisions that led to the respondents satisfying the criteria. Although statistical techniques can account for the systematic influences that affect the decision to visit the site or meet the criteria and therefore should be included in the survey, choosing to apply one of these techniques imposes a structure on individual decision making in the absence of theoretical guidance.

Other important problems or complications include incorporating the effects of substitute sites in the model, choosing a functional form, defining the appropriate choice set, allocating travel costs when trips are made for more than one purpose, accounting for differences in the amount of time spent on site, ensuring that choice of residence was not determined by recreational preferences, choosing and measuring the appropriate environmental variable, and dealing with the possibility that there is utility or extra-disutility derived from the time spent traveling to and from the site (in which event the travel time would have to enter into the utility function as well as into the time constraint and travel cost amount).

59 In the context of a recreation visit model, censoring and truncation may be understood as follows: If one samples randomly from a population, the number of visits for each respondent will either be 0 or an integer greater than 0. But the dependent variable of interest is "probability of visiting," and no observations of that, other than 0 or 1, are available, even though the model rests on the assumption that for each person in the sample such a probability exists. That sample is censored. If sampling is done at a site where by definition only visitors are encountered, and interest is in the number of visits per user, the sample will be truncated at one visit, no zero-visit observations being available.

Randall asserts that as a result of these and other persistent problems, the idea that researchers can "define a typical trip and specify its cost is prima facie implausible." (Randall, 1994, p. 91) Granting that it is plausible to assume that travel costs increase with distance and hence are "ordinally measurable," he concludes that the best that TCM can ever hope to achieve are ordinally measurable recreation benefits unique only up to a monotonic transformation.

In addition to these problems, TCM is rather limited in applicability. For instance, it is useful for measuring direct use values only where users expend significant time or money in getting to the resource and can be differentiated on this basis. In addition, measuring a change in some qualitative characteristic of the site (such as air or water quality), as opposed to simply valuing the existence of a site, requires the extension of the basic continuous-demand model, causing additional problems in approximating consumer's surplus[60]. As a result, attempts to value individual site characteristics led to the development of hedonic TCM, which attempts to integrate into TCM the methods of determining implicit prices for attributes of hedonic analysis in order to estimate values for site characteristics. Since its development in Brown and Mendelsohn (1984), hedonic TCM has been heavily criticized (largely as a result of studies that found negative marginal prices for attributes that should have been "goods"). These criticisms focus on the absence of a market mechanism to adjust the price of site characteristics to their equilibrium price (so that marginal implicit prices will necessarily reflect marginal values) and on whether the idea of paying more to acquire a higher level of some qualitative characteristic—which clearly makes sense in a one-time purchase—makes sense where the choice is number of trips to a particular site. Although there have been some attempts to salvage the methodology, it generally receives little attention. Since it is not particularly useful for project analysis anyway because it only generates general marginal implicit prices not specific to any one site, it will receive no further mention.

On the other hand, a third type of TCM, the random utility model (RUM), has become the preferred modeling strategy precisely because it can generate values for the attributes of a particular site or sites and also by treating individual behavior within the TCM framework (Soutukorva 2005). RUMs enjoy an additional advantage over continuous-demand models in that they are able not only to value the losses from eliminating an existing site but also to value the benefits of introducing a new site. The RUM label encompasses a range of econometric models that can be used to analyze preferences for different site characteristics by comparing the characteristics of the site chosen with the characteristics of possible substitutes (accounting for the characteristics of the

60 See Freeman (1993), pp. 459–460 and Bockstael et al. (1991), pp. 249–253 for discussions of these problems.

respondents)[61]. Because RUMs focus on the choice among substitute sites for any given trip as a function of the characteristics of the available sites, they are especially suitable when participants are choosing among a number of sites differentiated by quality, and they are generally employed to value changes in specific site characteristics, such as water quality, congestion, or fish catch rates (Freeman 2003). The distinction between RUM and continuous-demand models arises from the ways in which they regard time. Continuous-demand models examine demand for recreational sites over some period of time, such as a summer or a year, where people choose how many trips they are going to make to the different sites. RUMs examine demand for the sites each time a choice is made, so that people are effectively choosing which site to visit, if any (Freeman, 1993). Another implication of this difference is that continuous-demand models can directly explain the total number of trips an individual takes to a given site in a season, while RUMs cannot. As a result, a continuous-demand model is often appended to a RUM, with the latter explaining the allocation of trips over sites.

Although RUMs may be superior to continuous-demand models in their ability to deal with site characteristics and multiple sites, they are not immune to the difficulties inherent in modeling the complex recreation decision-making process. Three extensions of RUMs illustrate recent attempts to overcome these difficulties. An important limitation is the inability of these models to estimate an individual's number of trips in a particular period of time. To be able to factor this issue, RUMs are complemented by nested or count data models (Soutukorva 2005). The nested model restricts the nature of the relationships among alternative sites to avoid the assumption of the independence of irrelevant alternatives generally implied by RUMs. Nested models assume that individuals select their recreational experience by making a series of choices, such as the length of trip, region for the trip, and finally, a particular site within the region[62]. A second extension of RUM attempts to expand the analysis beyond the single-trip framework through the development of repeated and sequential choice models[63]. Finally, a number of researchers have attempted to address the failure of RUMs to generate quantity data by incorporating seasonality issues into the RUM framework. In any event, RUMs are widely viewed as the state of the art in TCM and they would be the model most applicable to evaluating water treatment projects because they can focus on a change in water quality at the site:

61 See Scarpa and Alberini (2005), Freeman (2003), Haab and McConnell (2002), Herriges and Kling (1999), Parsons and Kealy (1995), McConnell (1995), pp. 10–15; Bockstael et al. (1987), Bockstael et al. (1991), pp. 256–264; and Kolstad and Braden (1991), pp. 32–35, for detailed treatments of the theory and estimation of RUMs.

62 Kling and Thompson (1996) and Kaoru (1995) for examples of nested RUMs.

63 See Kling and Thompson (1996) for an example.

*Despite the fact that estimates of benefits from all versions of the RUM framework appear quite sensitive to strategic modeling decisions, enthusiasm remains high for continuing to base policy analyses on some type of RUM. The framework is attractive to practitioners because it consistently incorporates diverse sources of heterogeneous site characteristics. These characteristics provide the primary means to describe the status of the environmental resources supporting recreation uses. For example, in the case of recreational fishing, historic catch rates, emission loadings, contamination notices, as well as other proxy variables have served as the indicators of the quality of specific lakes, rivers, or areas along the coast that support this recreation. These would be difficult to include within [continuous] demand models.* (Smith, 1997, pp. 15–16)

*4.4.3. Averting Behavior*

The third type of indirect observed benefit estimation technique to be discussed here, averting behavior, infers a value for an improvement in environmental quality from changes in spending on ways to reduce the impact of the lower quality. This method is based on the household production theory of consumer behavior (Birol et al. 2006)[64]. The household produces consumption goods using an array of inputs, some of which are polluted. To avoid the adverse impact of the pollution, the households respond in different ways known as averting or defensive behaviors (Birol et al. 2006). In water, these may include boiling water for cooking or drinking, the installation of home filtration and disinfection systems, or the purchase of bottled water (Whittington 2008; Birol et. al 2006). The averting behavior method has been one of the most popular approaches to evaluating safe drinking water (Mi-Jung et al. 2002). The vast majority of the cases in water are in drinking water (see Pattanayak and Whitehead 2009; Whittington 2008; Yoshida and Kanai 2007; Rosado et al. 2006; Mi-Jung et al. 2002; Hagihara et al. 2004; Adote et al. 2000; McConnell and Rosado 2000; Whitehead et al. 1999; Abdalla et al. 1992; Gilman and Skillicom 1985 for applications) and fewer can be found on water quality (see Pei-Ing and Chu-Li 2001; Adote et al. 2000 for some applications). Averting-behavior techniques provide a relatively straightforward route to estimating indirect use values when the environmental threat or harm is known, and efficacious averting behaviors exist (Dickie 2003). Unfortunately, the results are often imprecise because

64 Adote et al. (2000) estimate an averting behavior model examining the choice between bottled, filtered tap, and unfiltered tap. The conclude that averting costs estimates using bottled water expenditures would lead to an overestimate of avoidance costs by about 12% (over water filters expenditures) and that water filter expenditures may accurately reflect averting costs.

the averting measures generally cannot exactly compensate for the loss. For example, the purchase of a water filter or bottled water is unlikely to provide health benefits so much as improved ambient water quality. In addition, people may employ more than one behavior to address a specific effect, such as boiling water and purchase bottled water to avoid an increased risk of waterborne disease. Generally, this method does not measure all the costs related to pollution affecting household utility hence, they provide a lower or upper bound estimate of the true cost of increased pollution (Birol 2006)[65].

Attempts to value environmental quality changes using averting behavior suffer from at least four other problems or complications. First, if conditions get too bad, some people are apt to simply relocate, thoroughly complicating the sampling and valuation processes. Second, the model depends upon people's subjective perceptions of the environmental deterioration and of the risk of harm to themselves, which do not necessarily directly correspond to changes in environmental quality. Third, the behavior often involves some form of discontinuous choice, such as a capital investment in the purchase of a water filter. Presented with this investment decision, people tend to initially resist the purchase, but when the decision to purchase is made, it is guided by long-term expectations as to future environmental conditions and therefore is difficult to interpret in relation to current conditions. Fourth, market imperfections, such as credit rationing, can constrain behavior. As a result of these problems, while averting-behavior values for changes in environmental quality may be relatively easy to obtain, they may only serve as upper or lower bounds for the true values[66].

