Chapter 17

CONTINGENT VALUATION

RICHARD T. CARSON
University of California, San Diego, USA

W. MICHAEL HANEMANN
University of California, Berkeley, USA

Contents

Handbook of Environmental Economics, Volume 2. Edited by K.-G. Mäler and J.R. Vincent
© 2005 Elsevier B.V. All rights reserved
DOI: 10.1016/S1574-0099(05)02017-6

Abstract

Value estimates for environmental goods can be obtained by either estimating preference parameters as "revealed" through behavior related to some aspect of the amenity or using "stated" information concerning preferences for the good. In the environmental economics literature the stated preference approach has come to be known as "contingent valuation" as the "valuation" estimated obtained from preference information given the respondent is said to be "contingent" on the details of the "constructed market" for the environmental good put forth in the survey.

Work on contingent valuation now typically comprises the largest single group of papers at major environmental economics conferences and in several of the leading journals in the field. As such, it is impossible to "review" the literature *per se* or even cover all of the major papers in the area in some detail. Instead, in this chapter we seek to provide a coherent overview of the main issues and how they fit together.

The organization of the chapter is as follows. First, we provide an overview of the history of contingent valuation starting with its antecedents and foundational papers and then trace its subsequent development using several broad themes. Second, we put forth the theoretical foundations of contingent valuation with particular emphasis on ties to standard measures of economic welfare. Third, we look at the issue of existence/passive use considerations. Fourth, we consider the relationship of contingent valuation to information on preferences that can be obtained by observing revealed behavior and how the two sources of information might be combined. Fifth, we look at different ways in which preference information can be elicited in a CV survey, paying particular attention to the incentive structure posed by different elicitation formats. Sixth, we turn to econometric issues associated with these different elicitation formats. Seventh, we briefly consider survey design issues. Eighth, we look at issues related to survey administration and extrapolating the results obtained to the population of interest. Ninth, we describe the major controversies related to the use of contingent valuation and summarize the evidence. Finally, we provide some thoughts on where we think contingent valuation is headed in the future.

Keywords

contingent valuation, nonmarket valuation, consumer decision making, public goods, environmental economics

JEL classification: Q510, D610, D120, H400

1. Introduction[1]

Few environmental goods are bought and sold in the marketplace. For economists to move beyond an analysis of the cost-effectiveness of providing a specified level of a particular environmental good it is necessary to have some way of estimating the value of providing different levels of the amenity relative to its cost. Such values are most naturally expressed in monetary terms although other metrics are possible. These value estimates can be obtained by either estimating preference parameters as "revealed" through behavior related to some aspect of the amenity or using "stated" information concerning preferences for the good. In the environmental economics literature the stated preference approach has come to be known as "contingent valuation," as the "valuation" estimate obtained from preference information given that the respondent is said to be "contingent" on the details of the "constructed market" for the environmental good put forth in the survey.[2] The focus of this chapter is on contingent valuation (CV).

Contingent valuation is an inherently more flexible tool than revealed preference techniques such as hedonic pricing and the household production function approach. This is because it is possible in principle to use CV to examine environmental goods and terms for providing them that are different from what has been observed now or in the past. It is also possible in principle to create CV scenario experiments that avoid many of the economic modeling problems that are common to most observational data. Contingent valuation is also the only approach that can generally be used to include what is usually referred to as the existence or passive use component of the economic value of an environmental good. Offsetting these advantages are the problems that can arise with surveys and the reluctance by some economists to rely on information obtained from surveys. Such reluctance, however, is not neutral in terms of its consequences for policymaking.

Rather than seeing an inherent conflict between revealed and stated preference techniques, it is more productive to view the two approaches as complementary but having different strengths and weaknesses. Indeed, it is sometimes possible and useful to combine the two approaches.

The great advantage of the revealed preference approach is, of course, that it is based on actual behavior. The difficulty is that the tie between that behavior and the environmental good of interest is often complex and estimation of the implied economic value

[1] This work draws heavily upon our earlier work, and in particular: Carson (1991, 1997a, 1997b, 2000, in press), Carson, Flores and Mitchell (1999), Carson, Flores and Meade (2001), Carson, Groves and Machina (2000), Hanemann (1992, 1994, 1995, 1999), Hanemann and Kanninen (1999) and Mitchell and Carson (1989).

[2] A number of other terms have been used for deriving economic values from stated preference information in surveys. These terms often refer to the specific format in which the preference information is elicited and include among others: binary discrete choice, bidding game, direct (open-ended) question, (multinomial) choice experiment, choice-based conjoint analysis, contingent ranking, double-bounded dichotomous choice, paired comparison, payment card, and referendum question.

placed on a change in the environmental good is highly dependent upon both the underlying theoretical model postulated and the nature of the econometric assumptions made. Less frequently recognized is that the observed behavior takes place in the context of a particular market structure that may be substantially different from that in which the policy change is going to occur.

Contingent valuation has the opposite set of characteristics. The tie between the underlying theoretical model and the information on preferences obtained in the survey is usually quite close. There are, of course, econometric assumptions to be made but these tend to be different in nature and relate to aspects of the distribution of economic values and the nature of the survey data collection effort.

CV surveys differ from other surveys on public policy issues in several important ways. First, a major portion of the survey is devoted to a description of the public good (or goods) of interest. Second, the elicitation of preference for the good is more extensive and nuanced than in a typical opinion survey. Moreover, it involves the elicitation of monetary (Hicksian) measure of welfare: maximum willingness-to-pay (WTP) to obtain a desired good not currently possessed, or minimum compensation (WTA) to voluntarily give up a good currently possessed. CV surveys have been used to value large discrete changes such as the introduction of a new public good, the value associated with substituting one good for another, or the marginal value associated with changing one or more attributes of an existing good. CV surveys are generally organized in the following manner which reflects current practice: (1) an introductory section identifying the sponsor and general topic, (2) a section asking questions concerning prior knowledge about the good and attitudes toward it, (3) the presentation of the CV scenario including what the project was designed to accomplish, how it would be implemented and paid for, and what will happen under the current *status quo* situation if the project were not implemented, (4) question(s) asking for information about the respondent's WTP/WTA for the good, (5) debriefing questions to help ascertain how well respondents understood the scenario, and (6) demographic questions. Mitchell and Carson (1989) provide a comprehensive overview of the issues involved in the design and analysis of CV surveys, and Bateman et al. (2002) provide a useful manual for the practitioner.

In spite of the apparent simplicity that some see in asking people whether they would be willing to pay some specific amount for a given item, contingent valuation has its drawbacks with regard to providing useful information for policymakers. Much of this stems from the fact that economists are not generally trained in the design and administration of surveys. Indeed, much of the usefulness of conducting a CV study has nothing to do with explicitly obtaining an estimate of monetary value.

There is often not adequate awareness among policymakers and many economists that the choice of the characteristics of the market constructed in the survey can, and generally does, influence the nature of the economic valuation estimates obtained. In this sense, the flexibility of contingent valuation is both a blessing and a curse in that, unless adequate attention is paid, it is possible to obtain estimates that are not directly tied to the relevant policy changes being considered by decision makers. A good CV survey lays out the current *status quo* level of the good of interest and the potential

change in that level, the manner in which that change could be supplied, and how it would be paid for by the agent if supplied. It does so in a way that is both acceptable to the technical experts at governmental agencies and understandable to the general public. This brief description of the initial product of a CV study should also make clear that what is typically being valued is a program to provide the environmental good and not the environmental good alone.

The empirical results from a CV study provide a wealth of information about the population of interest's value of the program. Some of this is in the form of aggregate information, such as the aggregate of WTP used in neoclassical benefit–cost analysis or a summary statistic like median WTP (which has a well-known voting interpretation). Contingent valuation itself is agnostic as to the appropriate summary measure that decision makers should use. Contingent valuation studies generally generate information about the broad shape of the WTP (or WTA) distribution and typically information on how that distribution varies with respondent characteristics such as, income, geographic location, and the nature of the use of the environmental good. The nature of this heterogeneity in economic values is often of great importance to decision makers and explaining it is a key task in a CV study.

It is hard to overestimate the central importance of contingent valuation to modern environmental economics. Work on contingent valuation now typically comprises the largest single group of papers at major environmental economics conferences and in several of the leading journals in the field. Carson (in press) provides a bibliography spanning fifty years with over six thousand CV papers and studies from over one hundred countries. As such, it is impossible to "review" the literature *per se* or even cover all of the major papers in the area in any detail. Instead, we seek to provide a coherent overview of the main issues and how they fit together.

The vastness of the literature on contingent valuation also suggests another point that cannot be made too strongly. Contingent valuation is a generic approach to collecting survey information about agents' preferences in the context of a constructed market situation. It is impossible to make any valid general statement about the properties of contingent valuation without specifying more completely the nature of the application and the quality of its implementation. And, it is impossible for a single experiment or statistical test to say anything of substantial generality concerning contingent valuation [Randall (1998)]. Such tests, however, may be quite informative as to the properties of a particular type of CV application in a particular context.

The organization of this chapter is as follows. First, we provide a historical overview of contingent valuation starting with its antecedents and foundational papers and then trace its subsequent development using several broad themes. Second, we put forth the theoretical foundations of contingent valuation with particular emphasis on ties to standard measures of economic welfare and specification of models. Third, we look at types of value including issues concerning existence/passive use considerations and consider the relationship of CV to information on preferences that can be obtained by observing revealed behavior. Fourth, we look at different ways in which preference information can be elicited in a CV survey, paying particular attention to the incentive structure

posed by different elicitation formats. Fifth, we turn to econometric issues associated with these different elicitation formats. Sixth, we briefly consider survey design issues. Seventh, we look at issues related to survey administration and extrapolating the results obtained to the population of interest. Eighth, we describe the major controversies related to the use of contingent valuation and summarize the evidence. Ninth, we look at the consistent of CV results with actual behavior. Finally, we provide some thoughts on where we think contingent valuation is headed in the future.

2. History of contingent valuation

2.1. Antecedents and beginnings

Even though economists have largely focused on market prices as the indicator of economic value, earlier writers such as Clark (1915) and Hines (1951) clearly saw that much of an individual's utility was driven by unpaid costs and uncollected benefits and that "market prices" did not exist for many of the more interesting quantities to economists. A theory of public goods developed through the work of economists such as Lindahl helped formalize the notion of an equilibrium set of shadow prices for public goods and the difficulties involved in obtaining them. Other economists such as Pigou began to articulate environmental harm as unpriced externalities that drove a wedge between social cost and private cost making market prices suspect for some purposes.

Bowen (1943) and Ciriacy-Wantrup (1947) were the first to propose the use of specially structured public opinion surveys to value what Bowen called "social goods" and Ciriacy-Wantrup "collective, extra-market goods" – goods such as "beautification of the landscape" (Bowen) or soil conservation (Ciriacy-Wantrup) that "cannot easily be sold to individual consumers and the quantities available to different individuals cannot be adjusted according to their respective tastes" (Bowen). Both Bowen and Ciriacy-Wantrup saw that a distinctive feature of these goods was that, while individuals would have their own distinctive demand curves for these goods, the aggregate demand curve is "obtained by adding the marginal rate of substitution (expressed in money) of the various individuals at each possible quantity of the social good (vertical addition)" (Bowen). The practical problem was to estimate the individual marginal rate of substitution curve "since it requires the measurement of preference for goods which, by their nature, cannot be subject to individual consumer choice." Bowen suggested voting as "the closest substitute for consumer choice" and he noted the possibility of using "polls, questionnaires, interviews" as a means to implement this: "If polls are based on a representative sample of the population, and if questions are put in the same way as if the entire citizenry were voting, the results can, of course, be interpreted in exactly the same way." Ciriacy-Wantrup (1947) covers the same ground and develops the same argument in the context of addressing the difficulties of measuring the benefits of soil conservation programs, and he reiterated his call for the use of the "direct interview method" in his influential book *Resource Conservation: Economics and Policy* (1952), which is often

considered the first text book on environmental and resource economics. One major obstacle to Bowen and Ciriacy-Wantrup's calls for the use of surveys to measure benefits of public goods was that they soon clashed with Samuelson's seminal paper [Samuelson (1954)]. In this paper, Samuelson points out the problem of potential strategic behavior when aggregating over individual agents to get the benefit of providing a public good. Samuelson notes: "It is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective activity than he really has." In the penultimate paragraph of the paper, Samuelson pursues this point further with specific reference to the use of surveys:

> One could imagine every person in the community being indoctrinated to behave like a "parametric decentralized bureaucrat" who reveals his preferences by signaling in response to price parameters or Lagrangian multipliers, or to questionnaires or to other devices. Alas, by departing from his indoctrinated rules, one person can hope to snatch some selfish benefit in a way not possible under the self-policing competitive pricing of private goods.[3]

At roughly the same time a different debate played out between Lester (1946) and Machlup (1946) over whether companies were fully profit maximizing and pricing at marginal cost. Lester's work drew heavily on survey responses from businesses in interviews which suggested that they were *not* consciously maximizing profit in the neo-classical fashion. Machlup's attack relied heavily on disparaging the credibility of survey responses. Milton Friedman later weighed in with an instrumental view that people may effectively act like the rational maximizer of economic theory but did not realize they were doing so. As such, survey responses might be meaningless; and Friedman's analogy of the professional pool player not knowing the underlying physics behind the shot has long become part of the folklore of economists.[4]

Samuelson and Friedman's distrust of surveys played itself out in a variety of different areas of economics and to distance applied economic fields such as marketing and transportation from the economic mainstream.[5] What has always driven economists back to the idea of trying to collect nonfactual information for individual agents

[3] In this regard, the early experiment of Bohm (1972) played a key role by suggesting that strategic behavior might not play as large a role as once feared. The concept of strategic behavior was to become a major theme in the use of CV but not one that ultimately blocked its use [Mitchell and Carson (1989)] as economic theorists [e.g., Groves (1973)] started to explore the incentives faced by economic agents in different situations in a much more nuanced manner.

[4] Boulier and Goldfarb (1998) provide an interesting account of this part of the debate on the use of surveys by economists. Even though Machlup's view prevailed at the time, Lester might well be judged the ultimate victor, as a substantial amount of the research in industrial organization in recent years has gone toward explaining the "anomalies" he first identified.

[5] There is, of course, some irony in that almost all of applied microeconomics relies heavily on the use of data collected with survey instruments, typically by government agencies or in the form of large panel data sets like the Panel Study on Income Dynamics. Little attention is currently given by economists to the design of surveys to collect such data even though this was once an active area of research [e.g., Lansing and Morgan (1971)].

is the pressing need for it and lack of any other good way to get the data.[6] Nowhere has this need been more pressing than with public goods, where the major impediment to performing a benefit–cost analysis in many fields of applied economics is the lack of monetary values for the outputs of government policies.

2.2. Early empirical development

The first area where the lack of monetary numbers for a key output of government projects was considered to be a major problem was outdoor recreation. Here, there were two driving forces. The first was land based and involved the U.S. National Park Service and the U.S. Forest Service. The number of people seeking to recreate on government land exploded in the post World War II era, leading policymakers to recognize the need to know what people wanted and how much they were willing to pay for it. The obvious way to do this was to survey the public. The National Park Service engaged Audience Research Inc., a well-known marketing firm (associated with Gallup surveys) that specialized in determining the market value of products (such as movies that had not yet been released), to find out what the public wanted with respect to its national parks. These surveys were instrumental in developing the U.S. national park system in the 1950s and 1960s. Money to run the National Parks was always tight and to explore this issue Audience Research (1958) asked one set of survey respondents about their willingness to pay a day use fee. This question can be seen as the immediate precursor to the use of contingent valuation as an input to the economic evaluation of a project.

The second driving force was water based and involved the U.S. Army Corp of Engineers and other agencies building the major water projects of the 1950s and 1960s. It was on these projects that benefit–cost analysis evolved into its modern form and became institutionalized as a key component of government decision making.[7] Water based recreation along with electricity and flood control were the key outputs of many water projects. Such projects looked considerably less attractive if recreation was unpriced. This led to considerable interest in developing methods to reliably place a monetary value on different types of outdoor recreation.

Davis (1963a) was the first economist to empirically implement a CV survey in his Harvard dissertation entitled "The value of outdoor recreation: an economic study of the Maine woods."[8] Surprisingly, Davis was unaware of Ciriacy-Wantrup's suggestion

[6] See, for instance, Arrow (1958), Schelling (1968), Mishan (1976), McFadden (1986), Blinder (1991), and Manski and Straub (2000).

[7] In particular, see U.S. Congress, Senate Committee on Public Works (1957), Eckstein (1958), and Krutilla and Eckstein (1958). Hanemann (1992) provides an interesting account of the key role of water projects in driving developments in benefit–cost analysis.

[8] Davis notes as earlier antecedents the 1958 Audience Research study noted above, a master's thesis by Bruce Stewart (1961) at the University of Maine that asked about willingness to pay for improvements respondents thought desirable, and unpublished work by Ullman and Volk (1961) at Washington University (St. Louis) that asked respondents about their willingness to travel farther to recreate at a new lake. None of these studies bears much resemblance to the Davis dissertation which even today seems remarkably "modern" in its conceptualization of many of the major issues involving the use of CV.

to interview people to measure values associated with natural resources.[9] Davis was, however, strongly influenced by Stanley Stouffer, one of the country's leading academic survey researchers, who at the time was teaching a course in survey methods in Harvard's Social Relations Department. After sitting in on Stouffer's course, Davis reasoned that it should be possible to "approximate a market" in a survey by describing alternative kinds of areas and facilities to make available to the public, and then simulate market bidding behavior."[10] This method he later wrote (1963b), "would put the interviewer in the position of a seller who elicits the highest possible bid from the user for the services being offered."

Davis's dissertation sparked considerable interest in the technique. His dissertation is remarkable from several perspectives. It is comprehensive in terms of its treatment of theoretical economic issues, survey design and sampling issues, as well as statistical issues related to the analysis and interpretation of the data collected. Davis foresaw many of the key issues that later CV researchers would grapple with and researchers currently working on CV studies can still benefit from the insights in his dissertation.

Davis approached contingent valuation from a public finance perspective and was concerned with the need, in the context of a benefit–cost analysis, to measure all of the benefits and costs in monetary terms.[11] To quote Davis (1963a, p. 3):

> The temptations to patch up a benefit–cost statement for recreation in federal projects notwithstanding, the conclusions about our inability to do this in the absence of market prices or their surrogates are incontrovertible. Indeed, if there are no market valuations available, then the outdoor recreation services produced by an investment must be lumped with the intangible results and subjected to the nonquantitative, verbal analysis. At best a marginal production cost can be assigned to outdoor recreation facilities and the valuation of benefits arrived through administrative and political process.

Davis later compared a CV estimate to a corresponding estimate based on the travel cost method [an indirect approach developed at roughly the same time at Resources for the Future by Clawson and Knetsch (1966)] and found the two approaches produced similar estimates [Knetsch and Davis (1966)]. This was to be the first test of "convergent validity" whereby, the estimates from two different valuation techniques expected to produce similar estimates are compared.

[9] Davis was aware of Ciriacy-Wantrup's view that public natural resource problems involved issues of external benefits and costs and cites a chapter by him in an edited volume to this effect.

[10] Davis, personal communication, June 16, 1986.

[11] Otto Eckstein was Davis's dissertation adviser. Davis credits John Kenneth Gailbraith with being "instrumental in getting the study launched as a thesis topic." Resources for the Future provided Davis with a dissertation grant in 1961 and employed him in Washington during the final year of working on his dissertation. While at Resources for the Future, Davis was influenced by Marion Clawson, Irving Fox, John Krutilla, Allen Kneese, and Jack Knetsch.

Influenced by Davis, Ronald Ridker (1967) used the CV method in several studies of air pollution benefits. Although the primary thrust of Ridker's work was to value reducing dirty and soot from air pollution by using the hedonic pricing approach [Ridker (1967), Ridker and Henning (1967)], it was his recognition that people might value air pollution because of its "psychic costs" that led him to include a couple of WTP questions in two different surveys he conducted in Philadelphia and Syracuse in 1965. These questions, which asked people how much they would be willing to pay to avoid "dirt and soot" from air pollution, lacked many of the refinements that were later developed for CV surveys. In reflecting on his experience with the survey questions, including the WTP questions, Ridker made an observation that prefigured the later developments in CV research:

> It now seems evident that a much narrower, deeper, and psychologically sophisticated questionnaire is required to measure and untangle the various determinants of cleaning costs. Such a questionnaire would require substantially more time and expenditure-perhaps three to four times greater-than went into this essentially exploratory study. Even then the probability of success may not be very high, for such a project would raise problems in survey design that have not yet been solved. [Ridker (1967, p. 84)]

Over the next few years several other economists associated with Resources for the Future followed Davis's lead and used the CV approach to value various recreational amenities. In 1969, Brown and Hammack (1972), [Hammack and Brown (1974)] sent a mail questionnaire to a large sample of western hunters that asked them how much they would be willing to pay for and willing to accept to give up their rights to hunt waterfowl.[12] The $30 price tag from this work was eventually adopted by state and federal fish and game agencies for determining the value of a waterfowl kill and for estimating the benefits of habitat purchases. This study is also noteworthy for its ecosystem orientation. Next, in 1970 survey, Cicchetti and Smith (1973, 1976a, 1976b) asked individuals who were hiking in a wilderness area how much they would be willing to pay to reduce congestion in the area from other hikers. Influenced by Davis and other work at Resources for the Future, the Department of the Interior's Fish and Wildlife Survey of 1975 contained a question that asked respondents how much costs would have to increase before outdoor recreation would be given up [Brown, Charbonneau and Hay (1978)].

Work was also beginning outside of Resources for the Future's direct sphere of influence. A number of researchers including Miles (1967), LaPage (1968), Sarja (1969), Pattison (1970), Beardsley (1971), Randall (1973), Berry (1974), Meyer (1974), Shechter, Enis and Baron (1974), Gluck (1974) all used some variant of CV techniques

[12] In 1967, Mathews and Brown (1970) had undertaken an earlier study to value sport fishing using a simple survey based approach.

to value recreation.[13] Perhaps the most noteworthy of these studies was Darling (1973). Darling, influenced by Davis's work, used personal interviews to ask about willingness to pay for the amenities of three urban parks in California. He compared these estimates to those derived from a property value model and found the estimates from the survey to be on average lower than those estimated using property values. Jack Sinden (1974) also began his work on placing monetary values on recreational amenities at about this time and endeavored to trace out indifference curves. This work later resulted in one of the first books on nonmarket valuation focused on outdoor recreation [Sinden and Worrell (1979)].

Turning back to valuing the amenities of pollution control pioneered by Ridker, in the summer of 1972 Alan Randall and his colleagues used CV to study air visibility benefits in the Four Corners area in the southwest [Eastman, Randall and Hoffer (1974), Randall, Ives and Eastman (1974)].[14] This study, which built directly upon the bidding game approach put forth by Davis, plays a central role in the development of contingent valuation and is discussed in more detail in the next section. Hanemann also studied pollution, this time water pollution, by asking a sample of people in 1974 how much they were willing to pay for improved water quality at beaches in the Boston area [Hanemann (1978), Binkley and Hanemann (1978)]. Hanemann compared his findings with those obtained by a generalized travel cost model based on a random utility model, again finding reasonable correspondence.

Accompanying all of this empirical work was substantial advances on the theoretical side. Weisbrod (1964) had shown the potential importance of option value, Krutilla (1967) the potential importance of existence value and Arrow and Fisher (1974) the potential importance of quasi-option value. The import of all three papers was to suggest that there were many potential economic effects that were not adequately, if at all, reflected in market prices. These effects were thought by many to be particularly important in environmental decisions where irreversibility, a growing information basis, and uncertainty were nearly always present. Currie, Murphy and Schmitz (1971) put forth a coherent view of Hicksian consumer surplus measures while Mäler's (1974) seminal book provided a coherent theoretical view of environmental economics as a whole and the key role environmental externalities played. Freeman's widely used 1979 text on measuring environmental benefits was moving toward the now standard triumvirate of hedonic pricing, travel cost models, and willingness to pay surveys as the approaches to

[13] Sarja (1969), a Finnish study on lake recreation, appears to have been the first European CV study, Pattison (1970), a University of Alberta's master's thesis appears to have been Canada's first CV study, Gluck appears to have been the first CV study from Oceania, while Shechter, Enis and Baron (1974), an Israeli study, appears to have been the first study done outside of an OECD country.

[14] The pace of European CV studies also starts to increase during this time period. The original British studies looked at noise [i.e., Plowden (1970), Roskill (1970), Plowden and Sinnott (1977)] took place during this time. There is also a Dutch study by Jansen and Opschoor (1973) that looks at airport noise. Other early American studies looking at pollution were Barrett and Waddell (1973) and Reizenstein, Hills and Philpot (1974) who focused on urban air pollution while Oster (1977) and Gramlich (1977) focused on water pollution in the Merrimack and Charles Rivers, respectively.

obtaining values for nonmarketed goods.[15] Hanemann's work on random utility models (1984a, 1984b) was to have a substantial influence on both contingent valuation and travel cost models and made it clear that both approaches shared the same underlying theoretical framework.

2.3. Health, transportation, and the allocation of public budgets

Outside of the environmental arena, there were two other areas where surveys began to be used for the purpose of evaluating economic policies: health and transportation.[16] Interest in health applications stemmed from dissatisfaction with the human-capital approach to valuing health programs that placed little value on programs that saved mainly old people or young children. Acton (1973) was the first to undertake a serious empirical application here.[17] Acton applied the method to valuing programs such as improving ambulance services that reduced the risk of dying from a heart attack. It appears he did not know about Davis's work, but rather, was influenced by the work of the 1950s and 1960s on choice under uncertainty. Like Davis, Acton had enormous insights into the problems faced in setting up a situation whereby respondents could give up money to reduce low-level risks. Acton realized the difficulties in getting across such risks and notes that his own approach "is to state each probability calculation in several equivalent forms for each person so that he can choose the form most meaningful to him." Many of Acton's findings carry over to today's health risk studies.

Michael Jones-Lee in a seminal theoretical paper (1974) derived many of the properties that consumer surplus measures with respect to risk reductions. Jones-Lee, influenced like Acton, by Schelling and Mishan's call to use surveys to look at willingness to pay for reducing risk to human life, did so in his book [Jones-Lee (1976)]. Activity on valuing health risk using surveys was slow after this, in part due to the opposition of health policymakers to the concept of making decisions on the basis of willingness to pay. Maximizing quality adjusted life years (QUALY) subject to a budget constraint became the dominant paradigm in health policy, in spite of the dubious nature of its ties

[15] This process would be completed by the time of Freeman's (1993) book on the same topic. The well-known advanced text on welfare economics by Just, Hueth and Schmitz (1982) was perhaps the first to effectively take this position from a consistent theoretical stance.

[16] There was also relevant early work in marketing, such as Fiedler (1972) and Johnson (1974). Marketing research, however, turned in a more *ad hoc* psychological direction until developments in random utility theory [McFadden (1986)] provided a firmer economic foundation. Louviere (1994) and Louviere, Hensher and Swait (2000) provide an account of later developments in this area.

[17] The use of surveys to value reductions in risks gained a boost from articles by two well-known economists [Schelling (1968), Mishan (1971)], who noted that the other techniques for measuring benefits in these situation were either unavailable or necessitated a large number of implausible assumptions. They concluded that there was everything to gain and nothing to lose from trying the survey approach of asking people for their willingness to pay. Acton notes as the only predecessors to his survey work a draft survey instrument by Schelling and an unpublished paper by a University of Southern California master's student [Palmatier (1969)]. See Melinek (1974), another early study looking at willingness to pay for risk reductions.

to standard welfare economics.[18] Opposition by economists to both the human capital and QUALY approaches led to an interest in estimating risk premiums (and the statistical value of life) from hedonic wage equations. This approach was pioneered by Thaler and Rosen (1976).[19] On the health valuation side, further CV advances were stalled until the 1980s when interest in morbidity effects [e.g., Loehman and De (1982), Berger et al. (1987), Magat, Viscusi and Huber (1988)] and growing recognition of the difficulties involved in hedonic wage equations to value some types of mortality risks, fueled renewed interest in the use of surveys [e.g., Jones-Lee, Hammerton and Philips (1985), Gerking, de Haan and Schulze (1988), Eastaugh (1991)].[20]

In later years, the stated preference research in the area of health economics would encompass a wide variety of other areas of interest to policy ranging from various drug therapy options [Johannesson and Fagerberg (1992)] and discount rates for treatment options [Ganiats et al. (2000)] to the valuation of pharmacy services [Reardon and Pathak (1988)] and willingness to pay to reduce time on waiting lists [Propper (1990)].

In transportation, Jordan Louviere was the primary initiator of using stated preference data to look at transportation issues. Louviere was a geographer with a background in psychology. His first work (1974) dealt with the choice of trout fishing location as a function of the site attributes.[21] In many ways this application can be seen as one of the early outdoor recreation examples with the focus on investigating how distance (travel time) and stream quality (trout per quarter mile) influence stream preference rather than valuing a fishing day per se. Louviere's later work shifted to modeling the determinants of transportation mode choice (e.g., bus versus car) and to a random utility framework [Meyer, Levin and Louviere (1978)].[22] While the original focus of Louviere's work was not on valuation, the shift to a random utility framework [Louviere and Hensher (1983)] made the valuation of marginal changes in attributes straightforward. Later work in transportation would concentrate on the value of time [Hensher and Truong

[18] Ironically, an economist played an important role in the early development of the QUALY concept, since Zeckhauser and Shepard (1976) was one of the seminal papers on QUALY. This work grew out of the interest in multi-attribute utility assessment (MAUT), which breaks the elicitation of a multivariate utility function down through the elicitation of a set of separate univariate utility functions for individual outcomes such as health status [Keeney and Raiffa (1976)].

[19] See Viscusi (1993) survey article in the *Journal of Economic Literature* for a discussion of estimating the statistical value of life from hedonic wage regressions.

[20] For a comprehensive review of the issues involved in measuring health impacts see Johansson (1995).

[21] The other early stated preference study in transportation was by Davidson (1973) who undertook a survey aimed at forecasting potential traffic on short takeoff and landing aircraft. While Davidson's work did not have the influence that Louviere would have on modeling stated preferences, it did highlight one of the major issues in transportation, the almost complete lack of information on the impacts of introducing a new form of transportation in a data set on revealed behavior concerning transportation choices.

