NON-MARKET VALUATION USING STATED PREFERENCES: APPLICATIONS IN THE WATER SECTOR

By Paul J. Metcalfe

APRIL 2012

Thesis submitted to the Department of Geography and Environment, the London School of Economics and Political Science, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work, with the following exceptions: chapter 3 is 95% my own work; and chapter 4 is 90% my own work. Chapter 4 is partly based on work conducted prior to the start of the PhD research period.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

The final word count of this thesis is 49,800.

Acknowledgements

A great number of people have helped me during the course of my research in a variety of ways. My particular thanks go to Bill Baker, Ian Bateman, Richard Carson, Kenneth Train and my supervisor Giles Atkinson, for intellectual stimulation, sound advice and critical review of my work. My deepest gratitude however, is reserved for my wife Alice, for her loving support and patience over the course of my studies, and in particular during the last few months when I needed it the most.

Abstract

This thesis is concerned with the application of stated preference methods to non-market valuation problems. It reviews the literature on the state of the art of the method, and applies the techniques to three applications in the water sector.

In the first application, estimates are presented of the value to households in England and Wales of improvements to the quality of water in the natural environment. The need for value estimates arises from the European Community Water Framework Directive, which drives water policy across the European Union. Area based values were generated to maximise the potential for subsequent policy incorporation and value transfer. These were found to vary from £2,263 to £39,168 per km^2 depending on the population density around the location of the improvement, the ecological scope of that improvement, and the value elicitation method employed.

The second application investigates the cost of drought water use restrictions to households and businesses in London. Estimates of willingness to pay for service quality increments often play an important role in the decisions of regulators and regulated companies in industries where consumers have little opportunity to exercise their preferences for service quality. The estimates presented in this chapter are particularly applicable to regulatory appraisals of water company investment expenditure and to planning applications for projects to improve the resilience of urban water supply systems.

The final application examines the reliability of values measured before an economic downturn for application during the downturn via analysis of near identical surveys conducted before, and during, the 2008-2010 economic recession. The main result is that the economic downturn led to lower willingness to pay when elicited via

the payment card contingent valuation method, but had no effect on values elicited via a dichotomous choice (i.e. referendum-type) contingent valuation question. Potential explanations for this finding are explored in light of the literature on closed-ended versus open-ended elicitation method comparisons.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BG	Bidding game
CAPI	Computer aided personal interview
CBA	Cost-benefit analysis
CM	Choice modelling
CS	Compensating surplus
CV	Contingent valuation
DB	Double-bounded
DCE	Discrete choice experiment
EC	European Community
ES	Equivalent surplus
GES	Good ecological status
GIS	Geographic information system
IIA	Independence of Irrelevant Alternatives
MB	Multiple-bounded
OE	Open-ended
OOHB	One-and-one-half-bounded
PC	Payment card
RP	Revealed preference
SASE	Standard annual shortage event
SB	Single-bounded
SP	Stated preference
TB	Triple-bounded
WFD	Water Framework Directive
WTA	Willingness to accept
WTP	Willingness to pay

1 Introduction

1.1 Background to the Study

Decision makers around the world appraise the merits of proposed policies, programmes, projects, regulations (hereafter simply "interventions") using cost-benefit analysis (CBA). In the UK, central government guidance states that "*all new policies, programmes and projects, whether revenue, capital or regulatory, should be subject to comprehensive but proportionate assessment, wherever it is practicable, so as best to promote the public interes*t." [HM Treasury, 2003, p.1] It further goes on to recommend that CBA be used for this assessment, and to encourage that as many of the significant impacts as possible are valued quantitatively [HM Treasury, 2003, p.4]. In the England and Wales water sector, the Water Services Regulation Authority (Ofwat) similarly, but only recently, requested that all investment proposals must be justified using CBA [Ofwat, 2008. p.18].

In many appraisals, including most if not all in the water sector, significant costs or benefits arise from an intervention's impact on non-market goods such as environmental quality, health, safety or the risks of a network service failure. The techniques of non-market valuation are applied in these cases, to provide contributing evidence to a specific CBA appraisal or to provide generic evidence, e.g. on the value of a prevented fatality, that is applicable to a range of appraisal contexts.

One of the principal non-market valuation techniques is the stated preference (SP) method. This method is based on the use of surveys to obtain measures of preferences with which to estimate the welfare effects of non-market impacts. Despite some resistance [e.g. Hausman, 1993; Diamond and Hausman, 1994], the method has developed a growing credibility as a consequence of an immense intellectual effort undertaken by environmental economists over the past four decades or so to appraise and improve the techniques employed. The approach is now commonly used to provide inputs to cost-benefit analyses, and applications of the method are accumulating rapidly [Carson, 2011].

Part of the reason for the technique's popularity is its flexibility - SP surveys are capable of valuing a much broader range of non-market impacts than revealed preference (RP) methods. It is simply not possible to obtain values using RP methods in all cases, even if one would prefer to, because many types of value leave no behavioural trace in any market.

1.2 Research Objectives and Contributions

The aims of this thesis are twofold: to contribute to the growing literature on the stated preference methodology for non-market valuation; and to present policy-relevant empirical valuation models for application in the water sector. The thesis addresses these aims through three core empirical research chapters. We describe the research objectives and contributions of each of these three core studies in more detail below.

1.2.1 Study 1 - An Assessment of the Non-market Benefits of the Water Framework Directive to Households in England and Wales

The first core study addresses both of the core aims of the thesis. It obtains an empirical valuation model directed towards an important policy area – the European Union Water Framework Directive (WFD), and it explores some general issues pertaining to the SP methodology for non-market valuation.

The European Community Water Framework Directive (WFD) [European Parliament, 2000] requires that all natural water bodies should reach the common minimum European standard of "Good Ecological Status" (GES) by 2015, except where to do so would entail disproportionate cost. Benefits estimates are valuable to policy makers in this context to appraise programs of improvements at regional or national levels, and to use in assessments, on cost-benefit grounds, of whether achieving GES will be disproportionately costly for individual water bodies. In such cases, applications for derogations can be made to allow for a longer time to achieve compliance or for a less stringent environmental objective to be adopted. Study 1 was designed to address both purposes simultaneously. In this regard it departs from most previous studies of water quality improvements which have sought to value either a whole program of improvements [Carson and Mitchell, 1993; Brouwer, 2008] or improvements to a localized area [e.g. Alam and Marinova, 2003; Bateman et al., 2011; Hanley, Bell and Alvarez-Farizo, 2003; Hanley, Wright and Alvarez-Farizo, 2006; Kontogianni et al., 2003; Kramer and Eisen-Hecht, 2002; Loomis et al., 2000].

At the core of Study 1 is the development of a model, based on a large-scale nationwide SP survey, for valuing national and regional programs of WFD improvements as a function of key attributes relevant to strategy setting at these levels. These attributes include measures of the geographic scale of the implementation program, the balance between improvements to the worst areas and improvements to raise the number of high quality sites, and the balance between improvements in densely populated areas and improvements in more remote locations.

An additional contribution of Study 1 is in its contribution to the literature on SP methodology. Willingness to pay estimates are known to be sensitive to elicitation methods and question order effects [Venkatachalam, 2004; Welsh and Poe, 1998]. The survey included three types of SP question, and varied the order in which they were asked across the sample, in order to test for these effects.

The contributions of the paper are thus threefold. We obtain a model via a robust large-scale SP survey for valuing national programs of improvements as a function of key attributes relevant to strategy setting at this level. Additionally, we derive a transferable value function that can be used for disproportionate cost assessment at the level of individual sites, and which can validly be summed over sites so as to obtain values for regional programs of water quality improvements. Finally, we explore the sensitivity of our estimates to elicitation treatment effects.

1.2.2 Study 2 – Willingness to Pay to Avoid Drought Water Use Restrictions

The second core study is also a policy-relevant empirical application in the water sector. In this chapter, we investigate the value of avoiding drought water use restrictions in London, UK, by means of an SP survey of households and businesses that sought to measure willingness to pay (WTP) for reductions in the chances, duration and severity of future restrictions. Estimates of the value of avoiding drought water use restrictions are important for appraisals of water utility investments to enhance service reliability, as inputs into regulatory incentive schemes for water utility performance, and in operational decisions during a drought period where there is a need to balance the costs of early less severe restrictions against the value of water saved. Results from the model are applied to a practical context: a planning inquiry concerning a desalination plant in East London. Based in part on the estimates derived here, the plant was approved and built, and began operating in June 2010.

1.2.3 Study 3 – The Sensitivity of Willingness-to-Pay to an Economic Downturn

The third core study of this thesis contributes to the literature on the SP method of non-market valuation by examining the reliability of reliability of valuations obtained before an economic downturn for application during the downturn. SP studies are typically performed at one point in time, with the results then used for decision making several months or even years later, an approach that is only reliable if values are stable over time. This assumption is doubtable given the onset of a recession. The study assesses the reliability of SP valuations via analysis of near-identical surveys conducted before, and during, the 2008-2010 economic recession. Each survey employed two elicitation techniques. The main result is that the economic downturn led to lower WTP when elicited via one method, but had no effect on values elicited by the other. The chapter explores potential explanations for this finding in light of the literature on elicitation method comparisons.

1.3 Outline of the Thesis

The remainder of this thesis is structured as follows. First, chapter 2 contains a review of the literature on non-market valuation using stated preferences, and on applications of SP valuations in the water sector. Next, chapters 3-5 presents, in turn, the three core empirical chapters described above. Chapter 6 then critically discusses the findings from the three core chapters as a whole, in the context of the literature. Finally, chapter 7 draws conclusions on the implications of the results for practitioners and policy makers, and suggests priorities for future research.

2 Literature Review

This chapter establishes the contextual framework within which the analysis in the core chapters is undertaken. It begins with a review of the SP approach to non-market valuation, starting with an overview and brief history of the method, and then proceeding to outline and review the body of theory established by the literature. Following this, the chapter contains a review of the literature on water sector non-market valuation applications relevant to the empirical work that follows. The chapter concludes with a summary of the key strengths and weaknesses of the SP method with particular reference to its application in the water sector, and it motivates the following core empirical chapters in light of the literature reviewed.

2.1 The Stated Preference Approach to Non-market Valuation

2.1.1 Overview

The SP method is based on the use of surveys to obtain data on preferences for valuing the impacts of interventions where, for one reason or another, those impacts are not traded in markets. The two broad families of SP are the contingent valuation (CV), and choice modelling (CM) formats. A survey based around the CV method contains three core parts (Mitchell and Carson, 1989): (i) a detailed description of the policy good being valued and the hypothetical circumstances under which it is made available to the respondent; (ii) questions which elicit the respondents' willingness to pay for the good(s) being valued; and (iii) questions about respondents' use of the good, or related goods, relevant attitudes, and demographic characteristics. The latter information is used in regression equations to check whether valuations vary with respondent characteristics as would be expected, conformance to expectation being a partial assurance of the construct validity of the survey instrument. If the whole study is well designed, the results can be generalized to yield values for the full target population.

In a CM-based survey, rather than describing a single policy good to be valued, a generic format is created to define a policy alternative as a set of attributes, one of which is typically its cost [Hanley, Mourato and Wright, 2001]. The levels of all the attributes are then experimentally varied across alternatives offered in a series of pairwise or multi-way comparisons. In the most prominent CM method, the discrete choice experiment (DCE), respondents are asked to choose which alternative is their most preferred in each choice situation. The DCE format allows the researcher to investigate the trade-offs that people are prepared to make between attributes. If one of the attributes is always set to show the money effect of the change, the willingness to pay for each attribute can be inferred from the trade-offs people make between amounts of each attribute and increments to the cost. In addition, any alternative can be valued relative to a baseline, as in a CV survey, by specifying the levels of the attributes to match the policy scenario in question.

2.1.2 Brief history

The use of surveys to value public goods was first proposed by Bowen (1943), and Ciriacy-Wantrup (1947, 1952). It was not until Davis (1963a, 1963b, 1964), however, that the method was used in a form resembling the current CV method in an empirical application. Many early applications, including Davis (1963a, 1963b, 1964) used CV to value outdoor recreation. Subsequent applications extended the approach far and wide, to goods as diverse as air pollution reduction [Ridker, 1967, Randall, Ives and Eastman, 1974], health and safety [Acton, 1973. Jones-Lee, 1974, 1976], and public provision of grocery price information [Devine and Marion, 1979]. More recently, discrete choice experiments have been added to the toolbox of SP practitioners, either as a replacement for CV methods or in combination. The DCE approach was developed by Louviere and Hensher (1983) and Louviere and Woodworth (1983), and has since spread to an equally diverse range of applications. The collected literature on SP valuation methods and applications has grown exponentially since the 1970s, and there now number over 7,500 related papers and studies from over 130 countries [Carson, 2011].

An established body of theory and recommendations has developed to guide researchers through the field of SP valuation. Prominent manuals, collections and surveys of the field include Freeman (1979, 1993, 2003), Cummings, Brookshire and Schulze (1986), Mitchell and Carson (1989), Braden and Kolstad (1991), the NOAA Guidelines – Arrow et al. (1993), Bateman and Willis (1999), Louviere, Hensher and Swait (2000), Bateman et al. (2002), Haab and McConnell (2002), Hensher, Rose and Green (2005), Carson and Hanemann (2005), Alberini and Kahn (2006), Kanninen (ed) (2007), and Hoyos (2010). Much of the review in the remainder of this section covers material presented in much more detail in these works, and in numerous other surveys. We begin with the conceptual and theoretical foundations underlying the method.

2.1.3 Conceptual and theoretical foundations

2.1.3.1 Measures of value

The value of an intervention is measured in welfare economics by one or more of the four Hicksian consumers' surpluses [Hicks, 1943]. These include compensating variation, equivalent variation, compensating surplus and equivalent surplus measures. Compensating variation is the change in money income needed to accompany the intervention when quantities consumed are free to change in order for utility to be the same in both periods. Compensating surplus is defined exactly the same except that quantities are held fixed. This measure is the most commonly used for a non-market valuation exercise, and is consistent with the notions of both "willingness to pay" (WTP) for a public good, and "willingness to accept" (WTA) for a public bad, e.g. pollution. The remaining two types of value measure are the equivalent variation and equivalent surplus measures. Equivalent variation is the amount the household would be willing to pay, or by which it would need to be compensated, in order to avoid having the intervention take place. Equivalent surplus is defined exactly the same except that, as with compensating surplus, quantities are held fixed.

These concepts can be formalised as follows. Let the utility function of a representative household be given as $U(X, Z)$, where U is the utility of the representative household; X is a composite private good; and Z is a non-market good to be valued. We assume that X and Z are valued positively, hence $\frac{\partial U}{\partial X} > 0; \frac{\partial Z}{\partial X} > 0$.

The dual of the utility function in (1) is the expenditure function $e(p, Z, U)$, where p is the price of the composite good X. The function $e(.)$ is the amount the household would need to spend to achieve utility level U given prices p and the level of the non-market good Z.

The compensating surplus (CS) for a change in Z from Z_0 to Z_1 is given by the difference in the expenditure functions associated with the two levels of Z, holding U constant at $U_0 = U(X_0, Z_0)$, i.e.:

(2.1) $$CS = e(p, Z_0, U_0) - e(p, Z_1, U_0)$$

If $CS>0$, then the change from Z_0 to Z_1 is valued positively, hence less expenditure is required when $Z=Z_1$ to achieve the same level of utility as when $Z=Z_0$. Expression (2.1) is therefore a WTP measure. By contrast, if $CS<0$, the change from Z_0 to Z_1 is valued negatively, hence more expenditure is required when Z=Z_1 to achieve the same level of utility as when Z=Z_0. Expression (2.1) is in this case a WTA measure.

The equivalent surplus measure of value with respect to the same change in Z is similarly given as the difference between two expenditure functions. The difference is that in this case, the utility is held constant at U_1=$U(X_1,Z_1)$:

(2.2) $ES = e(p, Z_0, U_1) - e(p, Z_1, U_1)$

If ES>0, then the change from Z_0 to Z_1 is valued positively, hence less expenditure is required when Z=Z_1 to achieve the same level of utility as when Z=Z_0. Expression (2.2) is therefore a WTP measure. By contrast, if ES<0, the change from Z_0 to Z_1 is valued negatively, hence more expenditure is required when Z=Z_1 to achieve the same level of utility as when Z=Z_0. Expression (2.2) is in this case a WTA measure.

A summary of the four Hicksian surplus measures of value, i.e. where quantity of the public good is held fixed, is given in Table 2.1 below.

Table 2.1: The Four Hicksian Value Measures

Utility baseline	WTP	WTA
Before intervention	CS+	CS-
After intervention	ES+	ES-

The question as to which measure is the appropriate one to use is largely context specific, and should foremost be determined by the existing property rights [Hanemann, 1999]. For example, if people are held to have a right to clean air, then their WTA compensation for the air being polluted is the appropriate measure of the value, or cost, of that pollution. If, on the other hand, an improvement, say to a local park, is being valued, and people are held not to have any property right to the improvement per se, then their WTP for the improvement is the appropriate measure of its value.

Notwithstanding current property rights, however, the influential NOAA guidelines [Arrow et al., 1993], recommended the targeting of WTP when conducting an SP valuation even where a WTA measure would, by reason of property rights, be more appropriate. The principal reason for this is that there is a large volume of empirical evidence to show that WTA estimates are typically higher than WTP estimates, and often much higher [e.g. Knetsch and Sinden, 1984; Coursey, Hovis and Schulze, 1987; Brookshire and Coursey, 1987; Kahneman, Knetsch and Thaler, 1990; Dubourg, Jones-Lee and Loomes, 1994; Lanz et al., 2010]. The key issue is whether the difference is true, or whether it is an artifact of the elicitation process. If the disparity is true, then it should be reflected in decision making; if it is an artifact then targeting the more conservative measure may be appropriate, as Arrow et al. (1993) suggest.

The literature on this matter is vast, and not yet settled, with new papers on the topic continuing to appear on a regular basis. On one side are many papers suggesting that large disparities can arise without bias, either due to substitution effects [Hanemann, 1991; Shogren et al., 1994] or due to reference-dependent preferences [Tversky and Kahneman, 1991]; on the other side are papers explaining the disparity with respect to preference uncertainty [Kolstad and Guzman, 1999; Sileo, 1995] or strategic behaviour [Magat, Viscusi and Huber, 1988]. In these cases, the true disparity, if it even exists, is magnified by features of the scenario. Attempting to correct for this disparity in the manner recommended by the NOAA guidelines may in this case be justified.

2.1.3.2 Values and motivations

Values arise, as we have seen in the above discussion, in relation to exchanges, and potential exchanges. Insofar as it is important solely to estimate the total economic value of an intervention via stated preferences, this is all that matters. The types of motivation that give rise to economic values have, however, been explored in the literature (significant contributions include Weisbrod, 1964, Krutilla, 1967, and Arrow and Fisher, 1974). There is now a common typology in circulation which decomposes "total economic value" into various descriptive components (see Figure 2.1 for a typical breakdown, based on Bateman et al., 2002). The most significant juxtaposition is between so-called "use" values and "non-use", also known as "passive use", values. The former can be measured by revealed preference techniques since they relate to uses which leave a behavioural trace even if only indirect; the latter, by definition, can only be measured by stated preference methods since there is no behavioural trace [Carson, Flores and Mitchell, 1999].

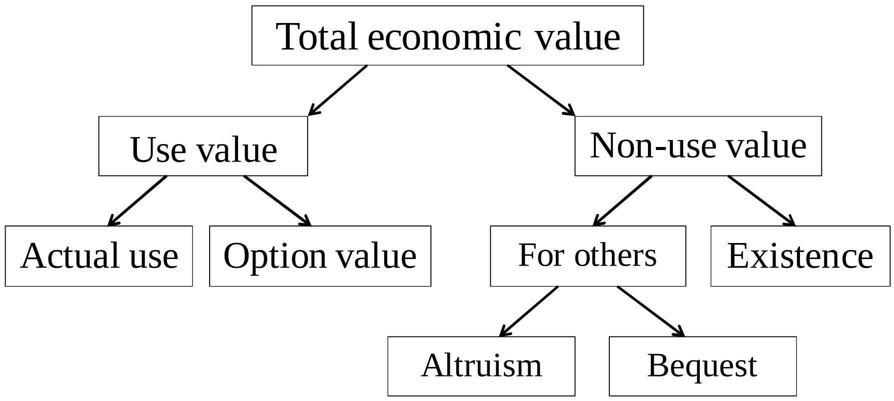


Figure 2.1: Total Economic Value

2.1.3.3 Aggregation and decision making

Historically, aggregating values as a way of guiding social decision making has been somewhat controversial. In the "old welfare economics" of Edgeworth, Marshall and Pigou, economists treated utility as a cardinal measure, and made social welfare comparisons on the basis of net utility gains or losses. Following Pareto (1896), however, a "new welfare economics" began that sought to explicitly distinguish between questions of *efficiency*, and those pertaining to *equity*. Economic efficiency was to be solely based on the Pareto principle, which states that allocations are efficient if and only if it is not possible to make anyone better off without making at least one person worse off. This principle requires only that allocations can be ranked by individuals; it does not require the strong assumptions of cardinality and interpersonal comparability.

Since many interventions lead to a mix of winners and losers, the Pareto principle on its own is often unable to determine any clear guidance. Economists are thus potentially left unable to draw any welfare conclusions. Two distinct approaches sprang from this impasse: one based on compensation criteria to allow the Pareto principle to be applied without requiring interpersonal utility comparisons (Kaldor, 1939, Hicks, 1939; Scitovsky, 1941, Samuelson, 1950); the second deriving social welfare functions to rank social states of affairs from explicit ethical premises (Bergson, 1938; Samuelson, 1947).

The "hypothetical compensation test" of Kaldor (1939) and Hicks (1939) remains the standard CBA decision rule used by economists today. It states that one should add up the compensating (Kaldor) or equivalent (Hicks) variations or surpluses for all affected parties, and then recommend the intervention if the sum is greater than zero. If the sum of compensating variations or surpluses is greater than zero, the winners could potentially compensate the losers and everyone would be at least as well off as before – an actual Pareto improvement. The fact that the compensation does not take place weakens the claim to ethical neutrality, but is seen by most as a pragmatic solution nonetheless. Later work by Scitovsky showed that the hypothetical compensation test could potentially lead to a paradox in which both an intervention, and its reversal, could both be justified [Scitovsky, 1941]. To rule out such situations, Scitovsky proposed the double compensation rule whereby for an intervention to be justified, not only would it need to pass a hypothetical compensation test, but the reversal of the intervention must also fail the hypothetical compensation test. Whilst the Scitovsky test appears to be a reasonable extra condition to attach to a recommendation, for various reasons in practice it has not been assimilated into CBA guidelines (e.g. HM Treasury, 2003).

The second broad approach, following Bergson (1938), has been to construct a social welfare function, based on the idea that economists should be able to advise on the policy implications of any reasonable set of ethical beliefs, whether or not they are widely held by society. In the practice of CBA, this approach is not too far a departure from the compensation approach - it still involves summing measures of value; this time, however the measures are weighted differentially according to income, with higher weights assigned to the lower income groups. CBA manuals occasionally recommend this approach as an extension to, or a sensitivity test against, the hypothetical compensation rule (e.g. HM Treasury, 2003).

No less controversial than the aggregation of values over people has been the aggregation of values over time. A range of approaches have been proposed in the literature and are applied around the world [Spackman, 2004; Zuang et al., 2007]. In the UK, since 2003, the recommended approach to discounting in public sector investment appraisal has been based on the social time preference rate, formulated by the Ramsey equation (Ramsey, 1928; HM Treasury, 2003):

(2.3) $r = \rho + \mu . g$

In equation (2.3), r, the discount rate, is derived as the sum of ρ, itself the sum of two elements: catastrophe risk, and pure time preference, and $\mu . g$, the elasticity of the marginal utility of consumption multiplied by the growth rate in per-capita consumption. In the UK, ρ is set to a value of 1.5%, based on Scott (1977); μ is set to 1.0, based on Pearce and Ulph (1995), Cowell and Gardiner (1999), and OXERA (2002); and g is set to 2% based on Maddison (2001) which examines UK data from 1950 to 1998. For short-term and medium-term projects and policies, the overall discount rate for the UK is therefore set at 3.5%.

More recently, attention has turned to the question of whether a declining discount rate should be used for long-term projects and policies, such as the mitigation of climate change. A body of evidence suggests that people do not discount at a constant rate themselves, but behave in a manner more consistent with a declining discount rate. Moreover, uncertainty about interest rates has also been shown theoretically to lead to a discount rate that declines over time [Weitzman 1998, 1999; Gollier, 2002]. Unfortunately use of a declining discount rate schedule can lead to time inconsistency in policy making, whereby policy makers have an incentive to alter their plans from previous periods [Strotz, 1956]. Whilst Henderson and Bateman (1995) argue that this is natural and not problematic; Hepburn (2003) shows that a naïve government employing a declining discount rate schedule may unwittingly manage a renewable resource into extinction. This problem is not easy to avoid [Heal, 1998]; however, the merits of a declining discount rate schedule have, in the UK at least, won the argument, with the approach enshrined in the Green Book manual for all public sector appraisal [HM Treasury, 2003].

Given an estimate of the discount rate, r, welfare weights (if any), w, and estimates of the monetary costs, C, and benefits, B, of an intervention to individuals i={1,..,N} in each period t={1,..,T}, the basic decision rule in CBA is to proceed if:

$$\sum_{i,t}(1+r)^{-t}w_i(B_{it}-C_{it})>0 \quad (2.4)$$

Expression (2.4) defines the CBA decision rule where the outcome is binary: "proceed" or "abandon". The alternative use of non-market values is in the context where the level of provision of the non-market good is to be optimised. In this case, as shown by Bradford (1970), the optimal level of provision is where aggregate marginal WTP is equal to the marginal cost of supplying the public good.

2.1.4 Elicitation techniques

At the core of any SP survey instrument is the questioning technique designed to elicit preferences, where the aim is to obtain the desired value measure without bias, with a good degree of robustness to procedural variations, and with a good degree of statistical precision. In this section, we begin by introducing the broad range of elicitation formats and the problems observed when implementing them in SP research. We then discuss the empirical regularities in valuations that have been observed when compared against one another. Finally in this section, we outline the theories that have been applied to explain these regularities – those based on the incentive properties of elicitation mechanisms, and those based on cognitive behaviour - and discuss how these theories fare against one another in the tests that have been performed to date.

2.1.4.1 Elicitation Formats

The most direct elicitation question is the open-ended (OE) format, which has the simple form “What is the maximum you would be willing to pay?” If stated values were invariant to the type of elicitation question, the OE format would be the preferred choice due to its resulting in a point measure of WTP, rather than a probability measure. Question formats do affect stated values, however, and although the OE method has been used successfully in some studies, [e.g. Hammack and Brown, 1974; Mitchell and Carson, 1986], it is a stylised fact of the literature that OE questions are generally very difficult for respondents to answer, and consequentially tend to result in many protest or outlying responses, [Desvouges, Smith and McGivney, 1983; Donaldson, Thomas and Torgerson, 1997].

A closely related alternative to the OE format is the payment card (PC) method, developed by Mitchell and Carson (1981, 1984). This technique involves showing respondents a card containing an array, or ladder, of values and asks respondents to pick the maximum value on the card that they would be willing to pay. This format has been found easier to respond to than the OE question (Mitchell and Carson, 1989), whilst still providing a highly informative measure of WTP from a statistical perspective. Response data from PC questions are sometimes interpreted as providing point measures of WTP, in which case they are equally as precise as OE data and are analysed in the same way. In other studies, following Cameron and Huppert (1989), the data are interpreted as corresponding to an interval, with the amount picked from the payment card treated as the lower bound of the interval, and the next highest amount on the card treated as the upper bound.

The bidding game (BG) technique, developed and used in the very first CV survey (Davis, 1964) and many times subsequently (e.g. Randall, Ives and Eastman, 1974; Brookshire, Ives and Schulze, 1976; Daubert and Young, 1981), involves asking respondents whether they would be willing to pay a specified cost for the intervention. If they say "yes", the interviewer then asks them if they would be willing to pay a marginally higher amount. If they say "yes" again, the bid is raised, and so on until they say no. The method thus resembles an ascending price auction. The last bid amount before the respondent says "no" is recorded as a point measure of willingness to pay. One practical disadvantage of this method is that it cannot be implemented as a postal survey, due to the need to wait for each yes/no answer before proceeding to the next bid level. A second disadvantage is that on many occasions estimated values have been found to be sensitive to the size of the opening bid – a phenomenon termed starting point bias (Rowe, D'Arge and Brookshire, 1980; Boyle, Bishop and Welsh, 1985).

The single bounded (SB) dichotomous choice (DC) question was developed by Bishop and Heberlein (1979), and quickly became the most commonly used CV elicitation technique, at least in part because it seemed to mirror real-world choice situations. The method involves asking respondents the question: "would you be willing to pay X?", where "X" is an amount of money that is varied over the sample. The referendum variant of the question asks: "if the program cost X would you vote for it or against it in a referendum?" The SB question has been found to be easy to answer, and, under certain assumptions, has desirable incentive properties as we will discuss below. For these reasons, use of this elicitation method was one of the core recommendations of the influential NOAA manual (Arrow et al., 1993). Unlike the OE and PC methods above, however, the responses to such a question do not give a point measure of WTP; they give only a single bound on the support of WTP. That is, if the respondent says “yes”, then one knows that WTP>X, and if he says “no” then one infers that WTP<X. The information contained in each data point is thus significantly weaker than the alternative methods, and larger samples are therefore usually needed for the same degree of precision. Unfortunately, estimates are also often highly sensitive to the choice of distributional assumptions made (Loomis, 1988; McConnell, 1990).

The double-bounded (DB) variant of the DC method was first proposed by Carson (1985) and Hanemann (1985), and implemented by Carson, Hanemann and Mitchell (1987). It involves asking a standard DC question, as before, but if the respondent says “yes”, a follow-up DC question is asked at a higher cost amount; and if the respondent says no the follow-up is asked at a lower cost amount. By double-bounding the range within which WTP is measured to lie, the information content of the data is substantially enhanced (Hanemann, Loomis and Kanninen, 1991). Unfortunately, however, many studies have since found that a given bid amount is significantly less likely to be accepted by respondents when that amount is presented in the follow-up than when it is the first bid (McFadden, 1994; Cameron and Quiggen, 1994; Herriges and Shogren, 1996; Alberini, Kanninen, and Carson, 1997; Bateman et al., 2001; DeShazo, 2002; Carson et al., 2003). These results imply that they cannot validly be interpreted as independently observed bounds around an underlying “true” WTP.

The DB, SB and BG methods are all nested in the general class of multiple-bound (MB) DC questions, which also includes, amongst others, the triple-bounded (TB) variant (Cooper and Hanemann, 1995), the "spike" model (Hanemann and Kriström, 1995) and the one-and-one-half-bound (OOHB) technique (Cooper, Hanemann and Signorello, 2002). The TB method was motivated as a means of obtaining further statistical precision on the measure of WTP elicited (Cooper and Hanemann, 1995; Langford, Bateman and Langford, 1996; Bateman et al., 1995, 1999, 2001). The method suffers from the same drawback as the DB method, however, which is that the successive responses cannot validly be treated as independently observed bounds around an underlying "true" WTP.

The "spike" model (Hanemann and Kriström, 1995; Kriström, 1997)was motivated by the observation that WTP distributions are likely to have a spike at zero, indicating indifference, and if this is the case, the usual distributional assumptions – log-normal, log-logistic, Weibull – would result in biased estimates of mean WTP. The spike format includes the question: "would you be willing to pay anything?" before a standard DC question to allow this density to be captured.

More recently, a one-and-one half-bound (OOHB) variant of the DC question has been suggested (Cooper, Hanemann and Signorello, 2002). With this method, rather than being told that a given amount is the true cost of the intervention, respondents are told that the cost is uncertain, but is known to lie in a certain range which they are then told. Respondents are asked if they are willing to pay the lower bound of this range, and if so, they are then asked if they are willing to pay the upper bound. Early tests of this method showed it to have much promise (Cooper, Hanemann and Signorello, 2002), and it was subsequently applied in many applications [Fernandez et al., 2004; Barreiro, Sanchez, & Viladrich-Grau, 2005; Nayga, Woodward, and Aiew, 2006; Powe, Willis, and Garrod, 2006; Cooper and Signorello, 2008.] Bateman et al. (2009) found, however, that responses were not invariant to whether the upper bound was asked first or the lower bound. This result implies that they cannot be validly interpreted as being independently observed bounds around an underlying "true" WTP.

Finally in the class of CV methods, a number of authors have argued for an approach which incorporates a degree of stated uncertainty [Li and Mattson, 1995; Ready, Whitehead and Blomquist, 1995; Wang, 1997; Loomis and Ekstrand, 1998; Welsh and Poe, 1998; Ready, Navrud and Dubourg, 2001; Flachaire and Hollard, 2007. These methods take several forms. In one version, proposed by Li and Mattson (1995), respondents are asked a standard DC CV question, and this is then followed up by a second question asking how certain the respondent was in making his choice. Answers to both questions are incorporated into the analysis in order to arrive at a revised WTP distribution. In an alternative version, developed by Ready, Whitehead and Blomquist (1995), respondents answer a polychotomous choice question where the respondent can state one of six responses to the stimulus scenario: "definitely yes", "probably yes", "maybe yes", "maybe no", "probably no" and "definitely no". The authors argue that this technique provides more meaningful information than a simple yes/no response.

Turning now to CM techniques, the principal formats include discrete choice experiments (DCE), contingent ranking, contingent rating, and paired comparisons [Bateman et al., 2002; Hanley, Mourato and Wright, 2001]. In a DCE, the respondent is asked to choose his most preferred alternative, from two or more options; in a contingent ranking exercise he ranks the alternatives in order of preference; in a contingent rating task he rates each alternative on a preference scale; and in a paired comparisons task he indicates his strength of preference between the two alternatives on offer on an intensity scale. Of these formats, only the DCE and contingent ranking formats are consistent with economic theory since strength of preference has no place in ordinal utility theory. Furthermore, respondents have been found in practice to often struggle to completely rank a number of alternatives. The DCE has therefore emerged as the most popular method of choice modelling for non-market valuation in recent years (Carson and Louviere, 2011).

The DCE approach was developed by Louviere and Hensher (1983) and Louviere and Woodworth (1983), based on the integration of discrete choice econometrics [McFadden, 1974; Manski and McFadden, 1981], attribute-based utility theory [Lancaster, 1966], and the conjoint methods used in marketing [Cattin and Wittink, 1982]. Early environmental applications of the DCE method include Adamowicz, Louviere and Williams (1994) and Boxall et al. (1996).

The distinctive features of CM formats, in comparison with CV, are that they: (i) allow for more than two options to be compared at a time; and (ii) involve asking a series of questions which vary not only according to the cost of a specified intervention, but also according to a set of attributes defining the outcome. These features give rise to two practical advantage over CV methods: values can be generated for marginal variations in individual attributes, as well as for a full intervention package of improvements to all attributes simultaneously; and fewer respondents are typically needed for a given degree of statistical precision in comparison with the SB DC methods.

2.1.4.2 Empirical comparisons

Since the earliest days of SP research, comparisons have been made between the results obtained from various elicitation methods, and there now exists a body of evidence containing many empirical regularities. One of these regularities is that the DC method tends to elicit higher WTP values than PC and OE methods, both for public goods [Bishop, Heberlein and Kealy, 1983; Sellar, Stoll and Chavas 1985; Kriström, 1993; McFadden, 1994; Bateman et al.; 1995; Holmes and Kramer, 1995; Brown et al., 1996; Hanley et al., 1998; Welsh and Poe, 1998; Bohara et al, 1998; Green et al., 1998; Cameron et al., 2002; Ryan, Scott and Donaldson 2004; Blaine et al., 2005], and for private goods [Johnson, Bregenzer and Shelby, 1990; Ready, Buzby and Hu, 1996; Lunander, 1998; Ready, Navrud and Dubourg, 2001]. In fact, even when actual money changes hands, the DC method still results in higher WTP than the PC or OE methods [Brown et al., 1996; Lunander, 1998; Champ and Bishop, 2006]. Studies finding no significant difference between methods are fewer in number [Frykblom and Shogren, 2000; Loomis et al., 1997; Reaves, Kramer and Holmes, 1999; Kramer and Mercer, 1997]; and only one study to our knowledge has found a DC WTP less than OE or PC WTP [Cadsby and Maynes, 1999]. The size of the differences vary in each case, but can be very substantial. Ready, Buzby and Hu (1996), for example, finds in a study to value food safety improvements that DC WTP was between 3.6 and 4.4 times as high as PC WTP.