Given these problems, it is not surprising that the averting-behavior method has not been used with anywhere near the frequency of the first two methods in either developed or developing countries. However, the one area where the methodology may have the most to contribute is in the analysis of water quality improvement projects. There are a number of reasons for this assertion. First, the perceived risks from poor water quality and the behaviors that can avoid these risks are relatively easily known and understood. Second, these behaviors are often within the means of developing-country residents. Third, the method may offer an inexpensive way to get at indirect use

65 In a recent study, Yoshinda and Kanai (2007) performed a comparison study of willingness to pay for drinking water quality in Japan by using the averting expenditures method and choice experiments. Estimated marginal willingness to pay to reduce chlorine was 14.9 yen for the averting behavior method and 6.8 yen for choice experiment suggesting that averting behavior may not be a lower bound of willingness to pay when compared to a choice experiment.

66 Whether these values serve as an upper or lower bounds depends on whether the behavior fully compensates for the environmental degradation, whether the behavior generates ancillary benefits, and whether it is the only averting behavior employed. See Abdalla et al. (1992) for a brief summary of the theoretical literature establishing averting-behavior values as WTP bounds.

benefits that are difficult to estimate with any other method. In this sense, the common practice of crediting wastewater treatment plants with reducing either the costs of downstream intake treatment of municipal drinking water or the costs associated with treating water for reuse in irrigation are aggregate applications of the averting-behavior method. Fourth, the results may be more easily communicated to, and related to, by decision makers than estimates generated by more esoteric or hypothetical methods (Dickie 2003). Finally, it may often be the cheapest of the available methods.

## 4.5. Hypothetical or Stated-Preference Methods

The hypothetical methods for benefit estimation include contingent valuation and conjoint analysis.

### *4.5.1. Contingent Valuation* [67]

#### *4.5.1.1. A Brief History*

The Contingent Valuation Method (CVM) was first proposed in theory by Ciricay-Wantrup (1947) to elicit market valuation of a non-market good (extra-market in Ciricay-Wantrup's terminology) by specifically mentioning that individuals should be interviewed and "asked how much money they are willing to pay for successive additional quantities of a collected extra-market good." If the individual values are aggregated, "the result corresponds to a market-demand schedule" (p 1189). This is the first indication of a recommendation to elicit the demand for an extra-market good through the application of a survey instrument. In 1963, Robert Davis was the first one to use the method empirically; he used questionnaire/interviews of 121 hunters and recreationists in the Maine area to estimate the benefits of outdoor recreation (Mitchell and Carson 1989). In the late 1960s, Ronald Ridker used the CVM in Philadelphia and Syracuse to estimate air pollution benefits (Mitchell and Carson 1989). The method gained popularity in the 1960s after it was recognized, in the economic literature, that option and existence values were important components of the total economic value in environmental economics and that CVM was the only available method capable of estimating these values (Venkatachalam 2004). By the late 1970s, CVM became a recommended method for determining project benefits for empirical resources such as environmental amenities. In 1979, the Water Resources Council recommended CVM in its "Principles and Standards for Water and Related Land Resources Planning". In early 1980s, the U.S. Army Corps

67 Given that the benefit estimates for the cost-benefit analysis application of the case study in this study are derived using the contingent valuation method, we will devote more attention in this section to the method.

of Engineers began to use CVM to measure project benefits and CVM was also recognized under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (Mitchell and Carson 1989). Contingent valuation is now used around the world by governments and multilateral development banks to assess a variety of investments[68]. There is an extensive literature on the use of the method and its application in different topics, including non-use values and, most recently, real estate (see Carson *forthcoming*; Alberini, Anna and James R. Kahn 2006; Biller et al. 2006; Birol et al. 2006; Zhongmin et al. 2006; Smith 2005; Carson and Hanemann 2005; Simons and Winson-Geideman 2005; Venkatachalam 2004; Haab and McConnell 2002; Whittington 2002; Mclean and Mundy 1998; Carson 1995; Hanemann 1994; Mitchell and Carson 1989 for reviews and applications). The popularity of the method does not necessarily translate into the effective application. Whittington (2002) stresses that

> *"There are three main reasons why so many of the contingent valuation studies conducted in developing countries are so bad. First, the contingent valuation surveys themselves are often poorly administered and executed. Second, contingent valuation scenarios are often very poorly crafted. Third, few CV studies conducted in developing countries are designed to test whether some of the key assumptions that the researcher made were the right ones, and whether the results are robust with respect to simple variations in research design and survey methods." (p. 323)*

#### *4.5.1.2. The Concept of CVM*

CVM is a stated preference technique used to measure willingness to pay for changes in the quantity or quality of goods or services, as well as the effect of covariates on willingness to pay (Haab and McConnell 2002). Since it may include nonuse values, it may be one's only alternative method for estimating use and nonuse values, especially for goods not traded in markets. Furthermore, the CVM is the only technique that measures Hicksian surplus directly, without requiring additional manipulation (Chapter 5 presents the theoretical underpinnings of CVM).

Under CVM a hypothetical scenario is created and individuals are surveyed through phone, mail or personal interview methods. To get a monetary measure of a scenario involving welfare change, individuals are asked the amount they would be willing to pay for the good (the welfare improvement) in question.

68 Biller et al. (2006) analyzed the 5,571 references in Carson's forthcoming publication and identified 250 country-specific CV studies undertaken in 73 different developing countries. Only 16 of those countries contributed to 57 percent of the CV studies. Seventy one percent of the studies addressed environmental issues. About one third of this 71% percent addressed brown issues such as air, water and soil pollution and coastal issues. Twenty-eight studies addressed water and sanitation (other 26 studies addressed a variety of combined issues).

CVM assumes that individuals respond the same way to a hypothetical situation as they do to a real scenario. If so, their willingness to pay will be a realistic money measure of the individuals' worth or utility gained or lost from changes in the availability or quality of non marketed environmental goods. Therefore, the most important task of CV analysis is the design of the questionnaires and survey procedures (Carson and Hanemann 2005; Haab and McConnell 2002). CVM is been widely used in environmental and natural resource economics and it is the most popular method for applied research (Hazilla 1999).

Contingent valuation approaches to project benefit estimation necessarily involve surveying samples of the population of interest. If the sample is representative of the population, the sample mean of willingness to pay per capita (or per household) can simply be attributed to everyone in the beneficiary population of size N, so total project benefits are obtained as the product of N and per capita WTP (Rodriguez 2000).

There are three basic survey designs: (a) open ended surveys where individuals are asked to state a willingness to pay; (b) closed end referendums where individuals are presented with a bid and respond yes/no binary decision (single response referendum); and (c) closed end double referendums where individuals are presented with a sequence of two payments to obtain binary decisions (double response referendum) (Mitchell and Carson 1989).

In the early years of CV, the method of payment elicitation was open ended. People were asked to reveal the specific monetary amount they would be willing to sacrifice for the provision of a non marketed good such as an improvement in ambient environmental quality. Obtaining a measure of central tendency from this kind of data was as simple as calculating the mean or median of the WTP values provided by the survey respondents. The econometric analysis involved was minimal, usually being confined to plausibility checks undertaken by split sample comparisons or by regressing the payment amounts on income and other socioeconomic variables to see if the signs on the parameter estimates in the relationship were consistent with prior expectations (e.g. WTP increasing with income).

All of this changed with the advent of the referendum format, which only asks if the respondent would or would not be willing to pay a specific pre selected amount. Under this format it is not possible to know the true WTP of any individual directly. Because those who answer in the affirmative might actually be willing to pay even more, and those who answer in the negative might be willing to pay something less, econometric techniques have to be brought to bear to somehow interpolate and infer an expected value or other central tendency measure from the dichotomous choice information. Simplicity of data analysis was sacrificed in the referendum method in order to construct what many felt was a more realistic choice game.

In consequence, the notion that contingent valuation experiments of the referendum type can reveal a unique number which accurately and unambiguously represents individual willingness to pay for water quality improvement is unrealistic. Rather, there are several possible numbers, each dependent upon the way the initial survey was designed and administered and the way the resulting raw data was passed through the summarizing econometric sieve and reconstituted in the form of a central tendency measure. In short, such estimates are always uncertain when we acknowledge the existence of many routes that potentially can be taken to get at them and the several decision alternatives present at each step along the way. This is not a counsel of doom, or a suggestion that CB analysis based on referendum CV not be undertaken. But it is a fact that any benefit estimate to a greater or lesser degree is always a product of the analyst's protocol and judgment, something respectable analysts recognize and communicate to the users of their results.

The potential for a negative estimate of the expected value of willingness to pay is only a special example of a more general issue with referendum CV, which is that the willingness to pay value extracted from the data can be heavily influenced by the methodological approach taken.

*4.5.1.3. Considerations in CVM*

There are basically two routes to analyzing referendum data. The one most frequently pursued by project economists involves several steps, beginning with the specification and statistical estimation of one or more probability models of individual choice, employing prior assumptions about the form of the inverse distribution, and the covariates belonging in the distribution which serve to change its location and shape across respondents. This is followed by the evaluation of conditional mean or median formulas derived from the choice model, which depend on its estimated parameters. After calculating individual specific means or medians, averages are taken over the entire sample to produce global central tendency measures. The parametric approaches can be particularly useful with the pattern of responses is well behaved hence; the estimates of willingness to pay are not especially sensitive to the choice of distribution for the unobserved random component of preferences or for the functional form of the preference function (Haab and McConnell 2002).