[22] In this, Louviere followed his dissertation advisor's footsteps with revealed preference data [Rushton (1969)]. Stated preference data proved to be better suited for this task when introducing new transportation modes or substantial changes from existing conditions. In part this was due to being able to control the attributes of choices that respondents saw through experimental design. Louviere and Woodworth (1983) is the seminal contribution with respect to the experimental design of such choice experiments.

(1985)] and the valuation of new forms of transportation such as alternative fuel vehicles [Brownstone, Bunch and Train (2000)].[23]

There was a third strand of work involving the use of surveys to elicit information about preferences for public goods. This involves asking respondents about how they would make trade-offs between multiple goods by allocating a budget. Some of the earliest work in this area was done by Peterson and Worrall (1970) who looked at tradeoffs involving access to different neighborhood services. Hoinville and Berthoud (1970) gave respondents a hypothetical budget and asked them to make tradeoffs involving travel time, road safety, vehicle pollution and vehicle congestion in London; and Pendse and Wyckoff (1972) looked at tradeoffs between different water resources.[24] The allocation game concept saw its full realization in Beardsley, Kovenock and Reynolds (1974) who looked at allocating resources to different national priorities, Strauss and Hughes (1976) who looked at substitution effects among budget categories, and Hardie and Strand (1979) who looked at how respondents would reallocate parts of the U.S. Forest Service's budget. Unfortunately, there were several reasons why this line of work never really took off. First, specification of a complete set of all the options was very tedious if done in sufficient detail for respondents to understand. Second, while calculation of marginal rates of substitution at the individual level was straightforward, aggregation across agents or calculation of willingness to pay measures was problematic. Third, the theoretical framework for this type of modeling never found the acceptance that the random utility approach that came to dominate most contingent valuation work did. From a theoretical perspective, Carson, Flores and Hanemann (1998) show that the valuation of a good is dependent upon the sequence in which it is valued (due to both substitution and income effects) making it difficult to meaningfully aggregate the independent budget allocations of different agents.

2.4. The existence value revolution

CV surveys were initially seen as having three distinct advantages. First, CV can obtain useful information where data on past consumer behavior had not been collected. Second, CV permits the creation and presentation of scenarios that provide new goods or changes in existing goods that were substantially outside the range of current consumer experience. Third, CV allows measurement of the desired Hicksian consumer surplus measure rather than its Marshallian approximation. For most economists, the major drawback to CV-based estimates was that they were based upon stated preferences rather than observed behavior.

The existence value revolution occurred when it was shown that willingness to pay, as measured in a CV survey, would include Krutilla's (1967) existence value. As such,

[23] See Hensher (1994) for a review of the stated preference work in transportation through the mid-1990s.

[24] Other studies in this vein were Mierheim (1974), an early German stated preference study which looked at parks, and O'Hanlon and Sinden (1978), one of the first studies to look at existence values associated with preserving an area (in New South Wales).

in many instances, non-CV measures of the value of an environmental amenity could be seen as defective because they did not include existence value, which might be a significant component of total value.

The most influential of the early studies that encompassed existence was the study by Randall, Ives and Eastman (1974) valuing visibility in the Four Corners area. Their effort was notable for, among other things, its theoretical rigor, its valuation of a good which could not be valued by alternative methods (such as travel cost and hedonic pricing), its use of photographs to depict the visibility levels being valued, and its experimental design whereby certain aspects of the bidding game (such as the payment vehicle) were varied systematically to see if they affected the WTP amount. Perhaps even more significant, the timely publication of their article on the study in the first volume of the new *Journal of Environmental Economics and Management* brought the method to the attention of a broader audience.

Randall, Ives and Eastman valued changes in air quality necessary to maintain scenic vistas in the Southwest. In this instance, however, the indirect valuation approach was not capable of being applied because all people in the area share the good equally (and hence, not bundled differentially into housing prices) and no expenditure of time or money is needed to enjoy it.

Randall, Ives and Eastman estimated what had been termed "existence value" by Krutilla (1967) in an influential paper. The novel element in Krutilla's framework was that existence values were not generally revealed by market purchases. He argued that some people care about environmental resources, such as wildernesses areas, irrespective of their desire to visit them. Krutilla had not measured existence values, but rather had recommended determining just how large they would have to be to tilt the decision in the other direction. He contended that the failure to include existence values in policymaking would likely entail too great a loss of existing environmental amenities and the provision of too few new environmental amenities. Other related concepts were soon enumerated (e.g., nonuse value, stewardship value, bequest value, option value) and eventually encompassed into a single term "passive-use value," first used in the 1989 court decision *Ohio v. Department of Interior* (880 F.2d 432, D.C. Cir.), which held that government trustees should include passive-use values in damage claims.[25]

What Krutilla had called existence values was part of what had previously been termed "intangible" values [Smith (1976)]. These were not well integrated into welfare economic theory and they were thought to be unmeasurable. The key to measuring them lies in the recognition that due to scarcity effects, any form of economic value can be assessed by creating a trade-off between money and the consideration in question. Monetary measures of economic value are implicitly defined by choices made subject to an income constraint and CV permits the construction of the appropriate trade-off choice. Economic value can be expressed in terms of any constraints and tradeoffs appearing in a choice scenario including time, other public goods, or private goods (money).

[25] A comprehensive review of the concept of passive use and passive use values is provided in Carson, Flores and Mitchell (1999).

2.5. Developments from the mid-1970s through the late 1980s

The period from the mid-1970s through the late 1980s saw a continuation of many of the themes of the early 1970s in terms of the types of goods valued by contingent valuation. Valuing outdoor recreation [e.g., McConnell (1977), Cocheba and Langford (1978)] remained the most popular use of CV. The types of recreation valued expanded considerably and now include such diverse applications as congestion in ski areas [Walsh, Miller and Gilliam (1983)] and diving from offshore oil platforms [Roberts, Thompson and Pawlyk (1985)]. Many studies were undertaken to value improvements in air quality [e.g., Brookshire, Ives and Schulze (1976), Loehman and De (1982), Tolley et al. (1986)] and water quality [e.g., Gramlich (1977), Greenley, Walsh and Young (1981), Mitchell and Carson (1986)].

The range of applications also broadened considerably. These included the benefits of reclaiming surface coal mining areas [Randall et al. (1978)], the value of decreased mortality risk from a nuclear power plant accident [Mulligan (1978)], losses associated with toxic waste dumps [Smith, Desvousges and Freeman (1985)], aesthetic benefits from forgoing construction of a geothermal power plant [Thayer (1981)], the benefits of the public collection and dissemination of grocery store price information [Devine and Marion (1979)], the production of local government statistics [Bohm (1984)], the benefits of government sponsored senior companion programs [Garbacz and Thayer (1983)], protecting endangered species [Samples, Dixon and Gower (1985)], preserving wild and scenic rivers [Walsh, Sander and Loomis (1985)], the willingness of farmers to sell development rights to farmland [Conrad and LeBlanc (1979)], the willingness of citizens to buy those rights [Bergstrom, Dillman and Stoll (1985)], and the benefits of government support for the arts [Throsby (1984)].

Many of these studies were considered exploratory with researchers concentrating on refining the CV method by identifying and testing for the possible biases that arose in its use, and on establishing its credibility by making comparisons between the benefits measured in CV studies and those measured for the same goods by one of the established techniques, such as the travel cost method. Much of the pioneering methodological work was conducted by Randall and his colleagues [e.g., Randall, Ives and Eastman (1974), Randall et al. (1978), Randall, Hoehn and Tolley (1981)] at Oregon State University, New Mexico State University, the University of Kentucky, and the University of Chicago, by Cummings, d'Arge, Brookshire, Rowe, Schulze, and Thayer at the universities of New Mexico and Wyoming and by Robert Mitchell and Richard Carson at Resources for the Future. Parallel theoretical work by many of the same researchers has established that CV data are generated in forms consistent with the theory of welfare change measurement [Randall, Ives and Eastman (1974), Mäler (1974), Freeman (1979), Brookshire, Randall and Stoll (1980), Just, Hueth and Schmitz (1982), Hanemann (1984a, 1984b), Hoehn and Randall (1987)].

The Water Resources Council (1979) published its newly revised "Principles and Standards for Water and Related Land Resources Planning" in the *Federal Register*. This important document set forth the guidelines for federal participation in project

evaluation which specified those methods that were acceptable for use in determining project benefits. The inclusion of contingent valuation as one of the three recommended methods (the other two were the travel cost and the unit day value methods) was a sign of contingent valuation's growing respectability.[26] The U.S. Army Corps of Engineers began to extensively use the contingent valuation method to measure project benefits during this time period. By 1986, the Corp had conducted almost twenty CV studies of varying degrees of sophistication and, as a consequence, published for Corp personnel the first government handbook on how to undertake contingent valuation studies [Moser and Dunning (1986)]. Contingent valuation was also recognized as an approved method for measuring benefits and damages under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (Superfund), according to the final rule promulgated by the Department of the Interior (1986).

Funding from the U.S. Environmental Protection Agency has played a particularly important role in contingent valuation's development. Agency economists recognized early that the prevailing methods of measuring benefits, such as the surrogate markets and the travel cost method, would be of limited use in valuing the benefits of pollution control regulations they would be responsible for implementing. In the mid-1970s the agency began to fund a program of research with the avowed methodological purpose of determining the promise and problems of the CV method.[27] At first, almost all of the CV studies funded by this program were designed to test various aspects of the method and to establish its theoretical underpinnings. As the method became better understood, and the agency's mandate to subject the proposed regulations to benefit–cost analysis was given sharper focus by the Reagan administration under Executive Order 12291 [Smith (1984)], EPA's interest shifted to ascertaining just how effectively contingent valuation could be used for policy purposes.

As part of this effort, in 1983 the EPA commissioned a state-of-the-art assessment of the CV method as a way to step back and reflect on past achievements, remaining problems, and future possibilities. A notable feature of this assessment was the involvement of a review panel of eminent economists and psychologists, including Nobel Laureate Kenneth Arrow and future Nobel Laureates Daniel Kahneman and Vernon Smith. During a conference at Palo Alto, California, in July 1984, these and other scholars who were actively involved in contingent valuation research offered their views about the method's promise as a means for evaluating environmental goods. The leading authors

[26] The Principles and Standards document, which was modified and expanded slightly in 1983, also enshrined the then-prevailing CV practice as the officially prescribed way to conduct CV studies of water project benefits. As late as 1986, contract administrators were known to have required researchers to use the by–then outdated bidding game elicitation technique because the Principles and Standards document had declared it to be the "preferred" elicitation format for CV studies.

[27] The Electric Power Research Institute also sponsored some of the early air visibility studies that used the CV method [Blank et al. (1978)]. EPRI continued to sporadically fund CV research on air quality issues. This led to the first sustained involvement by cognitive psychologists in issues related to the design of CV surveys [Fischhoff and Furby (1988)].

of the state-of-the-art assessment [Cummings, Brookshire and Schulze (1986)] concluded that although the method shows promise, some real challenges remained. They gave the highest priority for future research to the development of an overall framework based on a theory of individual behavior in contingent market settings which could serve as a basis for hypothesis testing.

CV's importance in the United States was raised considerably by presidential executive orders that required an assessment of the benefits and costs of all major new government regulations and reauthorization of existing ones [Smith (1984)]. Outside the U.S., CV was incorporated into OECD reports on measuring the economic value of pollution impacts [Pearce and Markandy (1989)] that in part reflected the rapidly growing use of contingent valuation in Europe, where the number of studies being done started to approach that of the United States. The early leaders in this effort were the Norwegians and Swedes who were later joined by the British.[28]

Near the end of this time period studies in developing countries began to be undertaken. Not surprisingly, the first of these studied outdoor recreation. Grandstaff and Dixon (1986) looked at outdoor recreation in Lumpinee, an urban park in Bangkok, Thailand and found that contingent valuation and travel cost analysis produced similar results, while Abala (1987) looked at the determinants of willingness to pay for Nairobi National Park in Kenya. Interest in this topic has continued ever since with researchers looking at both domestic recreation and foreign tourism with an ecotourism theme.

The major innovation in developing countries, however, was the valuation of basic environmental infrastructure projects. The World Bank [Mitchell (1982), World Bank (1987)], the Interamerican Development Bank [Ducci (1988)], and USAID [Whittington (1988)] began exploring the possibility of using contingent valuation as a tool for helping to evaluate projects. Applications followed involving water supply [e.g., Whittington et al. (1988)] and sewage treatment [e.g., McConnell and Ducci (1989)]. The use of contingent valuation surveys in developing countries was to become the growth area of the 1990s [Georgiou et al. (1997), Whittington (1998)].

2.6. The Mitchell and Carson book

The Mitchell and Carson book (1989) *Using Surveys to Value Public Goods: The Contingent Valuation Method* played a central role in defining the practice of contingent valuation. The book put forth a coherent theoretical framework and a now generally accepted typology of the different types of values potentially associated with the provision of a public good. It showed that they could all be measured in a total value framework using contingent valuation. Different elicitation formats were discussed and the issue of strategic behavior considered at length. The Mitchell and Carson book also put forth a comprehensive typology of all of the different types of biases and misspecifications that could occur in CV studies and showed how most of these could potentially

[28] Navrud (1992) provides an overview of early CV studies undertaken in the various European Countries.

be avoided by careful survey design. The book devoted considerable attention to survey design and administration issues and brought a rigorous survey perspective into the CV literature. The Mitchell and Carson book had appendices containing information on conducting CV experiments, the questionnaire from a national water quality study, and a brief review of over one hundred CV studies. By the time the Mitchell and Carson book appeared, contingent valuation had clearly ceased to be an "experimental" methodology. The number of CV studies was rapidly increasing and their use in government decision making was becoming commonplace.

2.7. Exxon Valdez

A major stimulus to the current CV debate was the enactment of U.S. laws that allowed for the recovery of monetary damages for injuries to natural resources. The focal point of this debate was the Exxon Valdez oil spill, where the state of Alaska's claim for damages was largely based upon a claim for the loss of passive use value [Carson et al. (1992, 2003)]. Potential liability for natural resource damages should increase a firm's precautionary activities and environmental restoration efforts. Firms facing such liability have responded by questioning whether passive use values should count and whether CV estimates of them were reliable.

Following the *Exxon Valdez* oil spill, the oil industry mounted an aggressive public relations campaign intended to convince policy makers that contingent valuation in any form was too unreliable to be used for any purpose. An important tool in their campaign was a set of contingent valuation surveys that their consultants conducted – on issues other than the Exxon Valdez oil spill – which showed CV in a poor light; these studies and the related critiques of CV were presented at an Exxon-sponsored conference in Washington DC in March 1992, and subsequently published in Hausman (1993).[29] Among the claims made were that CV results were insensitive to the scope of the good being valued, and were highly dependent on what else was being valued and on the way in which the valuation was elicited. It was also argued that the inclusion of passive use values represented a form of double counting. Essentially, the oil industry critique can be summarized as: (a) if contingent valuation did work it should work *always*, regardless of the circumstances and details of its application, and (b) if violations of economic theory are found in CV studies, there must be a flaw in the CV study rather than any problem with the received theory.[30]

Looking back on the debate after more than a decade, one senses that it had several positive aspects. It forced a much deeper consideration of how contingent valuation should be conducted, a deeper consideration of the underlying economic theory, and a

[29] Also see the American Petroleum Institution report by Cummings and Harrison (1992).

[30] There is, of course, some irony in that many of the key tenets of what is now often referred to as the behavioral economics revolution were first demonstrated in CV studies and declared anomalies of the method rather than of actual economic behavior.

reconsideration of whether some of the violations of economic theory that are commonly found in contingent valuation are also observed in other forms of economic behavior and could be evidence of a systematic shortcoming in the underlying theoretical model of behavior.

Immediately after the Exxon conference in March 1992, in the face of tremendous industry lobbying of the White House, the staff of U.S. National Oceanic and Atmospheric Administration (NOAA) convened a Blue Ribbon Panel co-chaired by two Nobel Prize winners [Arrow et al. (1993)] to consider whether passive use values should be included in natural resource damage assessment and whether they could reliably be measured by contingent valuation. The Panel held hearings in the summer of 1992 and reviewed a large volume of evidence. It concluded that passive use values should be included in natural resource damage assessment and that "CV studies can produce estimates reliable enough to be the starting point for a judicial or administrative determination of natural resource damages-including lost passive-use value." That Panel's report was also influential for the guidelines that it set forth for conducting CV studies in natural resource damage cases and for the methodological issues it identified as requiring further research.

The American Agricultural Economics Association organized a symposium on contingent valuation in its policy journal *Choices* in 1993 [Carson, Meade and Smith (1993), Randall (1993), Desvousges et al. (1993a)] while the American Economic Association organized a symposium on continent valuation in its policy oriented *Journal of Economic Perspectives* in 1994 [Portney (1994), Diamond and Hausman (1994), Hanemann (1994)]. These papers, in conjunction with the NOAA Panel report, set the research agenda for contingent valuation for the next decade.

2.8. Literature production

The CV literature grew rapidly after the NOAA Panel Report. Ironically, most of this growth was a response to the growing demand for more comprehensive benefit–cost assessments rather than having anything to do with natural resource damage assessments. The growth also reflected the increasing use of contingent valuation in other OECD countries as well as in developing countries. It also reflected an ever increasing variety of environmental amenities that were being valued and a growing interest in a wider range of methodological issues.

Figure 1 displays the production of CV literature by year from 1960 through 2000. Before 1960, there are only a few papers such as Ciracy-Wantrup's (1952) original proposal for doing contingent valuation and the 1958 Audience Research study for the National Park Service. For all practical purposes, the CV literature starts in the early 1960s with the work of Davis, and built slowly through the late 1960s and early 1970s. The first spike in the production of CV literature occurs in 1974, the year of the classic Randall, Ives and Eastman (1974) study in the *Journal of Environmental Economics and Management*. The presence of twenty-five papers in that year alone points to the active involvement of a number of different researchers and groups and the start of the

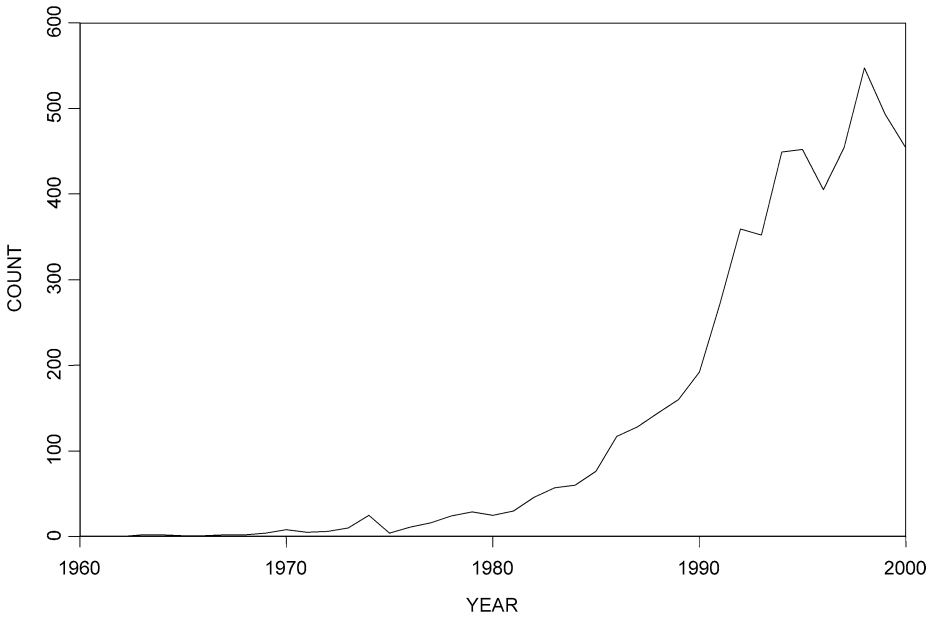


Figure 1. Contingent valuation literature production by year.

use of stated preference techniques in transportation. After a lull in 1975, the output of CV studies grew steadily, indeed at almost an exponential rate up until 1994. By 1994, the production of CV literature reached about 450 papers and studies. Through the end of the decade, the production of CV literature has remained in the 400 to 500 papers per year range.[31] The only exception to this is the upward spike in 1998 to around 550 papers. This appears to be in large part due to the First World Congress of Environmental and Resource Economists in Venice. The Venice World Congress resulted in a substantial increase in the production of papers in environmental economics with nonmarket valuation being one of the most popular themes.

Geographically, CV studies have now been conducted in twenty-nine out of the thirty current OECD members. Luxembourg is the only OECD member country for which a CV study was not found, although there were numerous reports to European Union

[31] The Carson (in press) bibliography contains citations to over 6000 papers and studies. There are approximately 250 citations each for the years 2001–2003. The bibliography for these years is more limited than in previous years and mainly lists published journal articles and papers presented at the major environmental economics meetings. The gray literature including government reports and university/research institute working papers and reports is not as well-represented for this time period as in earlier years. A rough comparison in terms of published papers from the 2001–2003 period to the 1994–2000 period suggests that the production of CV literature has remained in the 400 to 500 per year range.

institutions in Luxembourg that were based upon CV results. CV studies have also been conducted in 80 developing countries.[32]

Another indication of the importance of CV can be seen by looking at EVRI (Environmental Valuation Reference Inventory), a large online database currently being assembled for policy making purposes by Environment Canada, as a cooperative venture undertaken with the European Union, the U.S. EPA, the environmental protection agencies of Chile and Mexico, the World Bank, and the Economy and Environment Program for South East Asia. As of January 2005, the EVRI database (http://www.evri.ec.gc.ca/evri/) contained 742 studies based upon stated preferences, 385 studies based upon revealed preferences, and 215 studies based upon actual costs.

2.9. Major issues

One can divide the major issues surrounding contingent valuation into three groups. The first group involves the linkage between the survey response and economic theory. This has involved two key questions: what economic quantity should be measured by different questions and how can one tell whether the response to a CV survey corresponds to the predictions of economic theory. Sections 3, 4, and 5 of this chapter deal with the theory issues while the issue of assessing the congruence of CV results to economic theory has, as one might expect, generated controversies and is dealt with in Sections 9 and 10. To foreshadow some of that discussion, we simply note that CV results have, on several occasions, resulted in a reexamination of the relevant theory from first principles. What was first thought of as an anomaly or violation of economic theory was, instead, what one should have expected to see. Nowhere is this more true than in the current formulation of Section 5 on elicitation formats. In other instances, there are clear anomalies from the perspective of a narrowly defined self-interested maximizing economic agent in CV surveys, but deeper examination shows them to be ubiquitous in actual market transactions.

The second group of issues deals with designing a CV survey, administering it, and analyzing the resulting data. Sections 5 and 6 look at choosing a format for eliciting preference information and econometric estimation of WTP distributions. Sections 7 and 8 look at designing and administering a CV survey. Part of what emerges here is that there is a reasonable consensus about how to do a contingent valuation study if

[32] The list includes Argentina, Armenia, Bangladesh, Bahamas, Barbados, Botswana, Brazil, Bulgaria, Burkina Faso, Cambodia, Central African Republic, Chile, China, Columbia, Costa Rica, Cuba, Croatia, Dominica, Dominican Republic, Egypt, El Salvador, Ecuador, Estonia, Ethiopia, Gambia, Ghana, Georgia, Guatemala, Haiti, Honduras, Indonesia, India, Israel, Iran, Ivory Coast, Jamaica, Kenya, Laos, Latvia, Liberia, Lithuania, Madagascar, Malaysia, Mali, Micronesia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands Antilles, Nicaragua, Niger, Nigeria, Panama, Pakistan, Papua New Guinea, Peru, Philippines, Puerto Rico, Romania, Russia, Sierra Leone, Singapore, South Africa, Sri Lanka, St. Lucia, Taiwan, Tanzania, Thailand, Tunisia, Trinidad and Tobago, Uganda, Ukraine, Uruguay, Vanuatu, Venezuela, Vietnam, Zambia, and Zimbabwe.

there was no serious budget constraint. Disagreement occurs on what is the best way to collect the "appropriate" information at a reasonable cost. Decisions here are often hard because all parts of the policy evaluation need resources to better understand the implications of the choice faced by policy makers; and where time and money to do the analysis are both often quite limited.

The third group of issues related to how to successfully apply contingent valuation to value particular types of goods is not dealt with in this chapter. This at some level is both the "art" of contingent valuation and the specific knowledge that gets passed down. The history of contingent valuation is full of examples of survey scenarios and instruments that just did not work quite right. This continued until some researcher figured out how to explain or visually depict the good to be valued, and how to formulate a provision and/or payment mechanism in such a way that respondents understand what they are getting, believe it could be provided, and are satisfied that the government would enforce the payment if the good were provided.[33] Unfortunately, there is nothing in economic theory per se that provides this information, economists are typically not taught these skills in the course of their graduate education, and the issues involved in designing a CV survey are generally beyond those with which a conventional survey designer has experience. When someone finds an approach that works, it is often replicated and modified for valuing related goods. The best advice that we can offer here is to carefully read the studies (and most importantly their survey instruments) that try to value goods similar to the one of current interest and determine how well those survey instruments are working and where problems are seen to exist. Section 7 provides some guidance on evaluating how well a survey instrument is working.

3. Economic theory of contingent valuation

3.1. Economic welfare measures

Since CV uses surveys to measure an economic concept of value, we begin a brief review of the relevant economic theory (see Chapter 12 by Bockstael and Freeman in this handbook for a more detailed treatment).

The goal of a CV study is to measure an individual's monetary value for some item. We denote the item being valued by q; for now we will treat this as a single item – whether a single commodity or a single program involving some mix of commodities treated as a fixed group – and therefore q is a scalar.[34] Assuming the individual is a consumer and using essentially the same notation as Bockstael and Freeman, we assume

[33] For an interesting example involving the valuation of low level risks, one of the most difficult problems faced in environmental valuation, see Corso, Hammitt and Graham (2001).

[34] The valuation of multiple items is covered in Sections 5 and 6 under the discussion of choice experiments.

the individual has a utility function defined over the quantities of various market commodities, denoted by the vector x, and q, $u(x, q)$.[35] Corresponding to this direct utility function, we can write an indirect utility function, $v(p, q, y)$, where p is the vector of the prices of the market commodities and y is the person's income.[36] We make the conventional assumption that $u(x, q)$ is increasing and quasi-concave in x, which implies that $v(p, q, y)$ satisfies the standard properties with respect to p and y;[37] but we make no assumptions regarding q. If the agent regards q as a "good," $u(x, q)$ and $v(p, q, y)$ will both be increasing in q; if she regards it as a "bad," $u(x, q)$ and $v(p, q, y)$ will both be decreasing in q; and if she is indifferent to q, $u(x, q)$ and $v(p, q, y)$ will both be independent of q. We make no assumption regarding quasiconcavity with respect to q.

The act of valuation implies a contrast between two situations – a situation with the item, and one without it. We interpret what is being valued as a change in q.[38] Specifically, suppose that q changes from q^0 to q^1; the person's utility thus changes from $u^0 \equiv v(p, q^0, y)$ to $u^1 \equiv v(p, q^1, y)$. If she regards this change as an improvement, $u^1 > u^0$; if she regards it as a change for the worse, $u^1 < u^0$; and if she is indifferent, $u^1 = u^0$. The value of the change to her in monetary terms is represented by the two Hicksian measures, the compensating variation C which satisfies

$$v\big(p, q^1, y - C\big) = v\big(p, q^0, y\big), \tag{1}$$

and the equivalent variation E which satisfies

$$v\big(p, q^1, y\big) = v\big(p, q^0, y + E\big). \tag{2}$$

Observe that

$$\operatorname{sign}(C) = \operatorname{sign}(E) = \operatorname{sign}\big(u^1 - u^0\big). \tag{3}$$

If the change is regarded as an improvement, $C > 0$ and $E > 0$; in this case, C measures the individuals' maximum WTP to secure the change, while E measures her minimum WTA to forego it. If the change is regarded as being for the worse, $C < 0$ and $E < 0$; in this case, C measures the individuals' WTA to endure the change, while E measures her WTP to avoid it. If she is indifferent to the change, $C = E = 0$.

[35] One can develop the theory of a producer's WTP or WTA for a change in using the profit function in place of the indirect utility function, but this will not be pursued here.

[36] The income variable that applies here can be defined in different ways. For example, it could be the supernumerary income that is available to the individual after allowing for certain committed expenditures on market or nonmarket goods.

[37] That is, we assume $v(p, q, y)$ is homogeneous of degree zero in p and y, increasing in y, nonincreasing in p, and quasiconvex in p.

[38] The alternative is to represent it as a change in p. McConnell (1990) adopts this approach for a valuation question of the form. "Would you accept a payment of $\$A$ to give up your right to use this commodity for one year?" Let p^* be the choke price vector (i.e., a price vector such that, at these prices, the individual would choose not to consume the resource), and let p^0 be the baseline price vector. McConnell represents the change as a shift from (p^0, q, y) to (p^*, q, y).

To emphasize the dependence of the compensating and equivalent variation on (i) the starting value of q, (ii) the terminal value of q, and (iii) the value of (p, y) at which the change in q occurs, we sometimes write them as functions: $C = C(q^0, q^1, p, y)$ and $E = E(q^0, q^1, p, y)$. To simplify things, we will define the WTP function,

$$\mathrm{WTP}\big(q^0, q^1, p, y\big) = \begin{cases} C\big(q^0, q^1, p, y\big) & \text{if } C \geqslant 0, \\ -E\big(q^0, q^1, p, y\big) & \text{if } C \leqslant 0. \end{cases}$$

The WTA function $\mathrm{WTA}(q^0, q^1, p, y)$ is defined analogously. The goal of a CV study is to measure one or another of these valuation functions – either the entire function, or a particular point on the function. For simplicity, in the remainder of this section we assume that the change is an improvement ($C \geqslant 0$) and we focus on the measurement of WTP.[39]

Let $y = m(p, q, u)$ be the expenditure function corresponding to the direct utility function $u(x, q)$ and the indirect utility function $v(p, q, y)$; this should be increasing in u, and nondecreasing, concave and homogeneous of degree 1 in p. It is decreasing in q if q is desired, increasing in q if q is a bad, and independent of q if the individual is indifferent to q. In terms of the expenditure function, the compensating and equivalent variations are defined as

$$\begin{aligned} C &= m\big(p, q^0, u^0\big) - m\big(p, q^1, u^0\big) \\ &= y - m\big(p, q^1, u^0\big) \end{aligned} \tag{1′}$$

and

$$\begin{aligned} E &= m\big(p, q^0, u^1\big) - m\big(p, q^1, u^1\big) \\ &= m\big(p, q^0, u^1\big) - y.^{40} \end{aligned} \tag{2′}$$

It is also natural to impose the restriction that

$$\lim_{y \to 0} v(p, q, y) = -\infty \tag{4a}$$

or, equivalently, that, given any q' and any $y' > 0$, there exists no q'' such that

$$v\big(p, q'', 0\big) = v\big(p, q', y'\big); \tag{4b}$$

that is to say, the market goods x in the underlying direct utility function $u(x, q)$, taken as a group, are essential. This implies that $m(p, q, u) > 0$, and, from (1′), that

$$C < y. \tag{5}$$

[39] Note that in the valuation question considered by McConnell (1990), what is being measured is WTA, not WTP. McConnell emphasizes that a correct specification of a WTP or WTA function must *exclude* endogenous variables such as quantity demanded.