Comparing CV and CM methods, the most common finding is that CM WTP is higher than CV WTP [Hanley, Wright and Adamowicz, 1998; Hanley et al., 1998; Cameron et al., 2002; Bateman et al., 2006; Mogas, Riera and Bray, 2009]. The result appears to depend on the contextual details, however; for example, Foster and Mourato

(2003) and Christie and Azevedo (2009) both find that DCE WTP is higher than DC CV WTP for a large intervention, but that DCE WTP is less than CV WTP for a smaller intervention. Furthermore, a number of studies have found no significant difference [Jin, Wang and Ran, 2006; Colombo, Calatrava-Requena and Hanley, 2006; Tuan and Navrud, 2007] or a mixed set of comparisons, which depend heavily on the functional form used to model the DCE responses [Adamowicz et al., 1998]. One study only, in a contingent behaviour variant of CV and DCE, finds DCE WTP to be uniformly less than DC CV WTP [Boxall et al., 1996]. This last study was notable in this context for the fact that the CV question valued an improvement to one site only, whereas the DCE method was applied to valuing improvements in multiple sites simultaneously.

2.1.4.3 Theories and evidence

Explanations for the observed disparity between WTP elicitation methods, and for the various observed features of respondent behaviour more generally, fall into two broad groups: on the one side are authors who emphasize the strategic properties of the methods [Hoehn and Randall, 1987; Arrow et al., 1993; Cummings et al., 1997; Rondeau, Schulze and Poe, 1999; Carson and Groves, 2007, 2011]; on the other side are those who emphasize cognitive issues, such as anchoring [Green et al., 1998; Ariely, Loewenstein and Prelec, 2003], and learning [Plott, 1996; Braga and Starmer, 2005]. Those emphasizing the importance of the strategic properties of elicitation methods take it as given that people are able to arrive at well-formed valuations of the good by the time they are asked to reveal their WTP, so that the chief issue concerns whether they have the correct incentive to tell the truth. Those emphasizing cognitive issues are typically far less confident in respondents' ability to arrive at well-formed preferences, and focus on how features of the survey instrument might provide cues that influence respondents' value formation, despite their being neutral from a strategic perspective.

The incentive properties of elicitation methods have been exhaustively analysed by Carson and Groves (2007). Building on the results of Farquharson (1969), Green and Laffont (1978) and Carson et al. (1997), the authors argue that the SB DC question is compatible with truth-telling provided that the respondent believes the survey to be consequential, and that he will have to pay the amount he is told is the cost of the good. Surveys that are not considered consequential have no incentives at all, and so respondents cannot be expected to be motivated to reveal their true preferences, although they may do so anyway. Surveys where the consequences are perceived differently to how they are described will give rise to what may initially seem like anomalous behaviour. For example, respondents asked whether they would buy a new product have the incentive to say they would buy it even if they doubt that they would. This is because such a response can be expected to increase the likelihood that the new product would be provided, which can only have a positive value given that the respondent is then under no obligation to buy it.

No other question format is found to be incentive compatible by Carson and Groves (2007). In particular, choices involving more than two alternatives are not incentive compatible owing to the Gibbard-Satterthwaite theorem [Gibbard, 1973; Satterthwaite, 1975]. Sequences of questions such as in a choice experiment, or in MB DC questions are also not incentive compatible. The incentives arising from DB and MB DC questions depend on how respondents' beliefs about cost are formed. Carson and Groves (2007) suggest several possible processes: respondents may accept the new cost, but treat it as being uncertain; they may answer as though the true cost were an average of the current and the previous cost; or they may infer that the quality of the good has changed in line with the change in cost. Different beliefs will cause different response strategies to these questions, and it is not clear-cut which direction the bias will take.

Continuous valuation questions such as the OE and PC formats provide incentives to bias stated values for the intervention downwards because they allow respondents more discretion to attempt to bring about the result they most want. If the respondent expects his answer to influence how much he will actually pay, as well as whether the intervention will be provided, he will have an incentive to bid less than his true valuation. Specifically, if he expects the cost to be less than his true value, then he has an incentive to bid at the expected cost of the intervention. If he expects the cost to be more than his willingness to pay, then he has an incentive to bid zero. Both behaviours result in downwardly biased estimates from these elicitation methods. This line of argument also provides an explanation for the tendency for there to be an abundance of zero responses to OE and PC questions.

From a cognitive psychology perspective, by contrast, many authors have argued that the cost of the good as presented in a DC question signals its value, which "anchors" respondents' perceptions of what they would be willing to pay when unsure of their true valuations. As a consequence, DC results can be susceptible to influence from the survey design. The PC elicitation method, by contrast, is thought to be less susceptible to this sort of anchoring effect because respondents select their own WTP amount [Green et al., 1998; Jacowitz and Kahneman, 1995; Johnson and Schkade, 1989]. The evidence is mixed on whether the range of numbers printed on the payment card anchors WTP results. Dubourg, Jones-Lee and Loomes (1994) finds evidence that

PC CV WTP is affected by the range of values on the card; Rowe, Schulze and Breffle (1996) finds contrasting evidence that an exponential scale covering a sufficiently wide range causes no bias. DCE values for a sequence of policy changes are thought to be potentially biased because the act of answering multiple questions where many attributes vary encourages respondents to place less focus on the cost [Kahneman et al., 2006; Schkade and Kahneman, 1998].

Bateman et al. (2008) test between three differing conceptions of preference formation: (i) pre-existing and revealed via an incentive compatible SB DC question [as in Arrow et al., 1993]; (ii) learned through a process of repetition and experience [List, 2003; Plott, 1996; Braga and Starmer, 2005]; or (iii) internally coherent but heavily influenced by an initial arbitrary anchor [Ariely, Loewenstein and Prelec, 2003]. Both the first and last of these conceptions are rejected in favour of a model in which preferences converge towards standard expectations through a process of repetition and learning. This finding is important as it suggests that, despite the problems identified by cognitive psychologists, SP studies can work well to elicit values, but not by following the traditional guidelines of, e.g. Arrow et al. (1993). SP studies might do better to include several preference elicitation questions, and pay particular attention to the latter responses rather than treating any questions following an SB DC question as being subject to strategic bias and therefore invalid.

2.1.5 Survey design

A good SP questionnaire is typically structured as set out in Table 2.2 below [Bateman et al., 2002]. The main components: include an introductory section on usage and experiences; the main valuation section; a set of debriefing questions asking how valuation questions were made; and a section of demographic questions. It is also common to include a debriefing section at the end of the questionnaire to give interviewers an opportunity to comment on the respondent's understanding of the questionnaire and level of concentration shown, immediately after they have completed the survey.

Table 2.2: SP Questionnaire Structure

Section	Purpose / Content
Usage / Attitudes	To warm-up respondents towards the topic, and to gather evidence on relevant usage or experience, and on attitudes. Responses to these questions are often informative in their own right. They also serve as evidence for validity testing of the values obtained later in the survey.
Valuation	Explains the context for the valuation questions, including the institutional framework for delivering and paying for the good. Explains the attributes of the services in question, and how they vary, explains how to answer the questions, reminds respondents of their budget constraint, and then asks a series of questions designed to elicit their values for the good(s).
Debriefing	Questions asking for respondents' reasons for answering the valuation questions in the way they did. These questions serve as a validity check.
Demographics	Serves to allow for a check on the sample representativeness. Also, allows a check that valuations increase with income as expected. For households, typical demographics collected include gender, age, employment status, socio-economic grade, household size, membership of environmental organisations.

2.1.6 Analytical methods

The methods used to analyse SP data depend on the type of data that are obtained. For example, with OE response data, or with PC response data treated as OE data, an estimate of mean WTP is derived simply from the mean of the sample, and likewise for the median. Interval censored regression can be performed on PC or DC data, and is an ideal method for combining the two sources of data if obtained in the same instrument. For DCE data, the conditional logit model is usually employed [McFadden, 1974], although more recently, mixed logit approaches have become popular [Train, 2009]. In the following, we focus on the interval and the conditional logit models, which are the methods used in the empirical chapters of this thesis.

Interval frameworks are well suited to representing both DC and PC responses. The language of a PC question lends itself to an interval interpretation, with WTP lying between the amount indicated and the next highest amount labelled on the card [Cameron and Huppert, 1989; 1991]. Interval frameworks have also long been used to represent DC responses [Carson and Hanemann, 2005] with a "no" response indicating that WTP lies between zero and the amount asked and a "yes" response indicating that WTP lies between the amount asked and an upper bound reflecting financial resources.

The interval censored framework is straightforward to implement in a maximum likelihood context. Let y_n be an interval censored variable, which we wish to model as a linear function of explanatory variables x_n plus an i.i.d. error term ε_n with mean zero and variance σ^2. Then we have:

(2.5) $$\text{Prob}(y_n) = F\left(\frac{y_n^U - \boldsymbol{x_n\beta}}{\sigma}\right) - F\left(\frac{y_n^L - \boldsymbol{x_n\beta}}{\sigma}\right)$$

which implies the following log-likelihood:

(2.6) $$\text{LL} = \sum_{\text{n}} \log[\text{Prob}(\text{y}_{\text{n}})]$$

A distributional assumption is required for $F(.)$ to implement the estimation. The log-normal is a sensible choice because it ensures that WTP is non-negative (a problem with the normal) and it is straightforward to implement. Since the lower bound for some intervals is zero, the number "1" may be added to all lower and upper bound values before taking logs because the log of zero is undefined. This "1" is then subtracted in obtaining later estimates for mean and median WTP. In a panel context, where for each person, n, there is a PC and a DC response, indexed by t, we may let y_{nt} = log(1+WTP_{nt}) and define lower and upper bounds accordingly, where WTP_{nt} is the willingness to pay by respondent n, as elicited by question type t ($t \in$ {PC, DC}). $F(.)$ is then simply the standard normal cumulative distribution.

The log likelihood in (2.6) is based on the assumption that error terms are independent of one another. Independence is unlikely, however, when responses to both PC and DC questions are combined. To take account of within-person correlation between responses, one may also estimate a random effects panel version of the above model which involves decomposing the error term into an individual specific effect, u_n, assumed to be normally distributed with mean zero and variance σ^2_u and an i.i.d. normal variate with mean zero and variance $\sigma^2_{e.}$

Given a dataset of DCE choices, the conditional logit model is specified as follows. In the simplest case, the utility that customer n obtains from, or ascribes to, option i is represented as:

$$(2.7) \quad U_{ni} = \sum_k \beta_k x_{nik} + \gamma cost_{ni} + \varepsilon_{ni},$$

where x_{nik} is the level of the k^{th} attribute of option i presented to respondent n; β_k is the parameter reflecting the relative importance of attribute k on average for the population; $cost_{ni}$ is the cost facing n if she chooses option i; γ is the parameter reflecting the marginal utility of income; and ε_{ni} is a random error term.

The error term is assumed to be independently and identically Gumbel distributed, which implies that the probability that a respondent n will choose alternative i, when offered alternatives i and j, is given by the logit formula:

$$(2.8) \quad \text{Prob}(choice_n = i \mid x_{ni1}, x_{ni2}, .., x_{niK}, cost_{ni}) = \frac{e^{\sum_k \beta_k x_{nik} + \gamma cost_{ni}}}{e^{\sum_k \beta_k x_{nik} + \gamma cost_{ni}} + e^{\sum_k \beta_k x_{njk} + \gamma cost_{nj}}}.$$

Given a dataset of observed choices, the β and γ parameters in this model may be straightforwardly estimated by maximum likelihood.

In linear models like (2.7) above, we can divide the estimated coefficient on any other attribute by the estimated coefficient on cost to obtain the respondent's marginal willingness to pay for reduced levels of this attribute.

2.1.7 Validity and Reliability

The SP method of non-market valuation has been subject to some criticism over the years, by economists and psychologists [e.g. Kahneman and Knetsch, 1992; Hausman, 1993; Diamond and Hausman, 1994; Kahneman, Ritov and Schkade, 1999]. In this section, which concludes our review of the SP method of non-market valuation, we first examine the arguments put forward concerning the validity of results obtained from SP studies and review the evidence for and against them. Next, we review the methods recommended for establishing and testing validity by manuals such as Bateman et al. (2002). The third part of the section reviews the literature which examines the reliability of SP methods.

2.1.7.1 Challenges to the SP method

Challenges to the CV method have focused on the various ways that survey responses are not consistent with economic theory. The most prominent challenges have concerned: (i) lack of sensitivity to scope; (ii) "too small" income effects; (iii) large sequence and context effects; (iv) too large a disparity between WTP and WTA; and (v) starting point bias [Kahneman and Knetsch, 1992; Hausman, 1993; Diamond and Hausman, 1994; McFadden, 1994]. A variety of causes have been suggested for these effects: that lack of experience with the good in question means that preferences don't exist, and so necessarily cannot be divined [Diamond and Hausman, 1994]; that CV responses are actually attitude expressions, rather than economic values [Kahneman, Ritov and Schkade, 1999]; that preferences are labile and constructed over the course of an interview, rather than being in any sense stable and well-formed [Ariely, Loewenstein and Prelec, 2003].

Many of the early concerns regarding surveys per se have been partly addressed via further exploration within the neoclassical economics framework [Carson, Flores and Meade, 2001]. For example, an empirical review of a large number of SP studies found that the majority rejected tests of insensitivity to scope [Carson, 1997]. Furthermore, Rollins and Lyke (1998) argue that one should not expect perfect scope sensitivity due to diminishing marginal utility of existence values. Some concerns remain regarding scope sensitivity, however, particularly with reference to the valuation of small risk reductions [Beattie et al., 1998; Hammitt and Graham, 1999]. Some evidence suggests that visual aids help to bring risk reduction values in line with an economically reasonable degree of scope sensitivity [Corso, Hammitt and Graham, 2001].

Flores and Carson (1997) shows that the income elasticity of WTP is a fundamentally different measure from the income elasticity of demand, and so economic intuition regarding what constitutes a reasonable magnitude does not carry over from one to the other. Likewise, Carson and Mitchell (1995) and Carson, Flores and Hanemann (1998) show the importance of sequence and nesting effects in theory, and so demonstrate that they are not purely a survey artifact.

An accumulation of evidence suggests, however, that the neoclassical model of preferences itself may be inadequate [Kahneman and Tversky, 2000; Slovic and

Lichtenstein, 2006; Sugden, 1999], and that this is not specific to surveys [e.g. Bateman et al., 1997b]. Specifically, there is evidence that preferences are reference-dependent [Tversky and Kahneman, 1991; Bateman et al., 1997a], rather than defined on states of the world per se; and that preferences appear to be heavily influenced by framing and anchoring effects, to the extent that many authors view them as purely constructed [Slovic and Lichtenstein, 2006].

An alternative theory, which bridges the gap somewhat between the neoclassical economic viewpoint and the psychologist viewpoint, is that preferences are `discovered' via repeated market experience [List, 2003; Plott, 1996; Braga and Starmer, 2005]. The limited evidence to date that has sought to discriminate between these viewpoints in a survey context has found evidence supporting the latter 'discovered preference' position [Bateman et al., 2008]. The field is far from unified, however, on this fundamental question of how preferences are formed.

2.1.7.2 Validity testing

Procedures to test the validity of survey results are reasonably well established, even though the results may be interpreted differently depending on one's preferred theory of preference formation [Smith, 2006]. The main types of validity testing are content, or face, validity, and construct validity [Mitchell and Carson, 1989; Bateman et al., 2002]. Content validity tests are based on: the conformance of the survey instrument, implementation approach and analysis with best practice approaches recommended in the literature [e.g. Mitchell and Carson, 1989; Bateman et al., 2002]; and evidence from debriefing responses concerning how well the respondent understood the survey, believed the scenario and, as far as one can tell from this information, gave meaningful value responses. Construct validity tests examine the conformance of results with expectations, and with the results from related studies, e.g. RP studies. Typically, WTP is expected to increase with income, with the scope of the good, and with any attitude or use questions specific to the survey context; WTP is expected to decrease with distance from the good.

2.1.7.3 Reliability

Stated preference valuation studies typically are performed at one point in time, with the results then used for decision making several months or even years later. This approach is only reliable if values are stable over time, or are predictably different based on observable covariates. Fortunately, the weight of evidence suggests that this is often the case. A number of studies have administered similar questionnaires to independent samples at two points in time, and found that the estimated values, or valuation function, remained unchanged [Brouwer, 2006; Brouwer and Bateman, 2005; Carson and Mitchell, 1993; Carson et al., 1997; Reiling et al,. 1990; Whitehead and Hoban, 1999]; a second group of papers have performed a repeated survey on the same sample of respondents, and found reasonably high correlations between responses [Kealy, Montgomery and Dovidio, 1990; Loomis, 1990; McConnell, Strand and Valdes, 1998]. With one or two exceptions, the literature thus lends support to the application of values derived from historic contingent valuation surveys provided that reasonable adjustments are made for changes in observed determinants over the intervening period [Whitehead and Hoban, 1999].

There has been no study to date, however, which assesses the reliability of values taken before an economic downturn for application during the downturn. There are strong reasons to doubt whether WTP values, for e.g. environmental protection and improvement, remain valid following the onset of a recession. Even after controlling for current incomes, job security may be diminished, and concern for the environment and related policy areas may fall down the list of household priorities as a consequence. It is an open question whether these factors do indeed cause WTP values to fall.

2.2 Applications in the water sector

Stated preference techniques have been applied in the water sector since the earliest trials of the CV method in the 1970s (Gramlich, 1977; Hanemann, 1978; Daubert and Young, 1981; Greenley, Walsh and Young, 1981). Applications have focussed on valuing: in-stream water quality; in-stream flow levels; flood risk reduction; water demand for agriculture, industry and public water supply; drinking water quality; and water supply reliability. In line with the core chapters in this thesis, we focus in the following on the literature valuing in-stream water quality and water supply reliability.

2.2.1 Valuation of in-stream water quality

In-stream water quality is highest in its natural state, and deteriorates when impacted by diffuse and point sources of pollution. Poor water quality affects the aesthetic character of the water environment, curtails recreation opportunities such as boating, angling and swimming, and spoils the habitats of plants fish and animals. Curbing the levels of pollution via regulatory controls is costly, and yet the benefits of doing so can be substantial. For example, Carson and Mitchell (1993) found that the value of a nationwide U.S. improvement from poor quality (below "boatable" status) to good quality ("swimmable") status was $47 billion at 1990 prices. The optimal level of pollution is found where the marginal cost of abatement is equal to the marginal WTP for improvement, (or the marginal WTA additional pollution). The challenge for nonmarket valuation is to estimate this marginal WTP, or, as is often the case when improvements occur in sizeable discrete changes, to estimate WTP for the policy or project intervention causing the improvement.

Although RP methods can be, and have been, employed to estimate the benefits of water quality improvements [e.g. Bockstael, Hanemann and Kling, 1987; Leggett and Bockstael, 2000], the advantage of SP method is that it captures the non-use component of value, i.e. any willingness to pay for environmental improvement unrelated to its impact on the quality or availability of any use of the resource. In the case of water quality improvements, intuition suggests that non-use value may be significant, due to the fact that it affects habitats for fish and other animals. The evidence often confirms this to be the case [e.g. Walsh, Sanders and Loomis, 1985].

Many past studies have used SP methods to assess the benefits of in-stream water quality improvements. Mostly, studies have been conducted at the level of individual water bodies [Gramlich, 1977; Greenley, Walsh and Young, 1981; Smith and Desvouges, 1986; Walsh, Sanders and Loomis, 1985; Green and Tunstall, 1991; Choe, Whittington and Lauria, 1996; Farber and Griner, 2000; Georgiou et al., 2000; Loomis et al., 2000; Barton, 2002; Kramer and Eisen-hecht, 2002; Ready, Malzubris and Senkane, 2002; Alam and Marinova, 2003; Hanley, Bell and Alvarez-Farizo, 2003; Kontogianni et al., 2003; Hanley, Wright and Alvarez-Farizo, 2006; Mourato et al., 2006a; Nahman and Rigby, 2008; Eggert and Olsson, 2009; Birol and Das, 2010; Bradley, 2010)]. Olmstead (2010) and Dalrymple (2006) review this literature, and Van Houtven, Powers and Pattanayak (2007) performs a meta-analysis.

Fewer assessments have been made of the benefits of national or regional policies for water quality improvement such as the Water Framework Directive (WFD). DEFRA (2007) contains estimates of the potential benefits of the WFD in England and

Wales, but, in the absence of an original survey, this study relied on previous single-site valuation studies as a basis for their aggregate estimates. Similarly, Horton and Fisher (2006) relied on single-site valuation studies to compile aggregate estimates of the benefits of the environmental program which was embedded in the 2004 periodic review schedule of water industry investments in England and Wales. Unfortunately, a simple sum of the estimated values of individual water body improvements may substantially overstate the value of a wide-scale program of improvements because individual projects may be substitutes for one another, and aggregating a sequence of public good values is problematic where the goods are substitutes or complements [Hoehn and Randall, 1989; Willis, 2004]. A simple summation of individual public (or private) good values may lead to a serious overestimate of the true value of the whole sequence [Bateman et al., 1997b; Hoehn, 1991].

A smaller set of studies have avoided the problems of aggregating from sequences of individual goods by directly valuing whole programs of water environment improvements nationwide [Carson and Mitchell, 1993; Brouwer, 2008] This approach is adopted to estimate the benefits of aggregate water quality improvements in the US at a policy level. Within England and Wales similar approaches have been used for valuing the impacts of revisions to the EC Bathing Water Directive [Mourato et al., 2006b] and to value national programs for maintaining canals [Adamowicz, Garrod and Willis, 1995].

2.2.2 Valuation of water supply reliability

Water consumers value the reliability of their water supply as well as the water itself. Non-market valuation is needed to appraise infrastructure investments to maintain or enhance supply reliability. Indeed, the economically optimal water supply reliability will be such that the marginal cost of increased reliability equals the marginal WTP for that reliability. Of particular interest, given the focus of one of the core chapters of this thesis, is the value of supply reliability in drought conditions.

A range of studies have investigated WTP to avoid drought water use restrictions using SP methods. These include CV studies and DCE studies. The results display a wide array of values, as might be expected given the variety of experiences with restrictions around the world, the variety of scenarios being evaluated and the range of incomes of the surveyed populations. In the following, we review the key results from each study, grouped by method.

2.2.2.1 Contingent valuation studies

Using the CV method, Soto Montes de Oca and Bateman (2006), a Mexican study, valued two scenarios each comprising a package of risks to interruptions and also variations in water quality and pressure. A "maintenance" scenario, in which expected deteriorations to service would be avoided, was valued by households, on average, at 241 pesos (2001 pesos), equivalent to 164% of the current bill. An "improvement" scenario, which avoided the deteriorations and led to some improvements, was valued on average at 290 pesos, or 197% of the current bill. Genius and Tsagarakis (2006), a Greek study, included only one scenario – the elimination of all restrictions – and obtained an average value for this of €55.6 per household per year.

In each of the above cases, a lack of detailed information on marginal values with respect to the severity and duration of the restrictions would preclude a detailed comparison of asset strategies. One way of overcoming the CV method's limitation in respect of the number of scenarios that can be valued is to implement multiple split-sample versions of the survey instrument where certain attributes of the scenario are experimentally varied. This approach was adopted by Carson and Mitchell (1987), which used four versions of a CV survey of Californian households to obtain WTP for four improvement scenarios. The median values (in 1987 $) ranged from $83 per household per year to avoid the mildest set of restrictions (a 10%-15% shortage once every five years) to $258 per household per year to avoid the most severe restrictions (a 30%-35% shortage and a 10%-15% shortage every five years). In a similar study, Koss and Khawaja (2001) used seven versions of a CV survey of Californian households to obtain WTP for 14 improvement scenarios. The mean values in this case (in 1993 $) included a WTP of $144 per household per year to avoid a 10% shortage once every five years, and a WTP of $193 per household per year to avoid a 40% shortage once every ten years. Griffin and Mjelde (2000) also asked two questions of each respondent – WTP to avoid a current shortage, and WTP to reduce the risks of future shortages, but implemented multiple versions of the survey instrument in order to explore how values varied in response to changes in the frequency, severity and duration of restrictions across scenarios. Their results showed that respondents in seven Texan cities were willing to pay, on average, $25.34-$34.39 (in 1997 $) to avoid a current restriction on water consumption, depending on the extent of the shortage (10%-30%) and the duration of the restriction (14-28) days. They also found that respondents were willing to pay, on average, $9.76/month (or 25.6 per cent of their bill) to improve future reliability levels, a value that the authors argue is higher than one should expect given the results on WTP to avoid a current restriction.

Howe et al. (1994) applied a variant of the CV method in a survey addressed to households in three US towns: Boulder, Aurora and Longmont. Each survey included four valuation questions and so was able to obtain estimates for marginal improvements from each respondent. The survey focused on the value of the chance of a “standard annual shortage event" (SASE) corresponding to restrictions on outdoor water use for a period of three months. The survey asked each respondent four choices to obtain two measures of WTA, for differing sized increases in the chances of a SASE and two measures of WTP for reductions in the chances of a SASE. No information was obtained on the marginal costs of duration or the severity of the restrictions, however, and so the resulting valuation function was limited in the extent to which it could inform detailed comparisons of asset strategies. Results showed that households were willing to pay between $1.01 per household per month, for an improvement in the chances of a SASE from 1/300 to 1/1000, and $1.95 per household per month for an improvement from 1/10 to 1/60 to service reliability. (Different baselines corresponded to different locations of the household).

2.2.2.2 Discrete choice experiment studies

A smaller number of studies have adopted the DCE approach to the valuation of water service reliability: two in Australia [Blamey, Gordon and Chapman, 1999; Hensher, Shore and Train, 2006]; and two in the UK [Willis et al., 2002; Willis, Scarpa and Acutt, 2005]. Hensher, Shore and Train (2006) is the only DCE to date designed purely with the aim of obtaining measures of WTP for reducing the risks of water use restrictions. In this study, 211 households and 205 businesses completed a DCE with attributes including: the frequency with which drought water restrictions can be expected to occur {'once per year', 'once every 3 years', 'once every 10 years' and 'virtually none'}; the duration that water restrictions can be expected to last {'all year', 'all summer','1 month in summer' and 'no restrictions'}; the types of days that water restrictions apply { 'every day', 'on alternate days', and 'no restrictions'; and the level of water restrictions {six levels based on the restriction process adopted in the Australian Capital Territory}. This set of attributes and levels allowed for a very flexible valuation model for use in water resource investment planning. For example, the model showed that households were willing to pay on average AUS $ 11.95 (2003 AUS $) for a reduction in frequency from once every ten years to once every 20 years of "restrictions that matter", i.e. those that apply every day, last all year and are stage 3 or higher, where stage 3 implies "use of sprinklers not permitted, but hand held hoses and buckets in the morning and evening are allowed". Furthermore, residents were predicted to be willing to pay, on average, AUS$ 82.3 to have severe restrictions (level 3 or above) in place for a limited period or not all rather than all year given that the frequency of restrictions is once in every ten years.

Two of the remaining studies used the DCE to explore the wider environmental impacts of water supply enhancement strategies, rather than just their effects on restrictions. Blamey, Gordon and Chapman (1999) reports on a DCE study completed in Canberra, Australia, the aim of which was to investigate residents' preferences between alternative options for their water supply. Alternatives varied according to their cost, use restrictions and environmental impacts. The results suggest that residents were willing to pay AUS $10 (1997 AUS $), on average, to prevent a 10% reduction in water use under the status quo supply option, which would lead to a greater use of water restrictions. Willis et al. (2002) surveyed 412 households in Sussex, UK, to investigate households' preferences as between the environmental impacts associated with abstractions, water use restrictions and cost. The findings suggested that WTP to avoid water use restrictions was small, and in fact statistically insignificant at the 5% level. This finding may be partly due to the fact that only minor restrictions were evaluated: hosepipe bans, and interruptions of less than three days.

The final DCE study, Willis, Scarpa and Acutt (2005), was designed to value 14 distinct attributes of water and wastewater service provision, only one of which related to the frequency of restrictions. The study surveyed 1000 households and 500 businesses in Yorkshire, UK, and found that, on average, Yorkshire households were willing to pay £3.20 per year and Yorkshire businesses were willing to pay £16.90 per year to reduce the risk of experiencing a disruption event of "*2-3 months of no running water on the premises*" for a 250-year increase in the return period, e.g. from one occurrence in 500 years to one occurrence in 750 years.

2.3 Conclusions

The key strength of the SP method is its adaptability to any manner of situation where value evidence is needed. It can be used in cases where no RP methods are possible, to obtain valuations that are consistent with economic constructs, and that can therefore be applied in CBA appraisals. The CV method is adaptable enough to evaluate most kinds of discrete scenario change. The CM method is even more flexible, however, and can be used to develop models capable of predicting values at a number of policy-relevant margins for optimizing the supply characteristics of a public good.

Of course, adaptability is of no use if the method cannot be relied upon to yield valid measures of true preferences. On this matter, a broad review of the evidence suggests that validity inferences are necessarily study-specific. It is not the case that all SP studies are valid – far from it. It is equally untrue, however, that SP surveys universally fail to elicit meaningful measures of preference. The pursuit of validity in policy applications is an ongoing quest, with literature continually accumulating that challenges existing techniques and assumptions, and maps out the properties of the methods. Much research remains to be done to better understand the validity and reliability of SP values, and to derive new and improved techniques for future valuations.

Whilst generally consistent on broad principles that should be applied to ensure validity, e.g. on the layout of an SP questionnaire, and on the need for careful pre-testing to validate the survey instrument, the literature is by no means fully unified on all the methodological details. In particular, there are competing views in the literature concerning the way in which preferences are formed and revealed. Conducting applied non-market valuation studies in the context of an evolving and methodologically divided field challenges one to take a view on the most appropriate design approach for each task in hand. This has certainly been the case for the three applied valuation studies presented in this thesis, all conducted in the water sector.

The water sector has been an important case area for application of SP methods, and so the empirical studies in this thesis are by no means venturing into new territory in this regard. Chapter 3 builds on the empirical literature concerning in-stream water quality valuation with a specific focus on the value of improvements relating to the recently enacted Water Framework Directive, and moreover, by applying innovative techniques to address a clear policy need for valuation results to support the efficient implementation of the Directive in England and Wales. Chapter 4 builds on the literature relating to the benefits of water supply reliability by evaluating the benefits of improved resilience to drought in London, UK. The literature reviewed showed there to be a wide range of values obtained for supply reliability, suggesting that benefits are very sensitive to the context in which they are situated. The study in chapter 4 responds to a clear need for benefits estimates focussed on London households. Finally, the last of the core chapters – chapter 5 – seeks to add to the literature on the reliability of SP methods by addressing a topic previously unexplored in the literature - the sensitivity of stated values to an economic downturn.

3 An Assessment of the Non-market Benefits of the Water Framework Directive for Households in England and Wales

Abstract

Results are presented from a large scale stated preference study designed to estimate the non-market benefits for households in England and Wales arising from the European Union Water Framework Directive (WFD). Multiple elicitation methods (a discrete choice experiment and two forms of contingent valuation) are employed, with the order in which they are asked randomly varied across respondents, to obtain a robust model for valuing specified WFD implementation programs applied to all the lakes, reservoirs, rivers, canals, transitional and coastal waters of England and Wales. The potential for subsequent policy incorporation and value transfer was enhanced by generating area based values. These were found to vary from £2,263 to £39,168 per km^2 depending on the population density around the location of the improvement, the ecological scope of that improvement, and the value elicitation method employed. While the former factors are consistent with expectations, the latter suggests that decision makers need to be aware of such methodological effects when employing derived values.

3.1 Introduction

The European Community (EC) Water Framework Directive (WFD) [European Parliament, 2000] requires that all natural water bodies should reach the common minimum European standard of "Good Ecological Status" (GES) by 2015, except where to do so would entail disproportionate cost. This requirement is widely considered to be stringent and substantively different from most water quality standards that are based either on chemical assessments or the ability to support specific types of use. Achieving GES by 2015 will be technically demanding and expensive. It will require member states to restore many natural habitats for plants, fish and other wildlife by reducing pressures from over-abstraction, point and diffuse sources of pollution, non-native species, and from physical modifications such as dams, weirs and engineered channelling. The cost of achieving full compliance in England and Wales has been estimated to be £2.4 billion per year over a 43 year term [DEFRA, 2008], an amount that far surpasses the cost of any previous EC water policy directive.

Benefits estimates are valuable to policy makers in this context for two related purposes. First they can be used to appraise whole programs of improvements at regional or national levels, as a means to help decision makers set the overall scale of implementation of the directive. In addition, they can be used in assessments, on cost-benefit grounds, of whether achieving GES will be disproportionately costly for individual water bodies. In such cases, applications for derogations can be made to allow for a longer time to achieve compliance or for a less stringent environmental objective to be adopted. The present study was designed to address both purposes simultaneously. In this regard it departs from most previous studies of water quality improvements which have sought to value either a whole program of improvements [Carson and Mitchell, 1993; Brouwer, 2008] or improvements to a localized area [e.g.

Alam and Marinova, 2003; Bateman et al., 2011; Hanley, Bell and Alvarez-Farizo, 2003; Hanley, Wright and Alvarez-Farizo, 2006; Kontogianni et al., 2003; Kramer and Eisen-Hecht, 2002; Loomis et al., 2000].

At the core of our study is the development of a model for valuing national and regional programs of WFD improvements as a function of key attributes relevant to strategy setting at these levels. These attributes include measures of the geographic scale of the implementation program, the balance between improvements to the worst areas and improvements to raise the number of high quality sites, and the balance between improvements in densely populated areas and improvements in more remote locations. A key feature of the model is that it can also be used to value individual localized improvements as a function of the size of the area improved, the qualitative range of improvement and the population density of the area surrounding the water body. An advantage of this approach over typical benefits transfer methods for WFD disproportionate cost assessment [e.g. Bateman et al., 2011] is that the values obtained are fully consistent with the context of a nationwide program of simultaneous improvements. As a consequence they are not biased due to income and substitution effects which are liable to cause a discrepancy between the summed values from individual benefits transfers, and whole program valuations [Hoehn and Randall, 1989; Bateman et al., 1997b; Hoehn, 1991].

Data to estimate the model come from a large-scale nationwide stated preference survey employing three elicitation methods - a discrete choice experiment (DCE) and two contingent valuation (CV) questions. The DCE framework [Louviere, Hensher and Swait, 2000] is naturally suited to the development of multi-attribute valuation models of the kind required in our study. A number of studies have found, however, that the

DCE approach focused on multiple policy changes often elicits higher values for the same package of improvements than a contingent valuation (CV) study focused on a single policy change [Cameron et al., 2002; Foster and Mourato, 2003; Hanley et al., 1998]. A number of reasons have been put forth for this finding ranging from strategic behaviour to placing less weight on the cost attribute when it is varied simultaneously with other attributes to various types of learning behaviour. Most of these suggest that the relative values of non-cost attributes derived from a DCE can be considered reliable, but that total values, which depend on cost, are biased upwards. It therefore seems plausible that reliable estimates could be obtained by using relative values from a DCE and then calibrating their scale using CV estimates. We adopt this approach, and hence employ a CV component in addition to the DCE to obtain estimates of the value of a single large-scale program of WFD improvements with which to calibrate the DCE-derived estimates of individual attribute values.