The parametric route can quickly become quite complex, producing a wide array of central tendency estimates. It is not uncommon to find instances where predicted WTP can vary from low to high by a factor of two, five or ten with the same data (as we will see later in Chapter 6), depending on the analyst's choice of density function, the specification of the functional form of the indirect utility index and its arguments, and whether a mean, a truncated mean, or a median is used. In short, with referendum data there are a host of possible measures of central tendency of willingness to pay. Gauged by their frequency of use by practitioners, all of them might seem equally legitimate, but this is not a useful criterion (Chapter 5 will present the theoretical underpinnings of CVM of parametric and non parametric options).

A less frequently traveled but much easier route ignores covariates and does not specify any particular inverse distribution. Instead it uses all the data in pooled form (i.e. the marginal distribution) to produce nonparametric measures of central tendency. These measures of central tendency and dispersion of willingness to pay rely only on the notion that when an individual responses yes to a contingent valuation question, his/her willingness to pay is not less than the offered bid price (Haab and McConnell 2002).

The following list of requirements may be taken to define the essentials of a CV study. All CV studies (a) define the good or service of interest, (b) inform the respondents of a hypothetical change in the availability or level of the good or service to be valued, (c) describe the institutional structure by which the change is to be accomplished, along with the method of payment of the stated WTP, (d) ask respondents to reveal their WTP for this change, and (e) ask socioeconomic and demographic questions in order to relate WTP to these characteristics as an internal consistency check (Mitchell and Carson 1989).

The first two requirements, defining the good or service and informing respondents about the change that is to be valued, can be easier or harder, depending on such features of the problem as its familiarity to the respondents and its concreteness. Thus, a study of WTP to improve neighborhood conditions by installing sanitary sewers and getting rid of open sewerage ditches and drains could count on high levels of familiarity and concreteness among the affected residents. Asking those same people, as part of a larger area sample, about their WTP for changes in the quality of a large river flowing through a city might run into lack of familiarity. An even more unfamiliar setting would be evoked by WTP questions seeking to value a national policy that required every discharger to meet certain discharge standards for every part of every surface water body in a region or a nation to satisfy multidimensional ambient water quality standards.

One element of the survey problem this poses is how to make standards expressed in terms of technical measurements meaningful for respondents. Lay respondents are unlikely to have any idea of what different levels of dissolved oxygen or phosphorous mean. In terms of these measures, you cannot tell much by looking at a water body. And respondents are unlikely to have information on the quality parameters of bodies of water they have actually used recreationally.

In the 1980s, at Resources for the Future, as part of an EPA-funded study of WTP for the existing national policies aimed at improving water quality, this description and information problem was addressed by overlaying a "map" of scientific parameters of quality onto a set of activities, such as boating, fishing, and swimming, that could be carried on in contact with water of particular quality. The fruit of that effort, called the RFF Water Quality Ladder, is described in Annex 4.A.

The different variants of CV can be distinguished on the basis of exactly how they elicit WTP. Some studies ask an open-ended question about maximum WTP. Some engage respondents in a bidding game to determine this maximum. Some ask respondents to choose the value closest to their maximum from a set of values on a payment card. And some simply ask respondents whether they would be willing to pay a certain amount to support a program or policy change (Mitchell and Carson 1989). This last type of value elicitation method is commonly referred to as a "single-bounded dichotomous choice" if only one valuation question is asked or "double-bounded dichotomous choice" if two such questions are asked. Dichotomous choice, which is sometimes referred to as the "referendum format," has become the preferred CV approach because it is believed to be incentive compatible and immune to many of the biases attributed to the other methods (Carson and Hanemann 2005; Freeman 2003; Haab and McConnell 2002; NOAA 1993). However, the tradeoff for this perceived resistance to bias[69] is that respondents never actually state their maximum WTP. As a result, analysts must engage in a complex, somewhat controversial, exercise in econometrics to estimate some maximum WTP for each individual. It may be that the designation of dichotomous choice as the preferred approach is more of a testament to the depth of economists' general distrust of hypothetical methods and lingering suspicions of strategic responses than it is the result of a careful examination of the advantages and disadvantages of this format[70].

69 It should be noted that it is not altogether clear that dichotomous choice CV is any more immune to bias than the other methodologies. For example, see Boyle et al. (1997) for a study of anchoring bias in dichotomous choice CV.

70 See Randall (1997) and Carson (1997) for reviews of the recommendations made by the NOAA Blue Ribbon Panel.

As a result of this distrust and the use of CV to calculate damage in the Exxon Valdez oil spill, there has been an extensive and at times bitter debate over the validity of CV[71]. This debate has been fueled by a series of highly critical studies asserting that the results of CV methods lack internal consistency or are inconsistent with economic theory (Diamond and Hausman, 1994; Hausman, 1993). The response to this two-pronged attack has been either to challenge the assertions on the basis of their "understanding" of economic theory or CV, or to deflect the assertions by pointing out that although they may be valid criticisms of individual CV studies, they are not valid criticisms of the methodology in general (Carson, 1997; Mitchell and Carson, 1995; Carson and Mitchell, 1995; Hanemann, 1994; Smith, 1993). The results of attempts at "validating" (or "invalidating") CV results generally support this response. The validity "tests" can be roughly divided into three categories (Boyle 2003). Beginning with Bishop and Heberlein (1979), there have been numerous attempts at external validation by comparing the results of hypothetical studies with actual market transactions, the results of stated preference studies, or WTP experiments[72]. The results of these tests are mixed, and perhaps best described as open to interpretation (Smith 1997). A second approach has been to test the reliability or consistency of CV responses (essentially whether results can be replicated from one period to the next). These tests have generally shown CV results to be consistent (Boyle 2003; Diamond 1996; Carson et al. 1995). Finally, internal or content validation tests have been used to evaluate the elements in the design of the contingent valuation survey and data analyses are consistent with economic theory (Boyle 2003). The results of these tests also generally remain open to interpretation (Carson 1997; Smith 1997; Pearce et al. 1994). The consistent exercise to scrutinize the validity with CVM and its results are best described by Boyle (2003):

> *Many of the critics of contingent valuation appear to hold contingent valuation up to a criterion of perfection, but this is not realistic because perfection does not exist even in actual market decisions. The key is to consider where contingent valuation has been shown to work reasonably well and where there may be problems. (p 153)*

71 For an overview of the debate, see Carson and Hanemann (2005) and Portney (1994).

72 See Heinzen and Bridges (2008), Vossler and Kerkvliet (2003), Bennett et al. (2002), Randall (1997), Smith (1997), pp. 31–35, and Hoevenagel (1996) for discussions of some of these attempts.

*4.5.1.4. The NOAA Blue Ribbon Panel Report*

Another result of the debate is the emergence of efforts to describe what constitutes a state-of-the-art CV study, the most important of which are the recommendations made by the NOAA Blue Ribbon Panel (NOAA 1993). These recommendations are divided into general and specific guidelines, goals, and fatal problems.

The six general guidelines are the following: employ a professional sampling statistician to conduct probability sampling, minimize nonresponses, utilize face-to-face interviews instead of telephone and mail surveys, test for and assess the existence of interviewer effects, include a specified list of details of the study in the write-up and make the data publicly available, and carefully pretest the survey instrument and test whether respondents understood and accepted the instrument. This is followed by eleven more specific guidelines: where aspects of survey design and response analysis are ambiguous, opt for the choice that will generate a more conservative WTP measure; use WTP instead of willingness to accept; use the referendum format; provide an accurate description of the program or policy to be valued[73]; pretest the effects of photographs, if any; remind respondents of available substitutes; average values across independently drawn samples taken at different times to reduce time-dependent measurement noise (this is likely to be more of a problem in the case of environmental insults, e.g., an oil spill); provide a "no-answer" option in addition to "yes" and "no" options; ask for explanations as to why the respondent chose the option he or she chose; include questions that help to interpret the responses, such as income and knowledge of problem; and ensure that the instrument is not so complex that completing it is beyond the respondent's ability or interest.

The NOAA Report also presents four relevant goals for value elicitation surveys: remind respondents of their budget constraints; attempt to deflect emotive responses; ensure that respondents distinguish between steady-state and interim losses; and ensure that respondents are sensitive to the timing of the program or policy.

The NOAA Panel proposes five problems that, if present, would render the results of a CV study unreliable. These problems are: a high nonresponse rate to either the survey in general or to the valuation question in particular, inadequate

73 This guideline is easy to state but often it is not at all easy to live up to. One big problem is staying within the attention span and technical understanding of respondents. Thus, a table of water quality characteristics, such as dissolved oxygen, fecal coliforms, and turbidity, under with and without project conditions would probably leave lay respondents anywhere both bored and confused. But straightforward photographs are not in this case worth a thousand words (or numbers) because pollution is for the most part invisible. (Hence the reliance on dead fish to make the point in political and fundraising advertisements dealing with water quality.)

responsiveness to the scope of the benefits proposed, respondents' lack of understanding of the task, respondents' failure to "believe" the restoration scenario, and no explanations for responses that do not appear to reflect economically rational decision making, as when respondents do not refer to the cost or value of the program in explaining why they chose the option they did.

Although the recommendations have been described as a "thoroughly mixed bag[74]," their publication effectively constitutes the promulgation of a standard that makes CV prohibitively expensive for many applications. As a result, the qualified endorsement of CV by the NOAA Panel may turn out to be a rather Pyrrhic victory for CV supporters if it ultimately leads to a decline in the use of CV in developed countries.