[40] McConnell (1990) refers to (1′) and (2′) as *variation* functions and analyzes their derivatives with respect to p and y. He notes that these results can be used in a nonparametric way to compare predictions implied by a CV response model with actual behavior.

Note that from (2′), the equivalent variation (WTA) is *not* similarly bounded by y. We defer further discussion of the differences between WTP and WTA and the implications for estimation to Section 6.

We illustrate these concepts with a specific example, the Box–Cox indirect utility function:[41]

$$v_q = \alpha_q + \beta_q\left(\frac{y^\lambda - 1}{\lambda}\right), \quad q = 0, 1, \tag{6a}$$

where $\alpha_1 \geqslant \alpha_0$ and $\beta_1 \geqslant \beta_0$. This can be regarded as a form of CES utility function in q and y. The corresponding formulas for C is

$$C = \left(\frac{\beta_0 y^\lambda}{\beta_1} - \frac{\lambda\alpha}{\beta_1} + \frac{\beta_1 - \beta_0}{\beta_1}\right)^{1/\lambda}, \tag{6b}$$

where $\alpha \equiv \alpha_1 - \alpha_0$. McFadden and Leonard (1993) employ a restricted version of this model with $\beta_1 = \beta_0 \equiv \beta > 0$, yielding

$$v_q = \alpha_q + \beta\left(\frac{y^\lambda - 1}{\lambda}\right), \quad q = 0, 1, \tag{7a}$$

$$C = y - \left(y^\lambda - \frac{\alpha}{b}\right)^{1/\lambda}, \tag{7b}$$

where $b \equiv \beta/\lambda$. This nests many of the utility models used in the existing CV literature. It is somewhat flexible in that it permits a variety of income elasticities of WTP; the income elasticity of WTP is negative when $\lambda > 1$, zero when $\lambda = 1$, and positive when $\lambda < 1$.[42] For convenience, we will use (7a,b) to illustrate issues of modeling methodology.

3.2. From WTP to CV response

CV uses a survey to measure people's WTP or WTA for the change in q. The utility theoretic model of consumer preference outlined above provides the framework for interpreting the CV responses. Researchers use several different survey formats that involve asking people different questions; and the way one links the survey responses to the measurement of WTP or WTA is somewhat different in each case. In many cases, the response to the survey question is not itself a direct measure of WTP but yet a measure of WTP can be derived from the survey responses. The derivation typically involves a statistical analysis of the survey responses. In the framework of statistical modeling, it

[41] For simplicity, we suppress p and write the indirect utility function as a function of q and y: however, α_q and/or β_q would in fact be functions of p.

[42] It does not, however, satisfy (4a), (4b); this restriction must be imposed separately. Note also that there may be measurement issues with respect to y that may play a role in empirical econometric estimation.

is conventional to treat the survey responses as the realization of a random variable. It is necessary, therefore, to recast the deterministic model of WTP outlined above into a stochastic model that can generate a probability distribution for the survey responses.

The mapping from the deterministic model of WTP to a probabilistic characterization of survey responses involves two steps: (1) the introduction of a stochastic component into the deterministic utility model which leads to what is called a *WTP distribution*; and (2) the formulation of a connection between the WTP distribution and what we call the *survey response probability distribution* based on the assumption of a utility-maximizing response to the survey question. The WTP cumulative distribution function (cdf) will be denoted $G_C(x)$; for a given individual, it specifies the probability that the individual's WTP for item in question is less than x

$$G_C(x) \equiv \Pr(C \leqslant x),$$

where the compensating variation C is now viewed as a random variable.[43] The corresponding density function is denoted $g_C(x)$.

Step (1) above deals with how one goes about the formulation of C as a random variable (more on this later). Step (2) deals with the link between the survey responses and $G_C(x)$, which varies with the form of the survey question. We illustrate this here with two simple examples. The first is the *open-ended* question format: the respondent is asked "How much are you willing to pay for the change from q^0 to q^1?" Suppose the response is $\$A$? This means that, for this respondent, his value of C is $\$A$. Referring now to the underlying WTP distribution, the probability of obtaining this particular response from the individual is given by

$$\Pr(\text{Response to open-ended CV question is } A) = \Pr(C = A) \equiv g_c(A). \tag{8}$$

The second format is the *closed-ended, single-bound* discrete choice format: the respondent is asked: "Would you vote to support the change from q^0 to q^1 if it would cost you $\$A$?" Suppose the response is 'yes'. This means that for this individual, his value of C is some amount more than A. In terms of the underlying WTP distribution, the probability of obtaining a 'yes' response is given by

$$\Pr(\text{Response to closed-ended question is 'yes'}) = \Pr(C \geqslant A) \equiv 1 - G_C(A). \tag{9}$$

With the open-ended format, the response directly reveals the respondent's value of C; with the closed-ended format, it does not reveal the exact value of C but it does provide an interval in which C must lie. In both cases, however, we can forge a link between the WTP distribution and the response probability distribution.

How does one get a WTP distribution? Two approaches have been used in the literature. Historically, the first approach, adopted in the context of the open-ended CV format, was to directly specify a distribution for C. In effect, one writes C as a mean, $E(C) = \mu_C$ together with some white noise ε. The early literature adopted a pragmatic

[43] For now, we assume the change is regarded as an improvement, so that C measures WTP.

approach to the specification of μ_C and viewed it as a linear regression involving some covariates Z and their coefficients, $\mu_C = Z\gamma$. If the random term is additive and has a mean of zero, then

$$C = \mu_C + \varepsilon = Z\gamma + \varepsilon, \tag{10}$$

which is a linear regression model.[44] An alternative specification is that the logarithm of C has a mean of $Z\gamma$, or equivalently, that the error term is multiplicative.

$$\ln C = Z\gamma + \varepsilon. \tag{11}$$

If ε is a standard normal variate than (10) implies C is normally distributed. If the random term is multiplicative as in (11), then C is lognormally distributed. As an alternative to these somewhat *ad hoc* specifications for μ_C, one could also employ a utility theoretic specification of the compensating variation function as discussed above. Using an additive error term, this is

$$C = C\big(q^0, q^1, p, y\big) + \varepsilon. \tag{12}$$

In the case of the Box–Cox model (7), for example,

$$C = y - \left(y^\lambda - \frac{\alpha}{b}\right)^{1/\lambda} + \varepsilon. \tag{13}$$

The second approach introduces a random term directly in the utility function, appealing to the notion of random utility maximization (RUM). In a RUM model it is assumed that, while an individual knows her preferences with certainty and does not consider them stochastic or otherwise behave in a random manner, nevertheless her preferences contain some components that are unobservable to the econometric investigator and are treated by the investigator as random [McFadden (1974), Manski (1977)]. The unobservables could be characteristics of the individual and/or attributes of the items considered for choice and/or they can signify both variation in preferences among members of a population and measurement error or missing data. For now, we represent the stochastic component of preferences by ε without yet specifying whether it is a scalar or a vector, and we write the indirect utility function as $v(p, q, y; \varepsilon)$. Substituting $v(p, q, y; \varepsilon)$ into (1) yields $C(q^0, q^1 p, y; \varepsilon)$, which itself is a random variable. In a RUM model, while the individual's WTP for the change in q is something that she herself knows with certainty, it is something that the investigator does not know with certainty; and therefore, models as a random variable.

[44] Early examples of the regression analysis of open-ended CV response data can be found in Hammack and Brown (1974), Brookshire, Ives and Schulze (1976), and McConnell (1977). In closed-ended CV, the approach based on the specification of a WTP distribution was emphasized by Cameron (1988) and Cameron and James (1987) and is sometimes referred to as the Cameron approach; Duffield and Patterson (1991a) refer to it as the tolerance approach because of the close analogy between the WTP distribution and the tolerance distribution in bioassay.

The RUM approach proceeds by specifying a particular indirect utility function $v(p, q, y; \varepsilon)$ and a particular distribution for ε.[45] An example of a RUM version of the restricted Box–Cox model is

$$u_q = \alpha_q + \beta\left(\frac{y^\lambda - 1}{\lambda}\right) + \varepsilon_q, \quad q = 0, 1, \tag{14a}$$

where ε_0 and ε_1 are random variables with a mean of zero. Consequently,

$$C = y - \left(y^\lambda - \frac{\alpha}{b} - \frac{\eta}{b}\right)^{1/\lambda}, \tag{14b}$$

where $\alpha \equiv \alpha_1 - \alpha_0$, $b \equiv \beta/\lambda$ and $\eta \equiv \varepsilon_1 - \varepsilon_0$. The comparison of (14b) with (13) illustrates the difference between the two approaches to the formulation of a WTP distribution. Inserting an additive random term in the utility function (14a), leads to a random term that enters the formulas for C in a *nonadditive* manner. Thus, even if these disturbances were normal random variables, the distribution of C would *not* in general be normal.[46] For example, when $\lambda = 0$ one has the logarithmic model from Hanemann (1984b)

$$u_q = \alpha_q + \beta \ln y + \varepsilon_q, \tag{15a}$$

which yields

$$C = y\big(1 - \exp\big[-(\alpha + \eta)\big]/\beta\big). \tag{15b}$$

In this case, if ε_0 and ε_1 are normal, C has a lognormal distribution. Alternatively, the linear model ($\lambda = 1$) is written as

$$u_q = \alpha_q + \beta y + \varepsilon_q, \tag{16a}$$

which yields[47]

$$C = \frac{\alpha + \eta}{\beta}. \tag{16b}$$

This is a special case of the Box–Cox model in which C has the same form of distribution as ε; if ε_0 and ε_1 are normal random variables then C is normal as well. Moreover, if the term α/β depends on covariates through a linear regression of the form $\alpha/\beta = Z\gamma$

[45] This is sometimes called the Hanemann, or utility difference approach after Hanemann (1984b). As McConnell (1990) notes, the distinction between the Hanemann and Cameron approaches lies where the random term is appended.

[46] Note that one can also introduce a random term into a RUM model in a nonadditive manner by making a coefficient such as β in (7) random. It should also be noted that not all the parameters of the random utility function are necessarily identifiable from the WTP function. In the case of (14a,b), α_0 and α_1 are not separately identifiable, only their difference. Similarly, the CDFs of the additive random terms in the utility function ε_0 and ε_1, are not separately identifiable, only the CDF of η.

[47] It turns out in this case that $C = E$.

and letting $\varepsilon \equiv \eta/\beta$, one recovers the linear regression formulation in (10). In this case, the two approaches for specifying a WTP distribution lead to an identical result.[48]

Besides the Box–Cox model, another RUM which has featured prominently in the CV literature is the Bishop–Heberlein utility model.[49,50]

$$\begin{aligned} v\big(p, q^0, y; \varepsilon_0\big) &= y + \xi, \\ v\big(p, q^1, y; \varepsilon_1\big) &= y + \xi + \exp\left[\frac{\alpha + \eta}{\beta}\right] \end{aligned} \tag{17a}$$

which yields[51]

$$C = \exp\left[\frac{\alpha + \eta}{\beta}\right]. \tag{17b}$$

3.3. Statistical model of CV responses

The statistical model of the CV survey responses depends on the specific form of the question used in the survey. Here we focus on the closed-ended, single-bounded question format where the respondent is asked "Would you vote to support the change from q^0 to q^1 if it would cost you \$$A$?" As noted above, the probability that she answers "yes" can be represented in terms of the WTP distribution by[52]

$$\Pr(\text{Response is 'yes'}) = \Pr\big\{C\big(q^0, q^1, p, y; \varepsilon\big) \geqslant A\big\} \equiv 1 - G_C(A). \tag{9}$$

An equivalent representation of the response probability in terms of the RUM utility function is

$$\Pr\{\text{response is 'yes'}\} = \Pr\big\{v\big(q^1, p, y - A; \varepsilon\big) \geqslant v\big(q^0, p, y; \varepsilon\big)\big\}. \tag{18}$$

Given that $\mu_C = E[C(q^0, q^1, p, y; \varepsilon)]$, let $\sigma_C^2 = \mathrm{Var}[C(q^0, q^1, p, y; \varepsilon)]$, and let $G(\cdot)$ be the cumulative distribution function of the standardized variate $\omega = (C - \mu_C)/\sigma_C$.

[48] In general, for any given regression formulation of a WTP distribution, one can always find a RUM formulation which generates this distribution. In this sense, any given WTP distribution can be derived using either approach. However, except for (16a, b), for a given WTP distribution, the random term to be added in the regression approach and the random term to be inserted into the RUM model are different.

[49] The model was presented by Bishop and Heberlein at the AAEA summer meetings but was not actually described in the short published version of the proceedings [Bishop and Heberlein (1979)]. It is described in a subsequent working paper by Bishop and Heberlein (1980), and it first appeared in print in Bishop, Heberlein and Kealy (1983).

[50] While it was first thought [Hanemann (1984b)] that there exists no explicit RUM which could generate the Bishop–Heberlein WTP distribution (17b), Hanemann and Kanninen (1999) later showed that it is generated by (17a).

[51] In this case, too, it turns out that $C = E$.

[52] For now, we assume that the only other possible response is "no"; hence, Pr{response is "no"} = 1 − Pr{response is "yes"}. The treatment of "Don't know" responses is discussed in Section 6.

Switching from $G_C(\cdot)$ to $G(\cdot)$, we can re-write (9) as

$$\begin{aligned}\Pr\{\text{response is 'yes'}\} &= 1 - G\left(\frac{A-\mu_C}{\sigma_C}\right)\\ &\equiv 1 - G(-\gamma + \delta A) \qquad (9')\\ &\equiv H(A), \qquad (9'')\end{aligned}$$

where $\gamma \equiv \mu_C/\sigma_C$ and $\delta \equiv 1/\sigma_C$. This expresses the response probability in the form of a statistical model for a binary dependent variable as a function of a covariate, A, where δ is the coefficient on A.[53] In effect, (9) or (9′) constitute the integrability condition for single-bounded CV: a binary response statistical model representing the "yes" and "no" responses to a closed-ended CV survey as a function of the dollar amount in the survey is consistent with an economic model of maximizing behavior *if and only if it can be interpreted as the survivor function of an economic WTP distribution*.[54] The response probability model can be parametric or nonparametric depending on whether the WTP distribution whose survivor function is represented by the right-hand side of (9) has a parametric or nonparametric representation. The only requirement is that, since this is a survivor function, the graph relating a dollar amount A to the probability that the respondent would say "yes," to paying this amount, should be nonincreasing.[55] The graph of the response probability distribution can be viewed as a demand curve for the change in q, and the restriction is that this demand curve should not slope upward.

As an illustration of a parametric response model, the response probability distribution for the version of the Box–Cox model in (13) takes the form

$$\Pr\{\text{response is 'yes'}\} = 1 - \Pr\left\{y - \left(y^{\lambda} - \frac{\alpha}{b}\right)^{1/\lambda} + \varepsilon \leqslant A\right\}, \qquad (19)$$

whereas the response probability formula for the RUM version of the Box–Cox model (14) is

$$\Pr\{\text{response is 'yes'}\} = 1 - \Pr\left\{y - \left(y^{\lambda} - \frac{\alpha}{b} - \frac{\eta}{b}\right)^{1/\lambda} \leqslant A\right\}. \qquad (20)$$

In the linear model, where $\lambda = 1$, this simplifies to

$$\Pr\{\text{response is 'yes'}\} = 1 - \Pr\left\{\frac{\alpha+\eta}{\beta} \leqslant A\right\}. \qquad (21)$$

[53] A binary response model is a statistical model for a binary dependent variable which represents the probabilities that it takes each possible value as functions of covariates, of the form $\Pr(\text{response} = j) = H_j(A)$; these functions must return values in the interval $[0, 1]$, and they must sum to unity, conditions satisfied by (9″).

[54] The survivor function of a random variable $\widetilde{X}$ with the CDF $F_X(x)$ is $S_X(x) \equiv \Pr\{\widetilde{X} \geqslant x\} = 1 - F_X(x)$.

[55] This is because from (9′), $\partial \Pr\{\text{"yes"}\}/\partial A = -g_C(A) \leqslant 0$.

To be more specific, if η in (21) is a standard normal random variable, the response formula becomes a probit model[56]

$$\Pr\{\text{response is 'yes'}\} = \Phi(\alpha - \beta A). \tag{22}$$

If η is a standard logistic random variable, the response formula becomes a logit model[57]

$$\Pr\{\text{response is 'yes'}\} = \frac{1}{1 + \exp(-\alpha + \beta A)}. \tag{23}$$

With the Bishop–Heberlein RUM model, (17), the response probability distribution takes the form

$$\Pr\{\text{response is 'yes'}\} = 1 - \Pr\left\{\frac{\alpha + \eta}{\beta} \leqslant \ln A\right\}, \tag{24}$$

that resembles the Box–Cox RUM response model (21) except that $\ln A$ appears in place of A. If η in (24) is a standard normal random variable, the response distribution becomes a lognormal distribution

$$\Pr\{\text{response is 'yes'}\} = \Phi(\alpha - \beta \ln A), \tag{25}$$

while if η is a standard logistic random variable, the response distribution becomes log-logistic

$$\Pr\{\text{response is 'yes'}\} = \frac{1}{1 + \exp(-\alpha + \beta \ln A)}, \tag{26}$$

which is the form actually used by Bishop and Heberlein (1979) when they introduced the closed-ended single-bounded response format. If $(-\eta)$ has the standard extreme value distribution, the response distribution is a two-parameter Weibull distribution[58]

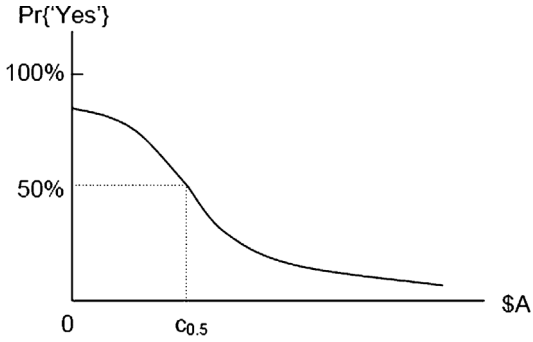
$$\Pr\{\text{response is 'yes'}\} = \exp\bigl[-\exp(-\alpha + \beta \ln A)\bigr]. \tag{27}$$

Figure 2 shows the graphs of the logit response model (23) and the log-logistic model (26); their shapes differ because, as $A \to 0$, the probability of a "yes" response converges on unity in Figure 2(b) but not in Figure 2(a).

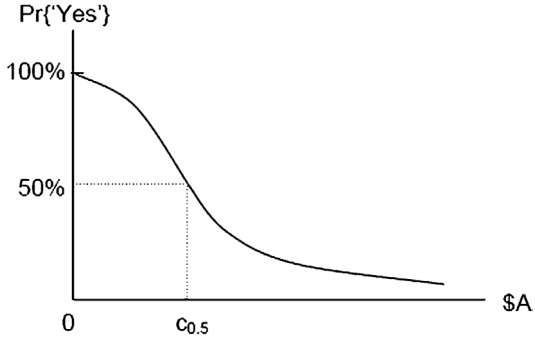
[56] The distribution of η is standard normal if one assumes that ε_0 and ε_1 in (16a) are i.i.d. normal with mean zero and variance 0.5. We use $\Phi(\,)$ to denote the standard normal CDF.

[57] The CDF of a logistic with scale parameter $\tau > 0$ and location parameter μ is $F(x) = [1 + \exp(-(x - \mu)/\tau)]^{-1}$; this has a mean and median of μ, and a variance of $\tau^2\pi^2/3$. In the standard logistic, $\mu = 0$ and $\tau = 1$. This logit model arises if the ε_q in (14a) are extreme value variates. The CDF of an extreme value variate with location parameter ς and scale parameter $\tau > 0$ is $F(\varepsilon) = \exp[-\exp(-\varepsilon - \zeta)/\tau]$; the mean is $\zeta + 0.5772\tau$ and the variance is $\tau^2\pi^2/6$. In the standard extreme value, $\zeta = 0$ and $\tau = 1$. If ε_1 and ε_0 are independent extreme value variates with separate location parameters ζ_1 and ζ_0 and a common scale parameter τ, $\eta \equiv \varepsilon_1 - \varepsilon_0$ has a logistic distribution with location parameter $\mu = \zeta_1 - \zeta_0$ and scale parameter τ. The standard logistic arises when $\zeta_1 = \zeta_0$ and $\tau = 1$.

[58] The two-parameter Weibull distribution with scale parameter $\theta > 0$ and shape parameter $\gamma > 0$ has a survivor function $S(x) = \exp[-(x/\theta)^\gamma]$. Setting $\gamma = 1$ produces the exponential distribution; setting $\gamma = 2$ produces the Rayleigh distribution. When $(-\eta)$ in (17a) is a standard extreme value variate, C in (17b) is Weibull with parameters $\theta = \mathrm{e}^{\alpha/\beta}$ and $\gamma = \beta$.



(*a*) Logit response model



(*b*) Log-logistic response model

Figure 2. Logit and log-logistic response models.

Whatever the specific formula, these parametric response probabilities provide the link to statistical estimation of the CV survey responses. Index respondents by i and bid values by k. Let N denote the total number of respondents and K the total number of separate bid values, and let N_k denote the number of respondents randomly assigned a bid value of $\$A_k$.[59] We use $\mathbb{1}_{i,\text{response}}$ as an indicator variable equal to one, if respondent i gives a specific response, and zero, otherwise – in this case, where the eligible responses are "yes" and "no;" this reduces to $\mathbb{1}_{i,Y}$ and $\mathbb{1}_{i,N}$. The most common approach

[59] The choice of the bid levels A_k and the determination of how many respondents are allocated to each bid N_k are discussed in Section 6.

to estimation in the CV literature is maximum likelihood.[60] Given that respondent i receives bid A_k, the probability that she responds "yes" is denoted $\pi_i^Y(A_k)$, where this might be given by one of the formulas listed above, and the probability that she responds "no" is denoted $\pi_i^N(A_k)$.[61] Assuming that the N respondents are allocated at random to one of the K bid levels, given the sample of respondents, the set of bid values, and the observed yes or no responses from each respondent, the likelihood of observing this pattern of responses is given by

$$L = \prod_{k=1}^{K} \prod_{i=1}^{N_k} \left[\pi_i^Y(A_k)\right]^{1_{i,Y}} \left[\pi_i^N(A_k)\right]^{1_{i,N}}. \tag{28a}$$

The corresponding log-likelihood function is

$$\ln L = \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left[1_{i,Y} \ln\left(\pi_i^Y(A_k)\right) + 1_{i,N} \ln\left(\pi_i^N(A_k)\right)\right]. \tag{28b}$$

By contrast, the nonparametric approach to the specification of a response probability model views this as an *unknown* function of the bid amount $H(A)$, whose value is observed only at a discrete set of points, namely A_k, $k = 1, \ldots, K$.[62] The function can be estimated only at those points, not at other points which are unobserved. The fully nonparametric response model, therefore, takes the form of a step function representing K constants $\pi_k^Y = H(A_k)$, corresponding to the K points at which $H(A)$ is observed; this graph is illustrated in Figure 3(a). The resulting nonparametric likelihood function is

$$L = \prod_{k=1}^{K} \prod_{i=1}^{N_{A_k}} \left[\pi_k^Y\right]^{1_{i,Y}} \left[1 - \pi_k^Y\right]^{1_{i,N}}, \tag{29}$$

where the parameters to be estimated are now the K constants π_k^Y, $k = 1, \ldots, K$.

Two points about the nonparametric approach should be noted. First, this approach treats each bid level as a separate experiment; it does not impose any structure across the bids. Second, by design the nonparametric estimator provides an estimate of the response probability distribution only at a set of points. It is necessary for some purposes, including the calculation of welfare measures (discussed below), to have an estimate of the response probability distribution at *other* points; this requires some form of extrapolation to "connect the dots" in Figure 3(a). Several different approaches have been used in the literature, including: (i) piecewise linear interpolation between each pair of

[60] Alternatives to maximum likelihood will be considered in Section 6. An excellent overview of likelihood based statistical procedures can be found in Pawitan (2001).

[61] Since these are the only two eligible responses, it follows that $\pi_i^Y(A_k) + \pi_i^N(A_k) = 1$.

[62] We adopt the convention that the bid values are indexed in ascending order, so that A_1 is the lowest bid, and A_K the highest.

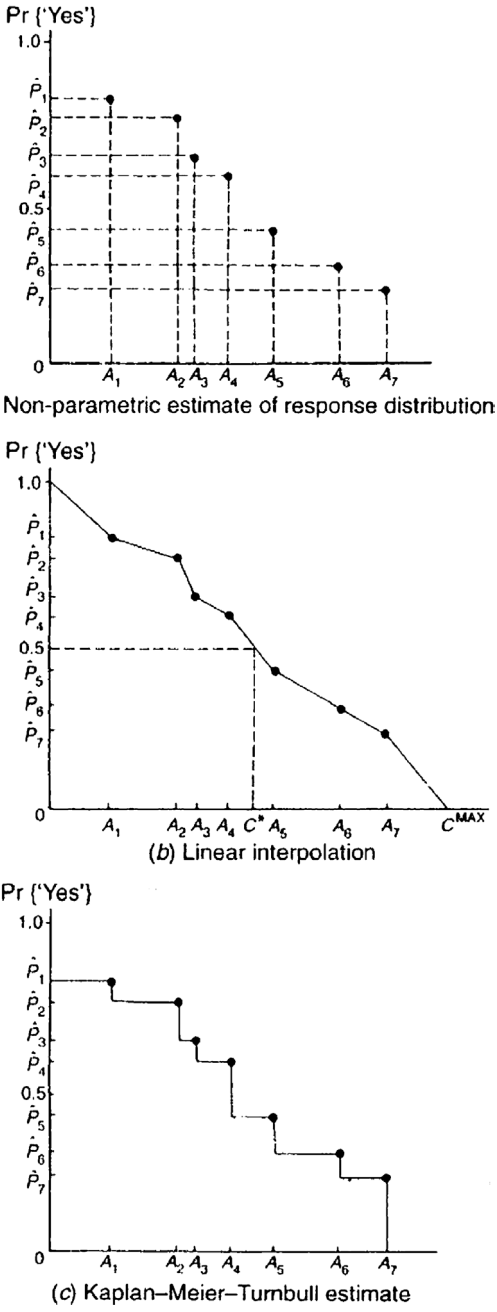


Figure 3. Nonparametric estimation of response model.

nonparametrically estimated response probabilities $[\hat{\pi}_k^Y, \hat{\pi}_{k+1}^Y]$ [Kriström (1990)]. This is illustrated in Figure 3(b). (ii) A more conservative extrapolation assumes that over the interval $[A_k, A_{k+1}]$, $\pi^Y(A) = \hat{\pi}_{k+1}^Y$, $k = 1, \ldots, K-1$, as illustrated in Figure 3(c) [Carson et al. (1992)]. This is a conservative assumption since, with $\pi^Y(A)$ monotone nonincreasing, we know that $\pi^Y(A) \geqslant \hat{\pi}_{k+1}^Y$ over the interval. (iii) Alternatively, a nonparametric smoothing estimator can be applied to the $\hat{\pi}_k^Y$'s; such as, the kernel estimator suggested by Copas (1983) if one is prepared to assume that the underlying WTP distribution is smooth.[63]

3.4. Summary measures of welfare

Having fitted a response model to the CV survey responses, what does one do next? The goal of the study was presumably to develop some summary measure of people's WTP for the change in q. How is this obtained from the fitted response model? The key to the interpretation of the fitted response model is the fact, noted in Section 3.2, that this is derived from an underlying WTP distribution, $G_C(x)$. By exploiting this relationship, one recovers the underlying WTP distribution from the fitted response model.

With the closed-ended, single-bounded format, as noted in (9), $\pi^Y(A) = 1 - G_C(A)$.[64] This makes it possible to derive welfare measures directly from the graph of the response probability distribution. For example, *percentiles* of the estimated WTP distribution can be read directly from the graph of the fitted response model. The θ-percentile of the WTP distribution C_θ satisfies

$$\theta = G_C(C_\theta) = 1 - \pi^Y(C_\theta), \tag{30}$$

i.e., there is a $100(1-\theta)\%$ probability that the individual would be willing to pay at least C_θ for the change in q.[65] In particular, the median estimated WTP is the quantity $C_{0.5}$, such that

$$0.5 = G_C(C_{0.5}) = 1 - \pi^Y(C_{0.5}), \tag{31}$$

i.e., there is a 50–50 chance that the individual would be willing to pay at least $C_{0.5}$. This was illustrated graphically in Figure 2 for the logit and log-logistic response models.

[63] The smoothing estimator could be applied to either the raw response probabilities or the PAV estimates; in the former case, one would choose the bandwidth parameter so as to ensure that the estimated response probability is nonincreasing.

[64] With other response formats, there still exists a relationship between the response model and the WTP, but it takes a different form.

[65] The analogs of C_θ in the biometrics literature are the effective dose $ED_{100\theta}$ and the lethal dose $LD_{100\theta}$. These are the dose levels or concentrations at which, on average, $100\theta\%$ of subjects respond (or succumb). The distribution of which they are the quantiles is usually called the tolerance distribution, analogous to the WTP distribution in the CV context. Much of the attention in the biometrics literature focuses on the medians ED_{50} and LD_{50}.

The analytic formulas for $C_{0.5}$ in these models are as follows. In the linear response model (21), if η has a symmetric distribution with a median (and mean) of zero, such as the standard normal or the standard logistic which generate the probit and logit response models (22) and (23), then

$$C_{0.5} = \frac{\alpha}{\beta}. \tag{32}$$

In the case of the Bishop–Heberlein model (17), if η has a symmetric distribution with a median (and mean) of zero, such as the standard normal or the standard logistic which generate the log-normal and log-logistic response models (25) and (26), then

$$C_{0.5} = \exp\left[\frac{\alpha}{\beta}\right]. \tag{33}$$

In the case of the nonparametric response model illustrated in Figure 3(a), while the definitions in (30) and (31) still apply, unless one of the estimated $\hat{\pi}_k^Y$ just happens to coincide with the quantile of interest, some form of interpolation is required; for example, in Figure 3(a), the point estimate of $C_{0.5}$ lies somewhere in the interval $[A_4, A_5]$. Any of the three approaches to interpolation mentioned above can be used. For example, Figure 3(b) shows the estimate of $C_{0.5}$ when one uses piecewise linear interpolation.