Willingness to pay estimates are known to be sensitive to elicitation methods and question order effects. Comparing across CV questions, many studies have found that dichotomous choice contingent valuation (DCCV) values exceed those obtained by open-ended formats such as the payment card contingent valuation (PCCV) approach [Venkatachalam, 2004; Welsh and Poe, 1998] to the extent that this is considered a 'stylized fact' of the CV approach [Carson and Groves, 2007]. Alternative lines of explanation for this divergence have been proposed in the CV literature, from a strategic behavior perspective [Carson and Groves, 2007] and from a cognitive psychology perspective [Green et al., 1998]. To test the sensitivity of our findings to elicitation method effects, we utilize both types of CV question and examine responses to them in a joint model. Additionally, several studies have demonstrated the importance of the order in which elicitation questions are presented [e.g. Bateman et al., 2008], and again alternative lines of explanation have been proposed in the CV literature with differing implications. We therefore vary the order of the elicitation questions asked using split samples to be able to isolate and test sensitivities to these effects.

The contributions of the paper are thus threefold. We obtain a model via a robust large-scale stated preference survey for valuing national programs of improvements as a function of key attributes relevant to strategy setting at this level. Additionally, we derive a transferable value function that can be used for disproportionate cost assessment at the level of individual sites, and which can validly be summed over sites so as to obtain values for regional programs of water quality improvements. Finally, we explore the sensitivity of our estimates to elicitation treatment effects.

The remainder of this paper is organized as follows. Section 3.2 outlines the design of the survey instrument, describes the survey administration and characterizes the sample obtained. Section 3.3 provides an overview of our approach to analysis. Section 3.4 describes the econometric models we estimate. Section 3.5 presents our main findings: first on the CV-derived benefits for a benchmark implementation of the WFD, then on the DCE-derived benefit function for valuing varied implementations of the WFD, and finally on our preferred model which combines DCE and CV results. Section 3.6 discusses our findings with respect to elicitation treatment effects. Section 3.7 concludes.

3.2 Survey Design, Administration and Data

The survey design and development for the present study broadly conform to best practices as set out in Arrow et al. (1993); Bateman et al. (2002) and Mitchell and Carson (1989). The description of the good was informed by a stakeholder survey, close work with a team of scientists, and a series of 12 focus groups involving members of the public. The survey instrument was extensively pre-tested with members of the public, via two phases of focus groups (eight groups in total), one phase of 30 cognitive interviews and two pilot surveys of 50 and 100 respondents respectively. The focus group work included tests of the language and concepts used by members of the public to understand and value the water environment; it examined how participants processed selections of visual and textual materials, and how people coped with exercises of varying types and degrees of difficulty; it explored the public's baseline perceptions of current status and drivers of change, their priorities in relation to types of value (eg use+non-use), types of environmental change, types of site, the importance of distance / locations of sites, and the contexts in which value is derived and on which it is dependent, eg timing of environmental change, substitutes and complements, and attitudes to uncertainties and responsibility.

3.2.1 Attributes and Levels

Policy scenarios for WFD improvements are characterized in the survey by the proportions of a respondent's Local area (within 30 miles), and of the National area (England and Wales), that will be High, Medium and Low quality in 2015 (8 years from the survey date), and in 2027 (20 years from the survey date). Table 3.1 gives definitions of the attributes used as they appear in our models, and the levels they take in the design. The table includes, for each attribute, the current proportions at each quality level, and the levels taken by the attribute in the CV and DCE designs.

Note that by including separate attributes for Local and National water bodies we are able to obtain values as a function of population density around water bodies. Prior work [Bateman et al., 2006] shows that individual valuations for spatially fixed environmental goods (such as those for water quality improvements) exhibit 'distance decay' in that they fall as the distance between that individual and the improvement increases. Given this then individual values within the 'Local area' should exceed corresponding per capita values in the 'National area'. This implies that per hectare values will be higher for improvements within densely populated areas than for those located in sparsely populated areas.

Table 3.1: Attributes and Levels

Attribute[a]	Definition[b]	Levels[c]		
		Current[d,e]	CV	DCE[f,g]
HighL8	Proportion at High quality in Local area at time=8 (in 2015)	9.0%	95% 75%	*HighL0* + 0.75(*MediumL0* - Δ*LowL8*) *HighL0* + 0.5(*MediumL0* - Δ*LowL8*) *HighL0* + 0.25(*MediumL0* - Δ*LowL8*) *HighL0* + 0.1(*MediumL0* - Δ*LowL8*)
LowL8	Proportion at Low quality in Local area at time=8 (in 2015)	58.6%	0	0 0.25*LowL0* 0.5*LowL0* 0.75*LowL0*
HighN8	Proportion at High quality in National area at time=8 (in 2015)	15.0%	95% 75%	*HighN0* + 0.75(*MediumN0* - Δ*LowN8*) *HighN0* + 0.5(*MediumN0* - Δ*LowN8*) *HighN0* + 0.25(*MediumN0* - Δ*LowN8*) *HighN0* + 0.1(*MediumN0* - Δ*LowN8*)
LowN8	Proportion at Low quality in National area at time=8 (in 2015)	44.0%	0	0 0.25*LowN0* 0.5*LowN0* 0.75*LowN0*
High20	Proportion at High quality in Local and National areas at time=20 (2027)	As now[h]	95%	95% 75%
Cost	Permanent increase in water bill and other household payments (£/hh/yr)	N/A	£5 £10 £20 £30 £50 £100 £200	£5 £10 £20 £30 £50 £100 £200

*Notes: **a** The quantities of High, Medium and Low quality always sum to 1, so Medium quality is omitted. **b** "Local area" refers to the area within 30 miles of the location of the respondent's interview and "National area" refers to the whole of England and Wales. **c** All environmental status levels were rounded to the nearest whole percentage point in the choice sets used. **d** Current condition levels shown here are those based on data used for the survey itself, rounded to one decimal place. Data are weighted for age, sex and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request. More* *recent data may suggest a different picture of current conditions in the water environment.* ***e*** *For attributes HighL8 and LowL8, current levels are sample mean values.* ***f*** *Terms ending in 0 refer to quality levels at time=0, i.e. current levels.* ***g*** *∆LowL8 ≡ LowL8 -LowL0, and ∆LowN8 ≡ LowN8 -LowN0.* ***h*** *Although "As now" was how the survey presented current conditions to respondents, a numeric value was needed to enter this attribute into the DCE choice models. This essentially involved a choice between HighL0 and HighN0. We chose to use HighL0 for statistical reasons.*

For the DCE, the levels of the future environmental status attributes are based on a "pivot design" methodology [Rose et al., 2008]. Pivot designs, which are common in transportation applications, take the respondent's baseline attributes levels as given and "pivot" off by assigning an increment to those levels to form new attribute levels for the DCE. The variables *LowL8*, *MediumL8*, and *HighL8* in the present case are generated from corresponding baseline conditions *LowL0*, *MediumL0*, *HighL0*, which are known (fixed) quantities and which vary according to respondent location. The generating functions for each level of each environmental attribute used for the main survey are shown in Table 3.1.

For the DCCV and PCCV questions, one half of the sample was offered a more extensive policy package than the other half in order to allow for analysis and testing of the sensitivity to scope of the CV values. In both scenarios, 95% of all water bodies are brought to High quality within 20 years with the remainder at Medium quality. The scenarios differ only with respect to the extent of improvement that occurs within the first eight years: in the "95% scenario", the full 95% is achieved within eight years; in the "75% scenario", 75% are brought to High quality in the first eight years, with the remaining improvement up to 95% High quality occurring between the eight-year and 20-year horizons.

The levels of the payment vehicle, *Cost*, for both the DCCV and the DCE questions were £5, £10, £20, £30, £50, £100 and £200, per household per year in extra water bills and other household payments. The amounts shown in the payment card for the PCCV question ranged from £0 to £1000 spread across 28 points distributed on an approximately logarithmic scale.

3.2.2 Survey Presentation

First, introductory questions on attitudes and use of the water environment were asked; then respondents were shown, in succession, two cards containing carefully developed descriptions of water quality at three colour-coded status levels. (Copies of these cards are contained in Appendix A.) The three status levels were assigned the labels "High quality", "Medium quality", and "Low quality", and the colours assigned to them were dark blue, mid-blue and light blue respectively. The three adopted status levels were linked to the WFD as follows: "High quality" corresponded to High or Good Ecological Status; "Medium quality" corresponded to Moderate or Poor Ecological Status; and "Low quality" corresponded to Bad Ecological Status. The first card contained generic descriptions of water quality at each status level while the second card gave illustrated descriptions specific to one of four water body types: rural river, urban river, lake, or estuary/coastal. Survey time constraints precluded the presentation of more than one type of water body per respondent. By randomly assigning respondents to different water body types, it was possible test for any effects caused by the particular water body type example shown. Statistical tests suggest no effect from the particular example water body the respondent saw so this issue is not discussed further.

Following the status descriptions, respondents were presented with two maps showing current water quality levels, colour coded to match the descriptions just shown. The first map showed the respondent's local area (within 30 miles of the location of the survey interview), and the second showed the whole of England and Wales. A pie chart was included on each map showing the proportions of the water environment in each status category. Examples of the maps shown are reproduced in the appendix to this paper.

The questionnaire then presented each respondent with a valuation exercise comprising: seven DCE questions each offering a choice between the status quo and two improvement alternatives; one DCCV question offering a choice between the status quo and one large-scale improvement alternative; and one PCCV question asking what amount on the card shown to them, or any amount in between, is the most they would be willing to pay, through increased water bills and other household payments every year to have the improvements shown. Included in the appendix are the valuation scenario, including statements to enhance consequentiality and the household's budget constraint, read out to respondents prior to their facing the valuation questions, and specific examples of the DCE, PCCV and DCCV questions.

3.2.3 Experimental Design

The experimental design for the survey was necessarily fairly complex in order to be able to test the range of treatments being considered, which included amongst other things: the CV scenario presented (75% or 95%), the DCCV cost amount offered (£5, £10, £20, £30, £50, £100, or £200), the combination of DCE choice profiles shown, and the order of elicitation questions (PCCV before or after the DCE; DCCV at the beginning middle or end of the DCE). In addition, survey instruments varied across sampling locations due to differences in current water status levels in the local area.

The design for allocating these treatments aimed to minimize the correlation between them and to achieve a good degree of balance across the sample. For the DCE design problem (i.e. the selection of combinations of choice profiles), this involved drawing choice sets (status quo plus two improvement alternatives) randomly, without replacement from the full factorial of every possible combination of attribute levels, excluding strictly dominated and practically impossible combinations, so that each choice card for each respondent was unique. Due to the large sample of unique option profiles, an experimental design created in this way should, with a large sample, provide a reasonable approximation to the full factorial itself, and so thereby be internally near-orthogonal. Compared to a main effects design, it is possibly less statistically efficient but it has the advantage of allowing estimation of lower order interaction terms which, given the variety of design issues under consideration, was considered a key requirement for the present study.

The remaining treatments were allocated independently from the above procedure, and were structured to ensure an even spread of treatments across each location sampled. To this end, by location, each instrument type (defined by its unique combination of water body type example, CV scenario, DCCV cost amount, and order of elicitation questions) was drawn with equal probability from the set of all instrument types, without replacement, so that no combination of treatments was allocated to more than one respondent in any one location. This procedure ensured that each instrument type was given an equal probability of selection overall, that sufficient numbers of certain key combinations would be present in the sample, that there would not be any clustering of treatments by location, and that orthogonality with respect to the DCE design would be preserved.

3.2.4 Survey Administration and Data

The study's target sample was developed as a set of 50 locations, with a target of 30 respondents for each. Locations were sampled in proportion to their population size, and respondents were recruited off the street from the busiest places in the area, with quotas set for age, gender and socio-economic characteristics. Additionally, in order to be in scope, recruits had to be responsible, solely or jointly, for paying the water bill and they had to live within 15 miles of the survey location so that the 30-mile radius map presented to them adequately represented what they would call their local area. An £8 incentive was offered to encourage participation. Although consideration of the range of treatments offered might suggest a need for a larger sample size, the emphasis on orthogonality in the experimental design ensured that all relevant comparisons, e.g. between question order treatments, could be tested without the need to control for all interactions with other treatment types. The target sample size of 1,500 was therefore expected to be more than adequate to estimate the models desired with reasonable precision.

In July 2007, 1,487 respondents were interviewed across the 50 sampling areas. Interviews were conducted face-to-face in a designated location by experienced professionals under the supervision of Accent Market Research using the computer aided personal interviewing (CAPI) technique. The interviews lasted an average of 32 minutes and the interviewers found good levels of understanding and attention were given to the questions.

A total of 165 respondents stated a PCCV WTP of £0 for the scenario, of whom 58 respondents were removed due to giving an invalid protest response. A further 23 were removed for giving no response at all to the PCCV question, and 17 were removed as outliers. Protest cases were identified by examining the verbatim follow-up responses to the elicitation questions; outliers were defined as those in the top 1% of the distribution of PCCV responses, which corresponds to all WTP amounts greater than or equal to £350 per household per year. The DCCV data show roughly 90% are in favor at the lowest amounts (with one small monotonicity violation at £10) dropping to about 40% in favour at £200, the highest amount asked. Interestingly, there were no non-responses to the DCCV or DCE questions, a result which may reflect the higher cognitive load of the more open-ended PCCV format. The total number of respondents removed amounts to 6.6% of the full-sample. Additional analysis reported in our technical report [NERA-Accent, 2007] examined the sensitivities of our main results to more liberal and more conservative, approaches for identifying and excluding protests and outliers. Our results are robust in a qualitative sense to the specific approach used.

Since believability of the scenario is crucial for obtaining valid estimates of WTP, we examined the verbatim responses to the PCCV follow-up question for evidence of any disbelief in the scenario presented. We expect that had respondents doubted that the improvements would take place, and had expressed this doubt by lowering their stated WTP, then they would have articulated this doubt when asked for the reasons underlying their stated response. From the verbatim follow-up responses, we identified only 8 people out of 1,389 in the analysis sample who indicated that they didn't believe the improvements would occur. This constitutes only 0.6% of the sample, which we take as evidence that disbelief in the scenario was not widely held. In addition to this evidence, we found that not a single person during the extensive pre-testing process expressed any doubt that the improvements would take place as described.

Table 3.2 presents population and sample characteristics, for both the raw sample and the analysis sample which excludes protests, outliers and non-responses. The analysis sample characteristics are almost identical to the raw sample characteristics. The raw sample appears to match the population reasonably well although clearly not perfectly. The sample contains a somewhat lower proportion of men than the population, and contains more people out of work, and a lower range of incomes. The sample also appears to be better educated than the population at large. In all the analysis that follows, the data are weighted using a three-way table of survey weights to match sample to population by age, sex and region, based on the 2001 UK Census.

3.3 Overview of Analysis

We analyze the data obtained from the survey as follows. First we combine the DCCV and PCCV responses using a single estimation technique - interval censored regression. This yields estimates of the value of the benchmark "95% scenario" for each question type, the effects of the question order on these estimates, and the effects of respondent covariates. Interval frameworks are well suited to representing both DCCV and PCCV responses. Cameron and Huppert (1989, 1991) have argued that the language of a payment card question lends itself to an interval interpretation, with WTP lying between the amount indicated and the next highest amount labelled on the card. Interval frameworks have also long been used to represent DCCV responses [Carson and Hanemann, 2005] with a no response indicating that WTP lies between zero and the amount asked and a yes response indicating that WTP lies between the amount asked and an upper bound reflecting financial resources. To be conservative and ensure consistency with a key assumption made about the PCCV data, we use an upper bound of £350 for the interval when a respondent said yes to the DCCV question, which is substantially higher than the largest amount used (£200). This does not rule out the possibility that larger WTP values are held by respondents, only that they were not observed in either our PCCV or DCCV data.

Table 3.2: Sample and Population Characteristics

	England & Wales Population[c]	Raw Sample[a]	Analysis Sample[b]
Age[d]			
18-29	19%	14%	14%
30-64	60%	65%	65%
65+	21%	21%	21%
N (=100%)	*40,246,981*	*1,486*	*1,388*
Sex[d]			
Male	48%	43%	43%
Female	52%	57%	57%
N (=100%)	*40,246,680*	*1,487*	*1,389*
Children? [e]			
Yes	29%	27%	27%
No	71%	73%	73%
N (=100%)	*21,660,682*	*1,487*	*1,389*
Education[f]			
Basic	31%	19%	19%
Medium	39%	42%	42%
High	30%	39%	39%
N (=100%)	*34,998,226*	*1,373*	*1,285*
Economic activity[f]			
Working	64%	53%	55%
Not working	29%	44%	43%
Student	7%	2%	2%
N (=100%)	*37,606,305*	*1,373*	*1,285*
Income (weekly)[g]			
Low (<£300)	30%	42%	42%
Med (£300-£1000)	53%	46%	46%
High (£1000+)	17%	12%	13%
N (=100%)	*18,823*	*1,060*	*1,009*

*Notes: **a** Base includes all respondents who answered the relevant question in the survey, unless indicated otherwise. **b** Base excludes from the raw sample the 98 respondents who failed to answer the WTP questions, or who were identified as protestors or outliers. **c** Stats are drawn from the 2001 UK Census except where indicated otherwise. **d** Base for population statistics = all individuals. **e** Base for population statistics = all households. **f** Base for sample statistics = respondents aged 18-74; base for population statistics = individuals aged 16-74. **g** Population statistics are drawn from the 2007/8 UK Family Resources Survey [Department for Work and Pensions, 2008]; Base = all households.*

Our next step is to analyze the DCE data, and we adopt the standard conditional logit model for this purpose [McFadden, 1974]. This model obtains distinct marginal, i.e. per percentage point, values of improvements from Low to Medium, and from

Medium to High, in Local and National areas. Again, we also examine question order effects and the effects of covariates as a test of the validity of the results, and report the range of values we obtain. The DCE-derived marginal values give rise to estimates of the value of the benchmark 95% scenario via inputting the degrees of improvement in each attribute that correspond to this scenario. We compare the results for the 95% scenario obtained from the DCE model with those from the CV model, and test the null hypothesis that they are the same. In line with findings from some other studies [e.g. Foster and Mourato, 2003], we find that the DCE results significantly exceed those derived from the CV model. The DCE results may be biased upwards, either because the act of presenting multiple packages to a respondent causes them to behave strategically, rather than accepting a choice at face value [Carson and Groves, 2007], or because the act of answering multiple questions where many attributes vary encourages respondents to place less focus on the cost [Kahneman et al., 2006; Schkade and Kahneman, 1998]. In the spirit of adopting a conservative approach to analysis, we scale the marginal values derived from the DCE so that the estimated value of the benchmark 95% scenario derived using the scaled DCE marginal values is equal to the CV value for this scenario.

The final step in our analysis is to invert the scaled DCE-based valuation function so that instead of measuring the value of national policies to households it measures the value of individual water body improvements as a function of the size of the area improved, the qualitative scope of improvement - Low to Medium, Medium to High or Low to High - and the density of the population surrounding the water body improved. Regional water body improvement programs such as River Basin Management Plans, and also water utility investment plans that impact on water quality in their area, can then be valued by summing the values over the water bodies improved.

By way of example, we present average values for improvements in a low population density region and a high population density region in England and Wales.

3.4 Econometric Models

3.4.1 Contingent Valuation Models

The interval censored framework is straightforward to implement in a maximum likelihood context. Let y_n be our interval censored variable, which we model as a linear function of explanatory variables x_n plus an i.i.d. error term ε_n with mean zero and variance σ^2. Then we have:

$$Prob(y_n) = F\left(\frac{y_n^U - \boldsymbol{x_n\beta}}{\sigma}\right) - F\left(\frac{y_n^L - \boldsymbol{x_n\beta}}{\sigma}\right),$$

which implies the following log-likelihood:

$$LL = \sum_n \log[Prob(y_n)]$$

A distributional assumption is required for $F(.)$ to implement the estimation. We chose the log-normal because it ensures that WTP is non-negative (a problem with the normal) and it is straightforward to implement. Since the lower bound for some intervals is zero, the number "1" was added to all lower and upper bound values before taking logs because the log of zero is undefined. This 1 was then subtracted in obtaining later estimates for mean and median WTP. In the panel context, where we have two responses per person, indexed by t, we thus let $y_{nt} = \log(1+WTP_{nt})$ and define lower and upper bounds accordingly, where WTP_{nt} is the willingness to pay by respondent n, as elicited by question type t ($t \in$ {PCCV, DCCV}). $F(.)$ is then simply the standard normal cumulative distribution.

The above log likelihood is based on the assumption that error terms are independent of one another. Independence is unlikely, however, when responses to both PCCV and DCCV questions are combined. To take account of within-person correlation between responses, we also estimate a random effects panel version of the above model which involves decomposing the error term into an individual specific effect, u_n, assumed to be normally distributed with mean zero and variance σ^2_u and an i.i.d. normal variate with mean zero and variance $\sigma^2_{e.}$

We present results for four models following this approach:

(CV1) $\log(1 + WTP_{nt}) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}; \boldsymbol{\alpha^{CV1}}) + e_{nt}$

(CV2) $\log(1 + WTP_{nt}) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}, \boldsymbol{Covariates_{nt}}; \boldsymbol{\alpha^{CV2}}) + e_{nt}$

(CV3) $\log(1 + WTP_{nt}) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}; \boldsymbol{\alpha^{CV3}}) + u_n + e_{nt}$

(CV4) $\log(1 + WTP_{nt}) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}, \boldsymbol{Covariates_{nt}}; \boldsymbol{\alpha^{CV4}}) + u_n + e_{nt}$

The first set of variables to enter the models is ***Scope***, which captures the degree of environmental improvement represented by the scenario presented for valuation. In the CV data, there is insufficient variation in the scope of the improvements offered with which to robustly identify separate values for Local versus National, and for extra High quality versus less of Low quality. In the case of improvements to the National area, for example, the data for the improvement option across the sample include only two values for High quality - 75% and 95%, and only one value for Low quality – 0%. Because of this lack of variation, a single scale – High quality Locally – was used to capture the degree of improvement in the CV models. Two variables are entered from this group, *Log %Change*, the log of the percentage change in High quality Locally that occurs within 8 years, and *T95 x Log %Change*, a variable which interacts *Log %Change* with an indicator for the 95% scenario treatment, *T95*, equal to one if the CV scenario results in an improvement to 95% in eight years, and equal to zero otherwise. Standard economic theory suggests the larger the change the more respondents should be willing to pay. Since initial water quality levels vary over the sample, however, two different respondents could be shown the same size change, and for one respondent it would represent 95% of the water in the Local area being of High quality while for another respondent it would represent 75%. If there is declining marginal utility in the spatial extent of the improvement, then a scenario resulting in 95% High quality Locally should be worth less than the improvement to 75% High quality Locally if the absolute size of the improvement is the same in each case. It is possible, however, that respondents only care about the magnitude of the actual change in which no effect should be found.

The second group to enter the CV models, ***Treat***, contains three variables. The first is *Payment Card*, an indicator for whether the observation relates to a response to the payment card question as opposed to a DCCV response. Given past empirical comparisons and the theoretical rationale put forward by Carson and Groves (2007), we expect to see this variable enter with a negative coefficient. The other two variables in this set are *PC x PC First*, a dummy equal to one if the observation is a PCCV response and the PCCV question was asked first, and *DC x DC First*, a dummy equal to one if the observation is a DCCV response and the DCCV question was asked first. These two variables do not have clear cut predictions; however, we present some possible interpretations of the findings in our discussion of the results in section 3.6.

The third group of variables, which enter models CV2 and CV4 only, is $\boldsymbol{Covariates_{nt}}$, a vector of respondent characteristics, such as income, education, use of the water environment and membership of an environmental club. Some of these covariates have a theoretical expectation, such as that frequent users of the water environment should be willing to pay more for its improvement than non-users.

Consistency of the results with such theoretical expectations is an important test of the validity of the results. We discuss our findings in relation to this test in section 3.5.

The terms $\boldsymbol{\alpha}^{1}$, $\boldsymbol{\alpha}^{2}$, $\boldsymbol{\alpha}^{3}$, and $\boldsymbol{\alpha}^{4}$ are the parameter vectors to be estimated for models 1 to 4 respectively. In models CV1 and CV2 the error term is e_{nt} , a normal i.i.d. variate with mean zero and variance σ^2_e. This implies the assumption that the response errors are uncorrelated within respondent. Models CV3 and CV4 relax this assumption via the inclusion of u_n as an additional error term, an individual specific effect assumed to be i.i.d. normally distributed over respondents with mean zero and variance σ^2_u.

3.4.2 Discrete Choice Experiment Models

We analyze the data obtained from the DCE using the conditional logit model [McFadden, 1974]. Let $choice_{nit}$ be a dummy variable equal to one if respondent n chose option i in choice situation t, and equal to zero otherwise. Respondent utility u_{nit} is composed of a deterministic component $v_{nit} = f(\boldsymbol{X}; \boldsymbol{\beta})$ plus an i.i.d. standard Gumbel error term ε_{nit} . Then we have the conditional logit probability expression:

$$Prob(choice_{nit}) = \frac{e^{v_{nit}}}{\sum_j e^{v_{nit}}},$$

where j indexes the alternatives in choice situation t. The above probability implies the following log-likelihood:

$$LL = \sum_n \sum_j \sum_t choice_{njt} \log\left[Prob\left(choice_{njt}\right)\right]$$

We estimate the following two DCE models within this framework:

(DCE1) $\quad v_{nit} = g(\boldsymbol{Scope}_{nit}, SQ_{nit}, \boldsymbol{Treat}_{nit}, Cost_{nit}; \boldsymbol{\beta}^{\boldsymbol{DCE1}})$

(DCE2) $\quad v_{nit} = g(\boldsymbol{Scope}_{nit}, SQ_{nit}, \boldsymbol{Treat}_{nit}, Cost_{nit}; \boldsymbol{Covariates}_{\boldsymbol{nit}}, \boldsymbol{\beta}^{\boldsymbol{DCE2}})$

The two models differ only insofar as DCE2 includes respondent covariates but DCE1 does not. The utility functions corresponding to the DCE models do not map neatly onto the willingness to pay functions specified for the CV models despite the appearance of the same variable sets ***Scope***, ***Treat*** and ***Covariates***. The richness of the data obtained via the DCE allows a richer specification of the value of environmental improvements than does the CV data. In particular, within the ***Scope*** group, we are able to include separate variables for each of the attributes *HighL8*, *LowL8*, *HighN8*, *LowN8*, and *High20*. For the CV data by contrast, there was insufficient variation to identify each of these scope variables separately and so a single scale – High quality Locally – was used to capture the degree of improvement.

The DCE models also include an alternative specific constant, labelled *SQ*, which indicates the status quo, or "no change" alternative which is present in each choice set. This variable captures the average preference for the status quo after allowing for the influence of the attribute level differences, modelled linearly. When such a variable is included in a choice model and enters the model with a positive coefficient it is typically interpreted as a status quo bias - an excessive preference for the status quo given the levels of its attributes in comparison with change alternatives. The opposite interpretation holds for a negative coefficient [Hartman, Doane and Woo, 1991].

The ***Treat*** group in the DCE models contains two variables, *PC First*, an indicator for whether the PCCV question was asked before the DCE questions, and *DC First*, an indicator for whether the DCCV question was asked before the DCE questions (in which case the PCCV question would have been asked after the DCE questions). These variables are entered into the model as interactions with *SQ*.

The payment vehicle variable *Cost* enters both DCE models linearly with coefficient β^{Cost}, which we interpret as minus the marginal utility of income. The final group of variables, ***Covariates***, appears in model DCE2 only. Respondent characteristics enter the DCE model via interactions with the *SQ* and *Cost* variables. When interacted with *Cost*, respondent characteristics impact on willingness to pay via their effect on the marginal utility of income; when they enter via an interaction with *SQ* they impact the probability of choosing an improvement scenario at all.

3.5 Findings

3.5.1 Contingent Valuation Estimates of Benchmark WFD Implementation Scenarios

Results from the interval censored regression models combining DCCV and PCCV responses are presented in Table 3.3. Model CV1 includes no respondent covariates and assumes independence of the error terms. The coefficient on *Log %Change* is of its expected positive sign in this model and significant at the $p < .01$ level. It is an elasticity, so it implies that a 1% improvement in the proportion of High quality improved, e.g. the difference between an improvement of 50%, such as from 25% to 75%, and an improvement of 50%*(1+1%), such as from 25% to 75.5%, results in a 0.73% increase in WTP. The interaction of the *T95* indicator with *Log %Change* has a negative sign ($p < .01$) suggesting that respondents are somewhat less willing to pay for a given change if it takes them all the way to 95% of the local area water being of the highest quality than if it takes them to 75%. The effect is just over 10% the magnitude of the main *Log %Change* coefficient. The interaction of this variable with the *Payment Card* indicator is positive ($p < .10$) and about half the magnitude of the original *T95 X Log %Change* suggesting that PCCV WTP for increased amounts of High quality water environment diminishes less than DCCV WTP over the range of spatial High quality density.

Turning to treatment effects, consist with prior expectations, *Payment Card* has a very significant ($p < .01$) downward effect on the WTP estimate. The coefficient of -0.514 implies that the PCCV treatment leads to an approximately 40% lower WTP estimate all else equal. The order effects are also significant. *DC X DC First* enters with a positive coefficient indicating that when the DCCV question comes first it results in higher DCCV estimates. By contrast, *PC X PC First* enters negatively, implying that PCCV WTP is lower when it is the first of the elicitation questions to appear. Taken together these findings suggest that PCCV WTP is substantially below DCCV WTP, and particularly when it occupies its typical first position.

Looking now at Model CV3, the results with respect to the coefficients on the experimental treatment variables are qualitatively and quantitatively very close to those of Model CV1. The main difference to note is the dramatic improvement in the log-likelihood between Model CV1 and Model CV3 that results from including an additional parameter, allowing for two rather than one error variance. The statistic ρ is equal to the fraction of total variance accounted for by the random individual effects. This takes a value of 0.605 in Model CV3, which is a further indication of the importance of these effects to the fit of the model.

Table 3.3: Interval Censored Models Combining DCCV and PCCV Responses

Variable	Mean[a]	CV1[a,b,c,d] Coef (Std Err)	CV2[a,b,c,d] Coef (Std Err)	CV3[a,b,c,d] Coef (Std Err)	CV4[a,b,c,d] Coef (Std Err)
Constant	1.000	1.121 (1.184)	-0.538 (1.160)	1.378 (1.151)	-0.471 (1.298)
Log %Change	4.315	0.731 (0.282)***	0.535 (0.269)**	0.676 (0.275)**	0.509 (0.305)*
T95 X Log %Change	2.188	-0.082 (0.024)***	-0.071 (0.023)***	-0.080 (0.024)***	-0.069 (0.025)***
PC X T95 X Log %Change	1.094	0.041 (0.022)*	0.044 (0.021)**	0.042 (0.016)***	0.043 (0.016)***
Payment Card (PC)	0.500	-0.514 (0.076)***	-0.806 (0.101)***	-0.562 (0.061)***	-0.817 (0.079)***
DC X DC First	0.092	0.268 (0.088)***	0.209 (0.087)**	0.189 (0.088)**	0.167 (0.086)*
PC X PC First	0.245	-0.400 (0.067)***	-0.375 (0.061)***	-0.361 (0.055)***	-0.346 (0.053)***
Log Income	4.325		0.215 (0.032)***		0.227 (0.037)***
Missing Income	0.265		1.162 (0.198)***		1.237 (0.220)***
Male	0.470		0.145 (0.048)***		0.155 (0.054)***
Child at Home	0.262		0.121 (0.052)**		0.120 (0.063)*
Wales	0.099		0.059 (0.093)		0.077 (0.108)
Water User	0.850		0.716 (0.217)***		0.685 (0.201)***
Pollution Control	0.859		0.826 (0.216)***		0.796 (0.199)***
P_Con. X Water User	0.730		-0.638 (0.230)***		-0.581 (0.217)***
Env. Club Member	0.271		0.228 (0.053)***		0.230 (0.063)***
Understood	0.865		0.298 (0.086)***		0.284 (0.084)***
Under X Not Concentrate	0.051		-0.219 (0.113)*		-0.209 (0.124)*
PC X Edu_High	0.197		0.440 (0.093)***		0.399 (0.074)***
PC X Edu_Med	0.194		0.196 (0.090)**		0.193 (0.071)***
σ		1.121 (0.019)***	1.048 (0.018)***		
σ_u				0.885 (0.027)***	0.791 (0.026)***
σ_e				0.715 (0.019)***	0.706 (0.018)***
ρ				0.605	0.556
Observations		2778	2778	2778	2778
Log Likelihood		-5276.3	-5121.2	-5101.1	-4983.4
Pseudo R^2		0.022	0.051	0.055	0.076

*Notes: **a** Results are weighted for age, sex and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request. **b** Models 1 and 2 are interval censored regression, with no assumed within-person correlation; Models 3 and 4 are interval censored regressions which do allow for within person correlation. The left hand side for each model is the pair {ly1,ly2},where ly1 is the log of one plus the lower bound of WTP and ly2 is the log of one plus the upper bound of WTP. **c** Standard errors are robust, calculated using the Huber-White estimator [White, 1980]; **d** Stars indicate p-value for 2-side t test:* p < 0.10, ** p < 0.05 *** p < 0.01.*

Models CV2 and CV4 in Table 3.3 include the same scope and treatment variables as Models CV1 and Model CV3 plus a number of respondent covariates. The first thing to note about these results is that the addition of these new variables does not change the signs of any of the experiment treatment variables. The most noticeable changes are in the magnitude of the *Payment Card* indicator which has jumped up substantially (for reasons noted below) and in the *Log % Change* variable which has fallen about 25% in magnitude and lost some significance (although it is still significant at the $p < .05$ level in Model CV2, and on the theoretically suggested one-sided test in Model CV4). The addition of the 14 respondent related covariates results in a large improvement in the log likelihood that is significant at the $p < .01$ level.

The signs and magnitudes of the coefficients in Models CV2 and CV4 are quite similar so we will discuss them jointly. The first respondent covariate considered is income. Since some of the sample refused to provide income information (27%), as is typical in surveys, we include two variables to model the income effect. The first, *Log Income*, is equal to the log of household income (£/week) for those that answered the income question and equal to zero otherwise. The second is an indicator variable, *Missing Income*, equal to one if income was not recorded for the respondent, and equal to zero otherwise. In combination, the coefficient on *Log Income* can be interpreted as the income elasticity of WTP for those that answered the income question, and the coefficient on *Missing Income* can be interpreted as the mean income effect of those who did not provide their income. The magnitude of the income elasticity of WTP is .28 in both Models CV3 and CV4 and significant at the $p < .01$ level. This elasticity tends to be smaller than its ordinary income elasticity of demand for theoretical reasons [Flores and Carson, 1997] and because measurement error in income tends to attenuate the coefficient toward zero.

Females tend to give lower WTP estimates ($p < .01$) which is to some degree offset by those with children at home tending to give higher WTP estimates ($p < .05$ in

Model CV2 and $p < .10$ in Model CV4). Lastly, with respect to demographic variables residents of Wales are WTP slightly more than those of England but this effect is not statistically significant at conventional levels suggesting that responses from England and Wales can be combined.

As expected, water users are WTP substantially more ($p < .01$) than those who don't use water under our broad definition of using water outdoors in England and Wales in the previous year. Likewise those who express a pro-environmental view with respect to pollution control are WTP substantially more ($p < .01$) than those who did not. An interaction between *Pollution Control* and *Water User* is negative and significant ($p < .01$) in both models. This term suggests that while the joint effect of these different variables is positive, it is sub-additive. Finally, being a member of an environmental club or organization (broadly defined) is associated with a moderate size increase in WTP which is significant at the $p < .01$ level.

Two variables related to interviewer assessment of the respondent during the interview are included. The first of these is an indicator of whether the respondent was seen as having understood the valuation questions. Those rated as understanding (86.5% of respondents) are WTP more than those who did not ($p < .01$). The other variable is an interaction of the understood indicator with the interviewer rating the respondent as not concentrating. The 4.8% of respondents classified as understanding but not concentrating are willing to pay less (about the same as those not understanding) with this effect being significant at $p < .1$ in both models. This pair of variables worked better than inclusion of both understanding and concentration indicators because of the high correlation between them.