#### *4.5.1.5. Advantages, Disadvantages, and Applications of Contingent Valuation*

Table 4.3 summarizes the key advantages and disadvantages of the CV method. The survival and growth of CV in spite of the attacks on its validity probably has more to do with its advantages in being able to address almost any policy question asked and being able to measure total economic value (TEV), than in the effectiveness of the responses to these attacks. In addition to nonuse values, indirect use values can be important and difficult to measure by revealed-preference methods. In any case, trying to arrive at TEV by adding up the results of separate revealed-preference methods applied to different routes to benefits is fraught with difficulty because of the good possibility of under- and overlaps. Finally, there is the problem of the "external data" required by the revealed-preference methodologies[75] and the advantages of conducting surveys in developing countries. On the advantages of CV, Dale Whittington, an experienced practitioner, observes:

> *There are some contingent valuation researchers (I count myself among them) that believe it is easier to administer high quality contingent valuation surveys in some developing countries than it is in industrialized countries. For example response rates are typically very high in developing countries, and respondents are often quite receptive to listening and considering the questions posed. Also interviewers are inexpensive relative to prices in industrialized countries. This allows CV researchers to use larger sample sizes and conduct more elaborate split-sample experiments. (Whittington 1998, p. 8)*

74 "These guidelines are a thoroughly mixed bag: Some are supported by the conventional wisdom distilled from a substantial accumulation of research, while others accept uncritically certain untested conjectures from recent literature." (Randall, 1997 p. 1490)

75 "The [CV], discrete choice and [TCM] techniques all require primary data collection. This perhaps explains their evident success in developing economy applications as, by and large; they avoid the use of secondary data sources." (Georgiou et al., 1997 pp. 114–115)

**Table 4.3. Advantages and Disadvantages of CV Method**

| Advantages | Disadvantages |
|---|---|
| Applicable to more environmental goods | Based on respondents' stated intentions |
| Directly (but not separately) measures nonuse benefits | Places respondents in unfamiliar decision setting |
| Directly estimates correct Hicksian welfare measure if questions correctly phrased | Depends on creation of understandable, plausible scenario |
| Can incorporate reliability and validity checks | Vulnerable to abuse in survey design |

*Source: Hoevenagel (1994b p. 252, Table 1).*

The use of CV in developing countries also presents a number of unique challenges. For one thing, many developing country economies may be only partially monetized, causing difficulties in translating values into monetary terms. This need not be an insurmountable obstacle, as shown by Shyamsundar and Kramer (1993), where WTP is stated in bushels of grain. A second challenge is presented by the translation of the survey instrument and responses into local languages or dialects[76]. Third, considerable attention must be paid to local institutional and cultural issues. The survey designer must be sensitive to the attitudes of local people and their perceptions about local, national, and international institutions. The difficulties of cross-cultural communications add to the complexities of conducting CV surveys in developing countries (Whittington 2004). Focus groups are a useful way to learn which payment vehicle, funding, and service delivery mechanism CV survey respondents are likely to trust. Also, respondents who are reluctant to say "no" to a question because of local mores pose interesting challenges for researchers utilizing hypothetical methods. Finally, since developing country applications rely almost exclusively on personal interviews, asking questions about what may appear to respondents or interviewers to be ridiculously high or low numbers in a personal interview is a concern.

Finally, the application of the methodologies to nonmarket benefit estimation in developing countries may be less a question of whether to use a revealed- or stated-preference methodology than whether the latter methods are worth the effort, given that the revealed-preference methods are not generally available. If so, a general comparison of the methodologies may be of limited use. However, Table 4.4 presents an attempt to provide one in a developing country context.

76 Loomis et al. (2002) tests the similarities of English and Spanish speaking households responses to a survey to estimate the benefits of two forest fire prevention programs. The authors found that while there were similar response rates, there were significant differences in the most frequent reasons provided for refusing to pay. Mean benefits reported by Spanish-speaking households were about one-third lower than English-speaking households.

**Table 4.4. Evaluation of Nonmarket Benefit Estimation Methods**

| Methodology | Validity and Reliability | Comprehensiveness | Completeness | Ease of Implementation |
|---|---|---|---|---|
| Dose–response | (2) | (0) | (3) | (2) |
| Hedonic market | (2) | (2) | (1) | (2) |
| Travel cost | (2) | (2) | (1) | (2) |
| Averting behavior | (2) | (2) | (1) | (2) |
| Contingent valuation | (2) | (4) | (4) | (3) |

*Source: Hoevenagel (1994b p. 263, Table 3).*

*Note: (0) represents a very low score, (1) a low score, (2) a moderate score, (3) a high score, and (4) a very high score.*

### *4.5.2. Conjoint Analysis*

Conjoint methods or conjoint analysis (CA)[77][78] are based on work in the sixties by mathematical psychologists and statisticians Luce and Tukey (1964) (Orme 2005). It was used in the 1960s and 1970s primarily in the field of marketing. In the 1980s, with the development of commercial software, the floodgates were opened (Orme 2005). In the 1990s we saw a decade of strong growth for CA in a variety of areas. By the year 2000 and beyond, the method is being used in many different fields, and we have seen an increase of applications for the valuation of environmental and natural resources (for a detailed historical perspective see Orme 2005; for some recent examples see Birol et al. 2008). CA encompasses a number of indirect hypothetical methodologies that are widely used by market researchers in the evaluation of new products and markets[79]. These multiattribute, preference elicitation techniques are based on the hedonic premise that commodities can be viewed as bundles of various attributes. In CA studies, the respondents are presented with a set of choices among a number of alternatives and they rank or rate a series of these bundles in which some or all of the different attributes are allowed to vary (Young 2005). From these rankings or ratings of the different bundles, marginal rates of substitution between the different attributes can be estimated. By including price as one of the attributes, these marginal rates of substitution can be translated into WTP for changes in attribute levels (Birol et al. 2008).

77 For comprehensive presentation of the theoretical foundations of CA see Gustafsson et al. (2007); Hensher et al. (2005); Train (2003); Louviere (1988); McFadden (1974).

78 This terminology is borrowed from the marketing and transportation literatures, at least partly because no alternative has been proposed. The authors that do not use this label generally confine their analysis (and terminology) to specific methodologies within this category, e.g., contingent choice, ranking, or rating.

79 "Conjoint analysis has become an increasingly popular approach to modeling consumer preferences for multiattribute choices. For example, over a decade ago, Cattin and Wittink (1982) estimated that more than 1,000 [CA] applications had been reported. [CA] has been used extensively in the marketing literature where it has proven especially useful in analysis of new products, market segmentation, or product differentiation." (Gan and Luzar 1993, p. 37)

Four different kinds of multiattribute elicitation formats are used in CA. The first of these is dichotomous or contingent choice (referendum CV), where respondents are simply asked to choose their most preferred alternative from two or more choices with differing levels of attributes. Some contingent choice studies force respondents to choose one of the alternatives and some allow respondents to reject all. Dichotomous choice CV is essentially a special case of dichotomous choice CA, where the study is limited to two alternatives, one of which is the status quo, and only two variables, price and the environmental quality variable, are allowed to vary. Relaxing these restrictions allows CA to emphasize tradeoffs among hypothetical alternatives over the purchase of an environmental amenity[80]. Some have argued that this change in emphasis deflects emotive responses and as a result is less likely to generate protest or symbolic responses (see Mazzotta et al. 1997).

A second form of CA, contingent ranking, asks respondents to rank a set of hypothetical alternatives from most preferred to least preferred. Each alternative is characterized by a number of attributes, offered at different levels across options (Foster and Mourato 2002).

A third form of CA, contingent rating, asks respondents to rate a set of hypothetical alternatives on a numerical scale. This approach does not involve a direct comparison of alternative choices so there is no formal theoretical link between the expressed ratings and economic choices (Hanley et al. 2002). The difference between ranking and rating is that the latter asks respondents to supply information about how much they prefer one bundle to another. Since the responses to a contingent rating survey contain more information than the responses to a contingent ranking survey, some authors have asserted that contingent rating is the superior exercise[81]. However, these assertions assume that the two methods are equivalent in terms of the "accuracy" of the preferences elicited, and this assumption is unproven.

A fourth type of CA is graded pair or pair wise rating. In graded pair surveys, respondents are shown two different alternatives and are asked to indicate their preference for one of the products by choosing a number within a

---

80 However, if the status quo is not one of the posited alternatives in the CA analysis, there may be some theoretical issues: "The contingent pair model differs from the conventional referendum model in that neither card's utility level necessarily matches the status quo utility V[Q, M]. This implies that compensated demands for Q derived from the contingent pairing model may not be well defined since they are not based on any reference utility level (Mitchell and Carson, 1989 pp. 85-86).

81 "The comparisons of models and WTP measures confirm the hypothesized advantages of the contingent rating approach, and demonstrate its appropriateness for valuing heterogeneous nonmarket goods." (Mackenzie 1993, p. 602)

set of numbers, say from 1 to 7, where 1 represents the strongest possible preference for one good and 7 the strongest possible preference for the other good. The exercise is then repeated a number of times with different hypothetical alternatives.

The choice of any of the different types of CA is still an issue of debate and research. Mackenzie (1993) compared empirically contingent rating, ranking and binary responses and argued that ratings provided informational efficiencies in econometric estimation over the other two response modes (Roe et al. 1996). Samir et al. (2005) when comparing the rating and the ranking method conclude that the utility model obtained with the rating method represents the preferences expressed better than the ranking method.