The alternative way to summarize the estimated WTP distribution is through the mean,

$$C^+ \equiv E\{C(q^0, q^1, p, y; \varepsilon)\} = \int_{C^{\min}}^{C^{\max}} c g_C(c)\, \mathrm{d}c, \tag{34}$$

where $C^{\min}$ and $C^{\max}$ are the lower and upper support of the estimated WTP distribution, and $g_C(\cdot)$ is the pdf corresponding to the estimated WTP distribution. With a parametric WTP distribution, the formula for C^+ varies according to the specific distribution selected. In the linear response model (21), for example, if η has a mean of zero, as with the standard normal or standard logistic, which generate the probit and logit response models (22) and (23),[66]

$$C^+ = \frac{\alpha}{\beta} = C_{0.5}. \tag{35}$$

In this case, mean WTP coincides with median WTP. However, in the case of the Bishop–Heberlein model (17), if η is standard normal or standard logistic which generate the log-normal and log-logistic response models (25) and (26), mean and median WTP are *different*. With the lognormal response model (25)

$$C^+ = \exp\left[\frac{\alpha}{\beta}\right] \exp\left[\frac{1}{2\beta^2}\right] = C_{0.5} \exp\left[\frac{1}{2\beta^2}\right] \tag{36}$$

[66] The reader should be warned that the formulas for C^+ given in (35), (36) and (37) violate the utility-theoretic bound on the WTP distribution even though these formulas are commonly used in the literature. See Section 6 for further discussion.

while with the log-logistic response model (26),

$$C^{+} = \begin{cases} \exp\left[\frac{\alpha}{\beta}\right]\Gamma\left[1+\frac{1}{\beta}\right]\Gamma\left[1-\frac{1}{\beta}\right] & \text{if } \beta > 1, \\ \infty & \text{if } \beta \leqslant 1. \end{cases} \tag{37}$$

In both cases $C^{+} \geqslant C_{0.5}$, which reflects the fact that these WTP distributions are skewed to the right.

In addition to analytic formulas for C^{+}, there is also a graphical representation which depends on the support of the WTP distribution. Suppose for now that $C(q^0, q^1, p, y; \varepsilon)$ has a lower support of $C^{\min} = 0$.[67] The graphical representation of C^{+} comes from a standard result about the relation between the mean of a positive random variable and the integral of its CDF [Parzen (1960)]:

$$\begin{aligned} C^{+} &= \int_0^{C^{\max}} \left[1 - G_C(A)\right] \mathrm{d}A \\ &= \int_0^{C^{\max}} \pi^{Y}(A)\, \mathrm{d}A \end{aligned} \tag{38}$$

that corresponds to the shaded area under the graph of the response function in Figure 4.

The graphical representation of mean WTP in (38) sheds some light on the reason why this can differ from median WTP. Whereas $C_{0.5}$ depends on the location of the response probability graph at a particular point, namely the 50% probability level, C^{+} depends on the location of the response probability graph *throughout its entire range*. When comparing any two response probability distributions, small differences in the right tail have essentially no effect on the median but they can effect the mean greatly.[68] This explains why the relation between mean and median WTP can vary with the

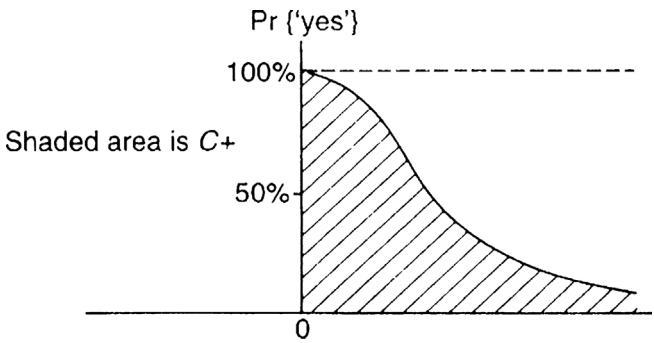


Figure 4. Mean WTP with nonnegative preferences.

[67] Whether or not this is an appropriate lower support is discussed further in Section 6.

[68] Boyle, Welsh and Bishop (1988) and Ready and Hu (1995) refer to this as the 'fat tails' problem.

specification of the WTP distribution. It illustrates the fact that in a RUM model, the specification of the stochastic component can have important economic implications.

In the case of the nonparametric response model illustrated in Figure 3(a), the calculation of C^+ is complicated by the fact that the nonparametric approach provides estimates of only a limited set of points within the support of the WTP distribution. This creates two problems. First, we need to "connect the dots" and flesh out the distribution at all *other* points within the range of observed bid values $[A_1, A_K]$. Second, we do not know what the response probability function looks like for amounts *lower* than A_1 or *greater* than A_K. For the first problem, three approaches to interpolation were noted above and two of these will lead to analytical expressions for C^+. In order to solve the second problem, one has to make some assumption about the unobserved portion of the response probability graph to the left of A_1 and to the right of A_K. A common approach is to fix the lower support of the WTP distribution by setting $\pi^Y(0) = 1$ (this will be revisited in Section 6), and to fix the upper support by assuming that there is some known amount $C^{\max}$, such that $\pi^Y(C^{\max}) = 0$. Combining these assumptions about the supports with the linear interpolation procedure (A), yields the graph depicted in Figure 3(b). Then, the nonparametric estimate of C^+ is obtained by applying (42) to this graph – i.e., by integrating the area under it. The resulting estimator of mean WTP corresponds to what is known in biometrics as the Spearman–Karber estimator $\hat{\mu}_{\mathrm{SK}}$.[69] Setting $A_0 \equiv 0$ and $A_K \equiv C^{\max}$, the formula for $\hat{\mu}_{\mathrm{SK}}$ can be written compactly as[70]

$$\hat{\mu}_{\mathrm{SK}} = \sum_{k=1}^{K+1} \left(\hat{\pi}_k^Y - \hat{\pi}_{k-1}^Y\right)(A_k + A_{k-1}). \tag{39}$$

An alternative and far more conservative approach to dealing with the upper and lower supports of the WTP distribution assumes that $\pi^Y(A) = \hat{\pi}_1^Y$ for $0 < A \leqslant A_1$ and $\pi^Y(A) = 0$ for all $A \geqslant A_K$. Combining these assumptions with the conservative interpolation (B) mentioned above, and integrating the area under the graph in Figure 3(c), yields an estimate of mean WTP [Carson et al. (1992)] that corresponds to what is known in biometrics as the Kaplan–Meier–Turnbull estimator, $\hat{\mu}_{\mathrm{KM}}$. This is given by[71]

$$\hat{\mu}_{\mathrm{KM}} = \sum_{k=1}^{K} \left(\hat{\pi}_k^Y - \hat{\pi}_{k+1}^Y\right) A_k. \tag{40}$$

Because of the conservative assumptions on which it is based, $\hat{\mu}_{\mathrm{KM}}$ should be viewed as a lower bound on the mean of the WTP distribution.[72] An alternative estimator is

[69] In the CV literature, this estimator was used by Kriström (1990).

[70] We are assuming here that the fitted response probability graph is nonincreasing.

[71] The nonparametric estimate of mean WTP proposed by Duffield and Patterson (1991a) is a slight variant of this.

[72] A potential drawback of the Kaplan–Meier–Turnbull lower bound on the mean WTP is that it is sensitive to the particular design points used in the following sense. The estimator is consistent in that as the possible

obtained by placing all of the density at the *upper* end of the intervals in Figure 3(c); this yields an upper bound on the estimate of mean WTP. This estimator bound the range from above and from below in which any estimate of mean WTP that is consistent with the observed data can fall.[73] The estimate of the Kaplan–Meier–Turnbull lower bound on mean WTP, in particular, is often very precisely defined if the sample size at each design point is fairly large.[74]

Given the difference between the mean C^+ and median $C_{0.5}$ as welfare measures, deciding which to use involves considerations of both statistics and economics. Suppose that the survey data came from repeated questioning of a single individual – while this is fairly impractical, one could imagine it happening. In that case, even though we were estimating a single individual's WTP, it still would be a random variable for us as outside observers because of the RUM hypothesis. The issue would then be just one of representation – what is the best way to summarize the probability distribution? The answer depends on the statistical loss function: with a sum-of-squared-errors loss function, the mean is the optimal measure of central tendency; with a sum-absolute errors loss function, the median is optimal. For this reason, the mean is more sensitive to skewness or kurtosis in the WTP distribution [Stavig and Gibbons (1977)]. This could be important because most RUM models with nonnegative preferences imply a skewed distribution of WTP. In that case, it can happen that the point estimate of the median is more robust, or has a much smaller sampling error than the estimate of the mean.

Now consider the more realistic situation where the survey data come from questioning different individuals in a population. In that case, the summary measure of the WTP distribution would be multiplied by the number of people in the population to produce an estimate of aggregate value. Thus, the choice of a summary statistic implies a particular approach to the aggregation of welfare across the population. The mean is equivalent to adopting the Kaldor–Hicks potential compensation principle. While the Kaldor–Hicks criterion is often invoked by economists, it can lead to logical inconsistencies and it has been severely criticized on ethical grounds [Little (1957)].[75] As a way to aggregate values, the median is equivalent to applying the principle of majority voting

support for the WTP distribution is filled in by using more and more design points, the Kaplan–Meier–Turnbull estimator increases toward the sample's WTP from below. An easy way to see the nature of the sensitivity to the design points chosen is to note that if the only design points used are small (relative to where most of the WTP distribution) lies, the Kaplan–Meier–Turnbull estimator cannot exceed the largest design point used.

[73] The upper bound on mean WTP is infinity as long as there is any fraction of the subsample receiving the highest design point is WTP that amount and no further restriction (discussed below) such as income is imposed as an upper bound on WTP.

[74] The formula for standard error of the Kaplan–Meier estimator is simply a linear combination of differences in the percent in favor at different design points and the magnitude of those design points.

[75] Critics of CV from outside of economics [e.g., Sagoff (1994)] are often in reality critics of benefit–cost analysis rather CV *per se*. It is CV's claim to encompass all "economic" value that makes the use of CV more troublesome to such critics.

(i.e., the change is desirable if a majority of the population would vote for it).[76] Using a lower quantile of the WTP distribution would correspond to super-majority voting; for example, it would correspond to requiring a two-thirds majority vote. It is known that majority voting rules do not satisfy Pareto efficiency; but they still may be considered ethically superior. In view of these welfare-theoretic implications, choosing a measure of central tendency is essentially a value judgment. Moreover, different circumstances may call for different welfare measure. Politicians tend to favor measures like the maximum amount a majority of the voters would approve, while in a natural resource damage assessment mean WTA is the natural measure because of the implied property right and legal requirement to restore all those who were injured to their original position. These are judgments that the researcher and policymakers must make.

4. Types of value/motives for valuation

So far we have made no assumption about the individual's motive for valuing q, or the nature of this value; whatever the reason why she cares for q, if she does care, this is reflected in her direct and indirect utility functions $u(x, q)$ and $v(p, q, y)$. In fact, the literature has developed an extensive set of potential motives and potential types of value [Mitchell and Carson (1989), Carson, Flores and Mitchell (1999)]. The standard typology distinguishes between existence value, bequest value, option value, and quasi-option value. Each of these can be explained in terms of a specific formulation of the individual's preference structure and a specific conceptualization of what is being valued by the individual. These conceptualizations are summarized below.

4.1. Existence value

The notion of existence value, also called passive use value, was first proposed by Krutilla (1967) who observed that people might be willing to pay for an environmental resource – say a national park – even if they knew that neither they nor their children would ever visit it because, as he famously put it, they "obtain satisfaction from mere knowledge that part of the wilderness in North America remains, even though they would be appalled by the prospect of being exposed to it." The underlying notion here is that these people value the park for motives unconnected with the possible use of the

[76] Calculation of the aggregate WTP estimate using majority approval is straightforward with a flax tax. This is no longer the case if the tax price can vary with agent characteristics like income. Werner and Groves (1993) look at this case by partitioning the population into groups and uses the average of the medians for each group. Suppose there are H groups; each group has a WTP distribution that has a mean C_h^+ and a median $C_{0.5,h}$. Werner and Groves' statistic, the average of the group medians, is $C^W \equiv (1/H)\sum C_{0.5,h}$. They establish that C^W generally lies between the overall population median $C_{0.5}$ and mean C^+, and that for a given sample size, while estimates of C^W do not have quite as low a variance as $C_{0.5}$, they have a considerably lower variance than C^+.

park. Hence, this value has also been called nonuse value, to distinguish it from an (active) use value that would be motivated by considerations of the enjoyment from using the resource in some fashion (e.g., visiting it). As we show below, although use and nonuse value are conceptually distinct, they are not mutually exclusive; they can both co-exist within the same individual.

Although several formalisms have been proposed in the literature for representing the concept of existence value, the one now most commonly used takes the following form. Suppose the direct utility function takes the specific form

$$u = u(x, q) = T\big[\bar{\phi}(x), q\big], \tag{41a}$$

where $T[\cdot, \cdot]$ is a bivariate function,[77] and $\bar{\phi}(\cdot)$ is a subfunction that aggregates the x's. Note that (41) involves a weak separability between the x's and q since the marginal rate of substitution between the consumption of any pair of market commodities, x_i and x_j, is entirely independent of the level of q. A consequence of the formulation in (41a) is that the ordinary demand functions for the x's are each independent of q; they take the form

$$x_i = h^i(p, q, y) = \bar{h}^i(p, y), \quad i = 1, \ldots, N, \tag{41b}$$

where the functions $\bar{h}^i(\cdot)$ are in fact the ordinary demand functions associated with the maximization of the subutility function $\bar{\phi}(\cdot)$ alone: maximizing $u(x, q)$ and $\bar{\phi}(x)$ subject to a budget constraint on the x's leads to exactly the same solution. The corresponding indirect utility function takes the form

$$u = v(p, q, y) = T\big[\bar{\psi}(p, y), q\big], \tag{41c}$$

where $\bar{\psi}(p, y)$ is the indirect utility function that corresponds to maximization of the subutility function $\bar{\phi}(\cdot)$ alone. While the person cares for q – it enters her direct and indirect utility functions via the $T[\cdot, \cdot]$ function – the presence of q in her utility function has *no* effect on her utility maximizing choice of the x's. It is in this sense that one could say that this individual values q for reasons *unconnected* with her valuation of the market goods x. In this case, q would be said to have a pure nonuse (existence, passive use) value.

Now consider a modified version of the above utility function

$$u = u(x, q) = T\big[\bar{\phi}(x, q), q\big], \tag{42a}$$

where, as before, $T[\cdot, \cdot]$ is a bivariate function and $\bar{\phi}(\cdot)$ is a subfunction. In this case, q enters the utility twice, once through its appearance in $\bar{\phi}(\cdot)$ and the other time as the second argument in $T[\cdot, \cdot]$. Here, the individual values q for two reasons: one reason is connected with her consumption of the x's and is represented by the interaction of x and q in $\bar{\phi}(x, q)$; the other reason is unconnected with her consumption of the x's and

[77] This is assuming that q is a scalar.

is represented by the appearance of q as the second argument of $T[\cdot,\cdot]$. In this case, the ordinary demand functions *do* depend on q: they take the form

$$x_i = h^i(p,q,y) = \bar{h}^i(p,q,y), \quad i = 1,\ldots,N, \tag{42b}$$

where the functions $\bar{h}^i(p,q,y)$ are in fact the ordinary demand functions associated with the maximization of the subutility function $\bar{\phi}(x,q)$ alone. The crucial implication of this fact is that revealed preferences based purely on estimation of the demand functions for market goods $\bar{h}^i(p,q,y)$, $i = 1,\ldots,N$, will recover only the subutility function $\bar{\phi}(x,q)$, but *not* the function $T[\cdot,\cdot]$. The indirect utility function associated with (42a,b) is

$$u = v(p,q,y) = T\big[\bar{\psi}(p,q,y),q\big], \tag{42c}$$

where $\bar{\psi}(p,q,y)$ is the indirect utility function that corresponds to maximization of the subutility function $\bar{\phi}(x,q)$ alone. Applying (1) to (42a), the total value that the individual places on a change in q, denoted C^{TOT}, is given by

$$T\big[\bar{\psi}\big(p,q^1,y-C^{\mathrm{TOT}}\big),q^1\big] = T\big[\bar{\psi}\big(p,q^0,y\big),q^0\big]. \tag{42d}$$

This has both a use value component, associated with $\bar{\psi}(p,q,y)$, and a nonuse value component, associated with $T[\cdot,q]$; the use component, C^{U}, satisfies

$$\bar{\psi}\big(p,q^1,y-C^{\mathrm{U}}\big) = \bar{\psi}\big(p,q^0,y\big), \tag{42e}$$

while the nonuse component C^{NU} would be defined as the difference

$$C^{\mathrm{NU}} \equiv C^{\mathrm{TOT}} - C^{\mathrm{NU}}. \tag{42f}$$

From a CV modeling perspective, the distinction between use and nonuse components of the individual's total value for a change in q could in principle be captured by formulating a CV response probability model based on a specification of a utility function that conforms to the structure of (42a,c), where the $\bar{\psi}(p,q,y)$ and $T[\cdot,\cdot]$ functions could be separately identified. The crucial feature of this structure is that prices and income interact in a manner that is partially separable from q.[78] However, the very simple specifications of the indirect utility function employed in the literature to date, in which commodity prices are not generally explicit variables, makes it unlikely that this can be done; and therefore, unlikely that the use and nonuse components of C^{TOT} can effectively be disentangled in the econometric estimation of CV responses if the *only* information elicited in the survey relates to C^{TOT}.

Several CV researchers have approached the problem by first asking respondents for their total value and then requesting that they allocate this total among several specific motives. When this has been done, it has been in the context of an open-ended CV elicitation format.[79] Suppose for the moment that the only motives are use and nonuse

[78] Entirely separable in the case of (41a,c).

[79] For an example see Walsh, Loomis and Gillman (1984).

value (the other components will be considered below). It seems unlikely that total value plus the elicited allocation between the two components could be used effectively to recover the specific underlying utility function – i.e., the separate components $\bar{\phi}(x, q)$ and $T[\cdot, \cdot]$. Aside from this theoretical consideration, there are several empirical objections to this type of decomposition approach. In terms of survey practice, it seems unwise to employ a top–down approach of first asking for a total, without identifying the specific individual items to be included in this total, and only afterwards specifying the individual components that the researcher has in mind and asking about them. On cognitive grounds, if people think holistically about an item, it may be difficult for them to produce a meaningful and reliable decomposition of this whole *ex post*. For both these reasons, a bottom–up approach seems preferable in which respondents are introduced to the separate components and asked about each of them, with the adding up occurring subsequently. In addition to the cognitive difficulty, Mitchell and Carson (1989), Carson and Mitchell (1991) and Cummings and Harrison (1995) point out that these decompositions are generally not unique since the portion of value allocated to a particular motivation may differ depending on the other elicited motivations and the sequence in which they are elicited. Moreover, given all of the potential motivations, no list could ever be exhaustive for *all* individuals.

4.2. Combining CV and RP data to measure use versus nonuse value

Given the structure of (42a–f) which generates the decomposition of the total value for the change in q into a use value component based on $\bar{\psi}(p, q, y)$, and a nonuse value component based on $T[\cdot, q]$; a more fruitful approach to identifying the two separate components of value is to collect two sets of information, one being total value from a CV survey and the other being revealed preference (RP) data on the demand functions for one or more of the x's. Suppose, for simplicity, there is sufficient revealed preference data to estimate a complete demand system for the x's. This would be combined with CV data on total value, using an assumed specification of the indirect utility function along the lines of (42a) to estimate a system consisting of

$$\begin{aligned}
x_1 &= h^1(p, q, y) = -\frac{\partial \bar{\psi}(p, q, y)/\partial p_1}{\partial \bar{\psi}(p, q, y)/\partial y}, \\
&\vdots \\
x_N &= h^N(p, q, y) = -\frac{\partial \bar{\psi}(p, q, y)/\partial p_N}{\partial \bar{\psi}(p, q, y)/\partial y}, \\
T\big[\bar{\psi}\big(p, q^1, y - C^{\mathrm{TOT}}\big), q^1\big] &= T\big[\bar{\psi}\big(p, q^0, y\big), q^0\big].
\end{aligned} \tag{43}$$

The advantage of this approach is that the revealed preference data enrich the CV data, the two sets of data are analyzed in a mutually consistent manner, and they permit the estimation of the separate components of total value; at least in this particular case,

where there are only use and nonuse components in total value. Pioneering studies following this approach include Cameron (1992) and Adamowicz, Louviere and Williams (1994).[80]

4.3. Contingent behavior

Before going on to consider other possible types of value, it is useful to take note of an approach that parallels the type of data combination in (43) but is restricted only to the estimation of use value in a pure revealed preference context. The goal of this approach is solely to estimate a demand function or system of demand functions; but the conventional data on prices and quantities consumed is supplemented by responses to a survey question with a discrete-response format question that is analogous to a CV question. The survey question focuses on demand behavior rather than willingness to pay *per se*; hence, this approach has been called contingent behavior by Chapman, Hanemann and Kanninen (1993).

The setting is typically that the researcher envisages a two-good demand system where one good is the particular commodity of interest (e.g., sport fishing at lakes and rivers in Wisconsin) and the second good is a Hicksian composite commodity of all other consumption.[81] The researcher has individual-level data on the price of the first good and the number of fishing trips taken by each angler and has used a survey to supplement this data. The survey question asks about circumstances under which the angler would stop fishing altogether (i.e., make his consumption of x_1 equal to zero). There are two alternative versions of the survey question. One version focuses on the price of the commodity p_1, and asks something like "If you had to pay an extra price per trip of Δ_p as an increment to the cost of fishing in Wisconsin, would you still keep fishing in Wisconsin, albeit perhaps at a reduced level, or would you stop fishing altogether?" The other version introduces the notion of annual permit, which entails an income reduction as opposed to a price increase; with this version, the question is something like "If you had to buy an annual permit costing Δ_y to be allowed to go fishing in Wisconsin would you buy this permit or would you give up fishing in Wisconsin altogether?" In either case, the response to the survey question is combined with data on the current number of fishing trips in Wisconsin taken by the angler, x_1, and the current cost of fishing there, p_1. Suppose the angler says that he *would* keep fishing despite the

[80] For recent review of the issues involved in combining the two sources of data see Azevedo, Herriges and Kling (2003). There exist a parallel literature in the transportation and marketing literatures [e.g., Ben-Akiva and Morikawa (1990)]. Swait and Louviere (1993) point out that there is no reason to expect the underlying variance parameters to be the same in the revealed and stated preference data sources; and, that this must be allowed for statistically when combining the two sources of data. Statistical issues related to this topic are further pursued in Hensher, Louviere and Swait (1999).

[81] There is a particular justification for focusing on this two-good model because it is only in this case that the Kuhn–Tucker conditions for a corner solution for the commodity of interest ($x_1 = 0$) can be represented fully in terms of the simple inequality on the demand function presented in (44); see Lee and Pitt (1986) and Hanemann (1985) for an elaboration of these conditions.

increment Δ_p or Δ_y. With the first version of the survey question, by combining the survey response with the travel cost data, the researcher knows the following about the individual's demand function for fishing

$$x_1 = h^1(p_1, p_2, y), \quad h^1(p_1 + \Delta_p, p_2, y) > 0. \tag{44a}$$

With the second version of the survey question, if the angler says he would still keep fishing in Wisconsin, the researcher knows that[82]

$$x_1 = h^1(p_1, p_2, y), \quad h^1(p_1, p_2, y - \Delta_y) > 0. \tag{44b}$$

The system of equations in (44a) and (44b) is the basis for estimation – for example, to formulate a likelihood function for maximum likelihood estimation – in a manner similar to that employed with (43).

4.4. Bequest value

In addition to existence value, Krutilla (1967) also introduced the concept of "bequest value" – some people would be willing to pay to protect a national park because they want to preserve it for their children and grandchildren. This motivation can be incorporated in the utility model used above in the following manner. As far as these people are concerned, the future generations' consumption is something that they care for but it is not something that they see themselves as controlling. Thus it enters their utility function in a manner analogous to q – something that affects their welfare but is exogenous to them. We can think now of q as being a vector with two separate components, $q = (q_1, q_2)$, where q_1 is the existence of the national park, which affects them either as visitors to the park (a use value) and/or for nonuse reasons associated with the significance of its existence value to them and where q_2 is the well-being of their children and grandchildren. Because of the bequest motive, the protection of the park now involves a shift in *both* elements of q. With this, the formalism in (1) and (2) carries over as the definition of the individual's WTP and WTA to protect the park. Thus, the bequest motive leads to a potential re-specification of the underlying utility function but does not otherwise change the formal theory of the CV response probability function.

4.5. Option value

In addition to the work of Krutilla (1967), the other foundational paper on extensions of use value was Weisbrod (1964) who, like Krutilla, adopted the premise that the natural

[82] The first version of the survey question leading to (44a) was proposed by Hanemann (1985) and applied by Chapman, Hanemann and Kanninen (1993). The second version of the survey question leading to (44b) was employed by Cameron (1992).

environment is not a conventional type of economic commodity and that some people's motives for valuing it differ from those for valuing private market goods. Weisbrod focused on uncertainty and what became known as "option value": some people who do not now visit a national park, say, may still be willing to pay money to protect it from destruction or irreversible damage because they want to preserve their option of visiting it in the future. Following the publication of Weisbrod's paper, there was extensive literature discussing the precise formulation of a formal utility-theoretic definition of the value that Weisbrod alluded to. Obviously, this involves a model of choice under uncertainty, unlike the model of certain choice employed thus far. A full and final resolution of the discussion on formalizing Weisbrod's concept was provided by Graham (1981, 1984). Cameron and Englin (1997) applied Graham's formalism for analyzing responses to a CV question on option value, and we follow their development here, using the context of protecting a national park from environmental damage as an example.

Two core features of Graham's model are state-dependent preferences and uncertainty about the state of world. The utility model of Section 3 is now adapted to incorporate these features.[83] There are two possible future states of the world; in one state, the individual is an active user (visitor) of the national park, while in the other state, the individual is not an active user but may or may not derive existence (or bequest) value from it. Conditional on the first state occurring, the individual's indirect utility function is $v_{\mathrm{a}}(q, y)$; conditional on the second state occurring, the individual's utility is $v_{\mathrm{n}}(q, y)$. As far as the park is concerned, that can either be environmentally damaged or undamaged; let $q = 0$ denote the situation where the park is damaged, and $q = 1$ is the situation where, presumably as the result of some policy intervention, the park is undamaged. Finally, π denotes the individual's subjective probability that he will in future be a user of the park, while $1-\pi$ is the probability of not being a user. Cameron and Englin allow for the possibility that this probability itself depends on the state of the park, so that $\pi = \pi(q)$; if the park is environmentally damaged, the probability of visiting it is lowered, $\pi(0) < \pi(1)$. In addition, there may be uncertainty that the environmental damage will occur: $P(1)$ is the individual's subjective probability that $q = 1$, while $P(0) = 1 - P(1)$ is the subjective probability that $q = 0$.

Following Graham (1981) and others in the theoretical literature on this topic, there is a distinction between option price and expected surplus. The expected surplus is derived as follows. If the individual knew for sure that he would be an active user of the park, his WTP to protect it from environmental damage would be S_{a}, where this satisfies

$$v_{\mathrm{a}}(1, y - S_{\mathrm{a}}) = v_{\mathrm{a}}(0, y); \tag{45a}$$

conversely, if he knew for sure that he would *not* be an active user, his WTP would be S_{n}, where

$$v_{\mathrm{n}}(1, y - S_{\mathrm{n}}) = v_{\mathrm{n}}(0, y). \tag{45b}$$

[83] For simplicity, we suppress the commodity prices p.

These are both *ex post* measures of WTP for $q = 1$ versus $q = 0$. *Ex ante*, the expected value of this WTP, known as the *expected surplus*, is

$$E\{S\} = \big[\pi(1)P(1) + \pi(0)P(0)\big]S_{\mathrm{a}} + \big[\big(1-\pi(1)\big)P(1) + \big(1-\pi(0)\big)P(0)\big]S_{\mathrm{n}}. \tag{45c}$$

Using Hirschleifer's (1965, 1966) extension of the von Neumann–Morgenstern expected utility theorem, the individual's baseline expected utility when no policy intervention is undertaken is

$$E\{V\}^{*} = \pi(0)P(0)v_{\mathrm{a}}(0, y) + \big(1-\pi(0)\big)P(0)v_{\mathrm{n}}(0, y). \tag{46}$$

The option price is the sure payment that, regardless of uncertain user status, the individual is willing to pay to ensure that the park is protected from environmental damage. It is the quantity *OP*, which satisfies the following equality:

$$E\{V\}^{*} = \pi(1)P(1)v_{\mathrm{a}}(1, y-OP) + \big(1-\pi(1)\big)P(1)v_{\mathrm{n}}(1, y-OP). \tag{47}$$

An empirical implementation of these measures requires explicit models for (i) $\pi = \pi(q)$, the subjective conditional probability of participation given the environmental status of the national park; (ii) $P(q)$, the subjective probability of the status of the park; and (iii) $v_{\mathrm{a}}(q, y)$ and $v_{\mathrm{n}}(q, y)$, the state-dependent utilities, which could be any of the utility models given in Section 3.

Cameron and Englin (1997) provide an empirical implementation of this framework to responses from surveys of the general population in four New England states in 1989 which had asked a discrete-response CV question "If acid rain damaged fishing in one fifth of all currently fishable high-altitude lakes in the Northeast, would you be willing to pay $\$X$ per year to prevent this?" They point out that because the survey question was not sufficiently explicit, there are two possible interpretations of the respondents' answers. One possibility is that respondents harbored *no* uncertainty regarding their future use condition – for example, they assumed that their condition now will carry over into the future, so that if they are a user now, they assume they will be a user in the future; while if they are not a user now, they assume they will not be a user in the future. In that case, their survey response is based on the surplus S_{a} or S_{n} – they answered "yes" if $S_j > X$, $j =$ a or n, and "no" otherwise. This would then be a standard discrete-response CV of the type discussed in Section 3. However, another possibility is that respondents were unsure of their future use status (and, perhaps also, the extent of future damage to the park) when they answered the survey question. In that case, their response would correspond to *OP* in (47), and they would have answered "yes" if $OP > X$, and "no" otherwise.