The final pair of variables in the model is a set of interactions between *Payment Card* and indicators for the middle and high education groups in our sample. The high education interaction is large, offsetting almost half of the negative payment card coefficient, and highly significant (p < .01). The interaction of the middle education group with the *Payment Card* indicator is substantially smaller and is less significant (p < .10 in Model CV2 and p < .05 in Model CV4). There were two somewhat surprising aspects of these two interaction terms. At first we included indicators for the middle and high education groups in our original modelling effort and they were significant predictors of WTP. We then added a number of interactions of the respondent covariates with the *Payment Card* indicator. Only two of the education interactions turned out to be strong predictors and when they were included, the indictors for high and medium education levels were no long significant on their own. This suggest that those with different education levels may be responding differently to a payment card with the response of higher education levels being much closer to that of the DCCV treatment.

Table 3.4 presents estimates of median and mean WTP from model CV3, our preferred model since it only includes the experimental design variables and accounts for the within respondent correlation), by question type and order. The estimates are obtained as follows. First, the predicted value of ln(1+WTP) is calculated for each member of the sample, conditional on the treatment pertaining to the cell shown in the table. For example, for the PCCV WTP values, *paycard* is set equal to 1; for the DCCV WTP values it is set equal to zero, etc. For all individuals the predicted value is also calculated as if they were shown the “95% scenario”. Median WTP is then calculated as exp(μ)-1, where μ is the sample average of predicted ln(1+WTP); and mean WTP is calculated as exp(μ +0.5(σ^2_u+ σ^2_e))-1, where σ^2_u and σ^2_e are as shown in Table 3.3.

Mean PCCV WTP is either £50.5 or £72.9, and mean DCCV is either £106.5 or £128.9 depending on the order of elicitation questions asked. The difference between PCCV is greatest when the PCCV question comes first and smallest when both PCCV and DCCV questions are preceded by the DCE. A useful comparison can be drawn between DCCV estimates from the model, and those based on the Turnbull non-parametric method [Turnbull, 1976], which calculates a lower bound on mean WTP by assuming all the density for each set of interval observations is at the lower bound of the interval. This effectively assumes the most conservative distribution that is consistent with the observed choices. For our data the Turnbull lower bound on mean WTP is £127.4 when the DCCV question is asked first and £106.7 when the DCCV question is not asked first. These estimates are approximately the same as those shown in Table 3.4. The reason why our model estimates are not higher than the Turnbull estimates is due to the fact that we have assumed a log normal distribution, and have applied the conservative assumption that none of our observed respondents who said “yes” would have paid more than £350 for consistency. It is something of a coincidence that both of these assumptions result in model estimates that are almost identical to the Turnbull estimates. The comparison suggests, however, that the DCCV estimates shown in Table 3.4 can be considered conservative. Higher estimates of mean WTP can be derived from the DCCV data with reasonable alternative assumptions. One of the contributions of this work is to bring the two approaches together in a single estimated model with the same set of assumptions imposed on both.

Table 3.4: CV WTP Estimates for Benchmark 95% Scenario, By Question Type and Order

Question Order	PCCV WTP £/hh/yr[a,b,c,d]			DCCV WTP £/hh/yr[a,b,c,d]		
	Median	**Mean**	**95% C.I.**	**Median**	**Mean**	**95% C.I.**
PCCV First	26.0	50.5	(47.4, 53.6)	55.3	106.5	(100, 113)
DCCV First	37.7	72.9	(68.4, 77.4)	67.0	128.9	(121, 136.8)
DCE First	37.7	72.9	(68.4, 77.4)	67.0	128.9	(121, 136.8)

*Notes: * Median WTP is calculated as exp(Xβ)-1 and mean WTP is calculated as* $exp(X\beta +0.5(\sigma^2_u + \sigma^2_e))-1$. ***a** Figures are calculated based on improvements from current (2007) water environment status levels, as presented in Table 3.1, to 95% High quality in Local and National areas by 2015, with the remainder at Medium quality. No further improvement occurs beyond this date.* ***b** WTP results are based on CV Model 3 coefficients presented in Table 3.3.* ***c** Estimates are based on £(July 2007).* ***d** Results are weighted for age, sex and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request.*

3.5.2 Discrete Choice Experiment Estimates

Table 3.5 presents DCE results estimation results for the two models described in section 3.4 above. Model DCE1 includes no covariate effects, except for a treatment effect to control for whether or not the PCCV question was asked before or after the DCE. The model is a reasonably good fit for the data. The (McFadden) pseudo-R^2 is 0.18, and coefficients all of the expected sign and statistically significant at least at the 5% level. The attributes *HighL8*, *HighN8* and *High20* enter positively, and *LowL8* and *LowN8* enter negatively, as expected. Furthermore, *Cost* is also negative and significant at $p<.01$.

Both models are linear in environmental improvement attributes. This functional form implies that the value of improving a water body depends on its current status and the status following the improvement, but not on the status of surrounding water bodies or the state of the national water environment generally. Non-linear forms were tested which allowed for, and found, diminishing marginal utility with respect to water environmental improvement, but these models did not significantly outperform the linear model in a statistical sense, and so the simpler linear model was adopted.

With respect to the nature and location of environmental improvements, the results imply that respondents prefer percentage point improvements from Medium to High quality over percentage point improvements from Low to Medium quality (since the coefficients on *HighL8* and *HighN8* are greater, in absolute value terms, than the coefficients on *LowL8* and *LowN8*, respectively). Additionally, respondents prefer a percentage improvement in the National environment more than a percentage improvement in their Local area (the coefficients on *HighN8* and *LowN8* are greater, in absolute value terms, than the coefficients on *HighL8* and *LowL8*, respectively). With regard to the latter finding, however, the size ratio of National to Local areas is approximately 20:1, and so the coefficients on *HighL8* and *LowL8* should be multiplied by 20 to draw a comparison with the coefficients on *HighN8* and *LowN8* in equivalent spatial terms. If this is done, local improvements are seen to be valued very much higher than non-local improvements. Thus, the results show that the typical person values local improvements substantially more than non-local improvements per hectare, which is as expected.

The *Status Quo* (*SQ)* indicator variable enters Model DCE1 with a negative coefficient, indicating that people would prefer an improvement alternative to the status quo after taking account the utility effects of the associated environmental improvements. Thus, rather than the more commonly cited "status quo bias", we find a general reluctance to stick with the status quo. The variable *SQ X PC First* enters the model positively, however, and with a coefficient 85% the size of the *Status Quo* coefficient. This implies that the *SQ* effect is almost wiped out if PCCV was the first elicitation question asked. That is, respondents are more likely to choose the status quo if the PCCV question had already been asked, than if it had not.

Table 3.5: DCE Estimation Results

Variable	Mean[a]	DCE1[a,b]		DCE2[a,b]	
		Coef	Std Err[c,d]	Coef	Std Err[c,d]
HighL8	0.340	0.915	(0.100)***	0.934	(0.102)***
LowL8	0.346	-0.615	(0.123)***	-0.658	(0.121)***
HighN8	0.399	1.128	(0.110)***	1.151	(0.111)***
LowN8	0.293	-0.918	(0.171)***	-0.944	(0.171)***
High20	0.605	0.423	(0.189)**	0.439	(0.186)**
Status Quo (SQ)	0.333	-0.364	(0.180)**	3.361	(0.560)***
SQ X PC Question First	0.163	0.311	(0.130)**	0.331	(0.134)**
SQ X Log Income	1.442			-0.459	(0.091)***
SQ X Missing Income	0.088			-2.208	(0.528)***
SQ X Water User	0.283			-0.590	(0.168)***
SQ X Pollution Control	0.286			-0.681	(0.164)***
SQ X Edu_High	0.131			-0.573	(0.159)***
Cost (£/hh/yr)	0.398	-1.185	(0.048)***	-1.474	(0.082)***
Cost X Male	0.188			0.201	(0.095)**
Cost X Edu_High	0.158			0.203	(0.098)**
Cost X Wales	0.041			0.595	(0.131)***
Observations		29169		29169	
Log Likelihood		-8769.83		-8440.64	
Pseudo-R^2		0.18		0.21	

*Notes: **a** Results are weighted for age, sex and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request. **b** The model is conditional logit; dependent variable is choice, a dummy equal to 1 if the option was chosen; **c** Standard errors are robust, calculated allowing for within-person correlation; **d** Stars indicate p-value for 2-side t test:* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$.*

The coefficients on the scope and treatment variables in model DCE2 are qualitatively the same, and quantitatively almost identical, to those found for model DCE1. The only substantial differences are for the *Status Quo* and *Cost* variables, and this is because these variables enter with interaction terms in DCE2. The interactions with *SQ* all enter negatively, and indicate that people were more likely to choose an improvement alternative if they had high incomes, were water users, held attitudes supporting pollution control efforts, and had a higher level of education. All these findings are consistent with expectation and so are supportive of the construct validity of the survey. The interactions with *Cost* indicate that, all else equal, men are willing to pay more than women for environmental improvements, those with a higher level of education are willing to pay more than others, and those living in Wales are willing to pay more than those living in England.

Table 3.6 shows marginal WTP figures for each of the environmental attributes, and corresponding WTP for the benchmark 95% scenario by question order. The 'parameter' column in this table introduces terminology that we refer to in subsequent text. The term $s^{DCE}_{HighL8}\Delta_{HighL8}$, for illustration, is the product of s^{DCE}_{HighL8} - the estimated marginal WTP, derived from the DCE model, for changes in the *HighL8* variable – and Δ_{HighL8} – the change in the *HighL8* variable. This change is also referred to via Δ_x in the column headings to the right of the 'parameter' column.

As anticipated, given the literature on DCE-CV comparisons [e.g. Cameron et al., 2002; Foster and Mourato, 2003; Hanley et al., 1998], the estimated values from our DCE model for the 95% scenario are substantially and significantly (p<.01) higher than those from the PCCV model and the DCCV model for all question orders, based on two-sided t-tests.

Table 3.6: DCE WTP Estimates for Marginal Changes in Variables, and for Benchmark 95% Scenario

Parameter	Marginal Effect (Δ_X=1%)		"95% Scenario"[a]		
	WTP (£/hh/yr)[b,c]		Δ_X[d]	WTP (£/hh/yr)[b,c]	
$s^{DCE}_{HighL8}\Delta_{HighL8}$	0.77	(0.6, 0.95)	86.0%	66.4	(51.3, 81.5)
$s^{DCE}_{LowL8}\Delta_{LowL8}$	-0.52	(-0.72, -0.32)	-58.6%	30.4	(18.5, 42.3)
$s^{DCE}_{HighN8}\Delta_{HighN8}$	0.95	(0.76, 1.14)	80.0%	76.2	(60.8, 91.6)
$s^{DCE}_{LowN8}\Delta_{LowN8}$	-0.77	(-1.07, -0.48)	-44.0%	34.1	(21.3, 46.9)
$s^{DCE}_{High20}\Delta_{High20}$	0.36	(0.05, 0.67)	86.0%	30.7	(3.9, 57.5)
$s^{DCE}_{SQ}\Delta_{SQ}$			-1	30.7	(1, 60.4)
$s^{DCE}_{SQ\_PC\ First}\Delta_{SQ\_PC\ First}$			-1	-26.2	(-47.8, -4.7)
TOTAL WTP (PCCV first)				242.3	(216.5, 268.1)
TOTAL WTP (DCE or DCCV first)				268.5	(241.8, 295.3)

Notes: ***a*** *Under the "95% scenario", 95% Local and National area is brought to High quality by 2015, with the remainder at Medium quality. No further improvement occurs beyond this date.* ***b*** *WTP results are based on DCE Model 1 coefficients presented in Table 3.5;* ***c*** *Estimates are based on £(July 2007).* ***d*** *Based on improvements from current (2007) water environment status levels, as presented in Table 3.1.*

3.5.3 Scaled WTP Estimates

We derive Low (PCCV)-scaled and High (DCCV)-scaled values for percentage point changes in Local High and Low quality, and National High and Low quality, by applying the formula below:

$$s^{CV}_{k0} = s^{DCE}_{k8}\frac{s^{CV}}{\sum_x s^{DCE}_{x8}\Delta^{95\%}_{x8}}\delta_8$$

In this expression, s^{CV}_{k0} is CV-scaled WTP for an instantaneous 1% change in dimension $k \in \{HighL, LowL, HighN, LowN\}$; s^{DCE}_{k8} is the corresponding DCE estimate from Table 3.6, measuring the value of an eight-year improvement path to an ultimate 1% change, s^{CV} is the CV estimate of WTP for the 95% scenario, as drawn from Table 3.4; $\sum_x s^{DCE}_{x8}\Delta^{95\%}_{x8}$ is the sum of the marginal DCE WTP estimates for *HighL8*, *LowL8*, *HighN8* and *LowN8* multiplied by the corresponding changes in those variables under the 95% scenario, as drawn from Table 3.6. The final term in the expression, δ_8, is the discount factor necessary to equate the value of an 8-year improvement path with an instantaneous change, i.e. $\delta_8 = \frac{1}{8}\sum_{t=1}^{8}(1+r)^{-t}$, for discount rate *r*.

The formula embeds a crucial step, which is to treat the s_{k8}^{DCE} parameters as representing relative values of *HighL*, *LowL*, *HighN* and *LowN*, into which the 95% scenario can be exhaustively decomposed. That is, in applying the formula we interpret the derived s_{k0}^{CV} values as estimates of WTP for a 1% improvement in the k^{th} value dimension with no deterioration thereafter. This step is innocuous if one is willing to impose, as is the case here, an exogenous discount rate. The s_{k0}^{CV} values are consistent with the 95% scenario CV values, in the sense that $\delta_8^{-1}\sum_k s_{k0}^{CV}\Delta_{k8}^{95\%} = s^{CV}$.

Based on the expression above, Table 3.7 presents PCCV-scaled and DCCV-scaled values for percentage improvements in Local High and Low quality, and National High and Low quality, for two discount rate assumptions, 3.5% - the "Green Book" rate used for UK public policy, and 7.0%. The figures for Low (PCCV)-scaled values are derived using s^{CV}=£50.5 per household per year, the lower of the PCCV estimates for the 95% scenario, which corresponds to the PCCV question having been asked first. For High (DCCV)-scaled values, s^{CV}= £128.9 per household per year, which is the higher of the DCCV estimates corresponding to the DCCV question having been asked first. We use the furthest apart estimates from each elicitation method in order to capture the full range of possible values, although we note that the range of values reported could be justifiably extended to incorporate sampling variation, as measured by the 95% confidence intervals reported in Table 3.4.

Table 3.7: Scaled WTP Estimates for Marginal Changes in Current Status

Parameter	Low (PCCV)-Scaled WTP (£/hh/yr)[a]		High (DCCV)-Scaled WTP (£/hh/yr)[a]	
	d.r.=3.5%	d.r.=7.0%	d.r.=3.5%	d.r.=7.0%
s_{HighL0}^{CV}	0.16	0.14	0.41	0.36
s_{LowL0}^{CV}	-0.11	-0.09	-0.28	-0.24
s_{HighN0}^{CV}	0.20	0.17	0.51	0.44
s_{LowN0}^{CV}	-0.16	-0.14	-0.41	-0.36

Notes: ***a*** *Estimates derived as discussed in the text of the paper, based on £(July 2007).*

The estimates presented in Table 3.7 allow valuation of programs of water environment improvements as a function of the geographic scale of the improvements, the extent of population around the area improved, and the change in quality afforded by the improvements. The final step in our analysis is now to derive the inverted valuation function so that instead of measuring the value of national policies to households it measures the value of individual catchment and water body improvements as a function of the size of the area improved, the qualitative scope of improvement - Low to Medium, Medium to High or Low to High - and the density of the population surrounding the area improved.

Table 3.8 presents the inverted function, which makes use of the s_{k0}^{CV} parameters presented in Table 3.7, plus two additional parameters, p and q. The parameter p is a local scalar equal to the population living within 30 miles of the water body in question divided by 1% of the area of a 30mile radius circle. Similarly, q is a national scalar equal to the national population, divided by 1% of the area of the country, including coastal areas.

Table 3.8: Water Body Valuation Function

Qualitative Scope of Improvement	WTP (£/km^2/yr)[a]
Low Quality to Medium Quality	$-(ps^{CV}_{LowL0} + qs^{CV}_{LowN0})$
Medium Quality to High Quality	$(ps^{CV}_{HighL0} + qs^{CV}_{HighN0})$
Low Quality to High Quality	$-(ps^{CV}_{LowL0} + qs^{CV}_{LowN0}) + (ps^{CV}_{HighL0} + qs^{CV}_{HighN0})$

Notes: ***a*** *p = (1/73.23)*(No. households within 30miles.) (NB 73.23=1% of a 30mile radius circle, measured in km^2; q=(No. households in nation/National area (including coastal area, in km2)/100). For England and Wales, q=12,883; s^{CV}_{k0} parameters to be drawn from Table 3.7*

Making use of the valuation function in Table 3.8 requires the use of GIS to obtain local population data for each water body. As an example, the average value of an improvement from low to medium quality in the Solway-Tweed river basin district - the lowest density district in England and Wales – is £2,870 per km^2 for the low (PCCV)-scaled estimate and £7,321 per km^2 for the high (DCCV)-scaled estimate. A similar improvement in the Thames river basin district – the highest density district in England and Wales – is valued at £8,911 per km^2 for the low (PCCV)-scaled estimate and £22,802 for the high (DCCV)-scaled estimate. These examples demonstrate the importance of local population to the values obtained.

3.6 Discussion of Treatment Effects

The effects of question type and order have been found to be both economically and statistically significant in this study, although the bounds they form are still likely to be useful for many policy purposes. The value of the benchmark 95% scenario, for example, was found to vary from £50.5 to £72.9 per household per year via the PCCV responses, from £106.5 to £128.9 per household per year via the DCCV responses, and from £242.3 to £268.5 via the DCE responses. All six of these values are significantly different from one another (p<.01).

Comparing across CV questions, the finding that DCCV values are higher than PCCV values is consistent with many previous findings [Venkatachalam, 2004; Welsh and Poe, 1998]. From a strategic behavioural perspective [Carson and Groves, 2007], the PCCV method is considered less robust than the DCCV method because it allows respondents more discretion to attempt to bring about the result they most want by giving answers that do not truthfully reflect their actual valuations. Under reasonable assumptions, this leads the PCCV approach to result in downwardly biased estimates. The DCCV approach, by contrast, is argued to be compatible with truth-telling due to the 'take-it-or-leave-it' nature of the question, provided that a stringent set of auxiliary conditions are met. Analysis of DCCV responses is more sensitive, however, to distributional assumptions and outliers. Further, some cognitive psychologists have argued that the DCCV method may signal a value for the good, which "anchors" respondents' perceptions of what they would be willing to pay when unsure of their true valuations. Open-end approaches like the PCCV elicitation method are thought to be less susceptible to this sort of anchoring effect because respondents select their own WTP amount [Green et al., 1998; Jacowitz and Kahneman, 1995; Johnson and Schkade, 1989]. (Although it should be noted that earlier suggestions were made that the particular range of amounts shown in the PCCV question can influence respondent answers.) Since these two perspectives have potential offsetting issues, the DCCV and PCCV approaches were both used to allow us to estimate a reasonable range of WTP for benchmark WFD implementation programs.

We also find that values are sensitive to the order in which the questions were asked, a result that is also consistent with many previous studies [e.g. Bateman et al., 2008] and behaviour in actual markets. In the present case, DCCV WTP is found to be higher if the DCCV question came first, and PCCV WTP is found to be lower if the

PCCV question came first. From the strategic behavioural perspective, the first scenario presented has special status since only the first scenario is free from the influence of prior scenarios. In contrast, various types of (non-strategic) hypothesized learning [e.g., Braga and Starmer, 2005; Plott, 1996] suggest that answers to later questions are likely to be more reliable than answers to earlier questions. In the present analysis we have not attempted to distinguish between the strategic and anchoring hypotheses, instead we have simply controlled for the order effects and reported the range of estimates we obtained.

3.7 Concluding Remarks

The research presented here on WTP for potential water quality changes meets a new substantive policy need in England and Wales. Results are based on a carefully designed and well tested stated preference survey that was implemented using a large in-person sample. The principal goal of the study was to develop a robust statistical valuation function capable of providing benefits estimates for national and regional programs of water quality improvements to meet the requirements of the Water Framework Directive (WFD), and to support the development of these programs by quantifying household's priorities with respect to the location of improvements and the types of improvements to be made. The results suggest that households in England and Wales value local improvements much higher than national improvements per km^2 of catchment, lake or coastal water improved, as expected, and value improvements from Medium quality (Poor/Moderate ecological status) to High quality (Good/High ecological status) substantially more than improvements from Low quality (Bad ecological status) to Medium quality (Poor/Moderate ecological status). Regionally averaged values for WFD improvements are found to vary from £2,263 to £39,168 per km^2 improved depending on where the improvement is made, the ecological scope of the improvement, and the source of the valuation estimate from within the range of treatments modelled.

The results are limited in three important ways. Firstly, the decision to focus on programs of improvements rather than on individually specified improvements meant that no information was given to respondents regarding which areas were to be improved except insofar as they were to be made in the local area, i.e. within 30 miles, or elsewhere. It is not hard to imagine that the range of values for individual water body improvements is likely to be substantial within these broad categories. For broad enough programs, errors in the values attributed to individual improvements are likely to cancel each other out. Considerable care should be taken, however, if using these results to make valuation estimates for one-off improvements. A second limitation of the results is that they only provide values for broad ranges of improvement. It is not strictly possible, for example, to use the results to value an improvement from Poor to Moderate ecological status because both status categories are embedded within the Medium quality level. For some purposes, this may be a significant restriction on applicability. The final limitation of the results is that the range of estimates reported with respect to elicitation treatments may be too wide for some policy purposes. Narrowing this range is likely to require taking a stance on the most preferred elicitation method and question ordering.

Despite these limitations, initial benefits estimates obtained from this study (as presented in the project's technical report NERA-Accent, 2007) have already been successfully used in several applications, and are likely to continue to find further policy uses in the years to come. The first application of the initial results was to contribute benefits estimates to a national impact assessment of the WFD in England

and Wales [DEFRA, 2008]. Following on from this, the results were used to support the development of, and appraise, all 11 regional River Basin Management Plans in England and Wales. (Reports are available at www.environment-agency.gov.uk/wfd.) Additionally, the initial results were applied to the appraisal of water utility investment programs in support of the 2009 water price review in England and Wales. The results presented in this paper, which have been revised since the initial policy applications, might usefully be applied in future to the second phase of River Basin Management Plans in 2015, and to the 2014 water company price review in England and Wales. Finally, with suitable adaptation the results would serve as a cross-check on the values of water quality improvement programs in other countries.

4 Willingness to Pay to Avoid Drought Water Use Restrictions

Abstract

Estimates of the value of avoiding drought water use restrictions are important for appraisals of water utility investments to enhance service reliability, as inputs into regulatory incentive schemes for water utility performance, and in operational decisions during a drought period where there is a need to balance the costs of early less severe restrictions against the value of water saved. We investigate the value of avoiding drought water use restrictions in London, UK, by means of a stated preference survey of households and businesses that sought to measure willingness to pay for reductions in the chances, duration and severity of future restrictions. Results from the model are applied to a practical context: a planning inquiry concerning a desalination plant in East London. Based in part on the estimates derived here, the plant was approved and built, and began operating in June 2010.

4.1 Introduction

Estimates of the value of avoiding drought water use restrictions are useful in a range of planning and decision contexts, including whenever investments are being appraised that aim to improve water supply resilience to drought induced shortage. Contexts for such appraisals include internal business planning, judicial reviews of land-use applications for supply augmentation projects, and regulatory reviews of business plans for price setting purposes. Estimates are also valuable as inputs into the design of performance incentive schemes for regulated utilities. This is because an optimal incentive scheme will be calibrated so as to reward or punish the regulated entity in proportion to the welfare consequences of its service levels. Additionally, estimates of the relative costs of restrictions of varying degrees of severity are useful in an operational context during a drought where there is a need to balance the costs of early less severe restrictions against the value of water saved.

A variety of methods are available to obtain the estimates needed, although they are not equally applicable across these decision contexts. Estimates may be obtained from realised costs attributable to drought restrictions [Ding, Hayes and Widhalm, 2010], from revealed preference methods such as demand functions estimated over periods that include drought restrictions [e.g. Woo, 1994; Roibás, García-Valiňas and Wall, 2007; Grafton and Ward, 2008], or via stated preference (SP) surveys that measure willingness to pay (WTP) for improved resilience to future droughts or willingness to accept (WTA) for lower resilience [e.g. Howe et al., 1994; Griffin and Mjelde, 2000; Hensher, Shore and Train, 2006].

In this paper, we investigate the value of reducing the risks of future drought water use restrictions in London, UK. The primary objective for this valuation was to appraise the benefits of a proposed desalination plant in East London – the Beckton plant – which would have a substantial impact on the chances of needing restrictions in future. A secondary aim was that the results could also be used to inform future water resource investment appraisals. Given that the most recent severe drought in London prior to 2006 occurred in 1975/76, and that the majority of properties in London are charged for water on an unmeasured basis, an SP study was the only feasible means of obtaining the estimates needed.

Specifically, we apply the discrete choice experiment (DCE) method [Louviere, Hensher and Swait, 2000; Bennett and Blamey, 2001] to obtain preference measures for London households and businesses. In a DCE, respondents answer a series of choice questions involving two or more alternatives, where the characteristics of the alternatives are experimentally varied so as to provide a rich source of data with which to estimate a valuation model. Such a model makes it possible to obtain estimates of WTP at all the relevant margins pertaining to decision making in regulatory and operational contexts. A disadvantage of the technique, in comparison with the more established contingent valuation (CV) method, is that it rests on the assumption that respondents treat each successive question as equivalent to an independent referendum, and do not carry over beliefs concerning the cost, or most likely outcome, from previous choice scenarios [Carson and Groves, 2007]. The limited empirical work on this topic has been unsupportive of this assumption [McNair, Bennett and Hensher, 2011]. We therefore pay attention to the validity of our findings by carrying out and reporting on a suite of validity tests.

The remainder of the paper is structured as follows. In section 4.2, we review the literature concerned with willingness to pay to avoid drought water use restrictions.

We then outline a model for incorporating welfare estimates of avoided restrictions into water resource asset optimisation in section 4.3. Section 4.4 discusses the survey design and administration, and data characteristics. Section 4.5 describes our econometric modelling methodology. Section 4.6 presents estimation results; section 4.7 reports on our appraisal of the validity of these results; section 4.8 applies the results to an appraisal of the benefits attributable to the Beckton plant, as conducted in 2006; and section 4.9 draws conclusions.

4.2 Survey of the Literature

A range of studies have investigated WTP to avoid drought water use restrictions using SP methods. These include CV studies and DCE studies. The results display a wide array of values, as might be expected given the variety of experiences with restrictions around the world, the variety of scenarios being evaluated and the range of incomes of the surveyed populations. In the following, we review the key results from each study, grouped by method.

4.2.1 Contingent valuation studies

Using the CV method, Soto Montes de Oca and Bateman (2006), a Mexican study, valued two scenarios each comprising a package of risks to interruptions and also variations in water quality and pressure. A “maintenance” scenario, in which expected deteriorations to service would be avoided, was valued by households, on average, at 241 pesos (2001 pesos), equivalent to 164% of the current bill. An “improvement” scenario, which avoided the deteriorations and led to some improvements, was valued on average at 290 pesos, or 197% of the current bill. Genius and Tsagarakis (2006), a Greek study, included only one scenario – the elimination of all restrictions – and obtained an average value for this of €55.6 per household per year.

In each of the above cases, a lack of detailed information on marginal values with respect to the severity and duration of the restrictions would preclude a detailed comparison of asset strategies. One way of overcoming the CV method's limitation in respect of the number of scenarios that can be valued is to implement multiple split-sample versions of the survey instrument where certain attributes of the scenario are experimentally varied. This approach was adopted by Carson and Mitchell (1987), which used four versions of a CV survey of Californian households to obtain WTP for four improvement scenarios. The median values (in 1987 $) ranged from $83 per household per year to avoid the mildest set of restrictions (a 10%-15% shortage once every five years) to $258 per household per year to avoid the most severe restrictions (a 30%-35% shortage and a 10%-15% shortage every five years). In a similar study, Koss and Khawaja (2001) used seven versions of a CV survey of Californian households to obtain WTP for 14 improvement scenarios. The mean values in this case (in 1993 $) included a WTP of $144 per household per year to avoid a 10% shortage once every five years, and a WTP of $193 per household per year to avoid a 40% shortage once every ten years. Griffin and Mjelde (2000) also asked two questions of each respondent – WTP to avoid a current shortage, and WTP to reduce the risks of future shortages, but implemented multiple versions of the survey instrument in order to explore how values varied in response to changes in the frequency, severity and duration of restrictions across scenarios. Their results showed that respondents in seven Texan cities were willing to pay, on average, $25.34-$34.39 (in 1997 $) to avoid a current restriction on water consumption, depending on the extent of the shortage (10%-30%) and the duration of the restriction (14-28) days. They also found that respondents were willing to pay, on average, $9.76/month (or 25.6 per cent of their bill) to improve future reliability levels, a value that the authors argue is higher than one should expect given the results on WTP to avoid a current restriction.

Howe et al. (1994) applied a variant of the CV method in a survey addressed to households in three US towns: Boulder, Aurora and Longmont. Each survey included four valuation questions and so was able to obtain estimates for marginal improvements from each respondent. The survey focused on the value of the chance of a "standard annual shortage event" (SASE) corresponding to restrictions on outdoor water use for a period of three months. The survey asked each respondent four choices to obtain two measures of WTA, for differing sized increases in the chances of a SASE and two measures of WTP for reductions in the chances of a SASE. No information was obtained on the marginal costs of duration or the severity of the restrictions, however, and so the resulting valuation function was limited in the extent to which it could inform detailed comparisons of asset strategies. Results showed that households were willing to pay between $1.01 per household per month, for an improvement in the chances of a SASE from 1/300 to 1/1000, and $1.95 per household per month for an improvement from 1/10 to 1/60 to service reliability. (Different baselines corresponded to different locations of the household).

4.2.2 Discrete choice experiment studies

A smaller number of studies have adopted the DCE approach to the valuation of water service reliability: two in Australia [Blamey, Gordon and Chapman, 1999; Hensher, Shore and Train, 2006]; and two in the UK [Willis et al., 2002; Willis, Scarpa and Acutt, 2005]. Hensher, Shore and Train (2006) is the only DCE to date designed purely with the aim of obtaining measures of WTP for reducing the risks of water use restrictions. In this study, 211 households and 205 businesses completed a DCE with attributes including: the frequency with which drought water restrictions can be expected to occur {‘once per year’, ‘once every 3 years’, ‘once every 10 years’ and ‘virtually none’}; the duration that water restrictions can be expected to last {‘all year’, ‘all summer’,‘1 month in summer’ and ‘no restrictions’}; the types of days that water restrictions apply { ‘every day’, ‘on alternate days’, and ‘no restrictions’; and the level of water restrictions {six levels based on the restriction process adopted in the Australian Capital Territory}. This set of attributes and levels allowed for a very flexible valuation model for use in water resource investment planning. For example, the model showed that households were willing to pay on average AUS $ 11.95 (2003 AUS $) for a reduction in frequency from once every ten years to once every 20 years of “restrictions that matter”, i.e. those that apply every day, last all year and are stage 3 or higher, where stage 3 implies “use of sprinklers not permitted, but hand held hoses and buckets in the morning and evening are allowed”. Furthermore, residents were predicted to be willing to pay, on average, AUS$ 82.3 to have severe restrictions (level 3 or above) in place for a limited period or not all rather than all year given that the frequency of restrictions is once in every ten years.

Two of the remaining studies used the DCE to explore the wider environmental impacts of water supply enhancement strategies, rather than just their effects on restrictions. Blamey, Gordon and Chapman (1999) reports on a DCE study completed in Canberra, Australia, the aim of which was to investigate residents’ preferences between alternative options for their water supply. Alternatives varied according to their cost, use restrictions and environmental impacts. The results suggest that residents were willing to pay AUS $10 (1997 AUS $), on average, to prevent a 10% reduction in water use under the status quo supply option, which would lead to a greater use of water restrictions. Willis et al. (2002) surveyed 412 households in Sussex, UK, to investigate households' preferences as between the environmental impacts associated with abstractions, water use restrictions and cost. The findings suggested that WTP to avoid water use restrictions was small, and in fact statistically insignificant at the 5% level. This finding may be partly due to the fact that only minor restrictions were evaluated: hosepipe bans, and interruptions of less than three days.

The final DCE study, Willis, Scarpa and Acutt (2005), was designed to value 14 distinct attributes of water and wastewater service provision, only one of which related to the frequency of restrictions. The study surveyed 1000 households and 500 businesses in Yorkshire, UK, and found that, on average, Yorkshire households were willing to pay £3.20 per year and Yorkshire businesses were willing to pay £16.90 per year to reduce the risk of experiencing a disruption event of "*2-3 months of no running water on the premises*" for a 250-year increase in the return period, e.g. from one occurrence in 500 years to one occurrence in 750 years.

4.3 Optimal Investment in Water Supply Resilience to Drought

Water utilities manage the capacity of their supply systems by building and maintaining abstraction, treatment, storage and distribution assets, and by investing in leakage reduction and active demand management practices such as metering or water efficiency campaigning. The welfare consequence of all this expenditure depends fundamentally on its effect on the system's capability to meet demand over the possible range of rainfall scenarios. A high level of service reliability for customers is achieved when the system is able to cope with extended droughts without the need for significant restrictions. High levels of service are clearly desirable to customers, but come at the cost of requiring more extensive supply investment. Optimal supply-demand planning therefore involves making a trade off between the costs of water shortages, including those costs borne by customers as a result of water use restrictions, and the costs of supply-demand investments.

Following Griffin and Mjelde (2000), we formalise these considerations as follows. Let aggregate water demand, *D*, be an increasing function of aridity, *a*; and let aggregate water supply, *S*, be a decreasing function of *a*, and an increasing function of investment, *I*. Over a certain segment of the distribution of *a*, supply is insufficient to meet demand, which causes a welfare loss that is a function of the size of the deficit. Accordingly, we specify the welfare loss function at time *t* as

$$(1) \qquad L_t(I,a_t) = \begin{cases} 0 & \text{if } D_t(a_t) \le S_t(I,a_t) \\ l_t\left(D_t(a_t) \le S_t(I,a_t)\right) & \text{if } D_t(a_t) > S_t(I,a_t) \end{cases}$$

The loss function incorporates a deterministic conversion from water shortfall into a usage restriction, and from this usage restriction to a welfare loss. Thus, greater shortages lead to more severe restrictions, which in turn lead to greater welfare losses. We also assume that it is given as a present value, i.e. it incorporates a discount factor.

Investment optimisation is based on minimising the present value sum of expected losses and investment costs, where the expectation is over the random variable *a*. Let $f_t(a_t)$ be the probability density of aridity; then expected losses at time *t* are

$$(2) \qquad E\left[L_t(I,a_t)\right] = \int_{a_t^0}^{\infty} L_t(I,a_t) f_t(a_t)\, da_t$$

where a_t^0 is the level of aridity for which $D_t(a_t) = S_t(I, a_t)$.

The optimisation problem can then be stated as

$$(3) \qquad \min_I \left[I + \sum_t \int_{a_t^0}^{\infty} L_t(I,a_t) f_t(a_t)\, da_t \right].$$

The first order condition to this problem is

(4) $$1 = \sum_t \int_{a_t^0}^{\infty} l_t'(\ldots) \frac{\partial S_t}{\partial I} f_t(a_t)\, da_t$$

The left hand side of equation (4) is the marginal cost of investment. The right hand side is its marginal benefit.

Appraisal of asset strategies within this framework thus requires the following inputs: (i) a measure of aridity that can serve as an input into demand and supply functions, and for which a probability distribution can be reliably derived; (ii) a probability measure of expected supply shortages over the range of aridity possibilities as a function of the supply capabilities of the assets in operation; (iii) a function to convert expected supply shortages into expected numbers of days of restrictions at each level of severity; and (iv) a function to convert expected numbers of days of restrictions at each level of severity into a monetary measure of welfare loss.

The focus of this paper is on the estimation of (iv). In section 4.7, we combine our estimates of the cost of restrictions with data obtained from Thames Water which allows us to estimate the benefits to customers in London of the Beckton plant.