CA is perceived as having a number of potential advantages over CV[82]. For one thing, respondents are not required to explicitly monetize environmental goods or services.

> *Finally, [CA] offers some significant practical advantages over [CV]. Respondents are generally more comfortable providing qualitative rankings or ratings of attribute bundles which include prices, rather than dollar valuations of the same bundles without prices. In de-emphasizing price as simply another attribute, [CA] minimizes many of the biases that can arise in open-ended [CV] studies when respondents are presented with the unfamiliar and often unrealistic task of putting prices on non-market amenities. (Mackenzie 1992 pp. 175–176)*

Stated differently, CA focuses respondents on marginal tradeoffs between attributes as opposed to stating a maximum WTP. A second advantage is that it allows a more detailed evaluation of the alternatives[83]. Like dichotomous choice CV, CA is believed to present a more realistic, familiar setting for respondents[84]

---

82 For additional discussions of advantages and disadvantages of these models see Adamowicz and Deshazo (2006), Orme (2005), Bennett and Blamey (2001).

83 "[CA] has the advantage of allowing for the valuation of both the product or program as a whole and the various attributes of the product or program." (Johnson et al. 1995, p. 2)

84 "Contingent referenda and contingent choices among specified attribute-price combinations both mimic familiar consumer decisions processes. While economists view prices and quantities as mathematically dual (and the conventional [CV] approach makes that duality explicit), consumer perceptions are unencumbered by this theoretical framework. Price is simply another attribute of the good in question. Surveys structured in accordance with such perceptions may avoid many of the protest and strategic biases that afflict [CV] studies." (Mackenzie 1993, p. 596)

"Among alternative [CV] structures, a principal advantage of the paired comparisons approach is that, in many cases, respondents find that choosing among alternative commodities is among the most natural and frequently experienced decision environments, compared to directly evaluating individual characteristics. For example, people have routine exposure to this kind of choice in their purchases of market goods, where they often choose between products that are similar, but not identical." (Opaluch et al, 1993, p. 47) and to be free from strategic bias (Birol et al. 2008)[85]. Finally, if, as some researchers believe (e.g., Johnson et al. 1995), the ability to accurately answer hypothetical questions about unfamiliar goods improves with reflection or examination of preferences, so that responding to a survey can be viewed as a dynamic learning process, then the greater number of elicited responses in CA may allow more room for this dynamic learning process to occur.

However, this last potential advantage is offset by a potential disadvantage—that people will tire of answering questions and as a result the accuracy of their responses will decline as the survey progresses[86]. Another disadvantage, potentially at least, is that CA "necessarily interposes in the environmental valuation process the step of estimating a utility function in circumstances where the functional form and arguments of the function are not clearly known." (Pearman 1994, p. 243) Further, the method requires the assumption that "the question sequence presented to each individual in a conjoint study can be treated as a panel of uncorrelated responses." (Smith 1997, pp. 55–56) There is no rule as to how attributes should be chosen, complicating the choice experiment (Campbell and Brown 2003). Also, CA "only identifies the preferred commodity or the relative preference of several commodities." (Roe et al. 1996, p. 158). Third, anyone attempting to apply or evaluate CA is handicapped by the very paucity of attempts to estimate nonmarket benefits with CA. This disadvantage is somewhat offset by the extensive use of CA in other applications, if for no other reason than that this use has led to the development of commercial software that significantly aids the design, administration, and interpretation of CA surveys[87]. Further, the survey instrument may be more complicated to administer and respondents to understand than other methods (CVM) (Young 2005, Farber and Griner 2000).

Although early evidence on CA is mixed, it seems clear that the indirect hypothetical methodologies will be the focus of much of the future development of nonmarket benefit estimation:

---

85 "There is also some evidence from the transport sector that [CA] diminishes the likelihood of strategic bias in responses, since, when offered trade-offs across a number of attributes, it is less clear how an individual might succeed in influencing policy by contriving answers to achieve certain ends." (Pearman 1994, p. 240)

86 "Because conjoint analysis requires answers to several trade-off questions, respondents may become bored or fatigued when answering questions. In these instances, responses to later questions may be of a lower quality." (Matthews et al. n.d., pp. 13–14)

87 For example, Sawtooth Software is a software development company exclusively focused on CA, SPSS and SAS have modules dedicated to conjoint analysis, Decision Pro has designed a Conjoint Analysis Model.

*In addition to willingness to pay, contingent purchases, and contingent policy referendum methods, we are now seeing an explosion of contingent ranking and contingent choice experiments. We can in the near future expect to see contingent resource compensation experiments in which the relative values of different kinds of resource services are compared directly, rather than mediated by money measures. This proliferation of methods is desirable, is being encouraged by the EPA-NSF research program in valuation, and can confidently be expected to continue. (Randall 1997, p. 1493)*

*Hypothetical market methods appear to have a substantial role to play in securing a fuller incorporation of environmental considerations into public decision making. For many environmental goods, they seem to offer what is effectively the only way forward. Within this context, [CV] and [CA] should be seen as variations on a theme. They share many of the same strengths and weaknesses; overcoming problems with one will often help overcome equivalent problems with the other. (Pearman 1994, p. 244)*

Whether this development will lead to or be accompanied by the application of CA methodologies to project analysis in developing countries is less clear. We are seeing more applications of the method in developing countries (see Othman 2004; Kamuanga et al. 2002) but none (so far) in the context of project analysis or in water quality. CA would seem to enjoy the same advantages as CV does over the revealed-preference (indirect observed) methodologies. In fact, the multiattribute nature of policy analysis may lend itself more readily to CA than CV. For instance, the description in McConnell (1995) of the typical water treatment project proposal as containing a number of alternatives with differing attributes could just as easily be describing a CA study. The challenge in molding such a scenario into a CA survey is to translate the highly technical or abstract attributes that characterize the different alternatives into attributes that are easily understood and meaningful to survey respondents. Where institutional concerns might bias the results of CV, CA might examine the preferences behind this bias by varying the institutional circumstances in different policy alternatives[88].

Although some assertions have been made as to the relative theoretical and empirical advantages of one group of methodologies over the other, this

88 Whittington describes circumstances where it is common for a substantial portion of the proceeds of World Bank loans to be diverted to local politicians, all of which does not go unnoticed by potential survey respondents. In the course of pretesting a survey, these respondents might well "say that they do not want their government to borrow money from the World Bank because they know that much of the funds will not be used for the intended purposes, and in this regard they are in fact correct. If the survey instrument were implemented in this form, the results could substantially underestimate households perceived benefits of the new water and sanitation system." (Whittington 1998, p. 5)

literature appears too embryonic to be relied upon to any great extent. For example, there are assertions that CA has theoretical and empirical advantages because it generates more information, but there is practically no evidence on the quality of the answers to the additional questions required to generate this information. Essentially the decision to employ CA as opposed to CV will likely turn on the perception of three factors: (a) whether respondents have difficulty in monetizing their preferences within the CV framework, (b) whether the additional information that can be gleaned from CA surveys is useful, and (c) whether the quality of responses is adversely affected by the more involved surveys utilized by the CA methodologies.

## 4.6. Measurement of Health Effects and Benefits

Despite their potential importance, few reliable studies have been reported that estimate the economic values for the health effects of poor water quality in developing countries. Most of the data available on benefits of clean water (improved water quality) relate to amenity benefits – recreational use and the like (Markandya 2004). Markandya and Murty (2004) use cost-benefit analysis to calculate costs and benefits of a river cleanup program for the Ganges in India. To estimate the health benefits of improvement in water quality, they apply a health production model. This seems to be the most recent study in a developing country (see Markandya and Murty 2004 for further details). There is one study that uses cost-benefit analysis to estimate the costs and benefits of There is however, a vast existing literature on the economic values of improving (quality) drinking water supply (Markandya 2004). The few studies that do attempt to estimate such values limit their analysis to costs of treating actual cases of sickness and forgone earnings resulting from morbidity or premature mortality rates that are based on dose–response functions (Sadoff 1996; Maimon n.d.; World Bank 1993; Margulis 1992; and Serôa da Motta et al., n.d.). This methodology appears to be the state of the art in developing countries in general.

However, despite their widespread use, for a number of reasons these studies are clearly inadequate as attempts to measure the economic value of improvements in health. First, the benefits considered are not all of the benefits that result from an improvement in health, although, not coincidentally, they are the only benefits that can be valued without resort to the nonmarket benefit estimation methods discussed in this chapter (Birol 2006; Young 2005)[89]. Since these benefits are

89 "Environmental pollution that impairs human health can reduce people's well-being through at least the following five channels: (1) medical expenses associated with treating pollution-induced diseases, including the opportunity cost of time spent in obtaining treatment; (2) lost wages; (3) defensive or averting expenditures

often large enough by themselves to support a project, they are often employed as rough lower bounds on the total health benefits.

Second, the human capital approach to valuing health implicit in this methodology carries with it a great deal of ethical or moral baggage:

> *Some of the implications of the human capital approach are unsettling. Because of discounting and the time lag before children become productive participants in the economy, the human capital approach places a much lower value on saving children's lives than on saving the lives of adults in their peak earnings years. And, because of earning differences by sex and race, the human capital approach places a lower value on saving the lives of women and nonwhites than on saving the lives of adult white males. Furthermore, the human capital approach assigns zero value to persons who are retired, handicapped, or totally disabled. (Cropper and Freeman 1991, p. 172)*

Finally, perhaps the most important problem with this approach for economists is that it defies microeconomic principles by failing to take into account human preferences.