5. Elicitation formats for stated preference information

Originally, economists tended to take one of two positions with regard to surveys (see Section 2): either survey respondents tell the truth, in which case useful information can

could be obtained from surveys, or they strategically misrepresent their preferences, in which case no useful information can be obtained. This all-or-nothing stance led to a paralysis among economists seeking to measure preferences using surveys until the empirical breakthrough by Davis (1963a) which produced "reasonable" and convincing survey results. This was followed up by a close comparison of the CV welfare estimate with one from a travel cost model by Knetsch and Davis (1966) which showed the two results were broadly similar. Further support was provided by Bohm's classic market experiment comparing "hypothetical" and "actual" payments, which suggested that the problems with a hypothetical survey were far less than economists had feared. As a result, the use of CV rapidly expanded and, as it did, a new issue presented itself. It became evident that different ways of asking preference questions yielded different estimates of willingness to pay.

The first CV surveys, starting with Davis (1963a), used what was called the "bidding game" format. In this format, respondents were asked a series of questions, along the lines of: "Would you continue to use this recreational area if the cost was to increase by $\$A$?" or "Would you be willing to pay $\$A$ for this item?" If the respondent said "yes" the question was then repeated with a larger value for A; if "no," it was repeated with a lower value for A. This continued until the response switched from "yes" to "no," or from "no" to "yes," thereby isolating a specific amount that was the most that the respondent was willing to pay. It was this final amount that was recorded by the researcher as the respondent's WTP. Researchers eventually began to focus attention on the starting bid utilized in the question sequence. Rowe, d'Arge and Brookshire (1980) found that the starting point had a sizeable influence on the final estimate of willingness to pay. However, Thayer (1981) found no such effect.

The lingering concern about the possible influence of the starting bid then led researchers to employ a simpler approach in which respondents were asked a single, open-ended question along the lines of "What is the most that you would be willing to pay for this item?" It was found that this sometimes produced a large number of "don't know" responses. To deal with this, Mitchell and Carson (1981) proposed the payment card approach, whereby respondents were given a card containing an array of numbers and asked to choose a number on the card (or any number in between) which best represented their maximum willingness to pay. Desvousges, Smith and Fisher (1987) tested these alternative approaches in the context of valuing water quality improvements and found that they produced different results: the lowest estimate of WTP was produced by a bidding game with a $25 starting point, followed by the direct (open-ended) question and the payment card, with the bidding game using a $125 starting point resulting in the highest estimate. In some comparisons, the $25 and $125 starting point bidding games differed statistically but in others they did not. In effect, these results fell somewhere between Rowe, d'Arge and Brookshire (1980) and Thayer (1981). The Desvousges, Smith and Fisher study was noteworthy for the care given to the investigation of how the different payment approaches affected the distribution of responses, something that had not been considered in the earlier literature.

Implicit in the elicitation formats considered so far was an effort to "isolate" the individual's exact maximum willingness to pay. Bishop and Heberlein (1979, 1980) took a different approach and sought to bound it rather than measure it exactly. They asked a single, closed-ended question along the lines of "If the cost to you was A, would you buy (vote for) this item?" Only one question was asked of each respondent, but the valuation amount A was varied across respondents, using random assignment with a set number of alternative values for A. In effect, this was analogous to a dose-response experiment in biometrics where A corresponded to the stimulus in the experiment and yes/no was the response; as illustrated in Figure 3, the plot of the percent of yes responses against the bid amount provided a useful response surface. This approach subsequently became known as the referendum, single-bounded, or binary discrete-response format.

The drawback with this approach was that it provided limited information about the respondent's willingness to pay, only that it was greater or less than A. Compared to the open-ended question format, one needed a large number of respondents to obtain an estimate of WTP with the same given precision. Carson (1985) showed that one could use a double sampling framework to ask a second binary discrete choice question conditional on the response to the first. If the respondent said "yes" to the first valuation question, the question was repeated with a higher value for A; if "no," it was repeated with a lower value for A. Hanemann (1985), Carson and Steinberg (1990) and Hanemann, Loomis and Kanninen (1991) were to further develop this notion into what is now known as the double-bounded approach because it has the potential to generate both an upper and a lower bound on the respondent's WTP.[84]

Many of the early CV studies asked respondents questions about several different changes in air quality [e.g., Randall, Ives and Eastman (1974)], outdoor recreation [Cicchetti and Smith (1973)], or water quality [Mitchell and Carson (1981)].[85] This was relatively straightforward to do, although concerns about impact of question order clearly existed [Mitchell and Carson (1986)]. The most natural way to extend this framework was to ask a series of paired comparisons, along the lines of "Which do you prefer – program 1, which involves paying A_1 and obtaining outcome q_1, or program 2, which involves paying A_2 and obtaining outcome q_2?" [86] The standard single-bounded question format can be viewed as a special case of a paired comparison with one alternative being the *status quo* – pay nothing and obtain nothing – and the other being the specific program in question. The paired comparison approach extends this by

[84] If two binary discrete choice questions could be asked, then a third could also be asked as proposed by Bateman et al. (2001).

[85] The usual intention here was to estimate a valuation function that gave WTP as a function of the level of the environmental amenity being valued and respondent characteristics. Perhaps the best know example is the national water quality valuation function in Carson and Mitchell (1993) that has been heavily used by the U.S. Environmental Protection Agency for a variety of policy purposes including valuing the benefits of the U.S. Clean Water Act. Carson, Mitchell and Ruud (1990) provide an example using health and visibility effects from air quality improvement programs. They fit a response surface design to the continuous WTP data elicited to allow interactions between the two program attributes.

[86] For a review of the approach see Brown and Peterson (2003).

permitting a sequence of paired comparisons involving multiple programs compared sequentially to a common *status quo* or to each other. In implementing this approach, one needs to choose the payment amounts A_j in such a way that they do not appear mutually inconsistent – i.e., a more extensive program should cost more than a less extensive program.

An alternative to a sequence of pairwise comparisons is a single, multinomial comparison, along the following lines: "The government is considering three possible programs: program 1, which involves paying A_1 and obtaining outcome q_1; program 2, which involves paying A_2 and obtaining outcome q_2; and program 3, which involves paying A_3 and obtaining outcome q_3. Alternatively, the government could take no action, which will cost you nothing but will lead to no change in the present situation. Which of these options do you prefer?" The responses can be analyzed in keeping with the random utility model proposed by McFadden (1974).[87] In the first environmental application [Carson et al. (1987), Carson, Hanemann and Steinberg (1990)] of what is now known as the choice experiment format, respondents were offered the possibility of purchasing varying numbers of fishing stamps that each allowed the right to catch and keep one Alaskan Kenai King salmon. The cost of the different options was randomly assigned and a no purchase option was offered. The major advantage of this approach, in addition to its improved statistical efficiency, is that it permits the researcher to estimate how changes in the individual attributes across the choice alternatives alter the respondents' choices and, hence, to value changes in individual attributes.[88]

An alternative to having the respondent select the best out of a set of multinomial choice alternatives is to have her rank the complete set of alternatives. This approach had been popular in the marketing and psychology literatures and was first implemented in the environmental valuation literature by Rae (1982) for air quality improvements in Cincinnati. However, the approach proved to be difficult and time consuming to implement as the number of choices and attributes grew. The method was shown by Ruud (1986) to produce results that seem to violate economic theory in the sense that respondents either had lexicographic preferences or ignored some attributes in ranking the alternatives.[89] In later years, it was not often implemented. Mackenzie (1993) revived

[87] For a review of this literature see Chapter 15 by Phaneuf and Smith in this Handbook or Herriges and Kling (1999).

[88] It is possible to ask multiple multinomial choice questions to gain even greater statistical efficiency. As explained below, a drawback of the approach, not recognized at first, is that for public goods (as opposed to private or quasi-private goods) only one configuration of the public good can actually be provided. This creates the possibility that truthful preference revelation is not a dominate strategy in the multinomial choice setting.

[89] This issue was explored in more depth in Hausman and Ruud (1987). If the set of alternatives is completely ranked and the usual i.i.d. error assumption made in a multinomial logit model holds, then it should be possible to explode the ranked ordered data into the set of choices implied [Chapman and Staelin (1982)]. Hausman and Ruud's analysis suggest that people pay less attention to alternatives ranked in the middle so that the i.i.d. assumption is violated. If it is necessary to model the "scale" factor associated with each level of ranking then much of the gain from the complete ranking exercise disappears.

the format to some degree by tying it to binary discrete choice CV and moving in the direction of using rating data.

Although it is hard for respondents to completely rank a large set of alternatives, it is not difficult for them to pick their most preferred and least preferred alternatives in what is called "best–worst" approach [Marley and Louviere (in press)]. Another approach asks respondents to give a "rating" which describes the intensity with which they prefer one alternative to another [Johnson and Desvousges (1997)]. Both these approaches are now seeing considerable application in the marketing literature. The best–worst approach maintains the tight link to random utility; however, the rated pairs approach veers off towards cardinal utility because it treats the rating as a measure of preference intensity.

5.1. Response probability models for alternative elicitation formats

As noted earlier, the statistical model of the CV survey responses depends on the specific form of the question used in the survey. The response probability models presented in Section 3.3 correspond to the single-bounded referendum format. For the open-ended question formats, whether the bidding game version, the payment card version, or the direct open-ended questions, the response probability model is a regression equation along the lines of (10) or (11).

In the case of the double-bounded format, the response probability model is given by a natural extension of (9) or (18). In this case, with two valuation questions, there are four possible response outcomes: (yes, yes); (yes, no); (no, yes) and (no, no). Let the dollar amount in the initial valuation question be denoted by A. If the response to that question is yes, this is followed with a second valuation question using a higher amount A_{U}; if no, it is followed up using a lower amount A_{L}.[90] Accordingly, the general formula for the various response probabilities is:[91]

$$\begin{aligned}
&\Pr(\text{Response is yes/no}) = \Pr(A_{\mathrm{U}} \geqslant C \geqslant A) \equiv G_C(A_{\mathrm{U}}) - G_C(A),\\
&\Pr(\text{Response is no/yes}) = \Pr(A \geqslant C \geqslant A_{\mathrm{L}}) \equiv G_C(A) - G_C(A_{\mathrm{L}}),\\
&\Pr(\text{Response is yes/yes}) = \Pr(C \geqslant A_{\mathrm{U}}) \equiv 1 - G_C(A_{\mathrm{U}}),\\
&\Pr(\text{Response is no/no}) = \Pr(A_{\mathrm{L}} \geqslant C) \equiv G_C(A_{\mathrm{L}}).
\end{aligned} \tag{45}$$

[90] It is worth commenting on how this differs from the bidding game format. The key difference is that, in the bidding game, the final response is conceptualized as being exactly equal to the respondent's maximum WTP, while in the double-bounded format the response is seen as yielding bounds on the WTP rather being than a direct estimate of WTP itself. Similarly, the payment card format can be conceptualized as yielding either an exact expression of WTP – as in Mitchell and Carson (1981) – or upper and lower bounds on WTP, which is the interpretation applied by Cameron and Huppert (1989).

[91] Another possibility is to ask a *third* valuation question, using a higher or lower bid amount that used in the second question, depending on the response; this "triple bound" format was considered by Bateman et al. (2001). As long as one views the final response as providing bounds rather than an exact measure of WTP, the response probability model takes the same form as (45).

With choice experiments, it is necessary to represent the response in terms of the random utility function rather than the willingness to pay function. Suppose there are K alternatives, each involving a cost of A_k and an outcome q_k; in a paired comparison, $K = 2$, while in a multinomial choice experiment $K > 2$. Suppose the individual selects alternative number 1; the probability of this response is given by

$$\begin{aligned}&\Pr\{\text{alternative 1 selected}\}\\ &\quad = \Pr\big\{v(q_1, p, y - A_1; \varepsilon) \geqslant v(q_k, p, y - A_k; \varepsilon),\ k = 1, \ldots, K\big\}.\end{aligned} \tag{46}$$

With ranking, suppose for simplicity that there are 3 alternatives in addition to the status quo (q_0, $A_0 = 0$), and suppose the respondent prefers all of the alternatives to doing nothing and ranks them in the order 1, then 2, and then 3. The probability of this response is given by

$$\begin{aligned}&\Pr\{\text{1 preferred to 2 preferred to 3 preferred to status quo}\}\\ &\quad = \Pr\big\{v(q_1, p, y - A_1; \varepsilon) \geqslant v(q_2, p, y - A_2; \varepsilon)\\ &\qquad \geqslant v(q_3, p, y - A_3; \varepsilon) \geqslant v(q_0, p, y; \varepsilon)\big\}.\end{aligned} \tag{47}$$

Analogous formulas apply for "best–worst" ranking format.

So far in this section, the focus has been a parametric representation of the response probability function. Just as with the single-bounded approach, the double-bounded format also lends itself well to nonparametric estimation of the response probabilities, in a manner parallel to that associated with (29) and Figure 3. It is now convenient to work with the probability of saying no rather than yes; let $\pi^N(A)$ denote the probability of saying no in a single-bounded survey. Considered nonparametrically, this can be represented by viewed as an unknown function of the bid amount that is observed only at a discrete set of points, namely A_k, $k = 1, \ldots, K$. We supplement them with two "artificial" bid amounts, A_0 and A_{K+1}. These are to be defined such that $\pi^N(A_0) = 0$ and $\pi^N(A_{K+1}) = 1$; a natural choice for the first is $A_0 = 0$ while for the second one chooses some amount that the researcher considers an upper bound on the likely WTP, $A_{K+1} = C^{\max}$, such that $\pi^N(C^{\max}) = 1$. Define the $K + 1$ constants $\omega_k = G(A_k) - G(A_{k-1})$, $k = 1, \ldots, K+1$, corresponding to the increase in the probability of a "no" response over the $K + 1$ intervals at which $G(A)$ is observed, including the two artificial bids. Observe that, for each $k = 1, \ldots, K + 1$,

$$G(A_k) = \sum_{1}^{k} \omega_j, \tag{48a}$$

while the ω_k satisfy the restrictions that

$$\omega_k \geqslant 0, \quad k = 1, \ldots, K + 1, \tag{48b}$$

which ensures that the $G(A_k)$ are monotone nondecreasing, and

$$\sum_{1}^{K+1} \omega_k = 1. \tag{48c}$$

The likelihood function analogous to (29) for the double-bounded response probabilities in (45) is given by

$$L = \prod_{k=1}^{K+1} [\omega_k]^{N_k}, \tag{49}$$

where the parameters ω_k, $k = 1, \ldots, K+1$, are to be estimated subject to the constraints in (48b,c). Turnbull (1976) developed the maximum likelihood estimator for (49) analogous to the Ayer et al. (1955) nonparametric estimator for (29). He shows that the first-order conditions for the likelihood maximization imply the pool-adjacent violators principle so that, if the observed sample $S(A_k)$ are decreasing in some region, one combines adjacent bids until the sequence is nondecreasing.[92] Given the Turnbull estimates of ω_k, the corresponding estimates of $S(A_k)$ are obtained recursively by repeated application of (48a).

Whichever response format is adopted, once the parameters of the WTP distribution and/or the random utility functions have been estimated, the welfare measures are calculated in the same way as described above in Section 3.4 in connection with the single-bounded format. The percentiles of WTP distribution are given by (3) and (31); the mean is given by (34) and (38); and, with the nonparametric Turnbull estimates of double-bounded data, the nonparametric Spearman–Karber and Kaplan–Meier estimates of the mean are given by (39) and (40).[93]

5.2. *The issue of incentive compatibility*

If all of these elicitation approaches produced statistically similar estimates of willingness to pay, the debate over the choice of format to use might have stayed focused on issues of statistical efficiency. Even that debate, though was being pulled in a different direction from the pragmatic perspective of survey administration. Conceptually, the simple direct question yielded the most information about an individual respondent's WTP for the good. This was perhaps the hardest elicitation format, however, to actually get respondents to answer in a survey and tended to result in large nonresponse rates and a much larger than expected number of zero responses. In any event, however, the estimates from different elicitation formats were not generally statistically equivalent and the differences between them were sometimes quite large.

One interpretation of this phenomenon is that respondents did not have well-formed/well-developed preferences for the environmental goods they were being asked

[92] This nonparametric approach was first applied to double-bounded CV data by Carson and Steinberg (1990).

[93] The Kaplan–Meier estimate is now popularly called the Turnbull lower bound on mean WTP [Carson et al. (1994), Haab and McConnell (1997)] after Turnbull (1976) who generalized the estimator to cover a wide range of possible data types.

about in CV surveys. The foundation for this viewpoint was the psychological literature which contended that preferences were "labile" and impacted by how questions were framed [Lichtenstein and Slovic (1971), Fischhoff, Slovic and Lichtenstein (1980), Tversky and Kahneman (1986)].[94] Preferences, in the sense thought of by economists as something, not necessarily fixed, but temporally quite stable [e.g., Stigler and Becker (1977)], were not seen as a valid concept by these psychologists.[95] Ironically, while framing effects can be shown to exist, preferences, at least at from the perspective of aggregate level valuation in a CV study, have been found to be quite constant over time periods of several years.[96]

An alternative explanation often advanced by economists, was that the environmental goods being valued were something new to respondents and the context in which they were being valued was novel, and as such respondent's lacked "market experience." The sensitivity to the way in which the elicitation question was asked and to other aspects of the scenario; such as, the payment vehicle, the order in which goods were valued, and the amount of information being provided, was seen as a specific issue for environmental valuation, rather than something that was problematic for economics as a whole.[97] As time went on, however, it became clear that the critique raised by psychologists and the burgeoning new field of behavioral economics had little to do with environmental valuation *per se* but was rather a critique of neoclassical consumer demand theory. In retrospect, this should not have been surprising. Many of the framing issues discovered by psychologists had long been exploited in marketing; and hence, were readily demonstrable with goods and choice contexts that were very familiar to consumers.

5.3. Some general results on incentives

Economists working on CV had occasionally paid attention to strategic incentives facing respondents. The first key insight was by Brookshire, Ives and Schulze (1976) who noted that strategic behavior would lead to a flatter WTP distribution, in the sense that there would be more low WTP amounts and more high WTP amounts if such behavior

[94] There have been occasional attempts [e.g., Fischhoff and Furby (1988)] to draw upon this literature to provide guidance on how to design reliable CV questions.

[95] Part of the debate here is semantic in nature. Economists don't really think that people walk around with a vector of willingness to pay for all possible goods in all possible contexts imprinted in their minds; even if formal theoretical models appear to implicitly make this assumption. Empirical models allow for the possibility of optimization errors and rational choice theory simply assumes that the errors made by agents are not systematic and predictable in advances. The heuristics and bias view of the world adapted by these psychologists has played a large role in influencing what is now termed behavioral economics [Camerer, Loewenstein and Rabin (2003)]. McFadden (1999a) provides an insightful discussion from the perspective of modeling choice data. Ironically, preferences, at least at from the perspective of aggregate level valuation in a CV study, have proven to be quite reliable over time periods of several years.

[96] For an example, see Carson et al. (1997).

[97] The best known attack on CV from the psychology perspective was Kahneman and Knetsch (1992) known widely for its "embedding" test. This issue is discussed at more length in Section 9.

existed, than in the true WTP distribution. The difficulty with this insight was that it required knowing the true WTP distribution. Brookshire, Ives and Schulze looked at this issue by providing respondents with information about the WTP distribution of the other respondents but this did not seem to have a large impact on subsequent responses.

Interest in exploring the implications of strategic behavior seemed to flounder until Hoehn and Randall (1987) and Mitchell and Carson (1989). Hoehn and Randall, who looked at binary discrete choice questions, made the pivotal jump to seeing strategic behavior as critically linked to the properties of the elicitation format used. Mitchell and Carson took the possibility of strategic behavior seriously in a variety of contexts and showed that some widely held views simply did not pan out. They explored the implications of the growing literature on incentive compatibility.

The implications of Hoehn and Randall's insight was far ranging. If the incentives for strategic behavior were tied to the elicitation format used, then different elicitation formats might well produce different results. Indeed, as Carson, Groves and Machina (2000) [hereafter CGM] were to argue in a comprehensive analysis of different elicitation formats, if different response formats did not yield different answers, then respondents were not acting as the standard maximizing economic agents they were assumed to be. This has turned on its head the criticism of CV that different elicitation formats result in different answers.

The CGM framework took, as its central core, maximizing agents and an n-person game theory approach. Survey questions can be divided into two types: inconsequential and consequential. In inconsequential questions, either agents do not care about the outcomes of the choices they are asked about or their responses have no influence on those outcomes. Economic theory has nothing to say about inconsequential survey questions and nothing can be learned from such a question about preferences in an economic sense. From this perspective, any effort to avoid strategic behavior by making a preference question "purely hypothetical" is destined to be ill-fated. Some respondents will not believe language to the effect that the survey is inconsequential as they observe a substantial effort being put into the administration of the survey and ask the reasonable question – to what purpose are the survey results likely to be used. If they do believe the language, then survey responses are uninterpretable from an economic perspective.

The starting point in the CGM framework is to note that the well-known theorem of Gibbard (1973) and Satterthwaite (1975) says that no mechanism with larger than a binary message space can be incentive compatible without restricting the space of allowable preference functions. The Gibbard–Satterthwaite theorem does not, however, say that all binary discrete choice questions are incentive compatible. For this to be true, further conditions must be imposed. Some of these are well known from the voting literature [Farquharson (1969)]. For the standard ballot proposition requiring binding plurality approval (e.g., majority or two-thirds in favor), the conditions for incentive compatibility are: (a) that a take-it-or-leave-it offer be made where the vote doesn't influence any other offer that may be made to agents and (b) that the payment mechanism be coercive in the sense that agents can be required to pay independent of their own vote if the requisite plurality is in favor. It should be noted that the requirement for a coercive

payment mechanism rules out voluntary contributions for public goods and voluntary purchases of private goods. The other provisions rule out the case where an agent may vote against a measure that is preferred to the *status quo* because a "no" vote might encourage an even better offer.

CMG show that the binding plurality approval vote provision of the typical ballot measure is a special case of a more general condition, that the probability that a specific alternative is provided is weakly monotonically increasing in terms of the percent of voters that favor it. This general condition, not the binding nature of the vote, is what is needed for incentive compatibility. CMG then show that the results of the vote can be considered probabilistically by policymakers, as long as that probability is greater than zero without altering the incentive properties of the mechanism. The resulting class of incentive compatible mechanisms includes binding referenda as well as advisory referenda. Finally, CMG rely on a result by Green and Laffont (1978) to show that an exogenously chosen sample of the population of interest can be substituted for an actual vote by all members of the population without altering the incentive properties of the mechanism. These conditions combined to show that an "advisory survey" has the same incentive properties as a binding referendum vote.

There are two main ways that incentive compatibility can be violated. The first keeps the binary discrete choice format but relaxes the take-it-or-leave-it nature of the question, or does not use a coercive payment mechanism. Many binary discrete choice CV questions fall into this category. Sometimes this is done by design, for example, when a voluntary contribution payment vehicle is used. Often, however, the problem stems from difficulty in posing a clear, credible take-it-or-leave-it choice.[98] Further, it is difficult to find a payment vehicle that is appropriate to the policy and coercive for all members of the population of interest. The second is to use some format other than a binary discrete choice.

Other preference elicitation formats generally fall into two categories: those that attempt to elicit a continuous WTP response, and those that ask more complicated discrete questions. Formats that directly elicit WTP (i.e., open-ended formats) are sometimes referred to as an equivalency or matching question because the respondent is asked to identify the maximum amount he would be willing to pay, which would render him indifferent whether or not the good is provided. The optimal response to this question depends upon the expected cost to the agent if the project is undertaken. The optimal response of an agent whose WTP is less than his expected cost is to give the lowest possible amount, typically zero.[99] This is true for almost any plausible belief about how

[98] In particular, it may be difficult to convince respondents that if the proposed alternative to the status quo is not approved that a different alternative will not be offered or that there may be the possibility of some third party paying all or part of the cost. See Richer (1995) for a CV example involving the Desert Protection Act which had been debated over many years in different forms.

[99] Popular variants of the direct WTP question include the bidding game and the payment card. Both of these elicitation formats back off of asking for actual WTP in favor of isolating WTP to a reasonably small interval. To some degree, these formats can also be seen as a series (potentially) long of repeated binary discrete

the government will aggregate/use the WTP response from the survey. The empirical evidence suggests that there is almost always a large group of zero responses. While it is possible that agents who give a zero response are truly indifferent, the most telling aspect of the empirical evidence is that there are typically few, if any, very small positive WTP responses. Conversely, for an agent whose WTP is greater than the expected cost, the optimal strategy depends upon his beliefs about the decision rule being used by the government and about the responses of the other agents. If the agent believes a strict benefit–cost decision rule is being followed and his true WTP is considerably above what he believes is the average of the other agents' WTP, then he may have an incentive to give the highest plausible WTP amount. However, extremely large amounts are not often observed, and when they are, they at most constitute a small fraction of the sample. While this may reflect the absence of nontruthful strategic behavior or a belief that very large WTP responses are not plausible, it likely reflects a belief that the government decision rule is not a strict benefit–cost one. Many other plausible rules, and in particular, those that have the government trying to reallocate costs toward or taking surplus from an agents (or types of agents) indicating WTP amounts higher than cost. The reported amounts from these agents should increase toward their true WTP from below as uncertainty over the expected cost increases.[100]

The simplest variant of asking a more complex discrete choice question is the double-bounded dichotomous choice format. In this case, the respondent is asked two closed-ended questions with the bid amount in the second question dependent on the response to the first question. The usual approach to analyzing this question [Hanemann, Loomis and Kanninen (1991)], as reflected in (45), assumes that the two responses are based on comparing the amounts asked about to true WTP. The empirical evidence suggests that respondents tend not answer the two questions in this way, although their two responses are clearly highly correlated [Cameron and Quiggin (1994), Herriges and Shogren (1996), Alberini, Kanninen and Carson (1997), León and Vazquez-Polo (1998)]. Nor should they answer in this way, because one of two things needs to be true for the second question to make sense. Either the cost is uncertain (which for a pure pubic good with a coercive payment vehicle will depress WTP relative to the single binary discrete choice question with a certain cost) or something about the good has to have changed between the first and second question – otherwise there would have been no

choice questions. This allows these formats to overcome to some degree some of the response problems associated with a straightforward question that directly elicits WTP. The payment card format further introduces uncertainty over what the cost of providing the good will be. This uncertainty can encourage more truthful preference revelation.

[100] One of the reasons that the payment card format seems to work is that it seems to induce the "right" amount of cost uncertainty. Unless extreme ranges of amounts are used on the payment card or there is a large gap between zero and the first positive amount, the payment card estimates seem reasonably robust to the particular amounts used. See Rowe, Schulze and Breffle (1996) for an experiment looking at the properties of the payment card.

need to ask the second question.[101] These considerations change the respondent's perception of what the second question means, and therefore affect her response. If there is uncertainty, the respondent may answer the second question on the basis of some type of weighted average between the first amount and the second. Alternatively, the respondent may conclude that the quality of the good or its probability of provision have changed commensurate with the change in cost; or that the government wants more information about WTP in order to extract surplus from high WTP-type agents and reallocate the cost toward them. CGM investigate the theoretical implications of such different beliefs. Most of these lead to "too many" no–no responses, but there is more ambiguity about other response patterns. Burton et al. (2003a, 2003b) provide an empirical investigation but it is unclear how generalizable results will be due to the dependence on particular beliefs about costs, good quality, and government use of the responses.

As one might expect, similar issues arise when one moves to even more complex choice formats. Consider a sequence of paired comparisons where respondents are asked whether they favored or opposed two alternative programs to provided specified levels of local air quality at specified costs amounts. While each of the questions could be answered independently, some respondents will ask how the government is going to pick a single level of air quality to provide for all local residents. This naturally causes the questions to be linked with almost any plausible belief about how the government might use the numbers. To see this, let the respondent assume a very simple rule, that the program getting the largest fraction in favor, in any paired comparison, will be the air quality level chosen. The respondent may now have an incentive to misrepresent responses to one or more of the paired comparisons; and indeed, may find it optimal to be in favor of a program that actually lowers utility in order to help prevent a program that would impose an even larger loss from being adopted.

A variant of the same issue occurs with the popular multinomial choice experiment format. Assume for a moment that there are three (or more) possible alternatives and the respondent assumes that the one receiving the largest plurality will be enacted. This can be seen as akin to a three candidate race with a plurality winner.[102] Only the two alternatives that the respondent believes will get the highest fractions in favor (independent of

[101] The one-and-a-half bound format introduced by Cooper, Hanemann and Signorello (2002) attempts to avoid the problem of changing the price unexpectedly when asking two valuation questions by framing the situation a little differently. The respondent is told up front that the cost is not known for certain, and can range from A_{L} to A_{U}. One of these amounts is selected at random, and he is asked if he would be willing to pay that amount. If this initial amount is A_{L} and the response is no, there is no follow up question; if the response is yes, the respondent is then asked about A_{U}. Conversely, if the initial amount is A_{U} and the response is yes, there is no follow-up question while, if the answer is no, there is a follow-up question about A_{L}. This generates a likelihood function similar to (45). Cooper, Hanemann and Signorello find that this retains most of the gain in statistical efficiency relative to the single-bounded approach, and it greatly reduced the discrepancy between the responses to the first and second valuation questions.

[102] In a three candidate race with a simple plurality winner, the candidate perceived to be in third place will always try to convince voters that there is no difference (in utility) between the two front runners and that there is a chance the candidate will win. In the more general formulation of this voting problem, these two factors are traded-off in utility terms against the possibility of casting a "wasted" vote.

the respondent's answer) are viable choices, so a respondent may pick other than his or her most preferred alternative. It is important to note here the respondent should not pick randomly but rather should pick between the two most viable alternatives. The common practice of asking a sequence of multinomial choice questions raises the same issue as does a sequence of paired comparisons about how preference information across the choices sets would be aggregated by the government.