4.4 Survey Design, Administration and Data

The survey was designed so as to be administered to separate household and business samples using the phone-post-phone method. With this method of administration, respondents are recruited by telephone, then sent a pack of show material by post, fax or email, and then re-contacted by telephone to complete the interview. The household and business samples were randomly selected from Thames Water's customer database, although larger businesses were oversampled in order to more precisely estimate total

WTP. The recruitment interview screened out those who were not responsible for paying the water bill for the property, those who worked in the water sector or the market research industry.

SHOWCARD Q
Summary of Water use restrictions

There are four levels of water restrictions – each level would include the measures taken in the lower level.

Level 1 includes advertisements asking people to save water. Water pressure will be lowered slightly in some places.

Level 2 includes Level 1 restrictions plus more advertisements and a ban on the use of sprinklers to water gardens.

Level 3 includes Levels 1 and 2 restrictions, plus bans on:

- the use of hosepipes for watering gardens
- water for parks, recreational and sports grounds, golf courses and racecourses, ornamental ponds and fountains
- car washes where water is not recycled
- operation of automatic flushing cisterns when buildings are unoccupied.

Level 4 restrictions are the most severe, and include all the measures in Levels 1, 2, and 3 plus:

- cutting-off the supply of water to households and businesses in rotation (for example every second day) or cutting-off the supply of water to households and businesses completely.
- water could only be obtained from standpipes (for example from a single tap at a hydrant on every block) or by local delivery of bottled supplies for drinking.
- Many businesses would need to shut down temporarily while the restrictions are in place.
- Emergency drought permits would also be sought to increase the take of water from the rivers. This could lead to further environmental damage.

Figure 4.1: Show Card Describing Restriction Levels

The design of the residential and business surveys was very similar. In each case, the survey was based around a DCE containing 12 choice situations per respondent, each requiring a choice between two service alternatives. Prior to the DCE, respondents were given some background information on London water supply issues, and on the various levels of water restrictions that operate in London. The respondent showcard describing these restrictions is reproduced as Figure 4.1.

After some additional preliminaries, including an explanation of the various ways of interpreting the chances of an event, the DCE began with an example choice, reproduced as Figure 4.2.

CHOICECARD X
Example

	PACKAGE A	**PACKAGE B**
In any year, the chance of Level 3 restrictions is:	1 in 10	1 in 40
When they are applied, Level 3 restrictions will last for:	3 months	9 months
In any year, the chance of Level 4 restrictions is: (Level 3 restrictions are always used first)	1 in 40	1 in 80
When they are applied, Level 4 restrictions will last for:	15 days	30 days
The total Water and Sewerage bill for the year will be:	£300	£330

Figure 4.2: Example Choice Card

The choice of attributes was informed by discussions with focus groups of household customers and in-depth interviews with business customers. The qualitative research suggested that restrictions at Levels 1 and 2 were of little concern to customers. Customers were more concerned about restrictions at Level 3, and much more concerned about restrictions at Level 4. The SP investigation therefore focused on the risk of restrictions at Levels 3 and 4.

The ranges of attribute levels used in the DCE are shown in Table 4.1. The table shows, for each attribute, the set of levels from which the design selected combinations to put to respondent in the DCE. The elements of each column are thus unrelated to one another. Attribute levels were selected to reflect current and target levels of service and were designed to allow for sufficient variation around these levels to allow for the calculation of customers' willingness to pay for the relevant security of supply improvements that the Beckton plant would provide. The bill levels were derived as multiples of the customer's actual annual bill, which was known from the sample database.

Table 4.1: Attribute Levels Used in Choice Sets

Attribute			**Levels**		
Probability of Level 3 restrictions	1 in 10	1 in 20	1 in 40	1 in 80	1 in 1000
Duration of Level 3 restrictions	9 month	3 month	1 month		
Probability of Level 4 restrictions	1 in 20	1 in 40	1 in 80	1 in 250	1 in 1000
Duration of Level 4 restrictions	90 days	30 days	15 days		
Total water and sewerage bill for the year	`=1.5*bill'	`=1.2*bill'	`=1.1*bill'	`=bill'	`=0.9*bill'

Choice sets were generated by randomly sampling option pairs, without replacement, from the full factorial design, and assigning a unique series of choices to each respondent. As noted in Hensher, Shore and Train, (2006), this approach provides for a greater amount of variation in the dataset as a whole than a design replicated, possibly in blocks, over the whole sample. Moreover, there is Monte Carlo evidence that suggests such designs often outperform fractional factorial designs of this kind [Lusk and Norwood, 2005]. Choice pairs were removed when all of the attributes from one package were better than or equal to those of the other package. In addition, combinations which were considered to be operationally unrealistic were also removed. These included packages in which the probability of Level 3 restrictions was less than the probability of Level 4 restrictions, and where the duration of Level 3 restrictions was less than the duration of Level 4 restrictions.

Following a pilot survey, fieldwork for the main survey collected responses from 302 London households and 152 London businesses. Resulting sample statistics for the business sample are given in Table 4.2.

Table 4.2: Business Sample Composition

	Sample	Population
Organisation size		
< 10 employees	46.7%	85.1%
11 - 50 employees	27.3%	11.8%
51 - 200 employees	18.0%	2.4%
201+ employees	8.0%	0.7%
Industry sector		
Agriculture, Forestry and Fishing	2%	0.3%
Mining and quarrying, energy, water supply and manufacturing	8%	5.7%
Construction	7%	5.6%
Distribution, hotels and catering, repair	25%	25.1%
Transport and Communication	5%	3.7%
Financial intermediation, real estate renting & business activities	18%	41.0%
Education and health	14%	5.8%
Public administration and services	21%	12.9%

Source: Population data taken from National Statistics, Inter-departmental Business Register, as cited in National Statistics (Winter 2004/05), "Region in Figures (London)", Chapter 3, Table 3.9, with data on business classifications in London in March 2003.

4.5 Econometric Models

Responses to the DCE are analysed using the logit model [McFadden, 1974]. The utility that customer n obtains from service option i is represented as $U_{ni} = \sum_k \beta_k x_{nik} + \gamma bill_{ni} + \varepsilon_{ni}$, where x_{nik} is the level of the k^{th} attribute of alternative i

presented to customer n; β_k is the parameter reflecting the relative importance of attribute k on average for the population; $bill_{ni}$ is the level of customer n's annual water bill under alternative i; γ is the parameter reflecting the marginal utility of income on average for the population; and ε_{ni} is a random error term. With this utility formalisation and assuming the error term is IID extreme value, the probability that a respondent n will choose alternative i, when offered alternatives i and j, is given by the logit formula:

$$\text{Prob}\left(choice_n = i \mid x_{ni1}, x_{ni2}, .., x_{niK}, bill_{ni}\right) = \frac{e^{\sum_k \beta_k x_{nik} + \gamma bill_{ni}}}{e^{\sum_k \beta_k x_{nik} + \gamma bill_{ni}} + e^{\sum_k \beta_k x_{njk} + \gamma bill_{nj}}}.$$

The β and γ coefficients in this model are estimated by maximum likelihood.

4.6 Estimation Results

4.6.1 Households

Table 4.3 presents our preferred model for household customers. This model represents utility as a linear function of *the expected number of days of restrictions* at each level, plus a linear income effect that varies for different income groups. The expected number of days restrictions at Level i is equal to the probability of Level i restrictions multiplied by the duration of Level i restrictions. It is natural that probability and duration should enter the model multiplicatively. The cost of an additional unit of probability depends on the duration of restrictions the probability relates to, and likewise the cost of an additional unit of duration depends on the probability of restrictions that the duration relates to. A specification containing each attribute as an independent variable would therefore not be economically sensible. The linear specification in expected durations was arrived at after considering and testing alternative non-linear specifications. (See NERA-Accent, 2006 for further details concerning this specification search.)

The pseudo-R^2 for the model indicates an acceptable fit for this type of model.[1] The coefficients on the expected number of days of Level 3 and Level 4 restrictions were negative, highly significant, and differed in size, indicating that respondents were much more concerned about Level 4 restrictions than about Level 3 restrictions. The income group coefficients were significant and negative, showing that all income groups preferred lower bills to higher bills.

Table 4.3: Choice Modelling Logit Estimates for Household Customers

Variable	**Definition**	**Results**
p3d3	Expected number of days of Level 3 restrictions per year; equals the probability multiplied by the duration of Level 3 restrictions.	-0.0165 (5.30)**
p4d4	Expected number of days of Level 4 restrictions per year; equals the probability multiplied by the duration of Level 4 restrictions.	-0.477 (8.38)**
£bill_inc1	Equal to annual water and sewerage bill, measured in pounds, for those respondents with income less than £20k. Equal to zero otherwise.	-0.0164 (11.21)**
£bill_inc2	Equal to annual water and sewerage bill, measured in pounds, for those respondents with income between £20k and £40k. Equal to zero otherwise.	-0.0073 (7.29)**
£bill_inc3	Equal to annual water and sewerage bill, measured in pounds, for those respondents with income greater than £40k. Equal to zero otherwise.	-0.0065 (7.78)**
£bill_miss	Equal to annual water and sewerage bill, measured in pounds, for those respondents with missing data on income. Equal to zero otherwise.	-0.0061 (6.15)**
Observations	Number of Observations (302 x 12)	3624
Respondents	Number of Respondents	302
Log-Likelihood	Measure of Goodness of Fit	-2328.99
Pseudo R^2	Measure of Goodness of Fit	0.07

Notes: Absolute value of z statistics in parentheses. "" stands for significant at 5% and "**" for significant at 1%. Dependent variable is "spchoice," a dummy variable indicating whether or not the respondent chose the alternative when offered. The model is estimated in binary logit form.*

[1] The pseudo-R^2 statistic is calculated as the difference between the log-likelihood of the model and the log-likelihood of a model containing only a constant term, divided by the log-likelihood of the model containing only a constant term.

Using these results for the utility function, we are able to calculate how much residential customers are willing to pay for water supply reliability. Our measure of supply reliability is the statistical expected day, calculated as the probability of a drought water-use restriction event multiplied by its duration. For example, if at the starting point there is a 0.1 chance of a restriction event in any year, and the likely duration of an event would be 100 days, then we calculate that there are 10 expected days of restrictions each year. A risk reduction of one expected day could be achieved by lowering the likely duration of an event to 90 days or by lowering the likelihood to 0.09 each year. Measured this way, we understand from Thames Water that the current reliability level for water service in the London area is around 1 expected day of Level 4 restrictions per year. Based on these results, London households, on average, are willing to pay £1.85 per year for each reduction of one expected day of Level 3 restrictions, plus £53 per year for each reduction of one expected day of Level 4 restrictions.

4.6.2 Businesses

Our preferred model for estimating the utility expressed in the London businesses' choices is shown in Table 4.4. As in the household model, utility is represented as a linear function of *the expected number of days of restrictions* at each level, plus a bill effect. In this case, the bill effect is included as a percentage of the business customer's current bill. The model groups business customers into three classes: the smallest businesses with fewer than 10 employees, mid-size businesses with between 11 and 200 employees, and the largest customers with more than 200 employees. Again, the linear specification in expected durations was arrived at after considering and testing alternative non-linear specifications. These included models with squared durations multiplied by probabilities, and models with interactions between level 3 and level 4 expected durations. None of the non-linear models outperformed the linear specification shown.

The business model has an acceptable fit. The coefficients on the expected number of days of Level 3 and Level 4 restrictions per year are negative, highly significant, and differed in size, confirming that businesses were more concerned about Level 4 than about Level 3 restrictions. The coefficients on the dummy variables for small and medium business size are significant and negative, confirming that willingness to pay to avoid supply restrictions, as a proportion of the annual water bill, increases with business size.

The largest customers exhibit very high willingness to pay to avoid restrictions, especially the severe Level 4 restrictions. This caused a difficulty with the estimation, because the largest customers appeared to have ignored the bill attribute when making their choices, indicating that the levels were probably set too low to encourage trading off between improved reliability and bill increases for these customers. (The maximum bill level used represented a 50% increase on respondents' current bills.) To overcome this difficulty, we omitted the bill attribute from the utility function for the largest customers, and imposed a cap on the extra annual amount large businesses would be prepared to pay to avoid one expected day of restrictions at 100 percent of their annual bill for Level 4 restrictions. Given the estimated marginal rate of substitution between Level 4 and Level 3 restrictions, this also corresponded to a cap of 6 percent of their annual bill to avoid one expected day of Level 3 restrictions for the largest customers.

Table 4.4: Choice Modelling Logit Estimates for Business Customers

Variable	Definition	Results
p3d3	Expected number of days of Level 3 restrictions per year; equals the probability multiplied by the duration of Level 3 restrictions.	-0.0207 (4.88)**
p4d4	Expected number of days of Level 4 restrictions per year; equals the probability multiplied by the duration of Level 4 restrictions.	-0.3715 (4.73)**
%bill_emp1	If no. of employees is less than 10, equal to annual water and sewerage bill as a percentage of current bill. Otherwise equal to zero	-1.3301 (5.32)**
%bill_emp23	If no. of employees is between 11 and 200, equal to annual water and sewerage bill as a percentage of current bill. Otherwise equal to zero	-0.5697 (2.34)*
Observations	Number of Observations (149 x 12)	1788
Respondents	Number of Respondents	149
Log-Likelihood	Measure of Goodness of Fit	-1199.69
Pseudo R^2	Measure of Goodness of Fit	0.03

Notes: Absolute value of z statistics in parentheses. "" stands for significant at 5% and "**" for significant at 1%. Dependent variable is "spchoice," a dummy variable indicating whether or not the respondent chose the alternative when offered. The model is estimated in binary logit form.*

From the utility function in Table 4.4, we estimated that on average, London businesses are willing to pay £48 per year for each reduction of one expected day in Level 3 restrictions, plus £845 per year for each reduction in one expected day in Level 4 restrictions. The unit of risk again is the statistical expected day, formed here just as for households.

Table 4.5 presents a summary of the WTP estimates for households and businesses.

Table 4.5: Household and Business Willingness to Pay Water Service Reliability

Value per Expected Day of Restrictions	Level 3 £ per customer year	Level 4 £ per customer year
Households	£2	£53
Businesses	£48	£845

4.7 Validity Appraisal

A number of measures were taken to test whether or not a survey has achieved valid measures of the preferences of the target population. Our analysis suggests that the questionnaire succeeded in eliciting meaningful statements of preferences from respondents, that results are consistent with prior expectation, and that they are reasonable in light of evidence from external sources. In the following we outline our findings on these matters, grouped into content validity, and construct validity appraisals.

4.7.1 Content validity

A survey is said to have high content validity if: 'the survey descriptions and questions are clear, reasonable and unbiased, … [such] that respondents are put in a frame of mind that motivates them to answer seriously and thoughtfully' [Schumann, 1996, p.77]. An examination of the responses given in a number of questions indicates that both residential and business consumers were able to provide sensible answers to the questions.

At the outset of the DCE, the vast majority of respondents (446/454) were able to provide articulate and rational reasons for selecting a particular package from the example choice question. This demonstrates that from the start of the choice exercises, individuals were able to understand the varying aspects of each package, compare the alternatives and make an informed selection. Furthermore, respondents in both household and business surveys were able to provide detailed and articulate explanations at the end of the DCE of how they went about selecting packages from the choice sets presented to them. In general, explanations coincided with the reasoning provided in the example set.

For example, a number of business respondents when asked to explain how they made their selections in the choice experiment explained that they were concerned about the impact of restrictions on their business and that additional costs were worth the decreases in potential harm. A few, like respondent 2100023 indicated that Level 4 restrictions would result in a total stoppage in activity,

"…would have to pay the bills at whatever the amount is. There is no way we could carry on without water."

A number of other respondents indicated that they could function with Level 4 restrictions but only for a very limited time period. Respondent 2100022 explained,

"Odds can live with but some of the longer spells would cripple us. We would have to spend the money."

Others, such as respondent 2040027 and 2120026 replied that choices were based on whatever option would allow them to continue to operate.

"(The) one that makes me close the least."

"Mainly trying to keep the business going."

These statements are typical and indicate that business respondents understood and were concerned with the impact restrictions would have on their ability to operate.

While not facing loss of production, residential respondents also expressed concern over the effect of restrictions on their daily activities. Again, verbatim responses demonstrate that individuals considered such impacts when making decisions about a willingness to pay for increased security of supply. A number of respondents, like 3030022, stated that Level 4 restrictions would be more than an inconvenience,

"… particularly level 4, that was more likely to be applied and we could not live with that level as it would make life more difficult."

Some respondents argued that they could live with Level 4 restrictions but could not withstand a lengthy period under such conditions. Respondent 3150052 explained,

"I just concentrated on the time of Level 4 restrictions. It's frightening to think that it would last three months."

Some respondents referred to specific aspects of their household that would make restrictions difficult. In particular, a few like 3110020, mentioned small children.

"So for example, Level 4 restrictions would be impossible for me with the kids…"

Like the business verbatims, these responses indicate that many residential customers determined that water restrictions would have a serious impact on their daily activities and demonstrate that they would be willing to pay to decrease the probability or duration of such restrictions.

Overall, these analyses demonstrate that the questionnaire and choice exercise was intelligible and was able to generate meaningful results from respondents.

4.7.2 Construct validity

Construct validity indicates whether or not the results vary across the sample data in line with expectations, and whether they are consistent with external evidence. Supporting evidence comes from the fact that the signs and magnitudes of the WTP measures are consistent with prior expectation, and that WTP varies with income and business size in line with expectation. Furthermore, results are consistent with the evidence from external studies reviewed in section 4.2.

In particular, the most closely comparable external study is Willis, Scarpa and Acutt (2005). This study surveyed 1000 households and 500 businesses in Yorkshire, UK, and found that, on average, Yorkshire households were willing to pay £3.20 per year and Yorkshire businesses were willing to pay £16.90 per year to reduce the risk of experiencing a disruption event of "*2-3 months of no running water on the premises*" for a 250-year increase in the return period, e.g. from one occurrence in 500 years to one occurrence in 750 years.

These results imply that, on average, residential customers were willing to pay between £18 and £107 per household per expected day reduction in Level 4 restrictions per year. [2] For businesses, the comparable range is £94 to £563 per business per expected day reduction in Level 4 restrictions per year.

Our results indicate that residential customers would be willing to pay £53 per year and business customers would be willing to pay £845 per year for one fewer day of expected Level 4 restrictions. Our main results for households thus sit comfortably within the range of comparable results derived from the Yorkshire Water study. For businesses, the differences in types of business between London and Yorkshire make it difficult to draw direct comparisons, although it certainly does not seem unreasonable that WTP by businesses in London might be significantly higher than those in

---

[2] For an improvement from 1/250 to 1/500 chance of a 90 day restriction, the change in expected number of days is equal to 1/500 * 90 =0.18. Implied WTP per expected day is then given by £3.20 / 0.18 = £17.78. At the top end of the reliability range considered - an improvement from 1/750 to 1/1000 - the change in expected number of days is equal to 1/3000 * 90 =0.03. Implied WTP per expected day in this case is given by £3.20 / 0.03 = £106.68.

Yorkshire. Our estimates are therefore generally consistent with those of Willis, Scarpa and Acutt (2005).

4.8 The Value of Improved Service due to the Beckton Plant

The primary purpose of the welfare estimates derived here was to contribute to an economic appraisal of the benefits of the UK's first desalination plant. Thames Water initially applied for planning approval for the Beckton plant in 2004. The application was approved, but the then mayor of London intervened, directing Newham Borough Council to overturn its decision. Mayor Livingstone's principal objection related to the fact that the desalination plant would emit large quantities of greenhouse gases. Thames Water appealed against the decision, and a public inquiry was held in 2006. The study presented in this paper was commissioned to provide Thames Water with supporting evidence at this inquiry that the Beckton plant would be beneficial to London.

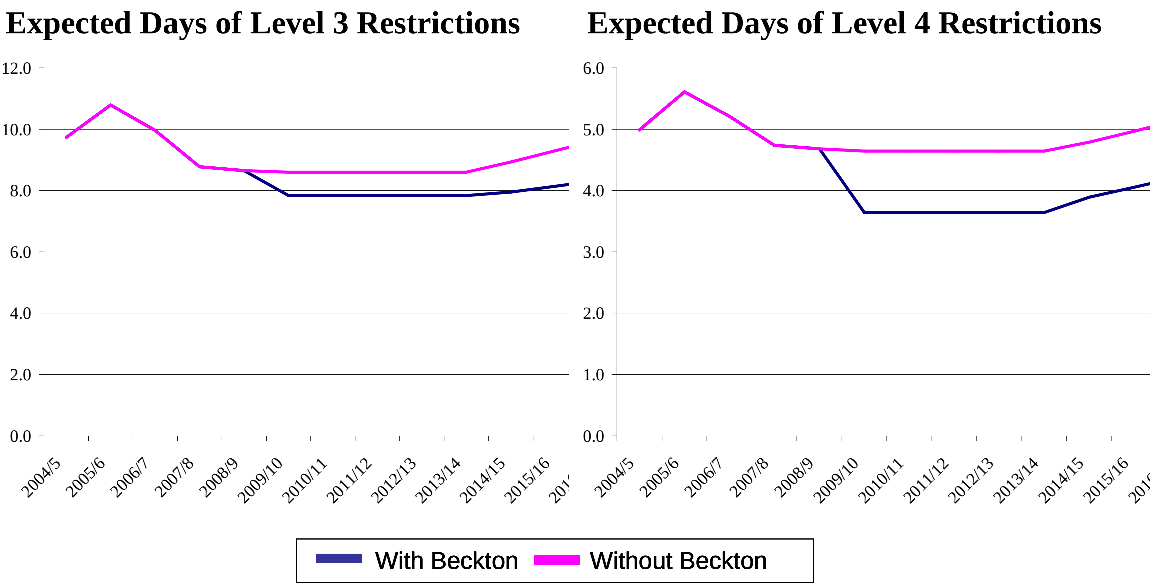
The contribution of the study was focussed on estimating the aggregate costs of additional water use restrictions resulting from not building the Beckton plant. The basis of our estimate was the difference in expected costs of restrictions between Thames Water's 2006 optimal asset strategy, which included the Beckton plant, and the expected costs of restrictions under the second best strategy which excluded the Beckton plant. This analysis provided an estimate of the reduction in the expected costs of water use restrictions associated with the Beckton plant or equivalent supply-demand balance improvements.

Estimates of the costs of restrictions to households or businesses resulting from not building the Beckton plant are calculated within the modeling framework outlined in section 4.3. Specifically, we calculate the present value of welfare losses due to restrictions: $\sum_t \int_{a_t^0}^{\infty} L_t(I, a_t) f_t(a_t)\, da_t$.

First, we separate out household and business costs and translate the above expression into the following more directly applicable formulation given our utility model specifications: $\sum_{t=0}^{T} \left(\frac{1}{1+r}\right)^t N_t \left(c_t^{L3} \Delta x_t^{L3} + c_t^{L4} \Delta x_t^{L4}\right)$, where $\left(\frac{1}{1+r}\right)^t$ is the discount factor used to bring future costs in year t into present value terms; N_t is the London population of households or businesses in year t; c_t^{L3} (c_t^{L4}), is the average willingness to pay of the London household or business population in year t for 1 expected day reduction of Level 3 (Level 4) restrictions; and Δx_t^{L3} (Δx_t^{L4}) is the difference in the expected number of days of Level 3 (Level 4) restrictions in year t between the cases where the Beckton plant is, and is not, included in Thames Water's asset strategy.

In 2006, Thames Water provided us with data on the expected numbers of days of restrictions at Levels 3 and 4 in each year for the next 20 years as a function of the assumed stock of assets in operation in each year. These data were derived by combining demand and supply forecasts as a function of aridity and assets in operation, converting supply shortfalls into numbers of days of restrictions at Levels 3 and 4, and calculating expected days by integrating expected days of restrictions at each level over an aridity probability distribution function based on 84 years of rainfall data.

Figure 4.3 plots the time series profiles of the expected numbers of days of restrictions at Levels 3 and 4, with and without the Beckton plant, based on the data supplied by Thames Water. The data are based on the assumption that the Beckton plant comes online in 2009.



Source: Analysis of data provided by Thames Water

Figure 4.3: Risk Profiles for Water Restrictions With and Without Beckton Plant

By applying the household and business valuation estimates from Table 4.5 to the reduction in supply restrictions that the Beckton water treatment plant would bring about in future years, and extrapolating the sample results to the full London population of households and businesses and summing over these, we estimate that London water customers value the increased reliability at £226 million in the first year of plant availability and about £3,521 million in present value terms over the life of the plant. This was many times the expected cost of the plant, which was estimated to be approximately £200 million. Partly as a consequence of the evidence obtained by this study, the planning inquiry overturned Mayor Livingstone's objection, and the Beckton plant was eventually constructed and began operations in June 2010.[3]

[3] Several sensitivity analyses were conducted in relation to these results. See NERA-Accent (2006), the technical report of this study, for details.

4.9 Concluding Remarks

This paper has presented estimates derived from an SP survey of the value of avoiding drought water use restrictions to households and businesses in London. Our analysis suggests that the survey instrument succeeded in eliciting meaningful statements of preferences from respondents, and that results are consistent with prior expectation, and with those from a comparable study [Willis, Scarpa and Acutt, 2005]. The findings indicate that customers attach a sizeable value to avoiding the most severe restrictions (including rota cuts to supply), but are much less concerned with lesser restrictions such as a hosepipe ban.

The principal output from the study was a quantitative model capable of providing welfare comparisons between asset strategies, given external data on the impact of those asset strategies on the expected numbers of days of restrictions over time. We applied our methodology and estimates to the appraisal of a desalination plant proposal in East London. The appraisal found that the benefits of the plant substantially exceeded the costs, and partly as a consequence, the plant was approved, and built, and began operating in June 2010.

Measures of WTP to avoid drought water use restrictions are useful in a range of contexts, not limited to the appraisal of a specific supply augmentation project. For example, the estimates presented in this paper were also used for water resource planning by Thames Water, and as evidence in an application for a drought order in June 2006 which would allow it to impose Level 3 restrictions in London, and thereby reduce the risk of needing Level 4 restrictions to curb demand later. (The application was subsequently withdrawn after the supply-demand balance improved considerably relative to expectation over July and August of that year.) It is well known amongst economists that scarcity-based pricing is a superior tool for rationing water during drought [Woo, 1994; Roibás, García-Valiňas and Wall, 2007; Grafton and Ward, 2008]. In many places however, including London, a majority of properties are not metered but rather are charged for water on an unmeasured basis. This precludes scarcity pricing, and so usage restrictions become the only means of rationing water. Measures of WTP to avoid drought water use restrictions are thus likely to continue to be useful despite the greater efficiency inherent in scarcity pricing.

5 The Sensitivity of Willingness to Pay to an Economic Downturn

Abstract

Stated preference valuation studies typically are performed at one point in time, with the results then used for decision making several months or even years later. This approach is only reliable if values are stable over time, an assumption which is doubtable given the onset of an economic downturn. We assess the reliability of values taken before an economic downturn for application during the downturn via analysis of near identical surveys conducted before, and during, the 2008-2010 economic recession. Each survey employed a dichotomous choice and a payment card contingent valuation question. Our main result is that the economic downturn led to lower willingness to pay when elicited via the payment card contingent valuation method, but had no effect on values elicited via a dichotomous choice (ie referendum-type) contingent valuation question. We explore potential explanations for this finding in light of the literature on closed-ended versus open-ended elicitation method comparisons.

5.1 Introduction

Stated preference valuation studies typically are performed at one point in time, with the results then used for decision making several months or even years later. This approach is only reliable if values are stable over time, or are predictably different based on observable covariates. Fortunately, the weight of evidence suggests that this is often the case. A number of studies have administered similar questionnaires to independent samples at two points in time, and found that the estimated values, or valuation function, remained unchanged [Brouwer, 2006; Brouwer and Bateman, 2005; Carson and Mitchell, 1993; Carson et al., 1997; Reiling et al., 1990; Whitehead and Hoban, 1999]; a second group of papers have performed a repeated survey on the same sample of respondents, and found reasonably high correlations between responses [Kealy, Montgomery and Dovidio, 1990; Loomis, 1990; McConnell, Strand and Valdes, 1998]. With one or two exceptions, the literature thus lends support to the application of values derived from historic contingent valuation surveys provided that reasonable adjustments are made for changes in observed determinants over the intervening period [Whitehead and Hoban, 1999].

There has been no study to date, however, which assesses the reliability of values taken before an economic downturn for application during the downturn. There are reasons to doubt whether WTP values, for e.g. environmental protection and improvement, remain valid following the onset of a recession. Even after controlling for current incomes, job security may be diminished, and concern for the environment and related policy areas may fall down the list of household priorities as a consequence. It is an open question whether these factors do indeed cause WTP values to fall, yet the answer has important implications for a wide range of policy applications.

The policy context in which the present study is situated is one such example. We conducted twin near-identical contingent valuation surveys of the customers of a large English water and sewerage company as part of the five-yearly regulatory price review process, one administered in June 2008 before the economic downturn, and one on a new sample conducted in June 2009, when the UK was deep in recession. Each survey included payment card (PC) and dichotomous choice (DC) contingent valuation (CV) methods to elicit WTP values. The data from these two surveys thus provide the opportunity to test and compare the sensitivities of both PC and DC WTP to an economic downturn. Only one previous study [Loomis, 1990] has assessed the comparative reliability of these alternative elicitation methods; thus this feature of the paper makes an additional contribution to the literature by providing this comparison in an important new context.

5.2 A Model to Assess Temporal Sensitivity of WTP

Willingness to pay is typically specified as a function of observed covariates. Partly, this is to demonstrate that WTP varies in line with expectation; partly it is to allow for a more accurate transfer of values from one site and/or time period to another. In the following, to lay out the framework in which we consider the sensitivity of WTP to an economic downturn, we focus on the distinction between observed and unobserved WTP covariates, ignoring the features of the good and study site as these stay the same.

Let WTP for individual *i* in time *t* be written as:

$$WTP_{it} = f(\boldsymbol{x}_{it}, \boldsymbol{\psi}_{it}; \boldsymbol{\zeta}) \tag{1}$$

where $\boldsymbol{x}_{it}$ is a vector of observed covariates, $\boldsymbol{\psi}_{it}$ is a vector of unobserved covariates and

$\boldsymbol{\zeta} = [\boldsymbol{\omega}^x, \boldsymbol{\omega}^\psi]$ is a vector of parameters. Note that $\boldsymbol{\zeta}$ is stable, that is, independent of *i* and *t*; all the variation over individuals and over time is captured by the two sets of variables $\boldsymbol{X}_t$ and $\boldsymbol{\Psi}_t$, where $\boldsymbol{X}_t = [\boldsymbol{x}_{1t}, \dots, \boldsymbol{x}_{Nt}]$, and $\boldsymbol{\Psi}_t = [\boldsymbol{\psi}_{1t}, \dots, \boldsymbol{\psi}_{Nt}]$.

Since $\boldsymbol{\psi}_{it}$ is unobserved, the following model is used as an approximation for estimation:

$$WTP_{it} = g(\boldsymbol{x}_{it}; \boldsymbol{\theta}_t) + \varepsilon_{it} \quad (2)$$

In (2), the unobserved covariates are no longer part of a deterministic function, and instead are captured by an error term, ε_{it}. Correspondingly, the functional form is changed from *f*(.) to *g*(.), and the associated parameter vector changes from $\boldsymbol{\zeta}$ to $\boldsymbol{\theta}_t$.

Estimation in time *t* typically relies on the identifying assumption that $E_i(\varepsilon_{it}|\boldsymbol{x}_{it}) = 0$. This is the case, for instance, when using OLS, tobit, logit, probit, or interval models, which are those most commonly employed to estimate valuation functions. The identifying assumption is generally invalid, however, if $\boldsymbol{X}_t$ and $\boldsymbol{\Psi}_t$ are correlated. Any correlation between the observed and unobserved covariates of WTP will cause the parameter vector $\boldsymbol{\theta}_t$ to be biased. Moreover, since the size of the coefficient bias depends on the unobserved data, and since this varies from year to year, the bias will itself vary from year to year. Only if the coefficients are unbiased, or if there is no substantial change in unobserved covariates, will the parameter vector stay stable from year to year.

In line with the terminology above, we assess the temporal reliability of WTP via the testing of two hypotheses:

(H1) $E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_1\right) = E\left(\overline{WTP}_2 \middle| \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_2\right)$, and

(H2) $E(\boldsymbol{\theta}_2 | \boldsymbol{X}_2) = E(\boldsymbol{\theta}_1 | \boldsymbol{X}_2)$

The first of these hypotheses states that average WTP is predictable given new data on observed covariates of WTP, but using a previously estimated model. The second hypothesis makes the stronger claim which is that the predictive model is stable over time. Given estimates of $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$, these hypotheses may be straightforwardly tested by standard statistical methods. In section 5.4 we discuss estimation methods. We discuss the tests employed, and their results in section 5.5.

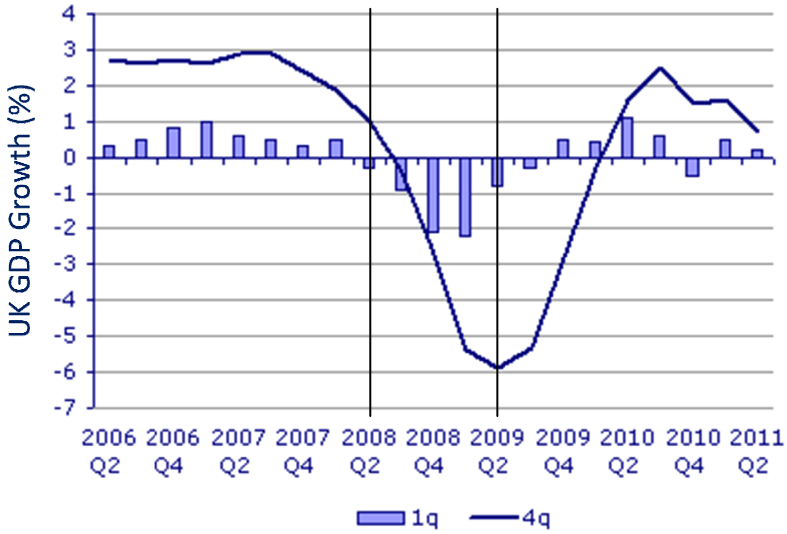
5.3 Survey Design, Administration and Data

Thames Water (TW) is the largest water and wastewater services company in the UK supplying 8.8 million water customers and 14 million wastewater customers in London and the South East of England. In June 2008, we implemented a survey to assess its household customers' WTP for the improvements in water and wastewater service due to TW's draft business plan for 2010-2015, and in June 2009 we used a very similar questionnaire to assess household customers' WTP for the slightly revised set of improvements in TW's final business plan, both of which were submitted to the economic regulator Ofwat as part of its five-yearly price review process for the England and Wales water sector. Our analysis suggests that customers are likely to view the extent of both sets of service improvements as about as sizeable as each other, and hence from here on we refer to them both as simply "TW's plan". The appendix to this paper contains a table showing the details of current service levels (as stated in the 2008 and 2009 surveys) and the service levels offered as a consequence of TW's 2008 and

2009 investment plans.[4] The recruitment method, introductory questions, valuation statement and elicitation methods were the same across both questionnaires.

The dates over which effects should be most effectively measured can be debated. For example, in June 2008, although a recession had not yet been declared, there were already some warning signs of economic troubles ahead which could have influenced WTP responses at the time. On the other hand, in June 2009 unemployment had not yet reached its peak and so there is also a case to be argued that this later date may not capture the full impact of the recession on WTP. Both arguments would tend to suggest that any effect we estimate, in respect of the sensitivity of WTP to an economic downturn, is a lower bound. Ideally, for the purposes of this research, three or four surveys would be conducted to track changes in WTP over the full course of the economic downturn, a prospect which was unfortunately not feasible. As Figure 5.1 shows, however, the survey dates are sufficiently situated in the economic cycle to have a good chance of capturing the effects we seek to examine.