There are alternatives to the human capital approach that explicitly consider preferences and are less susceptible to moral or ethical criticisms. The first alternative is to measure WTP through one of the observed methodologies. The applicable methodologies include averting behavior and hedonic wage and property models. Although there are a number of examples of studies employing these methods to estimate WTP for improvements in health in developed countries, there are few, if any, in developing countries. The difficulty of obtaining the data necessary to undertake these studies is, as already discussed, the most likely explanation for this absence. This constraint is particularly problematic for those revealed-preference methodologies that require an informational "infrastructure," such as a system for recording property transactions or keeping health care records.

Even if data constraints were relaxed, however, other practical problems cast doubt on the presumption that these methods offer an improvement over those based on dose–response functions. For example, averting-behavior studies often

---

associated with attempts to prevent pollution-induced disease; (4) disutility associated with the symptoms and lost opportunities for leisure activities; and (5) changes in life expectancy or risk of premature death. The first three of these effects have readily identifiable monetary counterparts. The latter two may not. Since reducing pollution may be beneficial to individuals because it reduces some or all of these adverse effects, a truly comprehensive measure of benefits should capture all of these effects. Measures based solely on decreases in medical costs or lost wages are inadequate because they omit major categories of beneficial effects." (Cropper and Freeman 1991, p. 166) involve discrete decisions, such as the purchase of bottled water or water filters (Abdalla et al., 1992), so that for the majority of people, the behavior cannot be pursued to the point where costs equal benefits, making it difficult to derive an average WTP (Whittington et al. 2008). Since the averting behavior method assumes that the individual is willing to pay up to the amount of expected damages to avoid them, it is not based on analysis of individual decisions (Young 2005). Joint products remain a problem for averting-behavior applications, for example, when bottled water is purchased both because it offers protection from disease and because it has improved taste. A key for all three methodologies is that individual knowledge and perception of risk (or more accurately, the perceived change in the level of risk on the basis of the averting behavior or choice of job or housing location) is required, as opposed to some more easily measured objective measure of risk[90]. Evidence from developed countries on perceptions of risk compared with objective measures is mixed, generally showing positive correlations, but the perceptions often substantially over- or understate objective measures of risk (Cropper and Freeman, 1991). Finally, it is also difficult to extend indirect examination beyond rather narrow sections of the population, such as a certain class of worker or particular neighborhoods.

On the other hand, the hypothetical methods do not share many of these difficulties and are being used with increasing frequency to measure health benefits in developed and developing countries. (See, for example, for CV: Carson forthcoming; Whittington et al. 2008; Birol et al. 2006; Young 2005; Carson and Hanemann 2005; Haab and McConnell 2002; Alberini et al., 1997; Choe et al., 1996; Gerking et al., 1988; Smith and Desvousges, 1987; Mitchell and Carson, 1986; and Jones-Lee et al., 1985. For CA: Birol 2006; Orme 2005; Louviere 2000; Desvousges et al., 1996; Krupnick and Cropper, 1992; Viscusi et al., 1991) Where the link between environmental quality and health problems is direct (such as between water pollution and waterborne illness) and the health problems are severe enough, dose–response methodology will likely continue to be the simplest approach to generating benefit estimates that are high enough to support project approval and that are relatively immune from questions about their "instrumental validity." However, where more precise numbers are needed, dose–response approaches are clearly inadequate and the observed methodologies can provide little additional comfort. In these cases, hypothetical methods will be the methodology of choice. The challenge here will be in constructing accurate but understandable descriptions of the alternative risk situations to be achieved.

90 "For market prices to convey information about individual preferences for risk reduction, individuals must be informed about the risks being valued. Furthermore, risks, as measured by the researcher, must correspond to individuals' risk perceptions at the time their market decisions were made." (Cropper and Freeman 1991, p. 187)

## 4.7. Special Techniques for Determining Benefits

The following sections discuss two particular techniques for determining benefits: the combination of observed and hypothetical methodologies, and benefit transfer.

### *4.7.1. Combining Observed (Revealed Preferences - RP) and Hypothetical Methodologies (Stated Preferences - SP)*

A number of researchers have shown that it is possible to use revealed preference and stated preference methods as complements (rather than substitutes) to improve upon the performance of observed and hypothetical methodologies[91] (Whitehead 2008; Azevedo et al. 2003; Freeman 2003). This joint estimation seeks to take advantage of the contrasting strengths of the different approaches while minimizing their weaknesses (Whitehead et al. 2008).

The first study published using this approach was Cameron (2002), who combined travel cost data with contingent valuation to estimate jointly both the parameters of the underlying utility function and its corresponding ordinary demand function via an illustrative example using an in-person survey of recreational fishermen from the Mexican border to the Louisiana State line between May and November 1987. Since then, several other studies have been conducted (see Pattanayak and Whitehead 2009; Whitehead et al. 2008; Yoshida and Kanai 2007; Birol et al. 2006; Rosado et al. 2006; Eom and Larson 2006; Azevedo et al. 2003 and 2000; Wu and Huang 2001; Whitehead et al. 2000; McConnell et al. 1999; Kling 1997; Huang et al. 1997; Adamowicz et al. 1994). Three main benefits can be identified when combining and jointly estimating RP and SP data: (a) allows an extension of the behavioral model beyond the limited range of historical experience, (b) breaks the multicollinearity characteristics from RP data (von Haefen and Phaneuf 2008), and (c) may mitigate hypothetical bias by combining hypothetical choices with real choice behavior (Pattanayak and Whitehead 2009; Whitehead et al. 2008; Rosado et al. 2006).

These combinations have occurred in two different ways. First, some researchers have attempted to "calibrate" the results from hypothetical surveys using revealed-preference results[92]. The problem with this approach would seem to be that the practical difficulties of implementing a revealed-preference approach constitute a real constraint that is in no way loosened by these combinations.

91 See Whitehead (2008) for a comprehensive discussion on joint estimation of revealed preference and stated preference data.
92 Smith (1997), pp. 43–53, provides an overview of the literature exploring the possibility of calibration for improving the results of benefit estimation studies.

The second approach to combining the methodologies is termed contingent activity:

> *[I]ndividuals can be asked how they would change the level of some activity in response to a change in an environmental amenity. If the activity can be interpreted in the context of some behavioral model, such as an averting behavior model or a recreation travel cost demand model, the appropriate indirect valuation method can be used to obtain a measure of willingness to pay. (Freeman 1993, p. 166)*

Although this approach overcomes the data difficulties, it does so at the price of incorporating the disadvantages of both types of methodologies with few of the advantages, i.e., it is hypothetical but requires the theoretic "tricks" of the revealed-preference methodologies to get at the values of interest. However, if a survey is conducted in such a way that stated- and revealed-preference methodologies may be employed to evaluate the responses, then these concerns may be somewhat dissipated:

> *Several of the valuation techniques typically use data from a household survey (for example contingent valuation, travel cost and hedonic property pricing methods). When a technique requires that primary data be collected with a household survey, it is often possible to design the survey to obtain the data necessary to undertake more than one valuation method. This approach is particularly useful in developing countries because reliable secondary data are rarely available for carrying out valuation work. (Georgiou et al., 1997, p. 107)*

However, this combination would add expense to survey design, administration, and interpretation (and might also increase the burden imposed on respondents by the survey, i.e., longer, more difficult surveys) without necessarily improving the estimates obtained and as a result may be difficult to justify.

The combination of revealed and stated preferences is relatively new and much research is still needed:

> *...combining data sources holds considerable promise for those who are willing to credit each method with strengths and weaknesses. By designing questionnaires to elicit revealed preferences and stated preferences data to take advantage of their respective strengths, and then combining the data sources in estimation, analysts may be able to provide policy makers with more efficient and accurate estimates of the value of public goods. This research agenda has only begun and will require significant effort on the part of environmental economists to identify the conditions under which improved welfare estimates can be obtained through such methods. (Azevedo et al. 2003, p. 535)*

### *4.7.2. Benefit Transfer*

The financial resources and time available for the careful evaluation of a specific site or policy are often limited (Young 2005). This is particularly obvious in the context of project analysis in multilateral development banks and in developing countries. Analysts have resorted to solve (at least partially) some of these limitations by developing the benefit transfer method, which relies in some way on evidence from previous research. Benefit transfer has been suggested as a way around the considerable time and costs involved in performing a study using one of the traditional techniques. Benefit transfer involves the transfer of monetary values or functions obtained from a valuation study or studies at one or more sites (the study site or sites) to an alternative site (the policy site)[93]. The practice with this method became common in the economic analysis of environmental regulation in the United States in the 1980s even before there was any rigorous testing of the validity of the practice[94] (Freeman 2003). This changed in 1992 with the publication of a special issue of *Water Resources Research* dedicated to benefit transfer[95] (Freeman 2003). There are three basic approaches to benefit transfer.

The first approach applies mean unit values estimated for the study site or sites to the policy site. The form of these unit values varies, depending upon the context. For example, the valuation of recreational sites typically generates a consumer's surplus or average WTP estimate that is expressed as the value of a "person-day" for each recreational activity at the site. This value is then multiplied by the change in such days forecasted for the policy site to arrive at an estimate of the aggregate economic benefits of the change in policy. This approach implicitly assumes that the consumer's surplus or WTP experienced on average by individuals at the study site(s) is equal to that which will be experienced at the policy site.