The multinomial choice question may not face the same problem when the good is a quasi-public or private one where different agents can, for example, go to different recreational sites or buy different computers.[103] But the larger issue of how the information from the question will be used still remains. Agents tend to see such questions as either a provision or pricing exercise. Response strategies tend to follow from that belief coupled with some notion of how the information will be used. More generally, agents should try to induce the supply of the good configured in their most desired manner at the lowest cost. There are some fairly simple strategies to signal an unhappiness with "price gouging" that imply not picking the most preferred alternative when its cost is seen as higher than it should be, in favor of picking something close to it which is available at a low price. Nontruthful preference revelation of this sort should cause violations of the independence of irrelevant alternatives condition.[104]

Asking respondents to completely rank the alternative simply allows the ranked alternatives to be "exploded" [Beggs and Hausman (1981), Chapman and Staelin (1982)] into a complete set of choice data. This amplifies the opportunities and incentives for misrepresenting preference information. The best–worst format of Marley and Louviere (in press) achieves a substantial part of the efficiency gain from a complete ranking while minimizing any additional incentives for misrepresentation. Another popular response format using "rated" pairs is difficult to analyze from a mechanism design perspective. In addition to inheriting all of the issues associated with a sequence of paired comparisons, respondents also need to make an assumption about how the ratings will be translated in any preference revelation exercise.

5.4. A typology of the incentive structure of elicitation formats

Figure 5 provides a typology of the structure of preference elicitation format incentives. At the top is a single binary discrete choice question that we have labeled "one

[103] CGM show that if respondents are offered K private/quasi-public goods as the choice alternatives and they believe that $K - 1$ will be provided, then the choice question is incentive compatible. That is because the worst off an agent can be is to get their second most preferred alternative made available. Effectively then, the choice question reduces to a binary one, where the agent's most preferred alternative is put up against an unknown alternative in a competition for which of the two goods will not be supplied. As such, the agent should always pick the most preferred alternative from the set.

[104] Standard tests of whether SP and RP data are consistent with each other [e.g., Swait and Louviere (1993)] are effectively tests of whether one of the data sources is either random or systematically different on multiple parameters, not tests of whether there is a systematic distortion in one of the parameters, typically the coefficient on price.

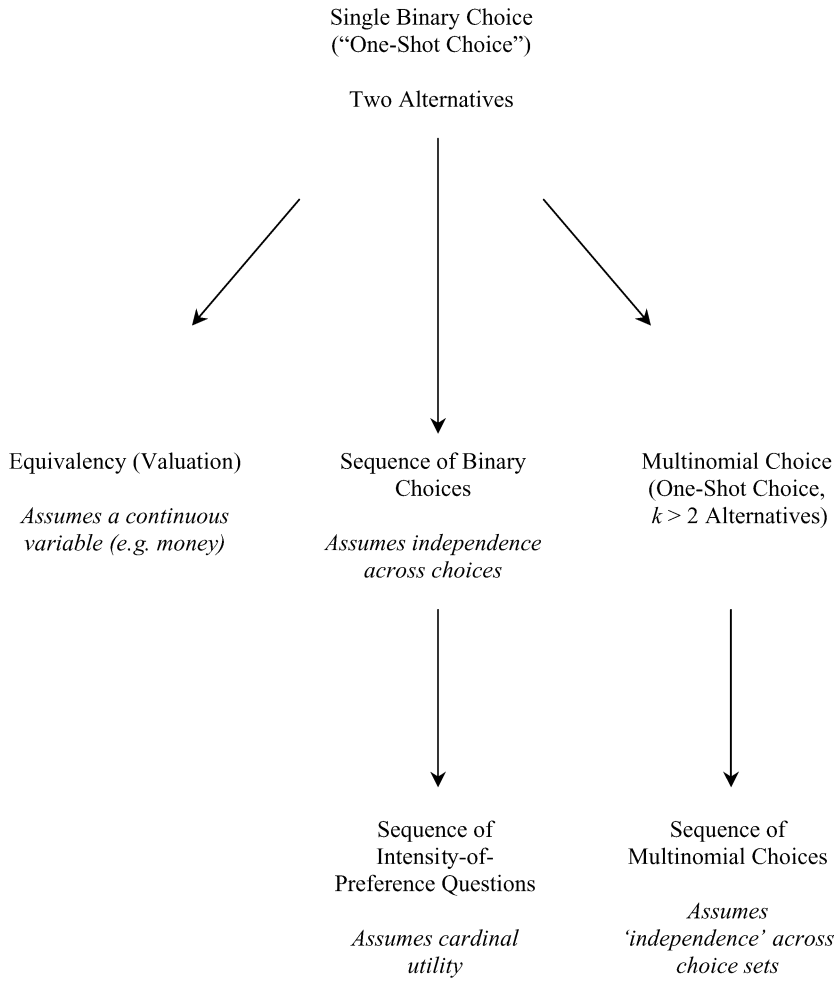


Figure 5. Typology of the structure of preference elicitation format incentives.

shot" since only one question is asked. Without the other conditions, such as a coercive payment mechanism discussed in the previous section, this elicitation format is not incentive compatible, in that truthfully revealing preferences may not be optimal for all agents depending upon their expectations. However, this typology illustrates the additional assumptions that need to be made for the subsequent elicitation format to maintain the same incentive properties of the original one-shot binary discrete choice question, whatever these may be. For example, the sequence of paired comparison format must assume independence across the pairs for this to hold. The continuous (direct WTP) response question must assume that there is a continuous underlying quantity

(typically money) that sets the change in utility from the program to zero and that the incentives for revealing a continuous response are the same as revealing a binary discrete choice response. The multinomial choice format must assume that moving from $K = 2$ to $K > 2$ choice alternatives does not alter the incentive structure. Moving to a sequence of such questions further requires the same assumption as for a sequence of paired comparisons: independent responses across choice sets. The rated pairs format requires cardinal utility; and for it to be equivalent to the sequence of paired comparisons, the revelation of the ratings (or other intensity) measure cannot alter the incentive structure.

5.5. Promoting incentive compatibility

To summarize, there are three views of the world: (a) people try to truthfully reveal their preferences irrespective of the incentives they face, (b) people tell the truth if there are no consequences associated with their answer, and (c) people only try to tell the truth when it is in their economic interest to do so. The empirical results in Carson et al. (2004b) and many other studies suggest that (a) and (b) are false.[105] That leaves (c) which is the perspective that one would if CV survey responses represented consequential economic actions by standard economic agents. The difficulty with (c) is that it makes the design of CV survey questions and their analysis much more challenging.

The additional assumptions required for the various elicitation formats in Figure 5 to be incentive compatible are often not likely to hold. However, the use of these other elicitation formats is motivated by the desire to obtain additional preference information in the form of tighter confidence intervals, or information about more possible configurations of a good, at a lower survey cost. The binary discrete choice format is potentially incentive compatible because it does not reveal much information. As such, there is an inherent conflict between the cost of obtaining preference information and the need to assess the performance of different formats (in light of the fact that they should produce different results).

One relevant question is what features promote incentive compatibility? Here one major factor appears to be cost, or more specifically, beliefs about cost. It has often been thought that survey respondents may state as their WTP amount what they perceive to be the cost of the item. In instances where there is a well-established price, a large faction of respondents simply give that number as their WTP amount. That sort of behavioral anomaly was once seen as an indication that people did not have well defined preference. The anomaly disappears once one realizes that the optimal response strategies generally pivot on the perceived cost of the good. Open-ended, direct questions are particularly problematic because many respondents simply said something akin to "tell me the cost" and "I will tell you whether I favor or oppose the project." Further, many

[105] It is important to note that this statement that some people don't always truthfully reveal their preferences, only that some will have an incentive not to and will not.

respondents believed that if the government has seriously worked out the details of how to implement the project then the cost had to be known. Promoting incentive compatibility will typically require that some sort of cost information be provided and that the statistical procedures, used correctly, are conditional on that information.[106]

Care needs to be taken in matching elicitation formats with the nature of the good being valued. The multinomial choice experiment is a natural one for quasi-public goods like recreation sites where the agent has the ability to choose a different good than another agent. It is often possible to use survey language to justify why different levels of a pure public good are being inquired about that may be natural in a sequence of paired comparisons but which makes no sense in a multinomial choice question.

More generally, there are three factors that contribute to incentive compatibility in a CV survey that are independent of the elicitation format used. The first of these is ensuring that all respondents see the survey as consequential. To accomplish this one needs to emphasize that policymakers want to know whether the public supports the project, given its cost. The second is to ensure that the good(s) being valued, including the different attribute levels and cost, are seen as plausible. The last and, perhaps, most important is to find the appropriate payment vehicle that can impose costs on all agents if the government undertakes the project. Voluntary contribution schemes, in particular, do not provide any way of asking respondents to make a real commitment with their survey response.

6. Econometric issues

6.1. Structure of WTP distributions

The key output of a CV study is an estimate of the WTP distribution for the good of interest.[107] From Section 3, two approaches are possible. One can adopt a parametric approach using some parametric functional form for the WTP distribution, relying on economic theory to guide the specification of the model, or one can adopt a completely nonparametric approach, letting the data speak for itself without imposing any assumptions about the nature of the data generating process. In either case, several issues of model specification arise.[108]

[106] In this sense, a variation on a payment card might be used to convey some sense of the range of possible costs. It is also possible to draw a connection between using a payment card in this way and the one and half bound discrete choice format proposed by Cooper, Hanemann and Signorello (2002).

[107] The discussion in this section is couched in terms of the WTP distribution for the good of interest. It could have been written more generally with some modifications to encompass a broader range of issues; such as the estimation of WTA distributions or the WTP distribution for changes in the level(s) of one or more attributes of a good.

[108] For information on the estimation of the chosen model, the reader may consult a standard graduate econometrics text such as Greene (2002) or a specialized text on the econometrics of nonmarket valuation such as Hanemann and Kanninen (1999) or Haab and McConnell (2002).

A useful way to characterize the issues is to think about the underlying shape of the WTP distribution, using the survival function representation of the WTP distribution and assuming some type of discrete choice elicitation format such as a binary discrete response, as in (9). As illustrated in Figure 2, the responses to the WTP question trace out the survival function for the WTP distribution. This representation illustrates some of the key decisions that the researcher has to make. The first decision is whether one should allow for *negative* WTP amounts. The second is whether there may be some number of respondents who have a *zero* WTP for the good. The third is how to ensure that the WTP distribution exhibits weak monotonicity as the monetary amount increases. The fourth is to determine how smooth the WTP distribution has to be away from zero. The fifth is how to deal with the right-hand tail of the response probability distribution – whether to impose some restriction on the WTP distribution as the monetary amount grows large and approaches income or wealth.

If one wishes to allow for negative WTP amounts, this implies that, for some negative amounts $\tilde{A} < 0$, $\pi^Y(\tilde{A}) < 1$; that is, the respondent would not necessarily accept the item even if he was paid to take it. This seems implausible for a wide range of environmental amenities. While it is easy to see the case where someone might not care one way or the other about say an improvement in air quality, say, it is hard to believe that someone would require compensation for an improvement in air quality in order to maintain their current level of utility.[109]

Assume for now that negative WTP is ruled out, which means that for all $\tilde{A} < 0$, $\pi^Y(\tilde{A}) = 1$. Next, there is the question of what happens at $A = 0$. Here there are two possibilities. One possibility is that $\pi^Y(0) = 1$; the other possibility is that $\pi^Y(0) < 1$, which implies that, even if the item costs them nothing, some people do not care one way or the other whether it is provided. In the latter case, the probability is $\theta \equiv 1 - \pi^Y(0)$ that people are indifferent to the item. Such a model is also said to have a "spike" at zero, the magnitude of the spike being given by θ. Given the treatment of the left tail, the next issue for the researcher is to ensure that the body of the response probability function exhibits weak monotonicity as the monetary amount increases.[110] If one uses a parametric response model, this will be satisfied automatically because weak monotonicity is a natural property of the survivor function. If one adopts a nonparametric approach resulting in the likelihood function (29), in the single-bounded case, or (49), in the double-bounded case, as noted earlier the maximum likelihood estimator automatically

[109] There are, of course, situations where some respondents may have negative WTP while others have a zero or positive WTP. Kriström (1997), looking at a change in traffic at an airport in Stockholm, shows that all three groups exist because of how the reduction in noise is traded-off against the reduction in convenience. An important case where one may see negative WTP values is in studies involving preservation versus development. Here it is not unusual to see the agents living far away from the site being considered for preservation having only zero or positive WTP values for preservation, while the WTP of some of those living close to the site may be quite negative because of the perceived connection between local jobs and development of the site.

[110] This restriction was noted earlier in footnote 9.

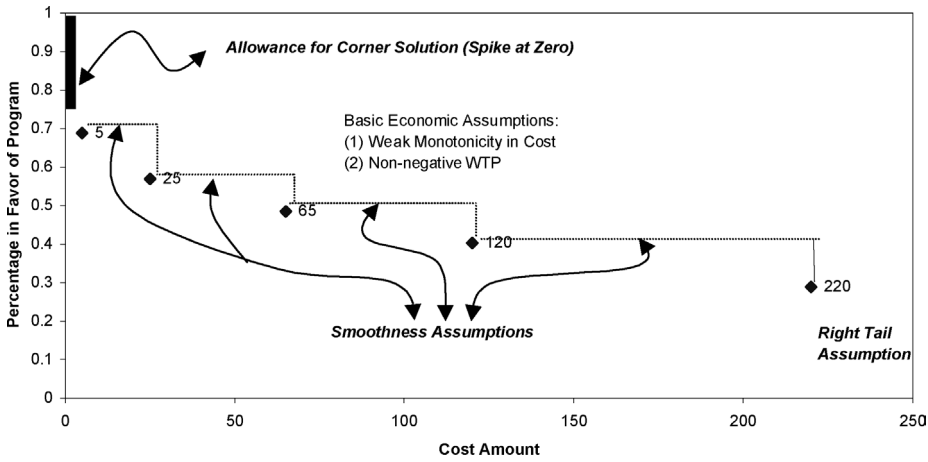


Figure 6. WTP distribution (survival representation).

imposes weak monotonicity through the pool – adjacent-violators algorithm of Ayer et al. (1955) and Turnbull (1976).

Figure 6 depicts these issues using data from a study valuing the prevention of oil spills along California's Central Coast [Carson et al. (2004a)], where respondents were randomly assigned to one of five monetary amounts ($5, $25, $65, $120, $220) in a single-bounded format. Here, the assumption has been made that no respondent *prefers* oil spills (i.e., has a negative WTP to prevent them), but the possibility of a "spike" at zero is allowed for.[111] For these data, the estimate of the Kaplan–Meier–Turnbull lower bound on mean WTP is $85.39. The standard error of this estimate is $3.90.

The final decision to be made is how to treat the right tail of the response probability distribution. This is highly influential with regard to the estimate of mean WTP which, as indicated in (38), depends crucially on the specification of $C^{\max}$, the upper support of the WTP distribution. It is natural to think of the researcher as imposing an estimate of this upper support before proceeding to calculate mean WTP. The most straightforward way to do this might seem to be simply asking some subsample of respondents about an extremely large amount. The difficulty here is that it is not credible to ask respondents about extremely high monetary amounts. Faced with an implausible cost amount, they will answer the question without reference to that amount and use what they think might be a reasonable amount.[112] One can fall back on economic theory which suggests WTP

[111] The percent in favor at the five design points in increasing order are 68.9% (219) at $5, 56.9% (216) at $25, 45.8% (241) at $65, 40.3% (181) at $120, and 28.9% (228) at $220 where the number in parentheses is the sample size at that design point.

[112] This problem is not unique to CV. Extremely high prices that come close to choking off demand are rarely, if ever, observed in data from actual markets. CV can help extend the range of prices available for the

has to be limited by wealth. Further, it might be reasonable to think that income, disposable income or some fraction of disposable income, are appropriate limits on the upper bound of an individual agent's WTP. Haab and McConnell (1998) and Carson and Jeon (2000) show that imposing such restrictions can rule out some otherwise plausible WTP distributions and dramatically reduces the sensitivity of the estimate of mean WTP to distributional assumptions which are consistent with the restriction imposed on the right tail of the WTP distribution.

6.2. Issues in fitting binary discrete choice models

The previous section has largely discussed the structure of the WTP distributions largely in terms of the nonparametric approach. There are a number of reasons why one might want to move away from this approach and fit a parametric or semi-parametric model. These include: (a) the desire to fit a particular utility function (see Section 3) to obtain a more precise estimate of mean WTP, (b) to reduce dependence of the estimate on the particular design points used, (c) to include continuous covariates, or (d) to incorporate information from more than one binary choice question, leading one to use parametric or semi-parametric estimators. In this section, we discuss several issues in the context of binary discrete choice models that can play a large role in the estimates obtained.

The first of these is the inclusion of a spike parameter (or other more complex way) of handling the possibility of a nontrivial fraction of the sample with a zero or close to zero WTP. Failure to include such a parameter often lies at the heart of many CV estimates that appear to be "too large" to be plausible. There is a simple intuition behind this result. Most of the parametric distributions fit to discrete choice CV data are proper survival distributions, such as the log-normal or Weibull. The location and scale parameters in these distributions are tightly linked. Fitting the fraction of the sample at or near zero can only be done by blowing up the scale parameter. Since mean WTP is a function of both the location parameter (the median in a log-normal and the 62nd quantile in a Weibull) and the scale parameter, the estimate of mean WTP can become greatly inflated. Ironically, an estimate suggesting a "fat right tail" problem is often blamed on the presence of so-called "yea-sayers" who answer "yes" without considering the specific amount they are asked about. However, fitting a spike parameter (which will include the nay-sayers) will often dramatically reduce the estimate of mean WTP by "dropping" the right tail. After doing so, a separate parameter of the possibility of yea-sayers is often no longer statistically significant.[113]

analysis but it cannot push that price range out beyond what most consumers would find to be plausible. The extreme sensitivity of consumer surplus measures to functional form/distributional assumption made about the far right tail is an intrinsic problem with any approach to economic valuation that looks at nonmarginal changes.

[113] There is much confusion in the literature about "yea-sayers." There is usually only a small fraction of "yes" responses to CV questions at reasonably high prices with effective payment vehicles. The strong tendency of double bounded questions to draw relatively few "yes–yes" responses if the highest design point is

The second is the choice of distributional assumptions. There are some choices that are reasonable and some that usually cause problems. The normal distribution generally implies some agents have negative WTP and as such should be generally be avoided unless negative values are thought possible. The log-logistic has very thick tails and a very wide range of plausible parameter values implies an infinite WTP. The Weibull generally has a shorter right tail than the log-normal and, in its "spike" configuration [Kriström (1997)], usually performs well. This specification should probably be the workhorse of applied CV analysis. More flexibility, as noted in Section 3, usually requires using some variant of a Box–Cox model or making a smoothness assumption and moving to a semi-parametric approach [e.g., Chen and Randall (1997), Creel and Loomis (1997), Cooper (2002)].

One of the main motivations of using parametric or semi-parametric models is the ease with which covariates can be used in the modeling exercise. This is particularly useful for assessing the validity of the CV exercise.[114] Specification of how the covariates should enter into the model can be difficult though particularly when there is a spike at zero. Werner (1999) provides an empirical example whereby income is negatively related to the probability that one has a positive WTP for preserving an area, but, conditional on having a positive willingness to pay, income is positively related to the magnitude of WTP. Simply entering income as a regressor in the standard way obscures the two offsetting effects. Haab (1999) provides a related analysis focusing on the issue of nonparticipants.

As one moves away from a single binary discrete choice question to either a double-bounded dichotomous choice format or a sequence of paired comparisons, there are two issues that should be taken into account. The first is that asking one respondent will generally result in less information from a statistical vantage point than asking the same questions to different respondents. This can be dealt with either by using an appropriate weighting scheme or by constructing one measure (such as the interval from a double bounded dichotomous choice question) from the set of questions asked. The second concerns the fact that each question asked may alter the incentive structure for subsequent questions. It is possible to test for these effects [e.g., Alberini, Kanninen and Carson (1997)] but correcting for them requires making assumptions about how the incentive/informational structure was altered.

both reasonably high and plausible runs counter to a suggestion that true yea-sayers represent a large fraction of the sample in most good CV studies. This is not to say though that problems do not exist in some individual CV studies. A large fraction of yes (or yes–yes) responses at the highest design point generally indicates a problem with the original choice of the design points or an incentive problem associated with payment vehicle. In the latter situation, it is not that the respondents are answering yes without paying attention to the amount, but that they answer yes because they want the good but will not have to pay for it. An approach such as the Turnbull lower-bound on the mean is quite robust against the sort of deviation in the right tail that a small fraction of yea-sayers might cause.

[114] For the simple purpose of describing the WTP distribution, an unconditional approach may work fine and avoids a number of modeling issues related to correctly specifying the conditional WTP distribution. However, policymakers generally want to see the conditional estimates as they care who the gains and losses accrue to.

6.3. Bid design

Collection of discrete choice data requires the use of a set of design points that represent the cost to agents, who are then randomly assigned to those design points. The choice of those design points can greatly influence how many observations are required for a given level of statistical efficiency.[115]

A good starting point is to think about the linear regression model with a single covariate and ask the question: if you had n observations and could run the experiment at two values of the covariate, what values would you chose and how many observations should you allocate to each in order to minimize the confidence interval on the covariate's parameter estimate? The answer is to pick the two values of the covariate as far apart as feasible and to allocate half the sample to each of the two values of the covariate. Some of this basic intuition carries over to binary discrete choice models which are effectively dose-response experiments where the "cost" is the "poison" and the percent of the population alive (those still in "favor") falls as the cost increases. This dose response model is nonlinear so that optimal design now depends upon having an estimate of the unknown model parameters. The optimal design, if the model's parameters were known, is generally to have the same number of design points as model parameters. The design points are not, however, placed at the far extremes of the (cost) stimulus variable because in contrast to the linear model there is too little density far out in the tails. Rather the design points are placed in the cost space in order to minimize some criteria. The most popular criterion is d-optimality which maximizes the determinant of the information matrix. This is a natural criterion when using maximum likelihood estimation and is the easiest to calculate [Cooper (1993)]. The d-optimal design, however, does not result in the smallest confidence interval for mean WTP as the estimator for that statistic is generally the ratio of two parameters and the statistical design that minimizes this quantity is known as the c-optimal design. Alberini and Carson (1990) and Alberini (1995) show that the c-optimal (fiducial) design can be substantially more efficient (on the order of 50%) than the d-optimal design under conditions relevant to CV studies. All three designs result in two design points being utilized if the underlying distribution is assumed to be fully characterized by two parameters and the design is not constrained to have more design points.[116] The designs differ in where the points are places with the d-optimal design generally resulting in placement further out in the tails. Kanninen (1993a) and Alberini (1995) provide comparisons of different design criteria for double-bounded estimators.

[115] Like survey design, experimental design is a topic not generally taught in economic departments. A classic text is Box, Hunter and Hunter (1978). For a more modern, comprehensive reference, see Atkinson and Donev (1992) or Wu and Hamada (2000).

[116] The design can be constrained to have more design points but note that forcing a design to have four design points will result in the second two design points being replicates of the original two design points, or arbitrarily close to them if they are forced to be distinct. If the distribution is assumed to be symmetric, an equal number of observations are generally assigned to corresponding design points on either side of the median. Asymmetric distributions can result in an asymmetric assignment of observations being optimal.

In practice, the main difficulty is that the model parameters are unknown when the design for the CV survey has to be derived. The natural thing to do is to use data from earlier development work such as pre-test and pilot studies. Kanninen (1993b) discusses the issues involved in sequential design which is the correct way to think about the issue. In general, the more uncertainty about the nature of the underlying WTP distribution, the more design points should be used; and it is possible to give a formal Bayesian interpretation to the design problem. The key issue though is that there is a clear tradeoff between the precision at which the distribution is pinned down at the design points and the number of design points. Alberini and Carson's (1990) work suggests that it is hard to justify more than eight design points and they show that the use of four to six design points which span the expected quartiles of the expected WTP distributions is usually preferable. This design produces estimates that are both reasonably efficient and robust to fairly large deviations between the expected and observed WTP distributions.[117]

6.4. Treatment of don't knows and protest zeros

All surveys generate "don't know" responses to some questions. Some of these don't know responses indicate a genuine "don't know" while other represent a desire not to put forth the effort needed to answer the question. Offering respondents an explicit "don't know" response tends to pick out the genuine don't knows while encouraging those who want to get through the survey quickly to effectively skip the question. The long standing wisdom in the survey research literature [Converse (1974), Schuman and Presser (1981)] was that attempting to obtain attitude/preference/knowledge information from people who do not actually possess it results in low quality answers from these respondents.[118] This view is embodied in the NOAA Blue Ribbon Panel on Contingent Valuation report [Arrow et al. (1993)] in the form of a recommendation that respondents be offered a "don't know" categorical response in addition to the usual favor/oppose options in a binary discrete choice question. There has long been a dissenting view [e.g., Bradburn and Sudman (1988)] which holds that, with some probing, many don't know responses can reliably be converted into clear opinions. Krosnick et al. (2002) review both the theoretical and existing empirical evidence and then look at nine experiments. They show in general that not offering an explicit "don't know" category does not harm the quality of survey responses and that offering an explicit "don't know" response reduces, often substantially, the amount of useful information that is collected.

Even without an explicit "don't know" response being offered, there will be some volunteered don't know responses and it will be necessary econometrically to decide how

[117] McFadden (1999b) shows that a much different design must be used involving spacing of a large number of design points over the support of WTP if one wants to be able to consistently estimate mean WTP without making parametric assumptions about the nature of the distribution.

[118] CV surveys differ from the classic public opinion survey where a respondent may be asked whether they favor a very briefly described policy that the respondent may not be familiar with. A good CV survey provides respondents with a considerable amount of information about a program being considered and then asks the respondent whether they favor the program.

to deal with them. There are three standard options available: (a) drop them, (b) treat them as "not favor" responses or (c) impute them using some type of model. Dropping them is equivalent to allocating them proportionate to the favor/nonfavor response pattern. This might be justified if one did not have any prior information about how "don't know" responses usually translate into favor/not-favor votes. In voting on ballot measures, it is clear that don't know responses in surveys taken prior to the election tend to disproportionately split with most becoming "not favor" responses. Carson et al. (1998) look at the issue specifically within the context of replicating the Exxon Valdez survey questionnaire [Carson et al. (1992)] and find that "don't know" responses (in the case where a don't know option is explicitly offered and in the case where it is not) tend to look like "not favor" responses.[119] A conservative approach, that will always tend to underestimate WTP, is to treat all don't know responses as "not favor" responses.[120]

A "protest zero" can be seen as a variant of a "don't know" response, in the sense that the respondent effectively said "no" but has given some type of response which suggests that it might not be a true zero.[121] Protest zeros were once routinely dropped from the analysis of CV data. This practice is being gradually abandoned because of growing evidence that most protest zeros tend to resemble "no" responses in discrete choice formats or "low" amounts in continuous responses formats. Further, an examination of the reasons for the protest often reveals that some aspect of the scenario is not fair, such as taxing a particular group of consumers in a particular way. Such reasoning does not really suggest a desire for the good but rather a desire not to pay for it. It is exactly the sort of additional behavior in a survey that someone who wanted to discourage the government from supplying the good should engage in. What seems clear is that these respondents would not "vote" in favor if given the opportunity. These agents may be willing to pay for the good under some quite different set of circumstances, but one cannot divorce the CV estimate from the scenario to get some "context" free value for the good.[122] Such a value in economic terms does not and cannot exist.

6.5. Treatment of respondent uncertainty

Belief that some respondents did not have well-defined preferences has led to a number of papers that propose some way to introduce an "extra" source of uncertainty

[119] Groothuis and Whitehead (2002) find a similar result for willingness to pay but not willingness to accept.

[120] This is also the cultural norm in some places where an affirmative response is required to indicate agreement with the proposition and formal disagreement with a proposition rarely voiced.

[121] For a serious attempt to statistically model don't know and protest responses see Strazzera et al. (2003). Some of these approaches rely on auxiliary questions. The key thing to remember here is that responses to such questions, while informative, can never been taken as definitive because of incentive issues and the difficulty of getting respondents to answer a question with an implausible cost.

[122] Seeing a large number of protest zeros being generated by particular aspects of a CV scenario can be important information for policymakers because it says something about one or more elements of the program that many respondents dislike. Protest responses are likely to be particularly sharp when a clearly coercive payment vehicle is used and as such protest responses *per se* do not represent a problem with a particular CV scenario.

into statistical models of CV responses and a large literature has grown up around this topic [e.g., Hanemann and Kriström (1995), Li and Mattsson (1995), Ready, Whitehead and Blomquist (1995), Champ et al. (1997), Dubourg, Jones-Lee and Loomes (1997), Ready, Navrud and Dubourg (2001)]. There can be several potential problems encountered with this approach. First, in the standard RUM model with a single observed choice, it is impossible to statistically distinguish Thurston's (1927) conceptualization of the error component in a choice model as representing a truly random element rather than the Manski–McFadden view that there is no true random component but rather aspects of the choice context that are not observed by the econometrician. As such, there is no room for an extra source of uncertainty without adding more structure to the model.

Second, identification of the extra source of uncertainty depends upon either assuming some particular specification for the utility function (and defining the deviation from that model as related to the extra uncertainty) or in having some extra piece of information that effectively reveals something about the nature of this extra source of uncertainty. The difficulty with assigning the deviation from an assumed model as the extra preference uncertainty is that there will almost always be such a deviation in any parsimoniously parameterized model. Use of an extra piece of information would be acceptable if there is a strong theoretical or practical basis for this or if empirically its introduction in a model resulted in small changes in the estimate of mean WTP and moderate increases in its confidence interval. However, models using an extra source of information tend to be *ad hoc* and the results tend to be very sensitive to specific assumptions employed.

This sensitivity has lead to increased reliance on the use of an extra source of information to identify the preference uncertainty. Typically this is obtained by asking a follow-up survey question that asks the respondent how certain they are of their answer or trying to calibrate the discrete choice answer to the estimate obtained from some other CV elicitation format. If a separate question is asked it is often posed on say a 7 (or 10) point scale with 1 labeled "completely certain" and 7 labeled "completely uncertain." There are a several potential difficulties. First, different respondents may use the same scale in different ways so that a "3" to one respondent may mean a "5" to another respondent. Second, asking for an assessment of certainty of the answer to a previous question may be taken as challenging the respondent as to whether he made the "correct" choice; there can be considerable heterogeneity in the response to such a challenge that has little to do with preference uncertainty. Third, there is inevitably some endogenity in the "certainty scale" response in that an agent's whose WTP is close to the particular bid amount is more likely to be uncertain than one whose WTP is far away from that amount. Fourth, it is not necessarily clear how one translates different levels of certainty into different purchase propensities.[123] A related issue arises when trying

[123] There may be some empirical regularity that allows the use of this certainty level in some type of calibration exercise. These exercises are popular in the marketing literature, where one can often observe repeated comparisons of survey results and actual market behavior. The difficulty here is keeping the marketing effort constant across the different comparisons or systematically accounting for differences in it.

to calibrate the discrete choice response obtained using a different elicitation format. If theoretically there should be differences between the two formats then there is no reason to make the two responses conform.