[4] We calculated the difference between the draft business plan (DBP) and final business plan (FBP) service improvement measures for each attribute, and used these to derive an index for the FBP based on the DBP and the current service level. If all proportional attribute improvements were given equal weighting by respondents, this approach determines that the FBP would imply "1% more" improvement than the DBP; i.e. probably a fairly trivial difference from the perception of respondents. Ideally we would use weights which match the relative values of the attributes rather than constant weights; however determining these weights is beyond the scope of this study.



Note: vertical lines represent survey dates

Figure 5.1: UK Gross Domestic Product Growth, 2006-2011

Our method of examining separate samples has the advantage over a repeated survey on the same sample in that it eliminates any potential for recall bias, wherein the respondent remembers his original responses and simply repeats his answers in the second survey. A disadvantage is that only differences in population statistics, e.g. mean and median, can be compared rather than individual comparisons. Since population statistics are usually all that are needed for policy applications, and since these can be compared robustly using standard statistical methods, we do not consider this a significant limitation.

The questionnaires each included a dichotomous choice (DC) contingent valuation question followed by a payment card (PC) question to elicit WTP for TW's plan. The payment vehicle was the annual water and wastewater bill increase; the levels for the DC question were drawn from the range {£5, £10, £20, £50, £100}; the payment card contained 30 numbers ranging from £0-£3000 on an approximately logarithmic scale. Many studies have found that DC values exceed those obtained by open-ended formats such as the PC approach [Cameron et al., 2002; Welsh and Poe, 1998] to the extent that this is considered a ‘stylized fact’ of the CV approach [Carson and Groves, 2007]. Loomis (1990) is the only previous study, however, to have compared empirically the intertemporal reliability of alternative elicitation methods. It resurveys the same sample nine months after the original survey, asking DC and open-ended (OE) CV questions on each occasion, and finds the correlation between responses to be around 0.6 for both elicitation methods. Given the similarity of OE and PC formats, we take this result as our prior that, in the absence of any further considerations, we would expect PC and DC to be equally sensitive or insensitive to an economic downturn. We test this assumption as part of our analysis.

The surveys in 2008 and 2009 were administered face-to-face by Accent Market Research using the Computer Aided Personal Interview (CAPI) method. Each sample was stratified to include representative proportions of respondents in London, in rural areas, and in urban areas outside of London, with an average of 20 interviews per location to ensure a dispersed sample. The average interview time was less than 30 minutes, and very few interviews took more than 40 minutes. The interviewers’ comments on and scoring of respondents suggest that they understood the survey well, maintained a good degree of focus, and gave the questions careful consideration. Almost universally respondents replied to a follow-up question by stating that the cost, and/or the value to them of the service improvements, was the reason for their WTP answers. A fairly low proportion of the sample (9%) were excluded due to giving inadmissible responses to either the DC or PC questions. This comprised a mix of protest cases, refusals or “don’t know” responses. A further 13% of the sample were excluded due to their failing to answer the income question. The final sample sizes are 257 for the 2008 survey and 275 for the 2009 survey.

A summary of the respondent characteristics in the 2008 and 2009 surveys is presented in Table 5.1, alongside indicative population counterparts. The samples are broadly comparable, although the 2009 sample is somewhat older, better educated, higher earning and less likely to be a member of an environmental club. In respect of environmental club membership, this may be due to a decline in membership in the population rather than differences in sample composition – we are unable to confirm this either way. Population values in most cases are unlikely to be fully reliable due to the length of time since the UK census was conducted (2001). The exception to this rule is the case of income data for the London and South East region, which are drawn from the annual Family Resources Survey (FRS) for the relevant years. Based on a large-scale UK government survey, the FRS data offer a reliable picture of how household finances changed in the UK between 2008 and 2009. As Table 5.1 shows, nominal earnings appear to have risen slightly, despite the onset of a recession. This is not altogether surprising since earnings, and employment, tend to lag behind output in the economic cycle. The small positive shift in the income distribution is reflected in the difference between the 2008 and 2009 samples that we obtained, however overall there are significantly more low income respondents in our sample than in the population, and correspondingly fewer earning high incomes. To correct for this we adjust the sample observations with weights so that the analytical results reflect the income distribution of the population of household customers. This also ensures that the difference in income between the two samples, when weighted, matches the difference in income for the population. For our analysis, we also deflate 2009 income data, PC WTP and DC cost levels to 2008 prices using the Consumer Price Index (CPI) in order that the data and all reported results are comparable in real terms.

5.4 Empirical Methods

We analyze the data obtained from the survey as follows. First we combine the DC and PC responses using a single estimation technique - interval censored regression - and estimate this separately using the 2008 and 2009 samples. Interval frameworks are well suited to representing both DC and PC responses. Cameron and Huppert (1989, 1991) have argued that the language of a PC question lends itself to an interval interpretation, with WTP lying between the amount indicated and the next highest amount labelled on the card. Interval frameworks have also long been used to represent DC responses [Carson and Hanemann, 2005] with a no response indicating that WTP lies between zero and the amount asked and a yes response indicating that WTP lies between the amount asked and an upper bound reflecting financial resources. To be conservative, we use an upper bound of £500 for the interval when a respondent said yes to the DC question, which is substantially higher than the largest amount used (£100). This does not rule out the possibility that larger WTP values are held by respondents, only that they were not observed in either our PC or DC data.

Table 5.1: Sample and Population Characteristics

	Population[1] (%)	2008 Sample (%)	2009 Sample (%)
Gender (2)			
Male	48.6	48.3	50.2
Female	51.4	51.8	49.8
Age (2)			
18-29	21.6	23.7	20.4
30-44	31.0	35.0	35.6
45-59	23.0	27.6	21.1
60-64	5.7	5.5	7.6
65-74	9.7	5.5	10.2
75+	9.0	2.7	5.1
Education (2)			
Primary	25.4	14.5	12.3
1-5 GCSEs/O-levels	16.1	25.4	18.4
5+ GCSEs/O-levels	20.5	13.3	15.7
2+ A-levels or NVQ3	10.1	15.3	17.6
First degree or higher	27.9	31.5	36.0
Employment Status (3)			
Working full-time (31+ hours)	42.9	47.6	46.0
Working part-time (<30 hours)	10.5	14.4	15.3
Self employed	9.3	4.8	7.7
Working and full-time student	2.8	1.6	2.3
Not working – seeking work	3.3	1.2	3.8
Not working – Full time student	5.3	4.4	2.7
Not working – retired	11.7	8.0	13.4
Not working – looking after home/family	6.8	10.8	6.9
Other	7.3	7.2	1.9
Weekly household income (4)			
Low (<£300)	22.3; 20.0	42.8	38.2
Medium (£300-£1000)	50.3; 52.4	45.1	44.4
High (>£1000)	27.4; 27.9	12.1	17.5
Environmental club membership (5)		19.8	15.3

Notes: N = 257 (2008 survey); N=275 (2009 survey). Base for each statistic includes the full sample unless indicated otherwise. (1) All population statistics are for the London and South East Government Office Regions combined. This region encompasses, and is somewhat broader than, the Thames Water supply area. (2) Source: Census (2001); (3) Source: Census (2001) (population aged between 16 and 74); (4) Source: Family Resource Survey (FRS); the first number in each pair is sourced from FRS (2008-09), representing the 12 months to March 2009; the second number in each pair is sourced from FRS (2009-10), representing the 12 months to March 2010; no adjustments have been made for inflation or other factors. (5) No population statistics available for environmental club membership in the region.

The interval censored framework is straightforward to implement in a maximum likelihood context. Let y_n be our interval censored variable, which we model as a linear function of explanatory variables x_n plus an i.i.d. error term ε_n with mean zero and variance σ^2. Then we have:

(3) $$\text{Prob}(y_n) = F\left(\frac{y_n^U - \boldsymbol{x_n\beta}}{\sigma}\right) - F\left(\frac{y_n^L - \boldsymbol{x_n\beta}}{\sigma}\right)$$

which implies the following log-likelihood:

(4) $$\text{LL} = \sum_n \log[\text{Prob}(\text{y}_\text{n})]$$

A distributional assumption is required for $F(.)$ to implement the estimation. We chose the log-normal because it ensures that WTP is non-negative (a problem with the normal) and it is straightforward to implement. Since the lower bound for some intervals is zero, the number "1" was added to all lower and upper bound values before taking logs because the log of zero is undefined. This "1" was then subtracted in obtaining later estimates for mean and median WTP. In the panel context, where for each person, *n*, we have a PC and a DC response, indexed by *t*, we thus let y_{nt} = log(1+WTP_{nt}) and define lower and upper bounds accordingly, where WTP_{nt} is the willingness to pay by respondent *n*, as elicited by question type *t* ($t \in$ {PC, DC}). $F(.)$ is then simply the standard normal cumulative distribution.

The log likelihood in (4) is based on the assumption that error terms are independent of one another. Independence is unlikely, however, when responses to both PC and DC questions are combined. To take account of within-person correlation between responses, we also estimate a random effects panel version of the above model which involves decomposing the error term into an individual specific effect, u_n, assumed to be normally distributed with mean zero and variance σ^2_u and an i.i.d. normal variate with mean zero and variance $\sigma^2_{e.}$ Estimation is performed using the *xtintreg* command in Stata (version 11), and details of the methods and formulae can be found in StataCorp (2009).

5.5 Results

We begin by presenting the (weighted) response distributions for the PC and DC questions in 2008 and 2009 surveys. Consistent with the results of previous studies [e.g. Welsh and Poe , 1998], Figure 5.2 shows that the DC distribution lies above the PC distribution at all cost amounts for each year, except at the £5 level for the 2008 sample. Comparing across years, we see that the PC response distribution for 2009 lies below the 2008 distribution across the entire support, whereas for the DC responses there is no clear systematic difference. To examine this further we turn to presentation of our interval models, from which we can derive comparable estimates of mean and median WTP for PC and DC methods for 2008 and 2009, and the standard errors around these estimates that allow for statistical testing of the differences between them.

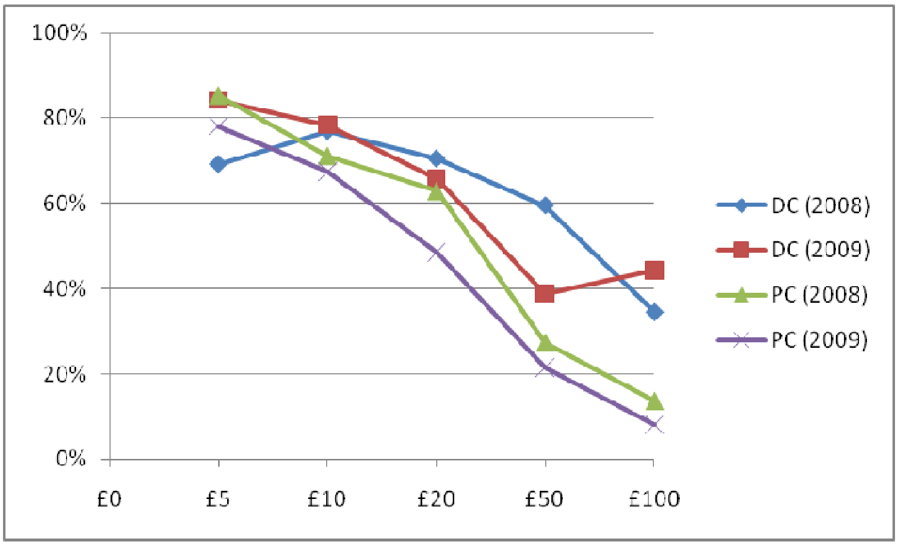


Figure 5.2: PC and DC Response Distributions in 2008 and 2009. Cumulative response frequencies offering at or above the WTP indicated amount, linearly interpolated between the DC levels used

Results from the interval models are presented in Table 5.2. The first model is for the 2008 survey sample. In this model, as anticipated, we see a significant ($p<.01$) negative coefficient on *Payment card*. The value of -0.381 indicates that PC WTP is around 32% lower than DC WTP all else equal. Turning to respondent covariates, income is positively associated with WTP ($p<.01$), again as expected. It enters in log form and so the coefficient on *Log income* is an elasticity; hence, the coefficient of 0.509 implies that a 10% increase in income is associated with a 5% increase in WTP. Membership of an environmental club enters the model as a dummy variable, with a positive coefficient ($p<.05$), and via an interaction with Log income which has a negative coefficient (p<05). The combination of these two coefficients indicates that members of environmental clubs tended to have higher WTP than non-members except for the highest income respondents. The parameter σ_u is the standard deviation of the random effects. The fact that this is significant ($p<.01$) indicates that the random effects are themselves jointly significant. Consistent with this finding, the coefficient on ρ is 0.416, which indicates that 41.6% of the error variance is accounted for by the random effects. This evidence provides strong support for the use of the random effects interval model, rather than the simpler pooled model which assumes independence of individuals' errors across the two elicitation methods.

In comparison with the 2008 model, all the coefficients in the 2009 model seem very different, suggesting a lack of transferability of the full 2008 combined PC and DC interval valuation function for use during the recession of the following year. The coefficient on Payment card is -0.708 in the 2009 model which is lower than in the 2008 model. Whereas in the 2008 model, PC WTP is around 32% lower than DC WTP all else equal; in the 2009 model PC WTP is around 52% lower. The income elasticity is also much lower in the 2009 model than in 2008, at least for those that are not members of environmental clubs. The effect of club membership generally, as a function of income, is very different in 2009 than in 2008 which suggests that the original combined PC and DC function was not particularly reliable. The one aspect of the original combined function which does remain stable is the error distribution, as measured by σ_u, σ_e and ρ. Thus the shape of the distribution, if not its conditional means and medians, remains stable despite the onset of the economic downturn.

Table 5.2: Interval Censored Models Combining DCCV and PCCV Responses

Variable	2008[a,b,c,d] Coef (Std Err)	2009[a,b,c,d] Coef (Std Err)
Constant	0.351 (0.560)	2.595 (0.512)***
Payment card	-0.381 (0.101)***	-0.708 (0.098)***
Log income	0.509 (0.089)***	0.138 (0.081)*
Club	2.888 (1.334)**	0.120 (1.352)
*Club*Log income*	-0.436 (0.205)**	0.084 (0.202)
σ_u	0.754 (0.076)***	0.765 (0.073)***
σ_e	0.893 (0.056)***	0.895 (0.054)***
ρ	0.416	0.422
Observations	514	550
Log Likelihood	-841.042	-872.740
Pseudo R^2	0.047	0.054

*Notes: **a** Results are weighted for income based on the UK Family Resources Survey for the relevant year. **b** All models are interval censored regressions allowing for within person correlation. The left hand side for each model is the pair {ly1,ly2},where ly1 is the log of one plus the lower bound of WTP and ly2 is the log of one plus the upper bound of WTP. **c** Standard errors are robust, calculated using the Huber-White estimator [White, 1980]; **d** Stars indicate p-value for 2-side t test:* p < 0.10, ** p < 0.05 *** p < 0.01.*

As set out in section 5.2, we assess the temporal reliability of WTP via the testing of the following two hypotheses:

(H1) $E(\overline{WTP}_2 | \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_1) = E(\overline{WTP}_2 | \boldsymbol{X}_2, \widehat{\boldsymbol{\theta}}_2)$, and

(H2) $E(\boldsymbol{\theta}_2 | \boldsymbol{X}_2) = E(\boldsymbol{\theta}_1 | \boldsymbol{X}_2)$

Hypothesis H1 states that predicted mean WTP in 2009 using new data ($\boldsymbol{X}_2$) but the original 2008 model ($\widehat{\boldsymbol{\theta}}_1$) is equal to our best estimate of actual mean WTP in 2009 based on both new data ($\boldsymbol{X}_2$) and a new model ($\widehat{\boldsymbol{\theta}}_2$). The second hypothesis, H2, makes the stronger claim that the predictive model is stable over time.

To test the stability of the 2008 valuation function (H2) we perform a Likelihood Ratio (LR) test to directly examine the suitability of the 2008 model coefficients for use in 2009. The 2009 equation presented in Table 5.2 is treated as the unrestricted model, and an equation also estimated on the 2009 sample but fixing all coefficients at the levels of the 2008 model, is treated as the restricted model. This LR test rejects the null hypothesis of transferrable coefficients ($p<.01$), hence the combined PC and DC 2008 model is not transferrable to 2009. This finding is consistent with the readily seen differences between 2008 and 2009 models shown in Table 5.2.

The test of model stability is stronger than is usually necessary for cost-benefit analysis. In most cases, estimates of mean and median WTP are all that are required for policy applications. This is the motivation for the test of hypothesis (H1) – which states that predicted 2009 mean WTP from 2008 model coefficients is equal to predicted 2009 mean WTP from 2009 model coefficients. Given the functional form of the model, and letting $\widehat{\boldsymbol{\theta}}_t = [\widehat{\boldsymbol{\beta}}_t, \hat{\sigma}_t]$ where $\widehat{\boldsymbol{\beta}}_t$ is the vector of coefficient estimates for time *t* and $\hat{\sigma}_t$ is the estimate of $(\sigma_u + \sigma_v)$ for time *t*, we can write:

$$E[\overline{WTP}_t | \boldsymbol{X_t}, \widehat{\boldsymbol{\theta}}_t] = E_i[(WTP_{it} | \boldsymbol{x}_{it}, \widehat{\boldsymbol{\theta}}_t)] = \exp\left(\widehat{\boldsymbol{\beta}}_t{}' \boldsymbol{X}_t + \frac{\hat{\sigma}_t^2}{2}\right) = g(\boldsymbol{X}_t, \widehat{\boldsymbol{\theta}}_t) \quad (5)$$

Then, following [*Whitehead and Hoban*, 1999], let the difference in WTP across time be

$$\Delta WTP_t = WTP_2 - WTP_1 \quad (6)$$

$$= g(\boldsymbol{X}_2, \boldsymbol{\theta}_2) - g(\boldsymbol{X}_1, \boldsymbol{\theta}_1) \quad (7)$$

$$= [g(\boldsymbol{X}_2, \boldsymbol{\theta}_2) - g(\boldsymbol{X}_2, \boldsymbol{\theta}_1)] + [g(\boldsymbol{X}_2, \boldsymbol{\theta}_1) - g(\boldsymbol{X}_1, \boldsymbol{\theta}_1)] \quad (8)$$

Table 5.3 presents this decomposition of WTP for the PC and DC predictions based on the estimated parameter vectors $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$, and the observed data $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The estimates are obtained as follows. First, the predicted value of ln(1+WTP) is calculated for each member of the sample, conditional on the treatment pertaining to the cell shown in the table. For the PC WTP values, *Payment card* is set equal to 1; for the DC WTP values it is set equal to zero. Mean WTP is then calculated as exp(μ +0.5(σ^2_u+ σ^2_e))-1, where μ is the sample average of predicted ln(1+WTP); and σ^2_u and σ^2_e are as shown in Table 5.2.

Looking first at the PC results, the table shows that mean 2008 PC WTP - that is, predicted mean WTP using the 2008 model parameter vector and the 2008 data - was £46.1 per household per year. In 2009, mean PC WTP fell to £34.0. Changes in observable determinants (***X***) caused a fall of £0.41 in PC WTP, although this difference is not significantly different from zero. The remaining £11.7 difference was caused by unobserved factors, and this difference is significantly different from zero ($p<.01$).

Now turning to the DC WTP results, we see that 2008 mean WTP was £68.0 per household per year, and in 2009 this rose to £69.9. Neither the difference attributable to changes in observed determinants (-£0.6), nor the difference attributable to changes in unobserved factors (£2.58) is statistically significant ($p>.10$). The implication of these results is that PC WTP is sensitive to the economic downturn but DC WTP is not.

Table 5.3: Decomposition of Willingness to Pay, by Year of Data, Year of Estimated Parameter Vector and Elicitation Method

Year of Data (X)	Elicitation Method and Year of Estimated Parameter Vector ($\hat{\boldsymbol{\theta}}$)			
	PCCV		DCCV	
	2008 ($\hat{\boldsymbol{\theta}}_1$)	2009 ($\hat{\boldsymbol{\theta}}_2$)	2008 ($\hat{\boldsymbol{\theta}}_1$)	2009 ($\hat{\boldsymbol{\theta}}_2$)
2008 (X_1)	46.1		68.0	
	(3.55)***		(5.20)***	
2009 (X_2)	45.7	34.0	67.4	69.9
	(3.40)***	(2.57)***	(4.98)***	(5.22)***
$g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_2) - g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_1)$	-11.7		2.58	
	4.27***		7.22	
$g(\boldsymbol{X}_2, \hat{\boldsymbol{\theta}}_1) - g(\boldsymbol{X}_1, \hat{\boldsymbol{\theta}}_1)$	-0.41		-0.60	
	4.92		7.20	

*Notes: Standard errors in parenthesis; standard errors are calculated using the delta method [*Greene*, 2003]; stars indicate p-value for 2-side t test:* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$.*

5.6 Discussion

The main findings are the following: (1) DC WTP is significantly higher than PC WTP in both survey samples; (2) the combined PC and DC valuation function as a whole was not found to be transferrable from 2008 to 2009; and (3) the onset of an economic downturn caused PC WTP to fall, while DC WTP remained unchanged. Finding (1) is consistent with the majority of the large number of studies that have compared DC and PC responses, as summarised in Champ and Bishop (2006). The second finding gives cause for concern in using a combined valuation function derived before a recession during a recession, but it is not a sufficient finding to warrant disregard of predicted population mean WTP. Indeed, finding (3) suggests that if you believe that DC WTP is the truth, then it is valid to predict mean WTP using a pre-recession valuation function, or just transfer the mean itself. We therefore focus our discussion on the implications of finding (3).

In light of the framework set out in section 5.2, we can infer from the findings that unobserved features of the downturn affected the PC responses but not DC responses. The principal unobserved features potentially affecting WTP are, we hypothesize, diminished job security and a less certain future income – current incomes are, we have seen, not substantially different between years in our sample. We may now explore the consistency of these factors and the observed finding of no change in DC WTP but a fall in PC WTP with explanations given in the literature concerning the "stylized fact" that PC and OE responses are typically lower, sometimes much lower, than estimates of WTP generated from DC responses.

A prominent view in the literature [Carson and Groves, 2007] explains the observed PC<DC relationship with reference to strategic response considerations. It is argued that the DC method is compatible with truth-telling provided certain stringent auxiliary conditions are met, namely that the DC question is asked before any other elicitation question, that the survey is constructed so as to convey the idea that respondents' answers will have a consequential impact on policy, and that respondents believe the scenario as presented to them, including the scope of the improvements and the cost they, and others, will have to pay. All three of these properties hold for the present study, and so it may be argued from this perspective that the DC WTP estimate is the truth. By contrast, under plausible belief structures – such as that the go/no go policy decision rule depends on summing respondents' stated PC WTP amounts, and that an individual's stated WTP amount is weakly correlated with the amount they will be required to pay should the policy be implemented – the PC method provides an incentive for respondents to understate their true WTP, either to minimize the chance that the policy goes ahead – stating a WTP of £0 when the cost is expected to be greater than true WTP – or to minimize the expected payment, by stating a WTP of the expected cost - sometimes rationalized as a "fair amount" - when the cost is expected to

be less than true WTP. Strategic considerations are thus predicted to cause respondents to understate their true WTP when offered the opportunity to do so.

For this view of the CV response process to hold, there would need to have been some change in incentives, or there must be some feature of a recession that causes respondent to become more strategically minded. The former condition can only be true if expectations of the true cost of the investment program had changed. Since there is no difference in the information given in the survey, it is unlikely that cost expectations could have changed between surveys. On the other hand, it is plausible that increased job/income insecurity might invoke a greater willingness to engage in strategic response behaviour. Unfortunately, however, we are not able to test this hypothesis with our dataset.

A different perspective suggests that the observed difference between PC and DC WTP is due to the certainty of respondents about their true WTP when they answer the questions [Ready, Navrud and Dubourg, 2001]. This view is backed up by supporting empirical evidence showing, firstly, that respondents are indeed less certain about their DC responses than they are about their PC and OE responses [Ready, Navrud and Dubourg, 2001; Welsh and Poe, 1998], and that fixing certainty levels resolves the discrepancy [Ready, Navrud and Dubourg, 2001; Welsh and Poe, 1998],. To be consistent with this perspective, there would need to be some feature of a recession that caused respondents to become less certain of their true WTP. This seems plausible to us, in that job insecurity might readily diminish certainty over WTP. This could cause there to be a wider uncertainty range, with a lower level of “certain” willingness to pay, but with no different a level at the top end of the range where respondents are “not sure” whether they would pay or not. Our results are thus also consistent with this explanation of the DC-PC difference.

5.7 Conclusions

Our main finding is that the recession caused PC WTP to fall, whereas DC WTP stayed the same. This result is statistically robust, hence the finding is probably not due to sampling variation. The principal explanations for the common finding that DC WTP>PC WTP – strategic behaviour and respondent uncertainty - are both potentially consistent with this result and hence we cannot say for sure why the recession caused PC WTP to fall, while leaving DC WTP unchanged. Consequently, since both perspectives yield differing conclusions regarding which is more trustworthy as a measure of true WTP, we cannot say for sure whether true WTP itself is sensitive to an economic downturn. Until future research addresses this uncertainty, researchers are encouraged to interpret our main finding in line with their own views on which theory correctly explains the DC>PC pattern.

6 Critical Discussion

In this chapter we summarise and discuss the results obtained from the three core studies presented in this thesis. Section 6.1 discusses the methods used to analyse the data; section 6.2 contains a discussion of core policy results; section 6.3 then discusses our findings in respect of treatment effects.

6.1 Discussion of Analysis Methods

The empirical analysis undertaken for this thesis has centred on the use of two techniques: the panel interval regression method, used to simultaneously analyse PC and DC CV response data in the empirical studies of chapters 3 and 5; and the conditional logit estimator, used to analyse the DCE data in chapters 3 and 4. In the following, I give my rationale for using these techniques and discuss their limitations.

6.1.1 Panel interval regression

The interval regression estimator has been widely applied to the analysis of CV data, both to PC and to DC responses. The studies in chapters 3 and 5 of this thesis, however, are the first to have applied a panel version of this estimator, which models PC and DC responses jointly. The core advantage of modeling the response data jointly is that it is possible to directly examine, and disentangle, the effects of elicitation technique. This is done on the basis of a null, or default, assumption that both techniques measure a latent “true” WTP value, and that the sources of variance in both sets of response data are the same, except for the treatment effects to be estimated and tested. Testing then proceeds by examining conditional mean differences in WTP due to the question type indicator, and differences in the conditional mean effects of other treatment variables, such as the scope of improvement, also due to the question type indicator. Thus covariate effects that are potentially confounding with the treatment effects are controlled for in an efficient way – allowing for joint estimation where we find no statistically significant difference in the influence of the variable due to the elicitation treatment. An additional benefit of the approach is that, under the null assumption, the efficiency of the estimates of scope effects, and covariate effects, on latent WTP is superior than could be obtained from using either one set of responses individually. This is simply because of the sample size is doubled in the panel model.

The principal limitation of the panel interval regression approach to modeling PC and DC response data is due to the well-known sensitivity of DCCV WTP results to the assumed shape of the latent WTP distribution. In the analysis presented in chapters 3 and 5, a log normal distribution was chosen. This distribution generally fits to PC and DC data reasonably well. Comparison with the Turnbull non-parametric lower bound estimator of mean WTP, however, suggested that the interval regression estimate of DC WTP may be somewhat conservative. Since it is usually regarded as desirable to adopt a conservative stance when estimating WTP for policy appraisal, following Arrow et al. (1993), this feature of the estimator is not altogether undesirable.

6.1.2 Conditional logit estimation

The conditional logit estimator [McFadden, 1974] is a well established tool for discrete choice analysis. Indeed it is by far the most common method used for such analysis [Train, 2009]. The core feature that distinguishes it from alternative methods such as nested logit, or mixed logit, is that it is based on the assumption of Independence of Irrelevant Alternatives (IIA). This assumption is equivalent to the statement that the unobserved components of utility are uncorrelated across alternatives.

The IIA assumption is potentially overly restrictive in some contexts, as it can result in unrealistic substitution patterns. For the purposes of DCE analysis of policy alternatives, however, as undertaken in the empirical chapters of this thesis, I would argue that the assumption is reasonable. This is because, in contrast to many of the cases in reference to which the assumption has been criticized, eg transport mode choice, differences between alternatives in a policy-focused DCE are fully captured by the levels of the hedonic attributes used to describe them. This means that there is no unobserved alternative-specific component to utility that could be correlated across alternatives.

A limitation of the conditional logit estimator, in comparison with mixed logit alternatives, is that it is only able to model heterogeneity in utility parameters as a function of observed covariates. Where mean values and possibly some limited segment analysis are all that are needed for policy appraisal, as they often are, this is not a serious limitation of the technique. There are potential advantages to be gained, however, from statistically exploring the heterogeneity of values across the population, based on unobserved as well as observed person-specific variance. Such analysis might provide important insights about the preferences of the target population. Future research, some of which is already underway, will explore the heterogeneity of values in relation to water quality and water supply reliability more fully using mixed logit techniques.

6.2 Discussion of Core Policy Results

6.2.1 The benefits of water quality improvements

In Study 1 (chapter 3), we estimated the value to households in England and Wales of improvements to the quality of water in the natural environment. The need for value estimates arose from the European Community Water Framework Directive (WFD), which drives water policy across the European Union. Area based values were generated to maximise the potential for subsequent policy incorporation and value transfer. These were found to vary from £2,263 to £39,168 per km^2 depending on the population density around the location of the improvement, the ecological scope of that improvement, and the value elicitation method employed.

The results were obtained from a carefully designed and well tested SP instrument implemented using a large in-person sample, and there is good supporting evidence to validate the results, both in terms of content validity and construct validity. A survey is said to have high content validity if: 'the survey descriptions and questions are clear, reasonable and unbiased, … [such] that respondents are put in a frame of mind that motivates them to answer seriously and thoughtfully' [Schumann, 1996, p.77]. Construct validity is indicated by whether or not the results vary across the sample data in line with expectations and prior research.

From a content validity perspective, our analysis suggests that the questionnaire succeeded in eliciting meaningful statements of preferences from respondents. Firstly, the vast majority of respondents answered the valuation questions as intended, with only a small proportion (6%) removed as protests or outliers. Additionally, the interviewers found good levels of understanding and attention were given to the questions. From the verbatim responses to a debriefing following the PC CV question, we identified only 8/1,389 respondents who indicated that they didn't believe the improvements would occur. This constitutes only 0.6% of the sample, which we take as evidence that disbelief in the scenario was not widely held. In addition to this evidence, we found that not a single person during the extensive pre-testing process expressed any doubt that the improvements would take place as described.

Supporting evidence of the construct validity of the results comes from analysis of the variation in WTP amounts across the sample. For each elicitation method, the WTP measures were found to vary as expected. Values were higher for greater amounts of improvement, based on both within-respondent and between-respondent scope comparisons. Values were higher, per hectare, for local versus non-local improvements, and for earlier versus later improvements. Comparing across respondents, WTP varied as expected with use of the water environment; attitudes towards paying for environmental protection and improvement; and with income. Against all these validity indicators, the results thus perform well, leading us to conclude that they are meaningful measures of the benefits of water quality improvements to households.

The results are, however, limited in three important ways. Firstly, the decision to focus on programs of improvements rather than on individually specified improvements meant that no information was given to respondents regarding which areas were to be improved except insofar as they were to be made in the local area, i.e. within 30 miles, or elsewhere. A second limitation of the results is that they only provide values for broad ranges of improvement. It is not strictly possible, for example, to use the results to value an improvement from Poor to Moderate ecological status because both status categories are embedded within the Medium quality level. The final limitation of the results is that the range of estimates reported with respect to elicitation treatments may be too wide for some policy purposes.

These limitations suggest some potentially fruitful avenues for future research. Firstly, a new study might attempt to explore how valuations vary across the population for different types of water body. The key factors to take into consideration in relation to the type of water body would be those pertaining to the recreational value of the site - based on the accessibility of the site, and the presence of complementary facilities and substitute recreational sites in the vicinity – and those pertaining to its value in the context of the natural environment – for example, whether it was indicated as a Site of Special Scientific Interest. Additionally, future research could explore how values vary according to the type of improvement. For example, it would be useful to investigate the relative importance of improvements to flow levels versus improvements to water quality. Here, perhaps it might be more worthwhile focusing on individual sites rather than on the whole country, in order to pay close attention to the qualitative change in question, rather than the spatial extent of the change.

6.2.2 The cost of drought water use restrictions

Chapter 4 investigated the cost of drought water use restrictions to households and businesses in London. The primary purpose of the welfare estimates was to contribute to an economic appraisal of the benefits of the UK's first desalination plant; however, the study was designed for maximum transferability of values with future applications in mind. The focus on transferability led to the use of a novel measure: the value of avoiding one expected day of water use restrictions per year. This measure combines the probability and duration of restrictions in an economically meaningful manner. It is equivalent, for example, to the reduction in the risk of a 30 day restriction event from

1/30 to 0. Households were found to value the avoidance of one expected day of severe restrictions at £53 per household per year; and of one expected day of lesser restrictions at £2 per household per year. The comparable results for businesses were, on average, £845 per business per year to avoid one expected day of severe restrictions and £48 per business per year to avoid one expected day of lesser restrictions.

A variety of evidence was examined to appraise the content and construct validity of these results. Content validity evidence was obtained by examining follow-up responses at two stages in the survey: after the initial example choice question, and after the completion of the DCE exercise. In both cases, the vast majority of households and businesses were able to provide articulate and rational reasons for their choice responses, which indicated that they had understood the questions well and responded meaningfully. There is also evidence of construct validity. The signs and magnitudes of the WTP measures are consistent with prior expectation, and WTP was found to vary with income and business size in line with expectation. Furthermore, results are consistent with the evidence from external studies; in particular the most closely comparable study - Willis, Scarpa and Acutt (2005).

Overall, the evidence thus suggests that the study succeeded in obtaining valid estimates of WTP from respondents. A limitation of this work, however, brought into sharp light by the results from the empirical study in chapter 3, is that DCE results can sometimes differ substantially from purportedly comparable CV results for the same package. A useful further study would involve re-implementing the survey instrument on a new sample, but with DC and PC CV questions added, which value a benchmark improvement package. If the results were to turn out quite similar, this would provide even stronger supporting evidence for the validity of the results derived in this chapter.

Given the results in chapter 3, however, one cannot be sure, in advance of the comparison, that this would turn out to be the case.

6.2.3 The sensitivity of WTP to an economic downturn

Finally, in Study 3 (chapter 5), we examined the reliability of values measured before an economic downturn for application during the downturn via analysis of near identical surveys conducted before, and during, the 2008-2010 economic recession. The policy motivation for this work arose in the context of the 2009 regulatory review of water prices in England and Wales. Most water companies had utilised SP survey results to estimate the benefits of their proposed investment plans. The surveys had typically been conducted before the onset of the recession in late 2008, however, and this gave the regulator, Ofwat, cause to doubt their reliability when the time came to determine prices.

Our main result was that the economic downturn led to lower willingness to pay when elicited via the PC CV method, but had no effect on values elicited via the DC CV technique. Clearly, an interpretation of this result requires consideration of the properties of the PC and DC methods, and we discuss these matters in section 6.3 below. In the remainder of this section we discuss the evidence concerning the validity of each year's individual survey results.

The surveys in 2008 and 2009 were each administered face-to-face using the Computer Aided Personal Interview (CAPI) method. The interviewers' comments on and scoring of respondents suggest that they understood the survey well, maintained a good degree of focus, and gave the questions careful consideration. Almost universally respondents were able to articulate a rational explanation of their choices and valuation responses, by stating that the cost, and/or the value to them of the service improvements, was the reason for their WTP answers. A fairly low proportion of the sample (9%) was excluded due to giving inadmissible responses to either the DC or PC questions. This comprised a mix of protest cases, refusals or "don't know" responses. This again provides evidence of the success of the valuation construct in eliciting meaningful data.

There are few indicators that one might expect to be correlated with WTP for water and wastewater service improvements. The only two measures expected a priori to be positively correlated with WTP were income, and membership of an environmental organisation. In both of these cases, WTP did indeed exhibit the expected direction of correlation. We conclude overall that the estimates obtained from each year's survey constitute valid measures of WTP.