The second approach does not make this assumption, but instead employs adjusted unit values. Adjusted unit values are simply mean unit values that have been systematically adjusted to account for any biases or differences between the sites that might affect WTP, such as differences in user demographics

93 For detailed reviews and applications of benefit transfer see Rolfe and Bennett (2006), Bergstrom and De Civita (1999), Desvousges et al. (1998), Barton (1999), O'Doherty (1995), Brookshire and Neill (1992), and Smith (1992).

94 Without being so named benefit transfer methods, there is evidence that these methods have been employed as early as the 1970s. Young and Gray (1972) "drew on published empirical estimates of price elasticities of demand for at-site residential water to derive corresponding at-source values for intersectoral water allocation comparisons." (Young 2005, pp. 153).

95 See Brookshire and Neill (2002) for an overview of the papers included in the special issue.

(between the populations making use, or valuing the environmental or resource change), the nature of the policy change, or the availability of substitute sites or services. Differences in demographics can include differences in income, tastes, and preferences (Freeman 2003). The ability to make these adjustments depends on the availability of enough original study information to produce the correction equations.

The third approach is to apply the demand curve estimated for the study site(s) to the policy site or in effect to transfer a benefit function from the study site(s) to the policy site[96]. This is considered the preferred approach because it is believed to better account for differences between the sites[97].

The advantages of benefit transfer obscure the extent of the obstacles to achieving believable estimates, i.e., that it is cheap and quick hardly inspires confidence in the results. First, the transferred benefits are only as good as the original study or studies that generated them (Wilson and Hoehn 2006). Second, applications are limited to the valuation of policies or projects that are closely analogous to existing policies or projects that were evaluated by a transferable study. Third, applications are limited to the valuation of sites that are closely analogous to existing sites that were evaluated by a transferable study. The net effect of these first three problems is that it is generally difficult to find a study that is suitable for transfer to the problem at hand[98]. However, even when a suitable study is found, transferring its results to the policy site is not an easy task because although the policy and study sites may be closely analogous, they will never be identical and the differences are sure to raise concerns about applying results across these differences. In addition, determining the geographical extent of the market at the policy site is difficult in the absence of information about local preferences.

The development of benefit transfer as an estimation methodology has proceeded along three tracks. First, researchers have attempted empirical evaluations of the validity of benefit transfer by taking two or more original studies of different sites, transferring the benefits of each to the other, and then testing to see if the transferred estimates are statistically different from the original ones for each site (e.g., Kristofersson and Navrud. 2005; Groothuis 2005; Smith et al. 2002; Kirchhoff et al. 1997; Downing and Ozuna 1994; Parsons and Kealy 1994;

96 For examples and expositions of benefit functions transfer, see Desvousges et al. (1999), Feather and Hellerstein (1997), Parsons and Kealy (1994), McConnell (1992), and Loomis (1992).

97 For details on how to use benefit transfer method in practice see Rosenberger and Loomis (2003) and Desvousges et al. (1998).

98 For examples of the requirements and difficulty of finding a transferrable study, see Bergstrom and Civita (1999), Desvousges et al. (1999), and Boyle and Bergstrom (1992).

Loomis 1992; Opaluch and Mazzotta 1992). In general, the result is that they are statistically different, and the conclusion is that benefit transfer should not be relied on to produce precise estimates[99]. More optimistically, the specific results of these studies may constitute the first steps in developing a protocol to make benefit transfer work with some reasonable degree of precision.

The second approach to developing benefit transfer involves the use of reviews or meta-analyses of segments of the benefit estimation literature to develop multistudy unit values or demand functions with some insight into the relationships of these functions or values with a wide range of explanatory variables[100]. These efforts have been hampered by the sensitivity of existing studies to the wide range of choices made by researchers as to model specification and econometric issues[101]. The conclusions of these studies have also been somewhat pessimistic, largely owing to the lack of studies generating the kind of data needed for this exercise.

The third way in which the development of benefit transfer has proceeded is in the implementation and reporting of benefit transfer studies. Although a large number of benefit transfer studies are performed, few are reported[102]. One area that is undoubtedly the subject of many of these studies, and in which there have been a couple of reported studies, is the estimation of health benefits[103]. Kask and Shogren (1994) summarize the problems and pitfalls in transferring health benefits:

99 "Our results imply that we ought to be skeptical of many efforts to transfer benefit estimates from one site, resource type, or environmental activity to another. Ultimately, the intended use of the benefit estimate determines whether benefit transfer is appropriate and provides adequate reliability. . . .However, when precision matters in the intended policy application, the appropriateness of benefit transfer is questionable." (Kirchhoff et al. 1997, p. 93)

100 "A potentially useful approach to the benefit transfer would be to pool the data from existing studies and apply multiple regression analysis. If the basic model specification is complete, that is, if it includes the relevant explanatory variables in the correct functional form, then it could explain the variation in benefits embodied in differences among the explanatory variables. The net benefit estimated for a site lacking data would then be predicted by inserting appropriate values of explanatory variables into the model fitted to data from other study sites." (Walsh et al. 1992 p. 707) See also Smith and Osborne (1996), Freeman (1995), and Smith and Kaoru (1990).

101 For illustrations of the importance of these effects, see Smith and Kaoru (1990) and Vaughan and Russell (1982).

102 "Benefits transfer, performed when research funding and time are scarce, is probably the most frequently used method of estimating the benefit of environmental improvement from some proposed policy; it is commonly used for quick analyses by and for U.S. federal and state agencies and on behalf of Indian tribes." (Eisworth and Shaw, 1997, p. 2381) Unit day values for recreation have long been required in many forms of cost-benefit analysis for U.S. water resources projects.

103 "This category [health benefits] is probably the easiest for making credible benefit transfers across locations, given comparable economic circumstances (comparison across affected populations with very different income levels or other socioeconomic circumstance is, in contrast, more difficult). Once atmospheric or other natural processes are taken into account (say, in the estimation of the effect of emission reductions on ambient air quality), one can presume to a first approximation that the health effects and the values that people place on avoiding them are reasonably similar across locations." (Kopp et al. 1997, p. 17)

*But transferring the benefits of reduced health risks presents a significant challenge. The challenge arises from the multiple sources of risk, the mortality and morbidity effects indicated by a variety of symptoms, the number of illness days, the long latency period between cause and effect, and an individual's ability to privately or collectively reduce the probability or severity of the risk. (Kask and Shogren 1994, p. 2813)*

Like much of the other literature on benefit transfer, these studies seem overwhelmed by the obstacles to generating reasonably precise benefit transfers.

As should by now be obvious, the gist of this literature is that, at least at this point, benefit transfer is only suitable for tasks where the need for accuracy is low. However, it can be useful in certain circumstances:

*If resource constraints are realistically considered, there are likely to be many cases in which benefit transfer methods can actually yield better estimates of economic values than can be obtained from studies specifically tailored to estimate the value of proposed changes at the policy site. This situation is most likely to arise when: (1) the study site is very similar to the policy site; (2) the policy change or project at the study site is very similar to that proposed at the policy site; (3) the valuation procedures used at the study site were analytically sound and carefully conducted; (4) time, financial resources, and (or) personnel available for analysis at the policy site are not sufficient to undertake a high-quality study. Also, benefit transfer methods may be particularly useful in policy contexts where rough or crude estimates of economic benefits may be sufficient to make a judgment regarding the advisability of a policy or project. (Pearce et al., 1994 pp. 172–173)*

Including in cost-benefit analysis:

*Transfer studies play two important roles in applied cost-benefit analysis: either as a preliminary scoping effort to help develop an original study that will ultimately be used for the policy analysis, or as the final policy analysis itself. How successful the transfer method is in either of these roles will be a function of the demands of the analysis task, the quality of the information available to meet that task, the level of complexity at which the information is transferred (for example, transferring a function versus transferring a scalar), and the required level of precision. (Desvousges et al. 1999 pp. 210)*

Although benefit transfer may not yet be capable of providing precise estimates, it is not clear that it should be discarded. Areas in which benefit transfer might prove useful include prefeasibility screening, assessment of environmental damage, and project valuation of global multiple works). The first steps to utilizing benefit transfer would be the creation of a database from which benefit estimates could be drawn and the adoption of a policy ensuring that original benefit estimates are performed and reported in such a way as to facilitate the future transfer of these estimates:

> *Research using primary data is the basis of benefit transfer via the provision of data and models needed to extend study site values to one or more policy sites. Thus, original investigations using primary data must not simply focus on the end result of estimating a value for the policy issue at hand. Original analyses using primary data, and reporting of these analyses, must reflect their future use as data for benefit transfer studies. These investigations of statistical relationships will help identify key variables and relationships for determining the suitability of a study site value for transfer, and perhaps, for adjusting study site values at the policy site to reduce potential biases in transfer estimates. (Boyle and Bergstrom 1992, p. 662)*

## 4.8. Concluding Remarks

The difficulties in applying the revealed-preference methodologies are well documented. Attempts to apply these methodologies in developing country contexts exacerbate these difficulties and create additional problems where data are scarce and related markets are, for one reason or another, less useful than in developed countries. The difficulties in applying the hypothetical methods are also well documented. However, these difficulties are to some extent ameliorated in developing countries. Perhaps more important, the hypothetical methods are, potentially at least, much more powerful techniques because they can generate an estimate of WTP for the exact policy or project envisioned. Still, the validity of the hypothetical methods is not accepted by all, and for some, the oft-repeated phrase "if you ask a hypothetical question, you get a hypothetical answer" remains the most apt summary (Azevedo et al. 2000; Bohm 1994).