Another way to approach these phenomena is through the scale parameter in a RUM setting. Some agents may have more uncertainty about their future economic status, a possibility under the Manski–McFadden view of RUM models, or, some agents may have a larger random element to their purchase behavior, a possibility under the Thurston view of choice models. Survey design can also be seen as having an influence on errors of optimization that is not uniform across respondents. Any of these possibilities induces heterogeneity in the error component. Two main ways to deal with this are to use a nonparametric estimator like the Kaplan–Meier–Turnbull that does not require an estimate of the scale parameter to obtain an estimate of mean or median WTP, or to formally model the error component as a function of variables reflecting the respondents and the circumstances of the preference elicitation.[124]

6.6. Continuous response CV data

The statistical analysis of data from CV formats eliciting continuous (open-ended) responses is in principle straightforward except for two main sets of problems. First, the incentive structure of the format tends to produce substantial distortions in the response obtained, as discussed in Section 5. Second, respondents tend not to give answers with tremendous precision like $11.14, but rather, tend to give more general and less specific responses like $10.

The incentive structure of open-ended CV surveys gives rise to a substantial number of "zero" WTP responses and few positive, yet very small, responses. There have been many attempts to sort the zero WTP responses into two types: true zero WTP responses and "protest" zero responses. While such efforts may be of some help in understanding the zero responses, they may never perform this classification task really well. This is because it is in the strategic interest of respondents who have already misrepresented their WTP to continue to do so.[125]

As Brookshire, Ives and Schulze (1976) first pointed out, strategic behavior in a CV question that obtains a continuous, or close to continuous response can flatten out the observed WTP distribution relative to the true WTP distribution.[126] This observation

[124] Louviere et al. (2002) consider a wide-range of issues associated with efforts to model differences in the scale component across agents and their implications for commonly used statistics such as WTP.

[125] This is also true of efforts preceding a WTP question (discrete choice or continuous response) to try to sort out those with a positive WTP from those with a zero WTP. It is in the interest of respondents who think that their WTP is less than what they expected the cost of the program to be to indicate a zero WTP initially and such a question may be answered with expected cost in mind rather than the "zero" or "trivial" cost that the researcher intended.

[126] In Brookshire, Ives and Schulze's original formulation higher bids increased the probability of provision without shifting the cost burden. This flattening may not occur on the upper end of the distribution if the respondent believes the government willing to shift the cost burden toward those types that indicate a high WTP for the good.

naturally leads to the proposal in Mitchell and Carson (1989) to use an α-trimmed mean where an α in the statistics is often set to 0.05 or 0.10 so that 5% or 10% of the observations are "trimmed" off before calculating the mean on the remaining observations. Because measurement error is common and generally not normal as often assumed, a 10% α-trimmed mean has even been shown to perform better [Stigler (1977)] than the ordinary sample mean even in estimating physical constants from experiments. However, one problem with the family of α-trimmed means is that they involve symmetric trimming, which can substantially alter the estimate of mean WTP if the true underlying distribution is asymmetric, as can often be the case with WTP data. There are two approaches to dealing with this problem. The first is to use a variant of an α-trimmed mean known as a windsorized mean; here, instead of dropping the trimmed observations before calculating the mean, the α-smallest observations are set equal to the smallest nonwinsorized observation and the α-largest observations are set equal to the largest nonwinsorized observation.[127] This will preserve much of the asymmetry in the WTP distribution as long as most of the value is not concentrated in far right tail. The second is to trim observations conditional on the observed covariates based on their likelihood of being an outlier using the approach proposed by Belsley, Kuh and Welsch (1980) and pioneered in a CV context by Smith and Desvousges (1986).[128]

6.7. Choice experiments

Choice experiments have become a very popular elicitation format in recent years. They generalize the binary discrete choice CV format by allowing for the valuation of changes in one or more of the attributes of the good of interest. The standard binary discrete choice CV question generally asks for a choice between a bundle of attributes, which could be labeled as the status quo good, and another bundle of attributes which could be labeled as the "alternative." One of the attributes of the alternative (usually cost) is randomly assigned to respondents and this allows the researcher to trace out WTP for the alternative relative to the status quo. The choice experiment format expands upon this by varying more than the cost attribute.

The simplest way to do this is to create two variants of survey instrument that differ only by changing one of the noncost attributes of the alternative, and administer each to statistically equivalent samples. Now two WTP distributions will be available that differ

[127] The Kaplan–Meier–Turnbull lower bound on mean WTP can be seen as a variant of a windsorized estimator for the highest design point, as it assumes the fraction estimated to be WTP the highest design point are willing to pay only that amount and no more.

[128] Because the incentive properties of the payment card are reasonably well known and generally conservative in the sense of being biased downward, if some light trimming of the right tail is done if a few large outliers are present, it may be a good format to use in studies where there is one program to be valued and it is not possible to use the large sample size required to obtain reasonable precision from a binary discrete choice question.

only by the difference in the attribute levels between the scenarios in the two different surveys. This variant of a choice experiment is sometimes referred to as a scope test [Arrow et al. (1993), Carson (1997a, 1997b)]. All of the parametric and nonparametric approaches [Carson et al. (1994)] can be used to test whether there is a difference in the WTP distributions corresponding to the difference in scope. Poe, Giraud and Loomis (2005) also show how the convolutions and bootstrap approaches can be applied to this case.

More than one level of an attribute can be valued simply by expanding the number of variants of the survey with one binary discrete choice question in each survey and the same clearly defined status quo in each one. Different attributes can also be valued in the same way either alone or in combinations. The difficulty with this approach is that it may become very expensive in terms of survey costs, as fairly large sample sizes are needed for each variant of the survey. The typical choice experiment gets around this problem by asking respondents a sequence of choice questions, each with a different alternative, or by asking a multinomial choice question featuring two or more alternatives to the status quo or both.[129]

In these models, the utility for a particular alternative using the standard RUM formulation is: $U_{ij} = V_{ij} + \varepsilon_{ij}$, where U_{ij} is the utility of the jth alternative to the ith agent, V_{ij} is the systematic part of the utility function, potentially observable to the econometrician subject to some sort of parameterization and ε_{ij} is an unobservable component. V_{ij} without loss of generality can be written as $f_i(X_j)$, where X_j is a vector of k attributes for the jth alternative. If one is prepared to assume that ε_{ij} are i.i.d. error terms that have a generalized extreme value distribution and that $V_{ij} = X_j\beta$, then the estimation task is straightforward, fit a linear in the parameters conditional logit model to the data.

However, these assumptions may be questionable in practice [Hensher, Rose and Greene (2005)]. Relaxing the independence part of the i.i.d. assumption on the error term leads to nested logit models or multinomial probit models which allow some type of correlation structure between the error terms from different alternatives. Relaxing the identical part of the model allows for different respondents to have different scale effects, which can play an important role in estimating the distribution of WTP for a change in an attribute.[130] When a respondent answers multiple choice questions, one

[129] Louviere, Hensher and Swait (2000) provide a comprehensive overview of choice experiments from multiple perspectives including environmental valuation, marketing and transportation. They provide a wealth of advice on experimental design and fitting statistical models. Reviews in the environmental valuation literature include Hanley, Wright and Adamowicz (1998), Adamowicz et al. (1999), Hanley, Mourato and Wright (2001), Bateman et al. (2002) and Holmes and Adamowicz (2003).

[130] The key thing to keep in mind here is that the parameters estimated in these models are actually β_k/σ where σ is the scale term. Obtaining an estimate for β_k requires an estimate for σ, which is usually parameterized as $-1/\beta_p$, where β_p is the coefficient on the cost term. Allowing σ to differ across either individuals or alternatives raises interesting issues with respect to determining the WTP distribution for a marginal change in the kth attribute. There is growing evidence that σ varies with demographic variables like age and education and some alternatives involve more uncertainty than others in a way that is not reflected in the measured X_j. DeShazo and Fermo (2002) show how the complexity of the choice task can influence the error component.

might also reasonably expect that the error component is correlated in some fashion across choice questions for the same respondent. Relaxing the $V_{ij} = X_j\beta$ assumption can take several forms. The first is to allow different agents to have different preference parameters. This can take the form of a random coefficients model [Layton (2000)] if one believes there is some reasonably continuous distribution of preferences, or a latent class model [Morey, Thatcher and Breffle (in press)] if one believes that there are a small number of consumer types. The second is to relax the linearity assumption. This can be done in two ways. The first is by allowing nonlinear transformation of the individual attributes comprising X_j. The second is by allowing interactions between the different attributes comprising X_j. Both of these raise the issue of whether the nonlinear transformation and/or interaction terms are statistically identified. What features of the attribute space are identified is a function of the experimental design used.[131]

What should be clear from this discussion is that it is far removed from the nonparametric Turnbull framework, where one could impose the very minimal assumptions of economic theory on the data and rely on a fairly simple random assignment of cost across respondents. The choice experiment framework with a nontrivial number of attributes and multiple questions asked of each respondent requires a large number of modeling assumptions in order to recover the parameters of interest. Results are often sensitive to those modeling assumptions, although statistical tests of fit can often be used to rule out some alternative specifications.

From Section 5 we know that moving away from the binary discrete choice framework, either by adding additional alternatives or asking about additional choice sets, can introduce incentives for nontruthful preference revelation. Because respondents do not pick randomly but, rather, want their most preferred bundle of attributes supplied at the lowest possible cost, these deviations from truthful preference revelation are likely to be reflected in picking an alternative in utility space close to the most preferred one. As such, one would expect the incentive structure of choice experiments to generate IIA violations.[132] Because nontruthful preference revelation relies on information about other alternatives and beliefs about how the responses will be used, the estimated parameters of the model may change as the respondent goes through a sequence of choice questions. This phenomenon has often been labeled as "learning" [about preferences]. From an incentive perspective, there may be learning but about incentive structure of the mechanism being used to elicit preference information rather than preferences themselves. It will generally be impossible to sort out which type of learning is going on.

An attraction of the choice experiment format to some is that it allows tradeoffs between different goods and between different configurations of the same good. There is a

[131] Statistical identification and statistical efficiency are two distinct concepts with the former being a necessary condition for the latter.

[132] Unfortunately, there are a multitude of other reasons for observing such violations so that their presence cannot be used as a test for nontruthful preference revelation, although the absence of IIA violations would imply that there were not incentive problems.

temptation to take this property of choice experiments and go a step forward and eliminate cost as an attribute. Particular in natural resource damage assessments, there is an interest in doing "resource to resource" scaling in the sense of replacing the damaged resource with some type of substitute. As Flores and Thatcher (2002) show, however, just because one is substituting other resources for the damaged ones, the lack of need for "money" in the model is largely an illusion. In order to be able to consistently aggregate welfare measures over heterogeneous consumers one needs at least one attribute that has a common metric. Money is the obvious attribute and there is not generally a good substitute.

6.8. Models with covariates

For many purposes, all that is needed is an estimate of the unconditional WTP distribution or statistics derived from that distribution. However, policy makers will often want to know various aspects of the conditional distribution; such as, whether there are WTP differences between different income groups and between residents of different geographic areas. In addition, one of the standard ways to look at the reliability of a CV survey is to construct a regression model to predict differences in WTP as a function of other variables in the survey; such as income, past recreational use, and various attitude and knowledge questions concerning the good. An equation with reasonable explanatory power and coefficients with the expected signs provides evidence in support of the proposition that the survey has measured the intended construct. If this is not the case, either the research team has failed to collect the relevant covariates in the survey, suggesting inadequate development work, or the WTP responses are random and completely useless. From an econometric perspective there are no special issues involved in estimating the WTP equation beyond those normally experienced with survey data.[133]

7. Survey design

The valuation scenario presented in the survey instrument lies at the heart of any CV study. The scenario must convey the change in the good to be valued, how that change would come about, how it would be paid for, and the larger context that is relevant for considering the change. It must do this in a way that is consistent with the underlying scientific/engineering reality and yet still be comprehensible to respondents who may

[133] Some of the typical issues that will arise are that while theory might suggest that a particular covariate is positively related to WTP, it rarely gives any guidance on the functional form. Many variables from survey data are categorical or ordinal in nature. Different respondents may use questionnaire scales in different ways. Endogeneity issues can arise with respect to the incorporation of covariates derived from debriefing questions. Measurement error is likely to attenuate the coefficients for many variables, and income in particular. Missing values will be present on some covariates and the pattern is not likely to be missing at random, so that dropping observations with missing regressors can create problems.

know little or nothing about the good in question. The respondent needs to be given enough information to be able to make an informed decision, but without being overwhelmed by the information. In general, this is not an easy task because few economists are trained in survey design. The best of the CV surveys represent the current state-of-the-art in survey design.

The development of the survey scenario is a key aspect of a CV study.[134] This is because it forces the sponsor of the survey to decide what the environmental good offered actually will provide to the public in terms that the public cares. In this regard, it is often surprising to find out that the policymakers seeking to evaluate policies do not know what the regulations they are contemplating will actually accomplish in terms of changes in environmental quality.[135] The effort by the survey team to get the interested parties to endorse the depiction of the good described in the CV scenario often exposes deep conflicts about what the proposed policy will actually do and, perhaps just as importantly, what the current baseline *status quo* is.

A CV survey should have what is known as "face validity." The environmental good and the circumstances under which it would be provided should be described clearly and accurately, and the tradeoff that the respondent is asked to make should be a plausible one. The information provided should be adequate for the decision the respondent is asked to perform but should not overwhelm the respondent with unimportant technical details. It is sometimes the case that a high level decision maker who has not been substantially involved earlier in the process of evaluating the policy at hand can learn more about the actual decision by reading the CV survey instrument than the various technical reports on the proposed project.

While it is generally impossible to provide concrete advice that is universally applicable, most good CV surveys contain the following: (a) an introductory section that helps set the general context for the decision to be made; (b) a detailed description of the good to be offered to the respondent; (c) the institutional setting in which the good will be provided; (d) the manner in which the good will be paid for; (e) a method by which the survey elicits the respondent's preferences with respect to the good; (f) debriefing questions about why respondents answered certain questions the way that they did; and (g) a set of questions regarding respondent characteristics including attitudes and demographic information.

A substantial amount of development work is usually required to produce a high quality CV survey instrument. Sometimes there is a substantial amount of prior work to draw on that resolves the major conceptual issues concerning the good and its provision.

[134] Further, with good development work through the pilot study phase, the survey team should have a reasonable idea of the results that are likely to be obtained from fielding the survey instrument to a large random sample of the population of interest.

[135] In some instances, there will be substantial scientific uncertainty about what the risks associated with a particular policy are. If there are a small number of distinct outcomes, then it is possible to value each outcome separately to see how sensitive the results are. Carson, Wilks and Imber (1994) provide an example involving proposed mining in the Kakadu Conservation Zone in Australia.

In such cases, there may still need to be local adaptation and careful modification for the inevitable difference between the good being valued in the earlier survey and the current one. In other cases, the type of good to be valued and/or the context in which the valuation is to take place will have been little studied. In this situation, conducting a successful CV study will require considerably more development work.

Development work typically includes focus groups and in-depth interviews to help determine the plausibility and understandability of the description of good to be provided and the context in which it is being provided. The task of translating technical material into a form understood by the general public is often a difficult one. Developing a useful CV survey instrument requires the research team to clearly define what the proposed project will produce in terms of outputs that people care about and in language they understand. Pretests and pilot studies will usually need to be conducted to assess how well the survey works as a whole. Some elements of the survey will usually be needed redesign to improve respondent understanding and the overall flow of the survey.

There are five key issues that will almost always need to be addressed during the development phase of a CV survey. The first is how much information to provide to respondents. Respondents need to be given enough information to make an informed decision but without being overwhelmed with information that they don't think they need. What information to give respondents will depend in large part on what information they already possess. The most difficult cases are where respondents are misinformed about key elements of the scenario and hold their views strongly.[136] Another difficult situation is where the information set held varies considerably across respondents with a substantial fraction of the population being very well informed and a substantial fraction of the population having little prior knowledge.

The second issue, one that always has to be addressed, concerns payment for the good. Here the payment vehicle has to be plausible and, as noted in Section 5, it has to be coercive in nature if incentive compatibility is desired. It is often hard to meet both criteria for every respondent. Another aspect of payment is whether it is a one time lump sum or recurrent payment. As a general rule, something which looks like a capital investment, such as setting aside a wilderness area, should use a lump sum payment mechanism while something like a water quality improvement that would disappear if there were not continued payments should use a recurring payment.[137]

[136] Likewise, prior experience with the good can be useful or detrimental when designing a CV scenario. It is useful when it reduces the amount of information that must be given to respondents and it is harmful when there are substantial differences between the current state of the good and its condition at the respondent's time of prior use. While some critics of CV claim that prior experience with a good is a necessary condition for CV to work, such a condition neither follows from economic theory nor accords with people acting as rational agents purchasing unfamiliar goods or voting on new political candidates. By contrast, when respondents are largely uniformed about the good to be valued, this is fairly straightforward for a good survey designer.

[137] If payment is on less than an annual time scale (e.g., a monthly utility) bill it may be helpful to also provide the cost to the respondent on an annual basis so there is no confusion and the recurring nature of the payment made obvious.

The third issue is the formulation of questions to help explain respondent WTP for the good. While it is certainly possible to design a good CV scenario without attention to this part of the survey instrument, it is always desirable to being able to explain differences in the WTP responses; and doing so, enhances faith in their reliability. In some instances, economic theory directly provides guidance and in others there are obvious relationships that should hold between attitudinal or resource use variables. A different set of relationships are likely to exist with indicators related to beliefs about the underlying scenario. Focus groups are a good source of information on what motivates different types of responses. In large samples substantial differences in WTP may be found across different demographic groups, although these relationships may not be linear ones.

As one might expect, respondents who do not believe that the good promised can be delivered with certainty are likely to discount how much they are willing to pay for the program to deliver the good. While it is a desirable goal to minimize divergence between the stated survey scenario and what respondents believe, there will always be some divergences. Debriefing questions can help the researcher in determining the implications of any such divergences.

The fourth issue that needs to be addressed in the design of the CV survey is making respondents feel comfortable with making either a "favor" or "oppose" decision.[138] In particular, respondents need to feel that while the technical details of the proposed program have been well-worked out, the implementation of the program is not at all a foregone conclusion and the public's input via the survey will play an important role in that decision.[139] This consideration needs to run throughout the CV survey. Before the response to the main valuation question is elicited, it is useful to briefly summarize why someone might be for or against the project. After obtaining the response to the valuation question and asking the standard set of debriefing questions and demographics including income, it may be useful to give respondents who favor the program an opportunity to change their answer.

The fifth issue that often must be confronted in the development work for a survey are the choice of stimulus, such as cost amounts and the range of other attribute levels. There are several considerations here. First, the levels of the attributes have to be plausible. This can be explored in focus groups but more subtle problems sometimes manifest

[138] The advice in this paragraph bears some resemblance to the use of "cheap talk" language in experiments advocated by Cummings and Taylor (1999) that tells agents to act as if the experiment was real and that people in hypothetical situations tend to over-estimate. While experiments can prove a fruitful venue for testing the efficacy of such interventions, there tend to be two main issues in evaluating their relevancy to CV surveys. The first of these is that good CV surveys should not be posed as "inconsequential," purely hypothetical questions because there are no theoretical predictions about what the response to such questions should be. As such results from experiments with a purely hypothetical (inconsequential) payment may be of very limited use. The second is that the "actual" monetary or behavior transaction in experiments held up as the benchmark often has incentive issues of its own.

[139] The danger in providing a very sketchy idea of a plan to provide a good is that the respondent may take the question posed as asking about whether the government should develop a concrete proposal on this topic or not. As such the respondent may not pay much attention to the cost stated in the survey because that amount is not thought to represent a firm estimate for the program.

themselves in pretest and pilots studies of the survey instrument administered under field conditions. Second, the choices of attribute levels need to provide for reasonable efficiency in statistical modeling. Generally, optimal design for attribute levels depend upon unknown parameters but reasonable inference about these parameter values can often be drawn from pretest and pilot study data. Third, there will often be experiments of one sort or another that will be performed in the course of the main study. Determination of the sample sizes to provide adequate power to test the hypotheses of interest will also require statistical information that can be taken from pretest and pilot studies.[140]

It is beyond the scope of this chapter to get into further details of how to conduct the survey research needed to design a good CV survey. CV surveys are among the most challenging surveys to design. They can be thought of as a structured conversation with a respondent; whereby, a large amount of information is conveyed and where the respondent is engaged in the task of providing preference information about a proposed policy change. Fortunately, there are a number of good books on designing and testing survey instruments [e.g., Bradburn et al. (2004), Presser et al. (2004), Tourangeau et al. (2000)].[141] There is also specific advice for designing CV studies in several books [Mitchell and Carson (1989), Louviere, Hensher and Swait (2000), Bateman et al. (2002), Champ, Boyle and Brown (2003), Alberini, Bjornstad and Kahn (in press)] and a pair of recent journal articles [Mitchell (2002), Whittington (2002)]. The latter article specifically focused on conducting CV studies in developing countries, which can pose some unique challenges.

8. Survey administration

8.1. Defining the population of interest

In much of economic analysis, the issue of defining the relevant population is routinely ignored because it is assumed to be obvious; in other words, those in the marketplace who buy the good of interest. The question of the relevant population becomes difficult, however, once it is asked: who are potential buyers of a good in a market or users of a

[140] It is all too common to see CV studies with sample sizes that are much too small to answer the question(s) being asked. Underlying WTP distributions are generally characterized by fairly large variances so that even studies using elicitation formats that obtain continuous responses need large sample sizes. Sample size requirements are further increased when discrete responses are collected. Collecting multiple responses from a single respondent can sometimes reduce sample size requirements but because of the correlated nature of such responses much less information is collected than often assumed. A basic reference on statistical power in experimental design is Bausell and Li (2002).

[141] Stanley Payne's (1951) book, *The Art of Asking Questions* is still invaluable reading for anyone thinking about how to structure a survey interview. Perhaps the best advice to readers is that they "will be disappointed if [they] expect to find here a set of definite rules or explicit directions. The art of asking questions is not likely ever to be reduced to some easy formulas (p. xi)."

government service under a different set of circumstances?[142] CV studies almost always have to face this issue. For example, an improvement in water quality may induce more households to fish.

There are two different perspectives on defining the population of interest in a CV survey. The first perspective is a legal/political one. The agency sponsoring the survey may only care about the welfare of particular subgroups such as taxpayers in their jurisdiction. The second perspective is based on a rough consideration of the cost and benefits of sampling from different groups. For instance, while there may be some value of a change in a local public good to an agent living far away, such agents may be few and far between. To concentrate the interviewing effort on such agents would neglect the agents near the resource who would experience the largest change in utility from the change in the resource.

8.2. Survey mode

There are a number of different ways in which a survey can be administered, including mail, telephone, and in-person.[143] There are four main considerations in choosing the mode of survey administration. The first of these has to do with how the characteristics of the sample of the population of interest are influenced by the survey administration mode. The second has to do with the nature of the stimuli that can be presented to the respondent. The third is the degree of control that the survey mode gives to the researcher over the order and consistency in which stimuli are presented to respondents. The fourth is the cost in term of money and time to obtain a sample of the desired size.

Some of these considerations may be more important for CV surveys valuing particular types of goods. For example, a survey trying to value visibility improvements in a national park will likely require a visual depiction of those changes; and hence, would be unsuitable for administration over the telephone. In-person interviews that take place at a recreation site may be the only feasible way to obtain a reliable set of survey responses from visitors to that site. Further, the prevalence of technology, such as the fraction of households possessing telephones or connected to the internet, or cultural norms that may differ across countries in ways that should be taken into account when deciding what mode of survey administration should be chosen for the CV survey.

[142] Defining the population of users is a contentious issue in litigation involving issues of potential harm from unfair trade practices. In marketing there are many examples (e.g., low cost airlines) where moving away from the usual surveying of current users to considering potential users who might be "in the market" if one or more attributes of the good were changed resulted in dramatically expanded sales.

[143] There is a very large literature on how to sample from the population of interest and undertake the physical administration of surveys. On sampling, the interested reader should consult a standard statistical text such as Cochran (1977) or Levy and Lemeshow (1999). On the administration of surveys, the interested reader should consult one or more of the following standard texts: Dillman (in press), Fowler and Mangione (1989), Groves et al. (2004), Lavrakas (1993), and Presser et al. (2004).

There are three classic survey administration modes: in-person with an interviewer, telephone and self-administered. These can be seen at some level as representing the survey questions being (a) read face-to-face by the interviewer to the respondent, (b) having the survey questions read to the respondent but not face-to-face, and (c) having no direct interaction with an interviewer when responding to the questions. The most common variant of a self-administered survey is a mail survey. Other variants include "intercept" surveys where a potential respondent is found at some location (other than their home or place of work) and given a survey to fill-out at that location[144] and more recently internet based surveys.

Hybrid variants of the three classic modes are also sometimes implemented. One can, for instance, do a telephone–mail–telephone survey [Smith and Mansfield (1998)] where respondents are recruited to take the survey (as well as providing responses to some questions), information (often visual) is mailed to them often with some self-administered questions and a follow-up telephone interview used to collect the answers to those questions and to ask additional questions. The use of this hybrid survey points to a key issue in the calculation of response rates. The standard way of calculating response rates will multiply the response rates at each stage together to obtain the final response rate.[145] As such, even if the response rate at each stage is considered to be quite reasonable by survey standards, final response rate will look fairly low. For the same response rate, the sample selection bias in this type of multistage survey is likely to be much lower than the equivalent selection bias in a single stage survey with the same response rate.

8.3. Sampling approaches

The key to drawing appropriate inference from any sample is that each member of the population of interest must have a known positive probability of being included in the sample. This implies either the existence of a list of the population or some way of generating that list. An example of the former is a database of fishing licenses if the population of interest consists of those that fished during a particular year. An example of the latter is using random digit dialing in a telephone survey. In the absence of a list or means of accurately generating a list, it may be necessary to do some type of enumeration before sampling from it. A different issue arises when the choices made by a respondent influence their probability of being included in a sample. An example here includes any type of on-site recreation survey [Moeltner and Shonkwiler (2005)]. The frequent characteristic of a choice-based sample is that it excludes people who are not currently undertaking an activity, which can be important when a main focus of the

[144] The intercept approach can also be used in conjunction with giving respondents a survey to fill out later and return by mail.

[145] Any type of internet panel with an initial stage of randomly recruited participants will have similar sampling properties.

policy change of interest is on the margin between some use and no use rather than solely focused on changing the amount of use by users.

Stratification of a sample increases its efficiency in that a small sample size is needed relative to simple random sampling for the same level of statistical precision. It does this by pre-determining the size of the subsamples from different strata. For example, in a national survey, states/provinces are natural strata, and drawing the correct proportion from each stratum eliminates the possibility of having a sample that comes disproportionately from one state.

The other common survey design feature is clustering, whereby multiple individuals are interviewed in fairly close proximity. Clustering in a sample reduces the efficiency of a sample relative to simple random sampling. It is often done in conjunction with in-person surveys to reduce interviewer travel time and costs. A variant of clustering is when one chooses (randomly) a relatively small number of locations at which to conduct an on-site survey. Clustering effects with preference data are not generally large if the number of clusters is reasonably large because of the degree of heterogeneity generally observed. However, measurements of other phenomena such as income or ethnicity can be highly correlated within a cluster which can give rise to large cluster effects.

In most instances, sampling is done "proportionately" so that each selected agent has the same weight. There are cases, however, where this is intentionally not the case. The usual example is where it is desirable to report estimates for some specific subgroup of the population of interest and a level of statistical precision beyond what their proportionate representation in the overall sample is desired. In such an instance, members of the subgroup can be over-sampled. The more common case occurs when the response rate to the survey differs systematically across identifiable subgroups of the population of interest. Here, the usual solution is to create sample weights so after weighting the data obtained that each subgroup is represented in the correct proportion in the weighted sample. Professional survey companies which administer the questionnaire usually supply the appropriate weights.

The departures from simple random sampling discussed in this section are generally corrected by some type of weighting scheme. For large samples with proportionate sampling and high response rates, the influence of stratification and clustering on response data from CV preference elicitation questions is typically small. This is not the case once one is dealing with small samples, nonproportionate sampling and/or choice-based samples. Standard sampling texts such as Cochran (1977) and Kish (1967) lay out the relevant details.

8.4. Nonresponse

With any survey, one gets nonresponse [Madow, Olkin and Rubin (1983), Groves et al. (2001)]. Nonresponse is of two types: failure to respond to the survey as a whole, and failure to respond to a particular question. The latter is often referred to as item nonresponse. The simplest case is where the response is considered "missing at random" [Little and Rubin (2002)]. A complete interview or response to a particular survey

question is missing at random if dropping it does not alter either the unconditional or conditional distribution of interest. If this is the case, then for the purpose of simple tabulation of responses on a particular question, then the missing interview or missing answer on a particular question can be ignored with the effective sample size simply being smaller.

The situation is more difficult when the response is not missing at random. There are two relevant cases here. The first one is where the missing at random assumption holds after conditioning on a set of observable covariates. An important instance of this condition occurs when different identifiable subgroups in a population respond to the survey as a whole or to a particular survey question at different rates. As noted in the previous section, one way to deal with this issue is to weight the data using the information on how the response rate is related to the covariates. A different (but to a large degree similar) way to deal with this issue is to impute the missing values using a nonparametric [Mitchell and Carson (1989)] or parametric framework [Little and Rubin (2002)]. The second is where the response is not missing at random after conditioning on the available covariates. This leads one to use some type of sample selection model along the lines pioneered by Heckman (1979). Sample selection bias is most likely to occur when potential respondents can decide, as is typically the case in a mail survey, whether to answer the survey questions after learning the main focus of the questionnaire. The nature of the selection bias in such cases may be very bimodal, with those with the highest and lowest values for the good most likely to respond.