The results presented from this study are limited in the extent that they can be used to infer the sensitivity of WTP to an economic downturn in general, due to the fact that the results may be dependent on the study context. In particular, the good being valued was a permanent programme of improvements, rather than a temporary change; and consistent with this, the payment vehicle was also represented as a permanent increase in a bill rather than as a one-off payment. It would not be unreasonable to argue that the sensitivity of WTP to an economic downturn might be much greater for a one-off payment for a temporary change than was found to be the case in the research reported here. Further research is therefore necessary covering differing contexts before firm conclusions can be drawn on the sensitivity of WTP to an economic downturn in general.

6.3 Discussion of Results on Treatment Effects

Two of the core chapters of this thesis have explored in a unique way the effects of alternative treatments, including elicitation method and question order effects. Study 1 utilised the PC, DC and DCE methods, and varied the order in which they were asked, to obtain a variety of preference data for subsequent analysis. Study 3 employed the PC and DC methods, and compared their relative sensitivities to an economic downturn.

We found in Study 1 that the DCE-derived measure of the value of a benchmark scenario exceeded the comparable DC measure, which itself exceeded the PC measure. Consistent with the latter comparison, Study 2 also found that DC measures of WTP exceeded PC measures in each year's survey. These findings are consistent with the review of the SP literature in chapter 2, which showed that this ordering of values by elicitation treatment was most commonly found amongst studies that compared them. [e.g. Welsh and Poe, 1998; Cameron et al., 2002].

Comparing across CV questions, the finding that DC values are higher than PC values is consistent with many previous findings [Venkatachalam, 2004; Champ and Bishop, 2006]. A prominent view in the literature [Carson and Groves, 2007] explains the observed PC<DC relationship with reference to strategic response considerations. It is argued that the DC method is compatible with truth-telling provided certain stringent auxiliary conditions are met, namely that the DC question is asked before any other elicitation question, that the survey is constructed so as to convey the idea that respondents' answers will have a consequential impact on policy, and that respondents believe the scenario as presented to them, including the scope of the improvements and the cost they, and others, will have to pay. All three of these properties can be argued to have held for Studies 1 and 3, and so it may be argued from this perspective that the DC

WTP estimate is the truth. By contrast, under plausible belief structures – such as that the go/no go policy decision rule depends on summing respondents' stated PC WTP amounts, and that an individual's stated WTP amount is weakly correlated with the amount they will be required to pay should the policy be implemented – the PC method provides an incentive for respondents to understate their true WTP, either to minimize the chance that the policy goes ahead – stating a WTP of £0 when the cost is expected to be greater than true WTP – or to minimize the expected payment, by stating a WTP of the expected cost - sometimes rationalized as a "fair amount" - when the cost is expected to be less than true WTP. Strategic considerations are thus predicted to cause respondents to understate their true WTP when offered the opportunity to do so.

A different perspective suggests that the observed difference between PC and DC WTP is due to the certainty of respondents about their true WTP when they answer the questions [Ready, Navrud and Dubourg, 2001]. This view is backed up by supporting empirical evidence showing, firstly, that respondents are indeed less certain about their DC responses than they are about their PC and OE responses [Ready, Navrud and Dubourg, 2001; Welsh and Poe, 1998], and that fixing certainty levels resolves the discrepancy [Ready, Navrud and Dubourg, 2001; Welsh and Poe, 1998].

The research presented in this thesis has not sought to test the validity of alternative treatments, and we have not argued for one method or another as being the sole route to the truth. Instead, we have sought to understand the range of estimates that is obtained by varying the elicitation method and question order. In doing so, we have utilized an innovative, in the context of SP research, estimation technique, to combine the data from DC and PC responses in a consistent way. This technique, the panel interval censored estimator, allows for correlation in unobserved determinants of WTP across elicitation methods, and was successfully applied in chapters 3 and 5. (Chapter 4 contained only DC questions for the CV component and so the technique was not applicable here.) The parameter included to capture the correlation in unobserved effects was highly significant in both studies, which supports the idea that the results from the two question types should be modelled jointly.

We also found in Study 1 that values were sensitive to the order in which the questions were asked, a result that is also consistent with many previous studies [e.g. Bateman et al., 2008] and behaviour in actual markets. In the present case, DC WTP was found to be higher if the DC question came first, and PC WTP is found to be lower if the PC question came first. From the strategic behavioural perspective, the first scenario presented has special status since only the first scenario is free from the influence of prior scenarios. In contrast, various types of (non-strategic) hypothesized learning [e.g., Braga and Starmer, 2005; Plott, 1996] suggest that answers to later questions are likely to be more reliable than answers to earlier questions. In the present analysis we have not attempted to distinguish between the strategic and anchoring hypotheses, instead we have simply controlled for the order effects and reported the range of estimates we obtained.

Finally, chapter 5 examines the relative sensitivity of PC and DC WTP measures to an economic downturn. Our main finding is that the recession caused PC WTP to fall, whereas DC WTP stayed the same. For the strategic behavioural view of the CV response process to hold, there would need to have been some change in incentives, or there must be some feature of a recession that causes respondent to become more strategically minded. The former condition can only be true if expectations of the true cost of the investment program had changed. Since there is no difference in the information given in the survey, it is unlikely that cost expectations could have changed between surveys. On the other hand, it is plausible that increased job/income insecurity might invoke a greater willingness to engage in strategic response behaviour. Unfortunately, however, we are not able to test this hypothesis with our dataset.

To be consistent with the differential certainty perspective, there would need to be some feature of a recession that caused respondents to become less certain of their true WTP. This seems plausible to us, in that job insecurity might readily diminish certainty over WTP. This could cause there to be a wider uncertainty range, with a lower level of "certain" willingness to pay, but with no different a level at the top end of the range where respondents are "not sure" whether they would pay or not. Our results are thus also consistent with this explanation of the DC-PC difference.

The principal explanations for the common finding that DC WTP>PC WTP – strategic behaviour and respondent uncertainty - are both potentially consistent with this result and hence we cannot say for sure why the recession caused PC WTP to fall, while leaving DC WTP unchanged. Nor can we use the result as evidence in support of one or the other theory.

Future research might lead to a convergence in views on what is the right elicitation approach to use. Until this happens, if indeed it ever will, we would recommend taking the pragmatic approach adopted in this thesis, of exploring the sensitivity of values to treatment effects and reporting the range obtained.

7 Conclusions

The main theme of this thesis concerns the establishment of non-market benefits valuation models for use in current and future policy applications. The research is focussed on the UK water sector, but the results have broader application: as evidence for value transfer studies in other countries or regions facing similar policy needs, and as further evidence concerning the reliability of the SP method. All three empirical studies in this report adopted survey design, implementation and analysis principles broadly consistent with the guidelines contained within the vast and growing literature on the SP approach. The response data were also found to perform well against the suite of validity tests recommended in the literature, as discussed in section 6. Notwithstanding these findings, however, we acknowledge that SP valuation is an evolving methodology, and the literature continues to challenge existing methods and assumptions.

Our main conclusions from this research are as follows. Household values for water quality improvements consistent with Water Framework Directive targets were found to vary from £2,263 to £39,168 per km^2, at 2007 prices, depending on the population density around the location of the improvement, the ecological scope of that improvement, and the value elicitation method employed. While the former factors are consistent with expectations, the latter suggests that decision makers need to be aware of such methodological effects when employing derived values. From Study 2, households were found to value the avoidance of one expected day of severe drought water use restrictions at £53 per household per year; and of one expected day of lesser restrictions at £2 per household per year. The comparable results for businesses were, on average, £845 per business per year to avoid one expected day of severe restrictions and £48 per business per year to avoid one expected day of lesser restrictions. Based in part on these results, the Beckton desalination plant in East London was approved and built, and began operating in June 2010. The final main conclusion we draw from this research is the cautionary note that economic downturns may lead to lower willingness to pay for water service improvements. Whether they actually do or not depends on which WTP elicitation method is closer to the truth. Our research suggests that WTP is sensitive to an economic downturn when elicited via the PC CV method, but that the downturn has no effect on WTP elicited via a DC (ie referendum-type) CV question.

For future policy applications we recommend that research should explore the sensitivity of values to elicitation treatments. Until there is significantly greater consensus regarding the process of preference formation and revelation under a wide range of treatments, it is important for consumers of the research, i.e. decision makers, to understand how sensitive SP valuations are to purportedly innocuous procedural variations. The challenge for the future of the SP method is to continue exploring and mapping out the multitude of ways in which survey design, implementation and analysis procedures impact on the value estimates obtained, and by so doing continue to refine the guidelines around which the majority of researchers can agree. The demand for high quality SP research will surely continue unabated all the while.

References

Acton, J. P. (1973) *Evaluating Public Programs to Save Lives: The Case of Heart Attacks*, Research Report R-73-02, RAND corporation, Santa Monica, California.

Adamowicz, W., Boxall, P., Williams, M. and Louviere, J. (1998) Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation, *American Journal of Agricultural Economics*, **80**(1), 64-75

Adamowicz, W., Garrod, G. D. and Willis, K. G. (1995), *Estimating the Passive Use Benefits of Britain's Inland Waterways*, Report to British Waterways, UK.

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

Alam, M. K., and Marinova, D. (2003) Measuring the Total Value of a River Cleanup, *Water Science and Technology*, **48**(7), 149-156.

Alberini, A. and Kahn, J. (2006) *Handbook on Contingent Valuation,* Edward Elgar, Cheltenham, UK.

Alberini, A., Kanninen, B. and Carson, R. T. (1997) Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data, *Land Economics*, **73**(3), 309-324.

Ariely, D., Lowenstein, G. and Prelec, D. (2003) “Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences, *Quarterly Journal of Economics*, **118**(1), 73-105.

Arrow, K., Colombatto, E., Perleman, M., Schmidt, C. (Eds.) (1996) *Rational Foundations of Economic Behavior*. Macmillan, London.

Arrow, K. and Fisher, A. (1974) Environmental Preservation, Uncertainty and Irreversibility, *Quarterly Journal of Economics*, **88**, 313-319

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Shuman, H. (1993) Report of the NOAA panel on contingent valuation. *Federal Register,* **58,** 4601-4614.

Barreiro, J., Sánchez, M. and Viladrich-Grau, M. (2005) How Much are People Willing to Pay for Silence? A Contingent Valuation Study, *Applied Economics*, **37**, 1233-1246

Barton, D. N. (2002) The Transferability of Benefit Transfer: Contingent Valuation of Water Quality Improvements in Costa Rica, *Ecological Economics*, **42**(1-2), 147-164.

Bateman, I. J., Brouwer, R., Ferrini, S., Schaafsma, M., Barton, D. N., Dubgaard, A., Hasler, B., Hime, S., Liekens, I., Navrud, S., De Nocker, L., Ščeponavičiūtė, R. and Semėnienė, D. (2011) Making Benefit Transfers Work: Deriving and Testing Principles for Value Transfers for Similar and Dissimilar Sites Using a Case Study of the Non-Market Benefits of Water Quality Improvements Across Europe, *Environmental and Resource Economics*, Forthcoming.

Bateman, I. J., Burgess, D., Hutchinson, W. G., Matthews, D. I. (2008) Learning Design Contingent Valuation (LDCV): NOAA Guidelines, Preference Learning and Coherent Arbitrariness, *Journal of Environmental Economics and Management,* **55**(2), 127-141.

Bateman, I. J., Carson, R. T., Day, B., Hanemann, W. M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozedemiroglu, E., Pearce, D., Sugden, R. and Swanson, J. (2002) *Economic Valuation with Stated Preference Techniques: A Manual,* Edward Elgar, Cheltenham, UK.

Bateman, I. J., Cole, M. A., Georgiou, S. and Hadley, D. J. (2006) Comparing Contingent Valuation and Contingent Ranking: A Case study Considering the Benefits of Urban River Water Quality Improvements, *Journal of Environmental Management,* **79**, 221-231.

Bateman, I. J., Day, B. H., Dupont, D. P. and Georgiou, S. (2009) Procedural Invariance Testing of the One-and-One-Half-Bound Dichotomous Choice Elicitation Method, *Review of Economics and Statistics*, **91**(4): 806–820.

Bateman, I. J., Langford, I. H., Jones, A. P. and Kerr, G. N. (2001) Bound and Path Effects in Double and Triple Bounded Dichotomous Choice Contingent Valuation, *Resource and Energy Economics*, **23**, 191-213.

Bateman, I. J., Langford, I. H., Turner, R. K., Willis, K. G. and Garrod, G. D. (1995) Elicitation and Truncation Effects in Contingent Valuation Studies. *Ecological Economics* **12**, 161–179.

Bateman, I. J., Langford, I. H., Rasbash, J. (1999) Willingness-to-Pay Question Format Effects in Contingent Valuation Studies, in Bateman, I. J. and Willis, K. G. (Eds.) (1999) *Contingent Valuation of Environmental Preferences: Assessing Theory and Practice in the USA, Europe, and Developing Countries*. Oxford University Press, Oxford, 511–539.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997a) A Theory of Reference-Dependent Preferences. *Quarterly Journal of Economics,* **112,** 479-505.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997b) Does Part-Whole Bias Exist? An Experimental Investigation, *Economic Journal,* **107**(441), 322-332.

Bateman, I.J. and Willis, K.G. (Eds.) (1999) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries,* Oxford University Press, Oxford, UK.

Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A. and Spencer, A. (1998) On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1 – Caveat Investigator, *Journal of Risk and Uncertainty,* **17**(2), 5-25

Bennett, J. (Ed.) (2011) *International Handbook on Non-market Valuation*, Edward Elgar, Northampton, MA.

Bennett, J. and Blamey, R.K. (2001) *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar Publishing Limited, Cheltenham.

Bergson, A. (1938) A Reformulation of Certain Aspects of Welfare Economics, *Quarterly Journal of Economics*, **52**(2), 310-334.

Birol, E. and Das, S. (2010) Estimating the Value of Improved Wastewater Treatment: The Case of River Ganga, India, *Journal of Environmental Management*, **91**(11), 2163-2171.

Bishop, R. C. and Heberlein, T. A. (1979) Measuring Values of Extra-Market Goods, *American Journal of Agricultural Economics*, **61**, 926-930.

Bishop, R. C., Heberlein, T. A. and Kealy, M. J. (1983) Hypothetical Bias in Contingent Valuation: Results from a Simulated Market, *Natural Resources Journal*, **23**(3), 619-633.

Blaine, T. W., Lichtkoppler, F. R., Jones, K. R. and Zondag, R. H (2005) An Assessment of Household Willingness to Pay for Curbside Recycling: A Comparison of Payment Card and Referendum Approaches, *Journal of Environmental Management*, **76**(1), 15-22.

Blamey, R., Gordon, J. and Chapman, R. (1999) Choice Modelling: Assessing the Environmental Values of Water Supply Options, *Australian Journal of Agricultural and Resource Economics*, **43**(3), 337-357.

Bockstael, N., Hanemann, W. M. and Kling, C. (1987) Estimating the Value of Water Quality Improvements in a Recreational Demand Framework, *Water Resources Research*, **23**(5), 951-960.

Bohara, A. K., McKee, M., Berrens, R. P., Jenkins-Smith, H., Silva, C. L. and Brookshire, D. S. (1998) Effects of Total Cost and Group-Size Information on Willingness to Pay Responses: Open Ended vs. Dichotomous Choice, *Journal of Environmental Economics and Management*, **35**, 142-163.

Bowen, H. R. (1943) The Interpretation of Voting in the Allocation of Economic Resources, *Quarterly Journal of Economics*, **58**, 27-48.

Boxall, P., Adamowicz, W., Swait, J., Williams, M. and Louviere, J. (1996) A comparison of stated preference methods for environmental valuation. *Ecological Economics*, **18,** 243-253.

Boyle, K. J., Bishop, R. C. and Welsh, M. P. (1985) Starting Point Bias in Contingent Valuation Bidding Games, *Land Economics*, **61**(2), 188-194.

Braden, J. B. and Kolstad, C. D. (1991) *Measuring the Demand for Environmental Quality*. North Holland, New York.

Bradford, D. F. (1970) Benefit-Cost Analysis and Demand Curves for Public Goods, *Kyklos*, **23**, 775-91.

Bradley, R. M. (2010) Direct and Indirect Benefits of Improving River Quality: Quantifying Benefits and a Case Study of the River Klang, Malaysia, *Environmentalist*, **30**(3); 228-241.

Braga, J., and Starmer, C. (2005) Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis, *Environmental and Resource Economics*, **32**, 55-89.

Brookshire, D. S. and Coursey, D. L (1987) Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures, *American Economic Review*, 77, 554-566.

Brookshire, D. S., Ives, B. C. and Schulze, W. D. (1976) The Valuation of Aesthetic Preference, *Journal of Environmental Economics and Management*, **3**, 325-46.

Brouwer, R. (2006) Do Stated Preference Methods Stand the Test of Time? A Test of the Stability of Contingent Values and Models for Health Risks when Facing an Extreme Event, *Ecological Economics*, **60**(2), 399-406.

Brouwer, R. (2008) The Potential Role of Stated Preference Methods in the Water Framework Directive to Assess Disproportionate Costs, *Journal of Environmental Planning and Management*, **51**(5), 597-614.

Brouwer, R. and Bateman, I. J. (2005) Temporal Stability and Transferability of Models of Willingness to Pay for Flood Control and Wetland Conservation, *Water Resources Research*, **41**(3).

Brown, T. C., Champ, P. A., Bishop, R. C. and McCollum, D. W. (1996) Which Response Format Reveals the Truth about Donations to a Public Good? *Land Economics*, **72**(2), 152-166.

Cadsby, C. B. and Maynes, E. (1999) Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence, *Journal of Public Economics*, **71**, 53-73

Cameron, T. A. and Huppert, D. D. (1989) OLS versus ML estimation of Non-market Resource Values with Payment Card Interval Data, *Journal of Environmental Economics and Management*, **17**(3), 230-246.

Cameron, T. A. and Huppert, D. D. (1991) Referendum Contingent Valuation Estimate - Sensitivity to the Assignment of Offered Values, *Journal of the American Statistical Association*, **86**(416), 910-918.

Cameron, T. A., Poe, G. L., Ethier, R. G. and Schulze, W. D. (2002) Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?, *Journal of Environmental Economics and Management*, **44**(3), 391-425.

Cameron, T. A. and Quiggin, J. (1994) Estimation Using a "Contingent Valuation Data with Follow-up" Questionnaire, *Journal of Environmental Economics and Management*, **27**, 218-234.

Carson, R. T. (1985) *Three Essays on Contingent Valuation*, Ph.D thesis, University of California, Berkeley.

Carson, R. T. (1997) Contingent Valuation Surveys and Tests of Insensitivity to Scope, in Kopp, R.J., Pommerehne, W.W. and Schwarz, N. (Eds.) (1997) *Determining the Value of Non-marketed Goods*, Kluwer Academic Publishers, Boston.

Carson, R. T. (2011) *Contingent Valuation: A Comprehensive Bibliography and History*. Cheltenham: Edward Elgar.

Carson, R. T., Flores, N. and Hanemann, W. M. (1998) Sequencing and Valuing Public Goods, *Journal of Environmental Economics and Management*, **36**(3), 314-323.

Carson, R. T., Flores, N., Meade, N. F. (2001) Contingent Valuation: Controversies and Evidence, *Environmental and Resource Economics*, **19**, 173-210.

Carson, R. T., and Groves, T. (2007) Incentive and Informational Properties of Preference Questions, *Environmental & Resource Economics,* **37**(1), 181-210.

Carson, R. T. and Groves, T. (2011) Incentive and Informational Properties of Preference Questions: Commentary and Extensions, in Bennett, J. (Ed.) (2011) *International Handbook on Non-market Valuation*, Northampton, MA: Edward Elgar.

Carson, R. T. and Hanemann, W. M. (2005) Contingent Valuation, in Mäler, K. G. and Vincent, J. R., (Ed.) (2005) *Handbook of Environmental Economics. Valuing Environmental Changes*. Vol. 2. Amsterdam: Elsevier, 821-936.

Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Conaway, M., and Martin, K. (1997) Temporal Reliability of Estimates from Contingent Valuation, *Land Economics*, **73**(2), 151-163.

Carson, R. T., Hanemann, W. M. and Mitchell, R. C. (1987) *The Use of Simulated Political Markets to Value Public Goods*, Discussion paper 87–7, Department of Economics, University of California, San Diego.

Carson, R. T. and Louviere, J. J. (2011) A Common Nomenclature for Stated Preference Elicitation Approaches, *Environmental and Resource Economics*, **49**, 539-559.

Carson, R. T. and Mitchell, R. C. (1987) *Economic Value of Reliable Water Supplies for Residential Water Users in the State Water Project Service Area.* Contract report for Metropolitan Water District of Southern California, Los Angeles, CA.

Carson, R. T. and Mitchell, R. C. (1993) The Value of Clean Water - the Public's Willingness-to-Pay for Boatable, Fishable, and Swimmable Quality Water, *Water Resources Research*, **29**(7), 2445-2454.

Carson, R. T. and Mitchell, R. C. (1995) Sequencing and Nesting in Contingent Valuation Surveys, *Journal of Environmental Economics and Management*, **28**, 155-173.

Carson, R. T., Mitchell, R. C., Hanemann, W. M., Kopp, R. J., Presser, S. and Ruud, P. A. (2003), Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill. *Environmental and Resource Economics,* **25**, 257-286.

Cattin, P. and Wittink, D. R. (1982), Commercial Use of Conjoint Analysis: A Survey, *Journal of Marketing*, **46**, 44-53.

Champ, P. A. and Bishop, R. C. (2006), Is Willingness to Pay for a Public Good Sensitive to the Elicitation Format?, *Land Economics*, **82**(2), 162-173.

Choe, K. A., Whittington, D. and Lauria, D. T. (1996) The Economic Benefits of Surface Water Quality Improvements in Developing Countries: A Case Study of Davao, Philippines, *Land Economics*, **72**(4), 519-537.

Christie, M. and Azevedo, C. D. (2009) Testing the Consistency between Standard Contingent Valuation, Repeated Contingent Valuation and Choice Experiments, *Journal of Agricultural Economics*, 60(1), 154–170.

Ciriacy-Wantrup, S. V. (1947) Capital Returns from Soil-Conservation Practices, *Journal of Farm Economics*, **29**, 1181-1196.

Ciriacy-Wantrup, S. V. (1952) *Resource Conservation: Economics and Policies*, University of California Press, Berkeley, CA.

Colombo, S., Calatrava-Requena, J. and Hanley, N. (2006) Analysing the Social Benefits of Soil Conservation Measures Using Stated Preference Methods, *Ecological Economics*, **58**, 850-861.

Cooper, J. C. and Hanemann, W. M. (1995), *Referendum Contingent Valuation: How Many Bounds Are Enough?*', USDA Economic Research Service, Food and Consumer Economics Division, Working Paper. May, 1995.

Cooper, J. C., Hanemann, W. M. and Signorello, G. (2002) One-and-One-Half-Bound Dichotomous Choice Contingent Valuation, *Review of Economics and Statistics*, **84**(4), 742-750.

Cooper, J. C. and Signorello, G. (2008) Farmer Premiums for the Voluntary Adoption of Conservation Plans, *Journal of Environmental Planning and Management*, **51**(1), 1-14.

Corso, P. S., Hammitt, J. K. and Graham, J. D. (2001) Valuing Mortality Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation, *Journal of Risk and Uncertainty*, **23**, 165-184

Coursey, D. L., Hovis, J. L. and Schulze, W. D. (1987) The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value, *Quarterly Journal of Economics*, **102**, 679-690.

Cowell, F. A. and Gardiner, K. (1999) *Welfare Weights*, STICERD Economics Research Paper, 20, London School of Economics, London, UK.

Cummings, R. G., Brookshire, D. S. and Schulze, W. D. (1986) *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Rowman and Allanheld, Totowa, New Jersey.

Cummings, R. G., Elliott, S., Harrison, G. W. and Murphy, J. (1997) Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy*, **105**, 609-21

Dalrymple, G. (2006), *Valuing the Water Environment: A Review of International Literature*, Report to Scottish Executive Environment and Rural Affairs Department, UK.

Daubert, J. T. and Young, R. A. (1981) Recreational Demands for Maintaining Instream Flows: A Contingent Valuation Approach, *American Journal of Agricultural Economics*, **63**(4), 666-676.

Davis, R. K. (1963a) Recreation Planning as an Economic Problem, *Natural Resources Journal*, **3**, 239-249.

Davis, R. K. (1963b) *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*, Dissertation, Harvard University.

Davis, R. K. (1964), The Value of Big Game Hunting in a Private Forest, in *Transactions of the 29th North American Wildlife and Natural Resources Conference*, Wildlife Management Institute, Washington DC.

DEFRA (2007) *Draft Partial Regulatory Impact Assessment of Environmental Quality Standards for Implementation of the Water Framework Directive in the UK*, Report, DEFRA, UK.

DEFRA (2008), *Overall Impact Assessment for the Water Framework Directive (EC 2000/60/EC)*, Report, DEFRA, UK.

DeShazo, J. R. (2002) Designing Transactions without Framing Effects in Iterative Question Formats, *Journal of Environmental Economics and Management*, **43**, 360-385.

Desvouges, W. H., Smith, V. K. and McGivney, M. P. (1983) *A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvement*. Research Triangle Institute. Report to U.S. Environmental Protection Agency, EPA-230-05-83-001, Washington, D.C.

Devine, D. G. and Marion, B. W. (1979) The Influence of Consumer Price Information on Retail Pricing and Consumer Behaviour, *American Journal of Agricultural Economics*, **61**, 228-237.

Diamond, P. A. and Hausman, J. A. (1994) Contingent Valuation: Is Some Number Better than No Number? *Journal of Economic Perspectives*, **8**(4), 45-64.

Ding, Y., Hayes, M. J. and Widhalm, M. (2010) Measuring Economic Impacts of Drought: A Review and Discussion, *Disaster Prevention and Management*, **20**(4), 434-446.

Donaldson, C., Thomas, R. and Torgerson, D. J. (1997) Validity of Open-ended and Payment Scale Approaches to Eliciting Willingness to Pay, *Applied Economics*, **29**, 79-84.

Dubourg, W. R.; Jones-Lee, M. W. and Loomes, G. (1994) Imprecise Preferences and the WTP-WTA Disparity, *Journal of Risk and Uncertainty*, **9**(2), 115–33.

Eggert, H. and Olsson, B. (2009) Valuing multi-attribute marine water quality, *Marine Policy*, **33**(2), 201-206.

European Parliament (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action, *Official Journal of the European Communities*, L327, 1-72.

Farber, S, and Griner, B. (2000) Valuing Watershed Quality Improvements Using Conjoint Analysis, *Ecological Economics*, **34**(1), 63-76.

Farquharson, R. (1969) *Theory of Voting*, Yale University Press, New Haven.

Fernandez, C., Leon, C. J., Steel, M. F. J. and Vasquez-Polo, F. J. (2004) Bayesian Analysis of Interval Data Contingent Valuation Models and Pricing Policies, *Journal of Business & Economic Statistics*, **22**(4), 431-442.

Flachaire, E. and Hollard, G. (2007) Starting Point Bias and Respondent Uncertainty in Dichotomous Choice Contingent Valuation Surveys, *Resource and Energy Economics*, **29**, 183-194.

Flores, N. E. and Carson, R. T. (1997) The Relationship between Income Elasticities of Demand and Willingness to Pay, *Journal of Environmental Economics and Management*, **33**(3), 287-295.

Foster, V. and Mourato, S. (2003) Elicitation Format and Sensitivity to Scope - Do Contingent Valuation and Choice Experiments Give the Same Results? *Environmental & Resource Economics*, **24**(2)**,** 141-160.

Freeman, A. M. (1979) *The Benefits of Environmental Improvement: Theory and Practice*, John Hopkins University Press (for Resources for the Future), Baltimore.

Freeman, A. M. (1993) *The Measurement of Environmental and Resource Values: Theory and Methods*, Resources for the Future Press, Washington D.C. Second edition, 2003.

Frykblom, P. and Shogren, J. F. (2000) An Experimental Test of Anchoring Effects in Discrete Choice Questions, *Environmental and Resource Economics*, **16**, 329-341.

Genius, M. and Tsagarakis, K. P. (2006) Water Shortages and Implied Water Quality: A Contingent Valuation Study, *Water Resources Research*, **42**, W12407.

Georgiou, S., Bateman, I. J., Cole, M. and Hadley, D. (2000) *Contingent Ranking and Valuation of River Water Quality Improvements: Testing for Scope Sensitivity, Ordering and Distance Decay Effects*, CSERGE Working Paper, GEC-2000-18, University of East Anglia, UK.

Gibbard, A. (1973) Manipulation of Voting Schemes: A General Result, *Econometrica*, **41**(4), 587–601.

Gollier, C. (2002) Discounting an Uncertain Future, *Journal of Public Economics*, **85**, 149-166.

Grafton, R. Q. and Ward, M. B. (2008) Prices versus Rationing: Marshallian Surplus and Mandatory Water Restrictions, *Economic Record*, **84**, S57-S85.

Gramlich, F. W. (1977) The Demand for Clean Water: The Case of the Charles River, *National Tax Journal*, **30**, 183-194.

Green, D., Jacowitz, K. E., Kahneman, D. and McFadden, D. (1998) Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods, *Resource and Energy Economics*, **20**(2), 85-116.

Green, J.R. and Laffont, J. J. (1978) A Sampling Approach to the Free Rider Problem, in Sandmo, A. (Ed.) (1978) *Essays in Public Economics*, Lexington Books, Lexington, MA.

Green, C. H. and Tunstall, S. M. (1991) The Evaluation of River Water Quality Improvements by the Contingent Valuation Method, *Applied Economics*, **23**(7), 1135-1146.

Greenley, D. A., Walsh, R. G. and Young, R. A. (1981) Empirical Evidence from a Case Study of Recreation and Water Quality, *Quarterly Journal of Economics*, **96**, 657-674.

Griffin, R.C. and Mjelde, J.W. (2000) Valuing Water Supply Reliability, *American Journal of Agricultural Economics*, **82**, 414–26.

Haab, T. and McConnell, K. (2002) *Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation*, Edward Elgar, Cheltenham, UK.

Hammack, J. and Brown, G. M. (1974) *Waterfowl and Wetlands: Toward Bioeconomic Analysis*, Resources for the Future, Baltimore.

Hammitt, M. and Graham, J. (1999) Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? *Journal of Risk and Uncertainty*, **18**, 33-52.

Hanemann, W. M. (1978) *A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement*, Dissertation, Harvard University.

Hanemann, W. M. (1985) Some Issues in Continuous- and Discrete-Response Contingent Valuation Studies, *Northeast Journal of Agricultural Economics*, 5-13.

Hanemann, W. M. (1991) Willingness to Pay and Willingness to Accept: How Much Can They Differ? *American Economic Review*, **81**, 635–647.

Hanemann, W. M. (1999) The Economic Theory of WTP and WTA, in Bateman, I. J. and Willis, K. G. (Eds.) (1999) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*: Oxford University Press, Oxford, UK.

Hanemann, W. M., Loomis, J. and Kanninen, B. (1991) Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation, *American Journal of Agricultural Economics,* 73(4), 1255-1263

Hanemann, W. M., and Kriström, B. (1995) Preference Uncertainty, Optimal Designs and Spikes, in Johansson, P. O., Kriström, B. and Maler, K-G (Eds.) (1995) *Current Issues in Environmental Economics,* Manchester University Press, Manchester, UK.

Hanley, N., Bell, D. and Alvarez-Farizo, B. (2003), Valuing the Benefits of Coastal Water Quality Improvements using Contingent and Real Behaviour, *Environmental and Resource Economics*, **24**, 273-285.

Hanley, N., MacMillan, D., Wright, R.E., Bullock, C., Simpson, I., Parsisson, D., Crabtree, B. (1998) Contingent Valuation versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland. *Journal of Agricultural Economics*, **49**(1):1–15

Hanley, N., Mourato, S. and Wright, R. E (2001) Choice Modelling Approaches: A Superior Alternative for Environmental Valuation? *Journal of Economic Surveys*, **15**(3), 435-462.

Hanley, N., Wright, R. E. and Adamowicz, W. (1998) Using Choice Experiments to Value the Environment - Design Issues, Current Experience and Future Prospects. *Environmental and Resource Economics,* **11,** 3-4**,** 413-428.

Hanley, N., Wright, R. E. and Alvarez-Farizo, B. (2006), Estimating the Economic Value of Improvements in River Ecology using Choice Experiments: An Application to the Water Framework Directive, *Journal of Environmental Management*, **78**(2), 183-193.

Hartman, R. S., Doane, M. J. and Woo, C. (1991) Consumer Rationality and the Status Quo, *Quarterly Journal of Economics*, **106**(1), 141-162.

Hausman, J. (Ed.) (1993) *Contingent Valuation: A Critical Assessment*, North-Holland, Amsterdam.

Heal, G. (1998) *Valuing the Future: Economic Theory and Sustainability*, Columbia University Press, New York.

Henderson, N. and Bateman, I. J. (1995) Empirical and Public Choice Evidence for Hyperbolic Social Discount Rates and the Implications for Intergenerational Discounting, *Environmental and Resource Economics*, **5**, 413-423.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2005) *Applied Choice Analysis - A Primer,* Cambridge University Press, New York.

Hensher, D. A., Shore, N. and Train, K. (2006) Water Supply Security and Willingness to Pay to Avoid Drought Restrictions, *Economic Record*, **82**(256), 56-66.

Hepburn, C. (2003) *Resource Collapse and Hyperbolic Discounting*, Oxford University Department of Economics, Discussion Paper.

Herriges, J. A. and Shogren, J. F. (1996) Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning, *Journal of Environmental Economics and Management*, **30**, 112-131.

Hess, S. and Daly, A. (eds.) (2010) *Choice Modelling: the State-of-the-Art and the State-of-Practice*, Emerald Group Publishing.

Hicks, J. R. (1939) The Foundations of Welfare Economics, *Economic Journal*, **49**, 696-712.

Hicks, J. R. (1943) The Four Consumers' Surpluses, *Review of Economic Studies*, **11**(1), 31-41.

HM Treasury (2003) *The Green Book: Appraisal and Evaluation in Central Government*, TSO, London.

Hoehn, J. P. (1991) Valuing the Multidimensional Impacts of Environmental Policy: Theory and Methods, *American Journal of Agricultural Economics*, **73**(2), 289-299.

Hoehn, J. P. and Randall, A. (1987) A Satisfactory Benefit-Cost Indicator for Contingent Valuation, *Journal of Environmental Economics and Management*, **14**, 226–247.

Hoehn, J. P. and Randall, A. (1989) Too Many Proposals Pass the Benefit-Cost Test, *American Economic Review*, **79**(3), 544-551.

Holmes, T. P. and Kramer, R. A. (1995) An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation, *Journal of Environmental Economics and Management*, **29**, 121-132.

Horton, B. and Fisher, J. (2006) Valuing the Environmental Benefits of Water Industry Investment in England and Wales, in Pearce, D. (Ed.) (2006) *Environmental Valuation in Developed Countries*, Edward Elgar, UK.

Howe, C. W. and Smith, M. G. et al. (1994) The Value of Water Supply Reliability in Urban Water Systems, *Journal of Environmental Economics and Management*, **26**, 19–30.

Hoyos, D. (2010) The State of the Art of Environmental Valuation with Discrete Choice Experiments, *Ecological Economics*, **69**, 2372-2381.

Jacowitz, K. E., and Kahneman, D. (1995) Measures of Anchoring in Estimation Tasks, *Personality and Social Psychology Bulletin*, **21**(11), 1161-1166.

Jin, J., Wang, Z. and Ran, S. (2006) Comparison of Contingent Valuation and Choice Experiment in Solid Waste Management Programs in Macao, *Ecological Economics,* **57**, 430-441.

Johansson, P. O., Kristrom, B. and Maler, K-G (Eds.) (1995) *Current Issues in Environmental Economics,* Manchester University Press, Manchester, UK.

Johnson, R., Bregenzer, N. S. and Shelby, B. (1990) Contingent Valuation Question Formats: Dichotomous Choice versus Open-ended Responses, in Johnson, R. and Johnson, G. (Eds.) (1990) *Economic Valuation of Natural Resources*, Westview Press, Boulder, Colorado.