Even for converts, the hypothetical methods remain a difficult enterprise, requiring significant expertise, particularly in terms of survey design and administration. There is always the temptation to say that while one specific CV method may be best for one instance, other methods may be better for other instances, and to recommend the cultivation of an understanding of the characteristics that determine which methodology works best in which instances. However, the resources required to develop the necessary expertise to design and administer "valid" CV studies may be so great as to exclude the development of expertise in the other methods, rendering this recommendation fine in theory but unrealistic in practice.

Perhaps what is needed is a broader view of the problem. While the goal is often to obtain the best "number" in the cheapest fashion, this framing may be too narrow to properly judge the contributions of the methodologies. In the end, the contributions of the hypothetical methods to an understanding of preferences for project alternatives may prove to be the deciding factor.

> *[C]ontingent valuation studies serve several functions. While the focus in this volume has been on the estimates of values produced, it is worth remembering that contingent valuation also reveals considerable amounts of information about what local people want. Contingent valuation can therefore be used as a vehicle for public participation in decision making. This is a factor in favour of contingent valuation, especially as we are recognising that many investments 'fail' because of a lack of consultation and assessment of local wants and needs. The high cost of a contingent valuation study therefore needs to be compared to the multiple benefits of the surveys carried out. (Georgiou et al. 1997, p. 115)*

In a developing country context we must take into considerations the weaknesses in institutions and in information (data). The evaluation of policy and investment decisions requires reliable data that allows the analysts to perform cost-benefit analysis. In many instances, secondary sources of information are missing, justifying the use of hypothetical methods. In the case of water quality, it is necessary to be able to capture non use values and this can only be captured with CVM and CA.

In the following chapter we will get a closer look at the two main routes for analyzing data from CVM, the method of choice for the case study to be presented in Chapter 6: parametric, in which we employ prior assumptions about the form of the inverse distribution and the covariates belonging in the distribution, and nonparametric, which ignores covariates and does not specify any particular inverse distribution.

## Annex 4.A. The Water Quality Ladder

Water quality can be described either in terms of the uses for which a particular body of water is suitable or in terms of the objective characteristics of the water itself. In turn, objective characteristics traverse a continuum from those that are readily perceptible to those that can only be detected by scientific measurement. In certain dimensions [e.g., visible phenomena such as the extent of algal growth; the clearness of the water; and the existence of suds, foam, or debris (Dornbusch 1971) and smells such as hydrogen sulfide from anaerobic conditions], people at large find it easy to perceive changes in water quality. However, some more important characteristics, such as dissolved oxygen content, cannot be detected by human sight and smell. Thus it is not surprising that people's ratings of water quality levels are likely to exhibit a less-than-perfect degree of association with any one or a combination of the several scientific measures of quality conditions (Binkley and Hanemann 1978). This poses a problem for benefit estimation because the existence of a positive willingness to pay for water quality improvement depends upon the ability of people to perceive water quality changes when such changes do in fact occur.

This problem has led previous investigators either to attempt to engineer the marriage of an objective water quality index (based on some weighted combination of scientific quality parameters) and a subjective index of publicly perceived quality (Bouwes and Schneider 1979), or to link subjective indices of public perception and expert perception (Dornbusch 1975).

In the early 1980s, researchers at Resources for the Future developed a water quality ladder that describes water quality primarily in terms of the uses for which the water is suitable, and secondarily in terms of a few obvious quality conditions (clearness, odor, debris, etc.). The use-based levels were located by indexing a set of five objective scientific water quality parameters, using a variant of the National Sanitation Foundation's Water Quality Index (WQI) (Booth et al. 1976; McClelland 1974) along with informed judgment.

A number of sources were consulted to ascertain the minimally acceptable concentration levels of five measurable quality characteristics associated with five potential uses of natural watercourses. These characteristics were fecal coliforms, dissolved oxygen, maximum 5-day BOD, turbidity, and pH[104]. The five quality measures were the only

104 Sources consulted include Thomann (1971), USGS (1978), Pickle et al. (1973), Davis (1968), Economics Research Associates (1979), Katz (1969), Dorfman et al. (1972), North Carolina Environmental Management Commission (1977), APHA, AWWA, FSIWA (1955), National Technical Advisory Committee (1968), NAS-NAE (1972), EPA (1976), Davidson et al. (1966), National Planning Association (1975).

ones for which numerical values could be obtained across all use classifications, a requirement dictated by the index approach. Particular attention was given to state water quality standards (North Carolina Environmental Management Commission 1977; Dorfman et al. 1972), because they report specific critical water quality parameters associated with a set (usually four or five) of descriptive water quality classifications. The consensus results for each quality level are summarized in Table 4A.1.

**Table 4.1. Consensus Water Quality Characteristics of Five Water Quality Classes**

| | Measurable Water Quality Characteristics | | | | |
|---|---|---|---|---|---|
| Water Quality Classification | Fecal Coliforms (No.100 ml) | Dissolved Oxygen (mg/1)a | 5-day BOD (mg/1) | Turbidity (JTU)b | pH |
| Acceptable for drinking without treatment | 0 | 7.0 (90) | 0 | 5 | 7.25 |
| Acceptable for swimming | 200 | 6.5 (83) | 1.5 | 10 | 7.25 |
| Acceptable for game fishing | 1000 | 5.0 (64) | 3.0 | 50 | 7.25 |
| Acceptable for rough fishing | 1000 | 4.0 (51) | 3.0 | 50 | 7.25 |
| Acceptable for boating | 2000 | 3.5 (45) | 4.0 | 100 | 4.25 |

*a Percent saturation at 85°F in parentheses.*

*b JTU, Jackson turbidity unit.*

In order to associate each of the five possible sets of scientific measures with a single-valued ordinate of the quality ladder, a truncated version of the National Sanitation Foundation Water Quality Index was used:

$$WQI = \prod_{i=1}^{5} q_i^{\hat{w}_i} \; /10$$

where:

$q_i$ = The quality of the ith parameter; a number from 0 to 100 obtained from the transformation functions for water quality measures in McClelland (1974).

$\hat{w}_i$ = The weight assigned to the ith parameter. The original weights (wi) reported in McClelland (1974) cover nine quality measures and

$$\sum_{i=1}^{9} = 1.00$$

The RFF adjusted weights cover a smaller number of measures, but were also required to sum to 1.0 via the transformation

$$\hat{w}_i = w_i \left( \sum_{i=1}^{9} = w_i \Big/ \sum_{i=1}^{5} w_i \right)$$

The resultant ladder appears in Figure 4A.1. It has been used in several CV studies of water quality benefits, including those reported in Mitchell and Carson (1981) and Smith and Desvousges (1986).

For example, the index value for the "acceptable for rough fishing" classification was developed as shown in Table 4A.2. Similar calculations for the remaining four classes yield the water quality ladder shown in Figure 4A.1.

**Figure 4A.1: The Water Quality Ladder**

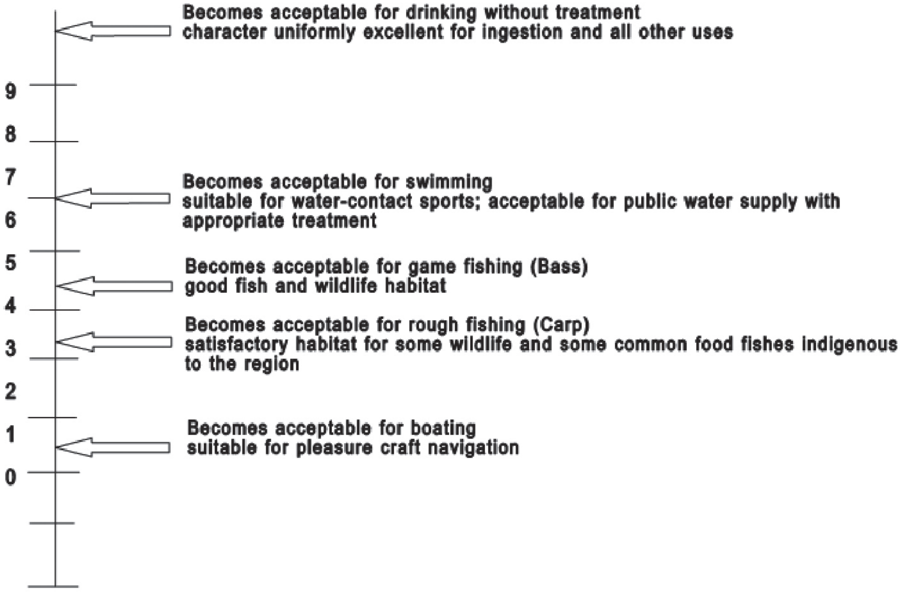



**Table 4A.2. Index Value for Rough Fishing**

| Characteristic | Value | Scaled Value (q1) | Weight ($\hat{w}_i$) | Weighted Scale Value ($q_i^{\hat{w}_i}$) |
|---|---|---|---|---|
| Fecal coliform | 1,000/100 ml | 20 | 0.242 | 1.985 |
| Dissolved oxygen | 51%a | 44 | 0.274 | 2.820 |
| Max. 5-day BOD | 3 mg/1 | 74 | 0.161 | 2.000 |
| Turbidity | 50 JTU | 38 | 0.129 | 1.599 |
| pH | 7.25 | 93 | 0.194 | 2.049 |
| Index $\left( \prod_{i=1}^{5} q_i^{\hat{w}_i} /10 \right)$ | | | | 4.5 |

*a Percent saturation at 85°F.*