Johnson, R. and Johnson, G. (Eds.) (1990) *Economic Valuation of Natural Resources*, Westview Press, Boulder, Colorado.

Johnson, E. J., and Schkade, D. A. (1989) Bias in Utility Assessments: Further Evidence and Explanations, *Management Science*, **35**(4), 406-424.

Jones-Lee, M. W. (1974) The Value of Changes in the Probability of Death or Injury, *Journal of Political Economy*, **82**, 835-849.

Jones-Lee, M. W. (1976) *The Value of Life: An Economic Analysis*, University of Chicago Press, Chicago.

Kahneman, D. and Knetsch, J. L. (1992) Valuing Public Goods: The Purchase of Moral Satisfaction, *Journal of Environmental Economics and Management*, **22**, 57-70.

Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1990) Experimental Tests of the Endowment Effect and the Coase Theorem, *Journal of Political Economy*, **98**, 1325-1348.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N. and Stone, A. A. (2006) Would you be Happier if you were Richer? A Focusing Illusion, *Science*, **312**(5782), 1908-1910.

Kahneman, D., Ritov, I. and Schkade, D. A. (1999) Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues, *Journal of Risk and Uncertainty*, **19**, 203-235.

Kahneman, D. and Tversky, A. (Eds.) (2000) *Choices, Values and Frames*, Cambridge University Press, Cambridge, UK.

Kaldor, N. (1939) Welfare Propositions of Economics and Interpersonal Comparisons of Utility, *Economic Journal*, **49**, 549-552.

Kanninen, B. (Ed.) (2006) *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Dordrecht: Springer

Kealy, M. J., Montgomery, M. and Dovidio, J. F. (1990) Reliability and Predictive-Validity of Contingent Values - Does the Nature of the Good Matter?, *Journal of Environmental Economics and Management*, **19**(3), 244-263.

Knetsch, J. L. and Sinden, J. A. (1984) Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value, *Quarterly Journal of Economics*, **99**, 507-521.

Kolstad, C. D. and Guzman, R. M. (1999) Information and the Divergence between Willingness to Accept and Willingness to Pay, *Journal of Environmental Economics and Management*, **38**, 66-80.

Kontogianni, A., Langford, I. H., Papandreou, A. and Skourtos, M. S. (2003) Social Preferences for Improving Water Quality: An Economic Analysis of Benefits from Wastewater Treatment, *Water Resources Management*, **17**(5), 317-336.

Kopp, R. J., Pommerehne, W. W. and Schwarz, N. (Eds.) (1997) *Determining the Value of Non-marketed Goods*, Kluwer Academic Publishers, Boston.

Koss, P. and Khawaja, M. S. (2001) The Value of Water Supply Reliability in California: A Contingent Valuation Study, *Water Policy*, **3**, 165–74.

Kramer, R. A., and Eisen-Hecht, J. I. (2002), Estimating the Economic Value of Water Quality Protection in the Catawba River Basin, *Water Resources Research*, **38**(9), 1182-1191.

Kramer, R. A. and Mercer, D. E. (1997) Valuing a Global Environmental Good: U.S. Residents' Willingness to Pay to Protect Tropical Rain Forests, *Land Economics*, **73**(2), 196-210.

Kriström, B. (1993) Comparing Continuous and Discrete Contingent Valuation Questions, *Environmental and Resource Economics*, 3(1), 63-71.

Kriström, B. (1997) Spike Models in Contingent Valuation, *American Journal of Agricultural Economics*, **79**(3), 1013-1023.

Krutilla, J. V. (1967) Conservation Reconsidered, *American Economic Review*, **57**, 777-786.

Lancaster, K. (1966) A New Approach to Consumer Theory, *Journal of Political Economy*, **74**, 132-57.

Langford, I. H., Bateman, I. J. and Langford, H.D. (1996) A Multilevel Modelling Approach to Triple-bounded Dichotomous Choice Contingent Valuation, *Environmental and Resource Economics*, **7**, 197–211.

Lanz, B., Provins, A., Bateman, I., Scarpa, R., Willis, K., and Ozdemiroglu, E. (2010), Investigating Willingness to Pay - Willingness to Accept Asymmetry in Choice Experiments', in Hess, S. and Daly, A. (eds.) *Choice Modelling: the State-of-the-Art and the State-of-Practice*, Emerald Group Publishing.

Leggett, C. G. and Bockstael, N. E. (2000) Evidence on the Effects of Water Quality on Residential Land Prices, *Journal of Environmental Economics and Management*, **39**, 124-144.

Li, C-Z and Mattson, L. (1995) Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation, *Journal of Environmental Economics and Management*, **28**, 256-269.

List, J. A. (2003) Does Market Experience Eliminate Market Anomalies? *Quarterly Journal of Economics*, **118**, 41–72.

Loomis, J. B. (1988) Contingent Valuation using Dichotomous Choice Models, *Journal of Leisure Research*, **20**(1), 46-56.

Loomis, J. B. (1990) Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques, *Journal of Environmental Economics and Management*, **18**(1), 78-85.

Loomis, J. B., Brown, T., Lucero, B. and Peterson, G. (1997) Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation, *Environmental and Resource Economics*, **10**, 109-123.

Loomis, J. B. and Ekstrand, E. (1998) Alternative Approaches for Incorporating Respondent Uncertainty when Estimating Willingness to Pay: The Case of the Mexican Spotted Owl, *Ecological Economics*, **27**, 29-41.

Loomis, J. B., P. Kent, L. Strange, K. Fausch, and A. Covich (2000), Measuring the Total Economic Value of Restoring Ecosystem Services in an Impaired River Basin: Results from a Contingent Valuation Survey, *Ecological Economics*, **33**(1), 103-117.

Louviere, J. J. and Hensher, D. A. (1983) Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event, *Journal of Consumer Research*, **10**, 348-361.

Louviere, J. J., Hensher, D. A., Swait, J. (2000), *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge.

Louviere, J. J. and Woodworth, G. (1983) Design and Analysis of Simulated Consumer Choice Experiments or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research*, **20**, 350-367.

Lunander, A. (1998) Inducing Incentives to Understate and to Overstate Willingness to Pay within the Open-Ended and the Dichotomous-Choice Elicitation Formats: An Experimental Study, *Journal of Environmental Economics and Management*, **35**, 88-102.

Lusk, J. and Norwood, F. B. (2005) Effect of Experimental Design on Choice-Based Conjoint Valuation Estimates, *American Journal of Agricultural Economics*, **87**(3), 771-785.

Maddison, A. (2001) *The World Economy: A Millennial Perspective*, OECD, Paris.

Magat, W. A., Viscusi, W. K. and Huber, J. (1988) Paired Comparison and Contingent Valuation Approaches to Morbidity Risk Valuation, *Journal of Environmental Economics and Management*, **15**, 395-411.

Manski, C. F. and McFadden, D. (1981) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

McConnell, K. E. (1990) Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation, *Journal of Environmental Economics and Management*, **18**, 19-34.

McConnell, K. E., Strand, I. E. and Valdes, S. (1998) Testing Temporal Reliability and Carry-over Effect: The Role of Correlated Responses in Test-Retest Reliability Studies, *Environmental and Resource Economics*, **12**(3), 357-374.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behaviour, in Zerembka, P. (Ed.) (1974) *Frontiers in Econometrics.* Academic Press, New York.

McFadden, D. (1994) Contingent Valuation and Social Choice, *American Journal of Agricultural Economics*, **76**, 689-708.

McNair, B. J., Bennett, J. and Hensher, D. A. (2011) A Comparison of Responses to Single and Repeated Discrete Choice Questions, *Resource and Energy Economics*, **33**, 554-571.

Mitchell, R. C., and Carson, R. T. (1981) *An Experiment in Determining Willingness to Pay for National Water Quality Improvements*, Report to the Environmental Protection Agency, Resources for the Future, Washington DC.

Mitchell, R. C., and Carson, R. T. (1984), *Willingness to Pay for National Water Quality Improvements*, Report to the Environmental Protection Agency, Resources for the Future, Washington DC.

Mitchell, R. C., and Carson, R. T. (1986) *The Use of Contingent Valuation Data for Benefit-Cost Analysis in Water Pollution Control*, Final Report to the Environmental Protection Agency, Resources for the Future, Washington DC.

Mitchell, R. C., and Carson, R. T. (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington, D.C.

Mogas, J., Riera, P. and Bray, R. (2009) Combining Contingent Valuation and Choice Experiments: A Forestry Application in Spain, *Environmental and Resource Economics*, **43**, 535-551.

Moran, D. and McMahon, P. (Eds.) (2001) *Economic Valuation of Water Resources: Policy and Practice*. Chartered Institution of Water and Environmental Management (CIWEM), London.

Mourato, S., Özdemiroglu, E., Atkinson, G., Newcombe, J. and de Garis, Y. (2006a) The Value of a Tidier Thames: Willingness to Pay to reduce Sewage Overflows, in Pearce, D. (Ed.) (2006) *Environmental Valuation in Developed Countries*, Edward Elgar, UK.

Mourato, S., Provins, A., Özdemiroglu, E., Georgiou, S. and Newcombe, J. (2006b) Is it Worth Revising the European Bathing Water Directive? A Choice Experiment, in Pearce, D. (Ed.) (2006) *Environmental Valuation in Developed Countries*, Edward Elgar, UK.

Nahman, A. and Rigby, D. (2008) Valuing Blue Flag Status and Estuarine Water Quality in Margate, South Africa, *South African Journal of Economics*, **76**(4), 721-737.

Nayga, R. M., Woodward, R. and Aiew, W. (2006) Willingness to Pay for Reduced Risk of Foodborne Illness: A Nonhypothetical Field Experiment, *Canadian Journal of Agricultural Economics,* **54**, 461-475.

NERA-Accent (2006) *The Cost of Water Use Restrictions*, Report for Thames Water, London, UK.

NERA-Accent (2007) *The Benefits of Water Framework Directive Programmes of Measures in England and Wales*, Report CRP 4b/c, DEFRA.

Ofwat (2008) *Setting Price Limits for 2010-15: Framework and Approach*.

Olmstead, S. M. (2010) The Economics of Water Quality, *Review of Environmental Economics and Policy*, **4**(1), 44–62.

OXERA (2002) *A Social Time Preference Rate for Use in Long-term Discounting*, Report for ODPM, DfT and DEFRA, UK.

Pareto, V. (1896) *Cours d'Economie Politique*, Vol. 2, F.Rouge, Lausanne.

Pearce, D. (Ed.) (2006) *Environmental Valuation in Developed Countries*, Edward Elgar, UK.

Pearce, D. and Ulph, D. (1995) A *Social Discount Rate for the United Kingdom*, CSERGE Working Paper, 95-01, School of Environmental Studies, UEA.

Plott, C. R. (1996) Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis, in Arrow, K., Colombatto, E., Perleman, M., Schmidt, C. (Eds.) (1996) *Rational Foundations of Economic Behavior*, Macmillan, London.

Portney, P. and Weyant, J. (Eds.) (1999) *Discounting and Intergenerational Equity*, Resources for the Future, Washington DC.

Powe, N. A., Willis, K. and Garrod, G. (2006) Difficulties in Valuing Streetlight Improvement: Trust, Surprise and Bound Effects, *Applied Economics*, **38**, 371-381.

Ramsey, F. P. (1928) A Mathematical Theory of Saving, *Economic Journal*, **38**, 543–59.

Randall, A., Ives, B. C. and Eastman, C. (1974) Bidding Games for the Valuation of Aesthetic Environmental Improvements, *Journal of Environmental Economics and Management*, **1**, 132-149.

Ready, R. C., Buzby, J. C. and Hu, D. (1996) Differences between Continuous and Discrete Contingent Valuation Estimates. *Land Economics*, 72**,** 3**,** 397-411.

Ready, R. C., Malzubris, J. and Senkane, S. (2002) The Relationship Between Environmental Values and Income in a Transition Economy: Surface Water Quality in Latvia, *Environment and Development Economics*, **7**, 147-156.

Ready, R. C., Navrud, S. and Dubourg, W. R. (2001) How do Respondents with Uncertain Willingness to Pay Answer Contingent Valuation Questions? *Land Economics*, **77**(3), 315-326.

Ready, R. C., Whitehead, J. C. and Blomquist, G. C. (1995) Contingent Valuation when Respondents are Ambivalent, *Journal of Environmental Economics and Management*, **29**(2), 181-196.

Reaves, D. W., Kramer, R. A. and Holmes, T. P. (1999) Does Question Format Matter? Valuing and Endangered Species, *Environmental and Resource Economics*, b(3), 365-383.

Reiling, S. D., Boyle, K. J., Phillips, M. L. and Anderson, M. W. (1990) Temporal Reliability of Contingent Values, *Land Economics*, **66**(2), 128-134.

Ridker, R.G. (1967) *Economic Costs of Air Pollution*, Praeger, New York.

Roibás, D., García-Valiňas, M. A. and Wall, A. (2007) Measuring Welfare Losses from Interruption and Pricing as Responses to Water Shortages: An Application to the Case of Seville, *Environmental and Resource Economics*, **38**, 231-243.

Rollins, K. and Lyke, A. (1998) The Case for Diminishing Marginal Existence Values, *Journal of Environmental Economics and Management*, **36**(3), 324-344.

Rondeau, D., Schulze, W. D. and Poe, G. L. (1999) Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism, *Journal of Public Economics*, **72**, 455-70.

Rose, J. M., Bliemer, M. C. A., Hensher, D. A. and Collins, A. T. (2008) Designing Efficient Stated Preference Experiments in the Presence of Reference Alternatives, *Transportation Research B*, **42**(4), 395-406.

Rowe, R. D., D’Arge, R. C. and Brookshire, D. S. (1980) An Experiment on the Economic Value of Visibility, *Journal of Environmental Economics and Management*, **7**(1), 1-19.

Rowe, R. D., Schuze, W. E. and Breffle, W. S. (1996) A Test for Payment Card Biases, *Journal of Environmental Economics and Management*, **31**(2), 178-185.

Ryan, M., Scott, D. A. and Donaldson, C. (2004) Valuing Healthcare using Willingness to Pay: A Comparison of the Payment Card and Dichotomous Choice Methods, *Journal of Health Economics*, **23**(2), 237-258.

Samuelson, P. (1947). Foundations of Economic Analysis, Harvard University Press, Cambridge, Massachussetts.

Samuelson, P. (1950) Evaluation of Real National Income, *Oxford Economic Papers*, 2(1), 1-29.

Samuelson, P. (1954). The Pure Theory of Public Expenditure. *Review of Economics and Statistics*, **36**, 387-389.

Sandmo, A. (Ed.) (1978) *Essays in Public Economics*, Lexington Books, Lexington, MA.

Satterthwaite, M. A. (1975) Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, *Journal of Economic Theory*, **10**, 187–217.

Schkade, D. A., and Kahneman, D. (1998) Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction, *Psychological Science*, **9**(5), 340-346.

Scitovsky, T. (1941) A Note on Welfare Propositions in Economics, *Review of Economic Studies*, **9**(1), 77-88.

Scott, M. F. G. (1977) The Test Rate of Discount and Changes in Base Level of Income in the United Kingdom, *Economic Journal*, **87**(346), 219-241.

Sellar, C., Stoll, J. R. and Chavas, J. P. (1985) Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques. *Land Economics*, **61**.

Shogren, J. F., Shin, S. Y., Hayes, D. J. and Kliebenstein, J. B. (1994), Resolving Differences in Willingness to Pay and Willingness to Accept', *American Economic Review*, **84**(1), 255–270.

Signorello, Giovanni, and Joseph C. Cooper, "Farmer Premiums for the Voluntary Adoption of Conservation Plans," FEEM working paper no. 27.2002 (2002).

Sileo, P. W. (1995) Intransitivity of Indifference, Strong Monotonicity and the Endowment Effect, *Journal of Economic Theory*, **66**, 198-223

Slovic, P. and Lichtenstein, S. (Eds.) (2006) *The Construction of Preference*, Cambridge University Press, Cambridge, UK.

Smith, V. K. (2006) Judging Quality, in Kanninen, B. (Ed.) (2006) *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Dordrecht: Springer

Smith, V. K. and Desvouges, W. H. (1986) *Measuring Water Quality Benefits*, Kluwer, Boston

Soto Montes de Oca, G. and Bateman, I. J. (2006) Scope Sensitivity in Households' Willingness to Pay for Maintained and Improved Water Supplies in a Developing World Urban Area: Investigating the Influence of Baseline Supply Quality and Income Distribution upon Stated Preferences in Mexico City, *Water Resources Research*, **42**, W07421.

Spackman, M. (2004) Time Discounting and of the Cost of Capital in Government, *Fiscal Studies*, **25**(4), 467-518.

StataCorp (2009), Stata: Release 11, Statistical Software (College Station, TX: StataCorp).

Strotz, R. (1956) Myopia and Inconsistency in Dynamic Utility Maximisation, *Review of Economic Studies*, **23**, 165-180.

Sugden, R. (1999) Alternatives to the Neoclassical Theory of Choice, in Bateman, I. J. and Willis, K. (Eds.) (1999) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, Oxford University Press, Oxford, UK.

Train, K. (2009) *Discrete Choice Methods with Simulation*, 2nd Edition, Cambridge University Press, Cambridge, UK.

Tuan, T. H. and Navrud, S. (2007) Valuing Cultural Heritage in Developing Countries: Comparing and Pooling Contingent Valuation and Choice Modelling Estimates, *Environmental and Resource Economics*, **38**(1), 51-69.

Turnbull, B. W. (1976) The Empirical Distribution Function with Arbitrarily Grouped, Censored, and Truncated Data, *Journal of the Royal Statistical Society*, **38**, 290-295.

Tversky, A. and Kahneman, D. (1991) Loss Aversion in Riskless Choice: A Reference-Dependent Model, *Quarterly Journal of Economics,* **106**, 1039-61

Van Houtven, G., Powers, J. and Pattanayak, S. K. (2007) Valuing water quality improvements in the United States using meta-analysis: Is the glass half-full or half-empty for national policy analysis?, *Resource and Energy Economics*, **29**(3), 206-228.

Venkatachalam, L. (2004) The Contingent Valuation Method: A Review, *Environmental Impact Assessment Review*, **24**(1), 89-124.

Viscusi, W. K., Huber, J. and Bell, J. (2008) The Economic Value of Water Quality, *Environmental and Resource Economics*, **41**, 169-187.

Walsh, R. G., Sanders, L. G. and Loomis, J. (1985) *Wild and Scenic River Economics: Recreation Use and Preservation Values*, American Wilderness Alliance.

Wang, H. (1997) Treatment of “Don’t Know” Responses in Contingent Valuation Surveys: A Random Valuation Model, *Journal of Environmental Economics and Management*, **32**, 219-232.

Ward, F. A. (1987) Economics of Water Allocation to Instream Uses in a Fully Appropriated River Basin: Evidence from a New Mexico Wild River, *Water Resources Research*, **23**(3), 381-392

Weisbrod, B. A. (1964), Collective Consumption Services of Individual Consumption Goods, *Quarterly Journal of Economics*, **78**, 471-477.

Weitzman, M. L. (1998) Why the Far Distant Future Should be Discounted at its Lowest Possible Rate, *Journal of Environmental Economics and Management*, **36**, 201-208.

Weitzman, M. L. (1999) Just Keep on Discounting, But…, in Portney, P. and Weyant, J. (Eds.) (1999) *Discounting and Intergenerational Equity,* Resources for the Future, Washington DC.

Welsh, M. P. and Poe, G. L. (1998) Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach, *Journal of Environmental Economics and Management*, **36**(2), 170-185.

White, H. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, **48**(4), 817-838.

Whitehead, J. C., and Hoban, T. J. (1999) Testing for Temporal Reliability in Contingent Valuation with Time for Changes in Factors Affecting Demand, *Land Economics*, **75**(3), 453-465.

Willis, K. G. (2004) *Environmental Benefit Assessment*, Report Ref. No. 04/EQ/01/1, UK Water Industry Research, London, UK

Willis, K. G., McMahon, P. L., Garrod, G. D. and Powe, N. A. (2002) Water Companies' Service Performance and Environmental Trade-offs, *Journal of Environmental Planning and Management*, **45**(3), 363-379

Willis, K. G., Scarpa, R. and Acutt, M. (2005) Assessing Water Company Customer Preferences and Willingness to Pay for Service Improvements: A Stated Choice Analysis, *Water Resources Research*, **41,** W02019.

Woo, C.-K. (1994) Managing Water Supply Shortage: Interruption vs Pricing, *Journal of Public Economics*, **54**, 145-160.

Zerembka, P. (Ed.) (1974) *Frontiers in Econometrics*, Academic Press, New York.

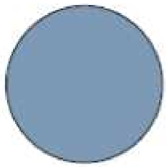
Zuang, J., Liang, Z., Lin, T. and De Guzman, F. (2007) *Theory and Practice in the Choice of Social Discount Rate for Cost-Benefit Analysis: A Survey*, Asian Development Bank, Manila, Philippines.

Appendix A Selected Show Materials (Study 1)

Figure 1. CARD 4a – Quality Levels

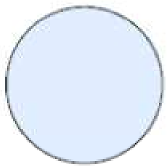
Dark Blue – quality is "High".

- There will be a diverse and natural range of plants, insects, fish, birds and other animals.
- Water will generally have the right degree of clarity and there will be no noticeable pollution.
- Water will generally be suitable for contact activities, such as rowing or wind surfing.



Mid-Blue – quality is "Medium".

- There will be plants, insects, fish, birds and other animals, but there will be some fish and other wildlife missing.
- Water will be slightly murky or discoloured in parts, and there will sometimes be visible pollution in some places, and some algal blooms.
- Water will be suitable for contact activities in some areas but not others.



Light Blue – quality is "Low".

- There may be limited or no plants or wildlife, or the water may be dominated by a single plant species.
- Water will generally be murky or discoloured, and may sometimes be bad-smelling in some places. There may also regularly be visible pollution in some places, and frequent algal blooms.
- Water will be unsuitable for contact activities.

Figure 2 CARD 4b – Lake

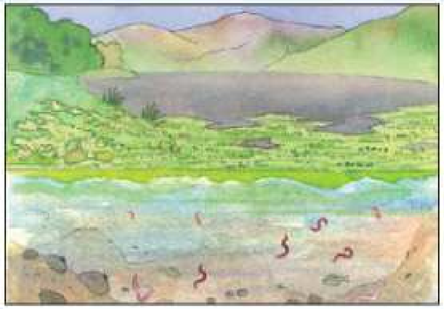
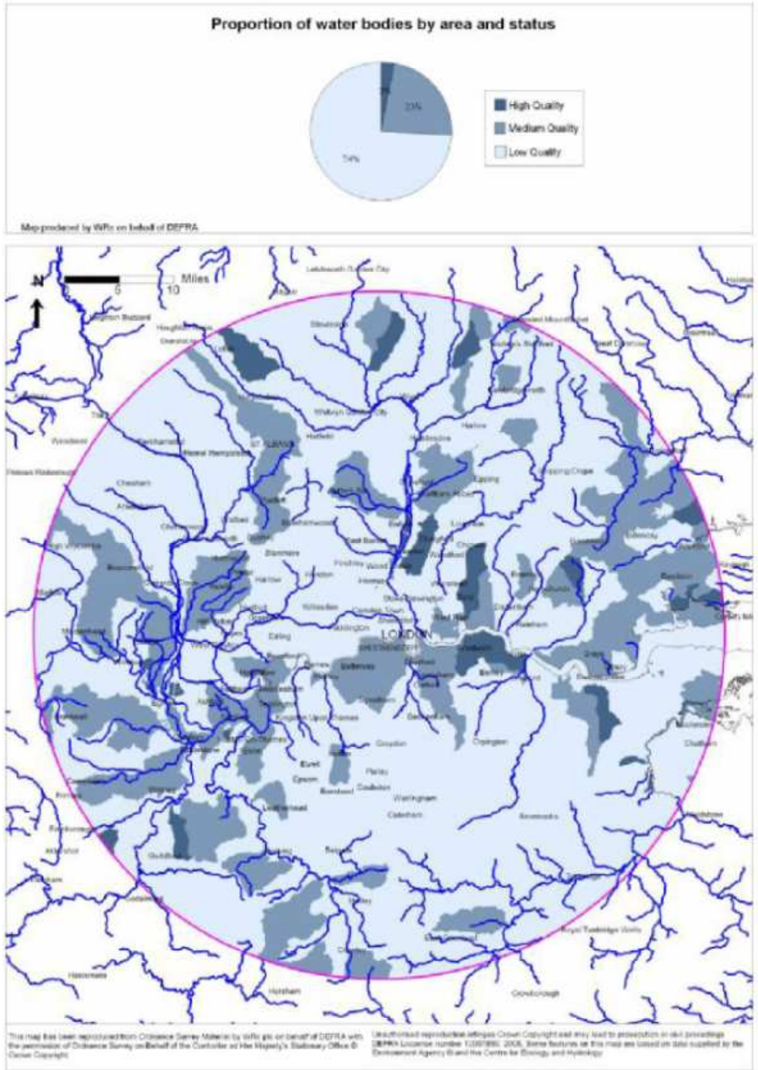
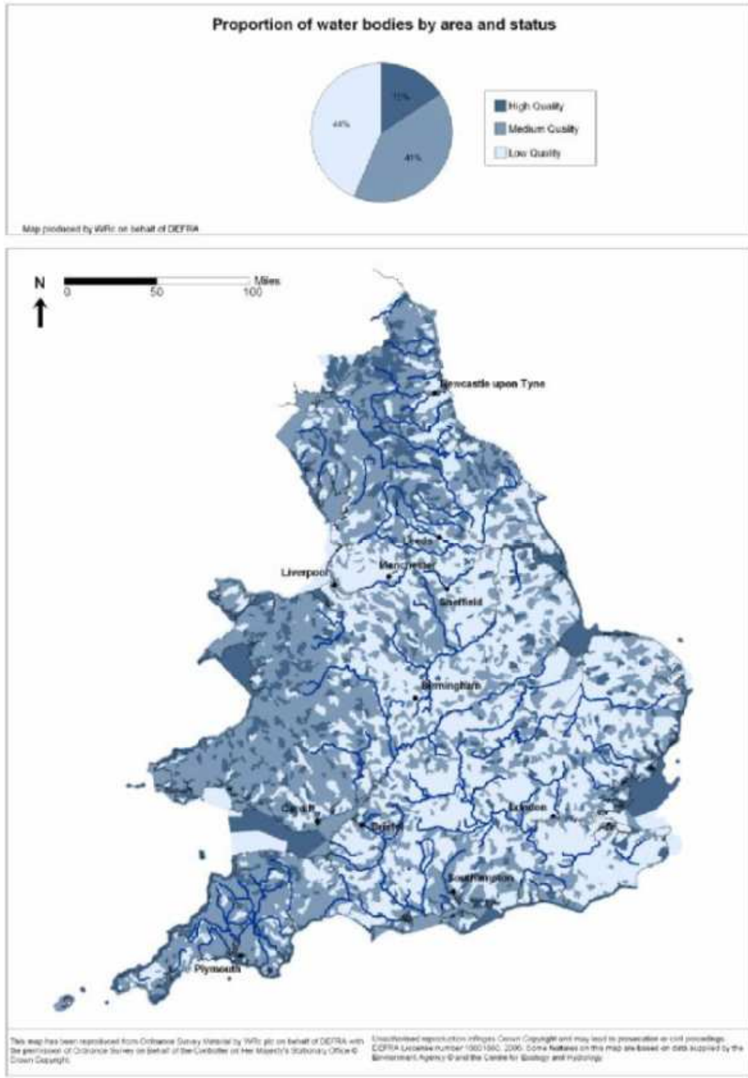
Quality	Illustration	Description
High Quality		A diversity of underwater plants, floating lilies, and tall flowering plants. Varied fish population, including trout and coarse fish. Insects such as dragonflies are present. Water with right degree of clarity and no noticeable pollution. Natural and seasonal variations in water levels. Suitable for contact activities.
Medium Quality		Some underwater and floating plants in shallow areas and around the lake. Some coarse fish and other animals present but limited. Insects are rare. Slightly unclear and occasionally discoloured water. Suitable for contact activities in some areas but not others.
Low Quality		Very few plants, except blanket weed, and very few fish or other animals, except worms and leeches. Cloudy, discoloured and possibly bad-smelling water. Unsuitable for contact activities.

Figure 3 Example Maps

Local Map



National Map



Valuation Context Statement

"Water quality is affected by pollution from households, farms and businesses, and climate change. Some works are needed just to prevent water sites from getting worse. The government's policy is that the polluter will have to pay for these works. This will make some every day products more expensive and will increase household water and sewerage bills too.

The government has estimated that these extra costs to each household, including yours, will be £10 per year, in terms of higher water and sewerage bills and higher prices on everyday products.

Improving the environment requires more cutting of pollution, which will make products more expensive and will further increase household water and sewerage bills.

I am now going to show you cards which have two or three options for water environment improvements. For all the options, steps will be taken so there will be no worsening of the water environment at any site, the most cost-effective works will be used, the money will be ring-fenced to make the improvements, and information will be made available to the public on progress towards the improvements.

It is important for us to get realistic choices from you regarding the values of these programmes, so before you make some real choices, please consider your household budget and all of the things that you and your household need or would prefer to spend your money on before you decide. Please also bear in mind that your water bill and other household expenses may change in future for other reasons not related to the water environment, and your income may also change in future. Your choices will influence how far to go with improvements, so will influence everyone's payment for improvements."

Figure 4 Example PCCV Card

High Quality	Medium Quality	Low Quality

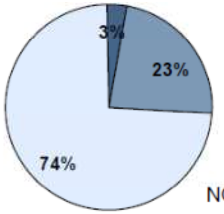
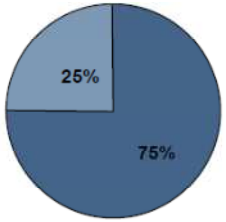
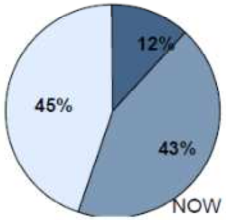
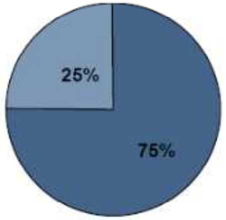
WILLINGNESS TO PAY CARD	Option A - No Change	Option B
Status of Local Area in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales and Local Area in 20 years time IN 2027	Same as Now	95%
Increase in your water bill and other household payments. Note: this payment will be added to the cost of avoiding any worsening of the water environment.	**No increase in water bills or other household payments**	**£ --------- per year** (continuing indefinitely)

Figure 5 Example DCCV Card

High Quality	Medium Quality	Low Quality

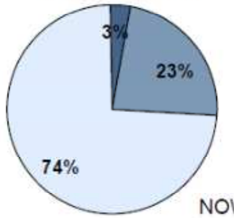
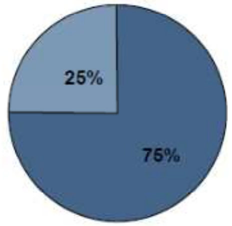
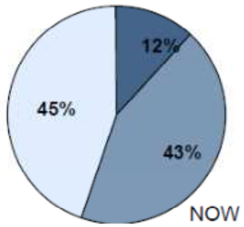
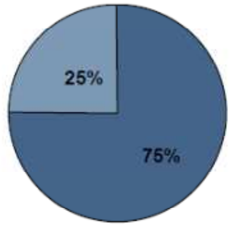
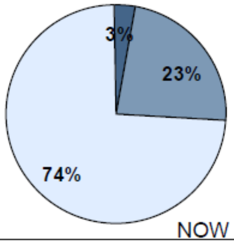
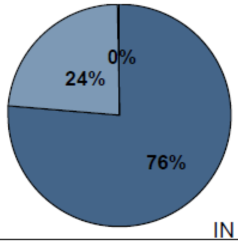
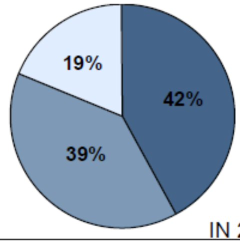
Choice Card 1	Option A - No Change	Option B
Status of Local Area in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales and Local Area in 20 years time IN 2027	Same as Now	95%
Increase in your water bill and other household payments. Note: this payment will be added to the cost of avoiding any worsening of the water environment.	**No increase in water bills or other household payments**	£ 30 per year
Remembering all the things that you could do with your money, which option would you choose, A or B		

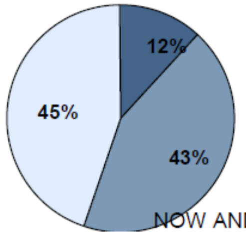
Figure 6 Example DCE Card

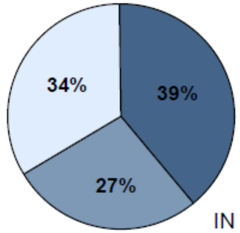
High Quality	Medium Quality	Low Quality

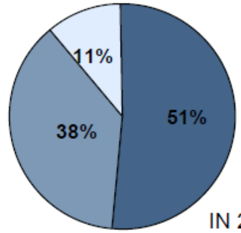
<table>
<tr><th>Choice Card 2</th><th>Option A - No Change</th><th>Option B</th><th>Option C</th></tr>
<tr><td>Status of Local Area
in 8 years time</td><td>

NOW AND 2015</td><td>

IN 2015</td><td>

IN 2015</td></tr>
<tr><td>Status of England and Wales
in 8 years time</td><td>

NOW AND 2015</td><td>

IN 2015</td><td>

IN 2015</td></tr>
<tr><td>Status of England and Wales and Local Area in 20 years time</td><td>Same as now</td><td>95%</td><td>95%</td></tr>
<tr><td>Increase in your water bill and other household payments.
Note: this payment will be added to the cost of avoiding any worsening of the water environment.</td><td>None</td><td>£ 20 per year</td><td>£ 30 per year</td></tr>
<tr><td></td><td>Choose Option A - No Change</td><td>Choose Option B</td><td>Choose Option C</td></tr>
<tr><td>Remembering all the things you could do with your money, which Option would you choose, A, B or C?</td><td></td><td></td><td></td></tr>
</table>

Appendix B Attribute Levels (Study 2)

Table B1: Water and Wastewater Service Levels by Scenario

	Current Service[a]	2008 Plan[b]	2009 Plan[c]
Tap Water Service and its Climate Change Impact			
Risk of severe water rationing (rota cuts and stand-pipes for up to 3 months)	Expected 1 in 87 years	Risk eliminated	Risk eliminated
Leakage from Thames Water pipes	25% of water lost[d] 27% of water lost[e]	20% of water lost (20% reduction)	24% of water lost (11% reduction)
Unplanned interruptions to water supply of greater than 6 hours	13,000 households have an interruption each year	10,000 households have an interruption each year (25% reduction)	9,000 households have an interruption each year (31% reduction)
Drinking water quality (complaints about taste, colour and smell)	1,600 complaints per year	1,500 complaints per year (6% reduction)	1,500 complaints per year (6% reduction)
Carbon dioxide emitted by Thames Water caused by tap water service	No change from current levels	Fall of 10% (out of total fall of 20%)	Fall of 10% (out of total fall of 20%)
Wastewater Service and its Impacts on River Water Quality and Climate Change			
Households affected by sewer flooding	2,300 households at risk each year[d] 1,620 households at risk each year[e]	1,700 households at risk each year (26 % reduction)	1,180 households at risk each year (27 % reduction)
Improved quality of rivers and estuaries	No improvements in river water quality	225 km of river has improved quality	368 km of river has improved quality
Households affected by smell from sewage treatment	23,000 households affected each year	7,000 households affected each year (83% reduction)	7,500 households affected each year (67% reduction)
Carbon dioxide emitted by Thames Water caused by wastewater service	No change from current levels	Fall of 10% (out of total fall of 20%)	Fall of 10% (out of total fall of 20%)

*Notes: **a** "Current" levels of service shown were the same in both 2008 and 2009 surveys, except where indicated with notes d and e. **b** This column shows the levels of the improvement plan shown in the 2008 survey. **c** This column shows the levels of the improvement plan shown in the 2009 survey. **d** Level shown in 2008 survey only. **e** Level shown in 2009 survey only.*