8.5. Sample size

The issue of sample size should be considered in the context of issues related to survey mode and the likely divergences from simple random sampling including stratification, clustering, and nonresponse issues.[146] The level of precision needed for the policy decision at hand and testing the hypotheses of interest should drive the sample size decision. The degree of precision necessary for the CV results to provide a useful input to the decision making process can vary substantially. In some instances, only a rough order of magnitude comparison between benefits and costs may be required, while in other instances, relatively small changes in an estimate may influence the preferred outcome.[147] This consideration should be reflected in the sample size chosen.

[146] Care should be taken in assessing all these issues due to the use of different conventions. In particular, how response rates in surveys are calculated varies considerably. For example, it is often the case in mail surveys that surveys returned due to "moved/bad addresses" are taken out of both the numerator and denominator before calculating the response rate. This would not be the case in a door-to-door survey as the new resident of the dwelling unit is eligible to be interviewed. Thus, there is some interaction between how the response rate to a survey is defined and the nature of any sample selection problem related to a deviation between the effective sample frame used and the population of interest.

[147] Further, where the cost estimate for the project is known before the CV survey is conducted, it may not be necessary to trace out the entire WTP distribution, only the part relevant to the decision at hand. This has implications for the design points used in the survey and the sample size required.

There are too many CV surveys that are carried out with a sample size that is inadequate to meet a study's objectives. There are areas where there are legitimate tradeoffs that can be made involving sample size. For instance, while it may be desirable to do in-person surveys, an internet-based survey with careful attention to matching the demographics of the population of interest may be able to obtain reasonable results at substantially lower cost. For the same cost, a mail survey with a high response rate may well be preferable to an in-person survey with a low response rate. Further, while it may be desirable to ask only a single binary discrete choice question, substantially more information can be obtained from either shifting to a different elicitation format such as a payment card or by asking multiple questions. While this may introduce some bias from an incentive perspective, Cameron and Huppert (1991) show that the use of a binary discrete choice without a large sample size and careful attention to the design points can produce estimates that have very high variances.

The standard warning that nonsampling errors may be important in any survey should be taken seriously. Groves et al. (2004) provides a good overview of thinking about surveys from a total error perspective.

9. Consistency of CV results with theoretical prediction

There are a number of interesting issues that CV results have raised about neoclassical economic theory which have prompted extensions or clarifications of that theory. CV surveys have sometimes revealed violations of neoclassical economic theory; however, similar violations can also be found in consumer behavior in actual markets. CV raises issues of incentive structures that are largely ignored in the conventional analysis of consumer demand but which dominate much of contemporary game theory. CV also raises some issues about the structure of markets and the aggregation of preferences which were of relatively little practical import until it was possible to use CV as an empirical measuring tool. The flexibility of CV takes the economist away from the role of using data passively and into the realm of survey and experimental design. Surveys and experimental design interact in interesting ways because of the possibility of doing random assignment, the key aspect of experiments, in the context of different variants of a survey instrument. Finally, CV raises issues related to consumer knowledge and understanding that are little discussed in the conventional context of consumer demand analysis.

9.1. Neoclassical theory and contingent valuation

One of the more interesting aspects of the history of CV is its role in motivating a much deeper understanding of the implications of neoclassical economic theory in the case of pure public goods. Careful examination of neoclassical economic theory helps to explain phenomena that at various times have been pointed to as anomalies, suggesting that CV surveys do not measure economic value [Hausman (1993)]. These include the

(a) magnitude of the divergence between WTP and WTA, (b) the magnitude of the income elasticity often estimated in CV studies, (c) the large effect the order in which goods are valued can play, and (d) the failure of the Diamond (1996) adding-up test to hold in many situations.

The key to understanding these issues was the unraveling of structure of demand for imposed quantity changes that was set in motion by Hanemann (1991). In that paper, the possibility of sharp difference between price changes and quantity changes emerged so that WTA could be infinity while WTP was a small finite number. Hicks himself seems to be aware that the structure of neoclassical theory was quite different. This is clearly reflected in the title of his classic article "The Four Consumer Surpluses" which today rather than equivalence and compensating variations and surpluses could have could have been characterized as WTP versus WTA crossed with small price changes versus imposed quantity changes. The inherited wisdom though was largely driven by Willig's seminal paper (1976) that argued that the difference between WTP and WTA was sufficiently small to be ignorable (for small price changes) where the qualification in parentheses was almost always dropped. Willig was careful not to include quantity changes in his analysis and it took quite some time for the full ramifications of the difference between the price change and imposed quantity changes cases to emerge.[148] Eventually, Cornes and Sandler (1996) was to show that pure public goods was simply a special case of rationed (i.e., imposed quantity change) goods.

9.2. WTP versus WTA

If total value, in an economic sense, can always be expressed in terms of WTP and WTA and the two measures differ substantially, either theoretically or empirically, the appropriate measure for a benefit–cost analysis depends upon the *property right*. The difference between the Willig and Hanemann theoretical results is that for a price change, an income effect alone governs the difference between WTP and WTA; and for a quantity change, both an income effect and a substitution effect together govern the difference. One of the earliest findings from CV studies was that WTP and WTA measures differed substantially [e.g., Hammack and Brown (1974)]. Based upon Willig's work, it was thought either one or both of the CV estimates were wrong, rather than understanding of the relevant theory being wrong. Work proceeded in several directions. One was to

[148] Randall and Stoll (1980) were the first to expand Willig's framework to consider the imposed quantity changes but the full import of the differences between the two cases was not clear until Hanemann (1991). The discovery of these differences first found in the context of CV studies, coupled with the initial belief that Willig's results were more widely applicable, caused some researchers to claim the observed CV results violated economic theory. Hanemann (1991), in a standard neo-classical theoretical framework, shows that for imposed quantity changes, WTP and WTA can be infinitely far apart due to interactions between income and substitution effects. Kahneman and Tversky (1979) had earlier put forth an alternative they called "prospect theory" where losses are valued more highly than gains. In the situation usually considered by economists, a single imposed quantity change, Kahneman and Tversky and Hanemann's models yield the same results.

show that large differences between WTP and WTA estimates were not an artifact of the survey context, but rather, consistently large differences were found in a variety of settings using actual transactions [e.g., Knetsch, Thaler and Kahneman (1990)].[149] Even financial assets such as junk bonds and over-the-counter stocks, when thinly traded, often show much larger bid (WTP)-ask (WTA) spreads than would be predicted by Willig's work. Another direction was to show that the WTA question format had a number of shortcomings, both from the perspective of its strategic incentives and of getting respondents to accept it as a legitimate framework for a policy choice. Still another direction was to suggest new theories outside the neoclassical framework [Kahneman and Tversky (1979)] and to show that within that framework [Hanemann (1991)] that the theory being applied failed to capture key aspects of the situation.[150] Much of the problem with the current framework may stem from its inherent static nature. Recent models that incorporate bargaining, information effects, transactions cost/experience, dynamic considerations and uncertainty, show considerable promise in being able to explain the magnitude of the divergence between WTP and WTA amounts [e.g., Kolstad and Guzman (1999), Kling, List and Zhao (2003), Zhao and Kling (2004), List (2000)]. The key implication of this divergence for applied policy work is that property rights can have a substantial influence on the magnitude of the welfare measure. This is true particularly when considering a reduction in an environmental service, where the common practice of substituting a WTP estimate for the desired WTA measure can result in a substantial underestimate, which, in turn, can have substantial policy implications [Knetsch (1990)].

9.3. Income elasticity of WTP

Drawing inferences about economic values from intuition regarding the demand for private goods, one expects to see a positive relationship between income and WTP, if the good being valued is a "normal" good. A frequent made claim, for which there is surprisingly little empirical support, is that most environmental goods are "luxury" goods.

[149] A striking feature is the standard empirical finding that WTA estimates based on actual transactions are usually much larger than CV WTP estimates for the same good. In particular, Horowitz and McConnell's (2002) review of the literature suggests that the ratio of WTA to WTP estimates is roughly the same for surveys and actual transactions. For a deep critique of the WTP–WTA divergence that is prepared to abandon substantial parts of neoclassical theory in the direction of some of the main tenets of behavioral economics and that recognizes that the issues involved are "not specific to CV surveys" see Sugden (1999).

[150] That a price change where the consumer is free to adjust is different from an imposed quantity change where the consumer cannot adjust seems obvious in retrospect. Indeed, it was clear to Hicks (1943) who first clearly developed the concept of utility constant welfare measures. Willig was also careful to specify that he was looking at price changes. This acknowledgement was largely left behind in the rapid incorporation of Willig's work in benefit–cost texts. Willig's work showing that WTP and WTA were close in most situations involving price changes and that the Marshallian consumer surplus measure lay between WTP and WTA. This justified the common applied practice of using the Marshallian consumer surplus as adequate approximation to the desired Hicksian measure.

If this were the case, one would expect the income elasticity to be greater than one. The usual empirical result from CV studies is to find a positive income elasticity of WTP substantially less than one for environmental commodities. This empirical result has been cited as evidence that contingent values are theoretically deficient. For instance, McFadden (1994) reporting on one of Exxon's studies notes:

> An economic interpretation of the results on the relationship of income to WTP in these experiments is that preservation of Selway-Bitterroot wilderness is a "necessary" good, with a low-income elasticity. However, it seems economically plausible that preservation would be a "luxury" good that for poor households is displaced by needs for food and shelter.

The problem is that the terms *necessary* (e.g., normal but not luxury) and *luxury* are defined in terms of the income elasticity of *demand*, not in terms of the income elasticity of WTP. Flores and Carson (1997) show that the two types of income elasticities are fundamentally different. The two income elasticities can be shown to be functionally related using the concept of a shadow or virtual price that responds to changes in the levels of rationed goods. WTP is found by simply integrating (summing) the virtual prices for infinitesimal quantity changes over the discrete quantity change in the rationed good of interest. The relationship between the ordinary income elasticity of demand and the income elasticity of WTP for two goods is given by:

$$\begin{bmatrix} \eta_1^{\mathrm{v}} \\ \eta_2^{\mathrm{v}} \end{bmatrix} = -\begin{bmatrix} \sigma_{11}^{\mathrm{d}} & \sigma_{12}^{\mathrm{d}} \\ \sigma_{21}^{\mathrm{d}} & \sigma_{22}^{\mathrm{d}} \end{bmatrix}^{-1} \begin{bmatrix} \eta_1^{\mathrm{d}} \\ \eta_2^{\mathrm{d}} \end{bmatrix} \frac{y}{e^{\mathrm{v}}},$$

where η_i^{v} are the (virtual) income elasticities of WTP, the σ^{d} are the cross price elasticities of demand, η_i^{d} are the ordinary income elasticities, and y/e^{v} is income divided by e^{v} which is equal to income plus the value of *all* public goods consumed expressed in monetary terms, which may well be quite large.

Flores and Carson's results show that for any fixed value of the income elasticity of demand, the income elasticity of WTP can differ significantly in magnitude and even sign. Thus, a good which is a luxury good in a demand sense may have a WTP income elasticity which is less than zero, between zero and one, or greater than one. If the matrix of cross-price elasticities is an identity matrix, the virtual price income elasticity is equal to the ordinary income elasticity of demand multiplied by a scale factor (the ratio of income to income *plus* the monetized value of all public goods), which must be less than one and probably substantially less. Thus, the income elasticity[151] of WTP is likely to be less than the corresponding income elasticity of demand.

[151] In empirical estimates of the income elasticity of WTP there also may be measurement issues associated with income that tend to bias the estimate toward zero. It is well-known that income is often poorly measured in surveys without a great deal of time and effort in the survey devoted to structuring this task. Further, disposable income after taxes and other fixed expenses such as housing and car payments is likely to be the relevant income measure. It is also likely that agents have mental accounts [Thaler (1985)] they consider when making decisions about environmental goods that may greatly reduce the relevant income available.

9.4. *Sequencing, nesting, scope insensitivity and the adding-up test*

We now turn to the relationship between CV estimates for multiple, possibly unrelated goods. Here, the context in which the CV exercise takes place is crucial. Two issues have received the most attention. The first involves the implications of adding together CV WTP estimates for different goods. The second involves the influence exerted on the estimated value of the good by the order in which it is valued as part of a sequence of goods. The two typical empirical findings turn on the same underlying theoretical issue: substitution and income effects.

The first empirical finding suggests that adding up what people say they are willing to pay for specific goods, each valued independently as the only change to the *status quo* (or equivalently valued first in a sequence), might easily exceed the income of some people. This strikes many nontechnically oriented CV critics as conclusive proof that CV estimates, if not complete nonsense, are gross over-estimates. However, Hoehn and Randall (1989) show theoretically why adding together independently derived WTP estimates for goods is likely to overstate the value of the set of goods taken as a package, and often grossly so. At an intuitive level, the reason is simple: each new public good the agent obtains reduces the agent's available income to spend on private goods. Further, if the public goods are substitutes for each other, then each one added to the package looks less desirable than when valued as if it were the only new addition to the stock of public goods. The problem should not be seen as residing with the original CV estimates, but with the analyst's incorrectly aggregating them without taking into account income and substitution effects.[152] The second standard empirical finding is that the value of a good falls, often precipitously, the later it is valued in a sequence of goods. Consider a stylized example reminiscent of some of the early work on air pollution valuation [Randall, Hoehn and Tolley (1981)]. A subsample of respondents in Chicago are willing to pay $100 for a specified air quality change in Chicago; and when offered an additional specified air quality improvement in the Grand Canyon, they are willing to pay $30 more. A different subsample of respondents for whom the sequence is reversed are willing to pay $60 for the Grand Canyon improvement and $70 for the Chicago improvement. Such a result may be disturbing to the policy maker who expects a good to have only one "true" value.

The standard economic explanation for this phenomenon is substitution and income effects. Carson, Flores and Hanemann (1998) show that if one assumes that the goods being valued are normal goods and (Hicksian) substitutes for each other, the value of a particular public good should be progressively smaller the later in a WTP sequence it is valued. An implication of this result is that the package of goods should be valued less than the sum of its independently valued constituents. The opposite effect occurs

[152] The problem of taking account of multiple changes has long been known to be troublesome in the benefit–cost literature [Just, Hueth and Schmitz (1982)] and is in no way specific to the use of CV estimates. The problem is often ignored in many benefit–cost applications due to the time and expense associated with determining the interactions between the different goods and the belief or hope that such effects are small.

in a WTA sequence; the later in a sequence the good is valued, the more highly it is valued.[153] Furthermore, the usual weak assumptions made concerning the curvature properties of utility functions effectively rule out the existence of a single context independent value for a particular public good.

CV critics counter that the sequence effects observed are too large because they contend the income effects should be small and goods such as those in the air quality example above are not close substitutes.[154] However, the CV critics' arguments about the likely magnitude of income and substitution effects are faulty because they are based on intuition derived from looking at price changes for private goods. Public goods are a special case of quantity rationed goods and, as a result, the focus should be on quantity space with an inverse demand system rather than price space with an ordinary demand system where consumers are free to choose their optimal consumption levels. Flores (1995) shows that the set of virtual price substitution elasticities that should lie behind the magnitude of any sequence effects is the inverse of the set of cross-price elasticities of demand upon which the CV critics' intuition appears to be based.

Consider the following set of compensated, cross-price elasticities of demand (σ_{ij}^{d}) taken from Deaton's (1974) well-known analysis of consumer demand in the United Kingdom. Good one is food and good two is clothing:

$$\begin{bmatrix} \sigma_{11}^{\mathrm{d}} & \sigma_{12}^{\mathrm{d}} \\ \sigma_{21}^{\mathrm{d}} & \sigma_{22}^{\mathrm{d}} \end{bmatrix} = \begin{bmatrix} -0.28 & 0.08 \\ 0.21 & -0.17 \end{bmatrix}.$$

Note that own-price (-0.28 for food and -0.17 for clothing) and cross-price elasticities (0.08 for the effect on food demand of a price increase in clothing and 0.21 for the effect on clothing demand of a price increase in food) in this example are all quite small. Thus, with respect to either good, the percentage change in demand will be small relative to the percentage change in either own price or the other good's price. Hence, particularly large context effects for price changes would not be expected. However, if one restricts choices, as is the case with environmental goods where the levels are usually collectively decided, a regime of partial rationing is in effect.

Rationing requires consideration of the inverse relationship – how the shadow or virtual prices for the rationed goods (food and clothing) respond to changes in the rationed levels of both of these goods. These measures of responsiveness, the virtual price substitution elasticities (σ_{ij}^{v}), are related inversely, as a system, to the compensated price

[153] The reason for this is that one is destroying existing public goods in a WTA sequence. As one goes farther and farther out in the sequence, fewer substitute public goods remain, and since the agent is compensated with money at each step in the sequence, income is increasing. Carson, Flores and Hanemann (1998) further show WTP for a good valued first in a WTP sequence is less than WTA for the good valued in any order in a WTA sequence.

[154] A typical example is Kahneman and Knetsch (1992). McFadden (1994) makes the argument in a somewhat different way as a conclusion based upon his empirical analysis of a CV data set from a wilderness area study: "These results indicate that either there are extraordinarily strong diminishing returns to preserving additional wilderness areas, or that there is a context effect that makes responses inconsistent with classical economic preferences."

elasticities [Madden (1991), Flores (1995)]. For the food and clothing example, the virtual price matrix of substitution terms is:

$$\begin{bmatrix} \sigma_{11}^{\mathrm{v}} & \sigma_{12}^{\mathrm{v}} \\ \sigma_{21}^{\mathrm{v}} & \sigma_{22}^{\mathrm{v}} \end{bmatrix} = \begin{bmatrix} \sigma_{11}^{\mathrm{d}} & \sigma_{12}^{\mathrm{d}} \\ \sigma_{21}^{\mathrm{d}} & \sigma_{22}^{\mathrm{d}} \end{bmatrix}^{-1} = \begin{bmatrix} -5.60 & -2.55 \\ -7.19 & -9.33 \end{bmatrix}.$$

The same demand system cross-price elasticities which implied fairly small increases in demand of one good when the price of the other good increases (an 8% increase in food demand accompanying a 100% price increase in clothing and a 21% increase in clothing demand accompanying a 100% price increase in food), now implies very large reductions (255% and 719%, respectively) in WTP if a unit of the other good has already been provided first in the WTP sequence. This example with private goods shows that one need not resort to explanations of inconsistent preferences or goods with peculiar characteristics to predict quite large context effects with respect to public good values.

Substitution effects are sufficient to drive the sequence effects observed in CV studies. Income effects, however, are likely to play a role as well. CV critics argue that since respondent WTP is usually just a small fraction of income, income effects should be small. Much of a household's income is already committed so that available discretionary income is much smaller, particularly if payment is required over a short time period. Further, income is known to be poorly measured in general population surveys [Sudman and Bradburn (1982)]. These sources of measurement error probably bias estimated income effects downward.

CV critics such as Kahneman and Knetsch (1992) respond that if sequence effects are indeed large, then CV estimates are arbitrary because they can be manipulated by the choice of the sequence order. Kahneman and Knetsch's statement is applicable to economic analysis in general which, if done correctly, is context specific,[155] as value in an economic sense is always a relative rather than absolute concept [Debreu (1959)]. Even more to the point is Flores's (1999) demonstration of a formal equivalence between the agenda control problem and WTP sequences for a set of public goods. As agenda control is a central issue in public choice [Mueller (1989)], it would have been surprising to see the use of CV somehow avoided it.[156]

Closely related to the issue of sequencing and often confused with it are the issues of nesting and scope [Carson and Mitchell (1995)].[157] Nesting occurs when one good can be considered a proper subset of another. This can occur in two distinct ways. One

[155] For specific examples of context effects on consumer surplus measures for price changes for private goods, see Lave (1984) or Randall and Hoehn (1996). The economic intuition behind such effects is the loss in consumer surplus for a specified increase in the price of chicken that is likely to be much larger if there are also large concurrent increases in prices of beef, pork, and fish.

[156] At least partially for reasons of agenda control, benefit–cost analysts are almost always told by policy makers to focus on whether a proposed program potentially improves social welfare rather than being asked how to globally maximize societal well-being.

[157] The term "embedding" was originally used by Kahneman and Knetsch (1992) in their well-known paper to refer two distinct phenomena. One involved sequencing of goods where neoclassical economic theory, as discussed above, predicts the same good should have different values depending upon the order in which it

situation is where different items are being valued and these can be considered in different combinations; the other is where different numbers of units of the *same* item are considered. An example of the first case is where one values some specified quantity of forests, fish and birds. This good nests the same quantities of forests and fish, which in turn nests the same quantity of forests. In the second case, the large quantity nests the smaller one. The key to sequence effects involving nested goods is the degree of substitution or complementarity between the separate items; as noted above, large sequencing effects here are consistent with neoclassical theory.

When two goods, one of which nests the other, are valued in the same order, there is a clear theoretical prediction that the larger good should be valued the same or higher than the smaller good, if both the nested good and its complement are seen as desirable. The test of this proposition has become known as the scope insensitivity hypothesis. Again, the key here is the nature of individual preferences. If scope insensitivity cannot be rejected, there could be two possible economic explanations. One possibility is the inability to perceive an *effective* difference in quantity; just as an individual might not notice the difference between 50 grams of sugar and 51 grams, so too she might not consider the difference between a 250 page novel and a 300 page novel to be a meaningful difference. Another explanation is satiation: additional units of the item have zero marginal utility. A less extreme and probably common variant of this is that the environmental amenity is subject to sharply declining marginal utility [Rollins and Lyke (1998)]. If so, fairly large sample sizes will be needed to adequately test the scope insensitivity hypothesis; and more generally, power is often a concern in empirical tests of this hypothesis since it is meaningless to fail to reject differences of plausible magnitude.

The empirical evidence is that there is some sensitivity to scope for a wide range of goods. Carson (1997a, 1997b) reviews a fairly large number of studies where two nested goods were valued separately (typically for policy purposes and a desire not to contaminate the response to different scenarios) and finds that the hypothesis of scope insensitivity is usually rejected. There are, however, clear examples where it is not. These examples seem to fall into a few key areas. First, the use of voluntary payments seems to be associated with scope insensitivity, which should not be surprising given the incentive structure associated with this type of mechanism. Second, instances where there is an attempt to value a good such as preserving a particular species but where it is clear that any effective protection plan has to involve preserving a larger set of species. The issue here appears to be plausibility of the provision mechanism. A different variant of this issue sometimes occurs when respondents believe that a "small" good is much more likely to be successfully provided than a "larger" variant of the good that nests the smaller one.[158] The third is where there are multiple metrics on which the good can be

is valued. The other involved the relationship between the value of two goods where one of the goods in the language of Carson and Mitchell (1995) nests another other. Carson and Mitchell (1995) recommend that the term "embedding" not be used because of ambiguity in its meaning and relationship to theoretical predictions.

[158] Ironically, this problem is often caused by trying to keep the survey language for the two goods as close as possible, whereas it often takes more effort in terms of survey design to get the perceived probability of provision for the larger good equal to that of the smaller good.

measured. The best known example here is the Desvousges et al. (1993b) chapter of the Hausman (1993) volume which finds respondents having roughly the same WTP to save the lives of 2,000 birds, described as much less than 1% of the population; 20,000 birds, described as less than 1% of the population; and 200,000 birds, described as about 2% of the population. If respondents were concerned about the percentage impact, rather than the absolute impact, these difference would seem trivial.[159] Finally, the valuation of changes in low level risk seems to always present substantial difficulties. This should not be surprising, though, given the difficulty people have in actual markets dealing with these issues.

Another context-related consistency test, termed an adding-up test, has been proposed by Diamond (1996). At an abstract level, the test follows from satisfying duality properties that are commonly assumed [and commonly violated, e.g., Bateman et al. (1997)] in other areas of applied microeconomics. The test requires that a sequence of specified changes add-up to the set of changes taken as a package. There are two practical difficulties with the test that come to light in trying to operationalize it using multiple subsamples of respondents. The usual approach to structuring the CV survey questions involves asking at least one of the subsamples to "pretend" that they had already received a specified good and paid a specified amount for it. It may be difficult, however, to get respondents to take such an exercise seriously. The other involves making the substitution assumption implicit in Diamond's illustrative example that respondents are indifferent between a program which prevents some number of existing birds from being killed and a hatchery program which produces the same number of new birds [Smith and Osborne (1996), Kopp and Smith (1997)]. Substitute children for birds and the implication of this assumption becomes striking.[160]

10. Consistency of CV results with actual behavior

Distrust of survey results often leads economists down the path of asking "How do CV results correspond with actual behavior?" A little reflection here, however, suggests that the question is ill-posed, at least to the extent that asking this question usually implies that a divergence would indicate a problem with CV. Ideally, one would like to see CV

[159] A recent replication of the bird survey from Desvousges et al. (1993b) using just 1% versus 10% of the population finds significant scope sensitivity; Hanemann and Felipe Vasquez (2005).

[160] The difficulty with the test can also be illustrated with a simple example involving hotel rooms where the substitution issue becomes sharply focused. Consider three scenarios. In scenario 1, your maximum WTP for a shabby hotel room (the only one available) in a city you are interested in traveling to is $50. In scenario 2, you have already made a nonrefundable payment for the shabby hotel room and your employer for some reason offers to reimburse you for the $50 you paid for this room so your income is back to the same level as in scenario 1. Now a room becomes available at a 5-star luxury hotel and your maximum WTP for it is $200. In scenario 3, you have the opportunity to make a reservations that includes both the shabby hotel and the luxury hotel. For Diamond's adding up test to work, you need to be willing to pay $250 for this opportunity in scenario 3.

estimates consistent with actual behavior when theory suggest they should be consistent and divergent when theory suggest that they should be divergent.

It may be useful to get the two main divergent cases out of the way first as they seem to be the examples CV critics often focus on. The first is the case of comparing actual voluntary contributions to survey derived estimates of likely contributions.[161] The second is the case of comparing willingness to pay or purchase rates for private goods to estimate of the same quantity based on surveys.[162] Sometimes the "revealed" behavior is taken from actual markets and sometimes from economic experiments.[163]

Comparing actual contributions to survey-based estimated suggests that that the survey estimate is higher than the actual contributions and often by a very considerable factor. This is what one should expect. The actual contributions should grossly underestimate actual willingness to pay because of the free-rider problem that has long been at the core of issues surrounding the provision of public goods. Less obvious is that the estimate from the survey should over-estimate true WTP. That is because if the survey is consequential as noted in Section 5, the main effect that indicating a willingness to contribute can have is to encourage an actual (voluntary) fundraising drive to be mounted. That effort must be taken in order for the agent to have the opportunity to get the good for free by having others pay for it through their voluntary contributions. Thus, all that can be said is that actual contributions should over-estimate true WTP and the survey responses should underestimate it.[164]

Private goods do not suffer the free-riding problem that is seen with public goods. As such, many CV proponents and critics have thought private goods represented the "best case" situation for comparing market and survey-based estimates. What this perspective failed to take into account is the incentive structure that respondents face in the CV survey. A "yes" response indicates a willingness to purchase the new good, and as such encourages the firm to produce the good and offer it to consumers. If the good is produced, the consumer can decide whether to purchase the good. The agent's response to the survey might still be informative in the sense that a "no" response to a particular amount might be interpretable as an actual no, but a "yes" response only translates to a "maybe" under the right future conditions. As such, a survey for a new private good traces out "potential" demand not "actual" demand. While this is useful information from a marketing perspective, the expected result is again the observed result: surveys tend to over-estimate, often by a substantial fraction, the percent who

[161] For examples, see Duffield and Patterson (1991b), Seip and Strand (1992), and Champ et al. (1997).

[162] For examples, see Infosino (1986), Neill et al. (1994), and Cummings, Harrison and Rutström (1995).

[163] In the economic experiments, agents are said to have "home-grown" preferences/values rather than having those assigned by the experiments. See [Harrison and List (2004)] for a review of this issue.

[164] Chilton and Hutchinson (1999) correctly point out that this bound doesn't necessarily hold if one switches the provider of the good (from say a respected nonprofit organization to the government) and agents who provide the good. This issue is potentially important if agents care about the type of provider even if the coercive nature of the payment mechanism is unchanged.

actually purchase the good. Contrary to popular wisdom, the private goods case should be seen as a worst-case rather than best-case situation for comparing actual behavior to survey-based estimates.

There are a couple of important caveats to this statement that surveys should overestimate purchase behavior with respect to new goods. The first is that this overestimation finding is typical with respect to those that purchase over some fairly short time horizon. Again, this is what one would expect. Over time, the fraction of purchases is likely to converge toward the survey estimate as the "right circumstances" happen for more and more agents.[165] The diffusion of information about the good, as well as the marketing effort, can influence this horizon. The second is that the expected result from a marketing survey that is interpreted as a "pricing" exercise rather than a "new goods provision" exercise will be the opposite. Respondents will be more price sensitive, that is they indicate they are less willing to purchase at a given price, than agents actual do in the marketplace [Brownstone and Small (2005)]. Again, this is a common finding in the marketing and transportation literature when dealing with an existing rather than a new product.[166]

One can turn to two situations where CV is heavily used where it is possible to make comparisons of survey-based estimates with revealed behavior. The first of these is with quasi-public goods, such as outdoor recreation, where there are other techniques such as the averting behavior, hedonic pricing, and travel cost analysis that can be used with different types of revealed behavior. Carson et al. (1996) perform a meta-analysis of 616 comparisons of CV estimates to revealed preference (RP) estimates from 83 separate studies and find that the mean ratio of CV to RP estimates is 0.89 with a 95% confidence interval of [0.81–0.96], suggesting that CV estimates in the case of quasi-public goods are on average a bit lower than RP estimates and fairly highly correlated (0.78).[167] What is perhaps just as interesting is that there is a clear publication bias, in that, two types of studies are most likely to get published: those with a CV/RP ratio close to 1 and those with very large CV/RP ratios.

Mechanism design theory suggests that a consequential binary discrete choice question with a coercive payment mechanism and a take-it-or-leave-it offer should be incentive compatible. Long-standing evidence in the polling literature suggests that surveys

[165] This is consistent with the results of a major World Bank review [Water Demand Research Team (1993)] of CV estimates of water system hook-ups in developing countries where CV estimates initially over-estimated the percent of households subscribing to water services. However, the CV estimates were reasonably accurate and not systematically biased forecasts of behavior over long time horizons.

[166] It is interesting that without a coherent theory of how people respond to surveys, the two marketing survey literature results, over-estimate for new goods and "too" price sensitive for existing goods, have coexisted in the form of simply "forecast" problems using surveys.

[167] There are more estimates than studies because some studies valued multiple goods or used a different statistical approach to value the same good. Giving equal weight to each study rather than each comparison, results in an average CV/RP ratio of 0.92, where the confidence interval [0.81–1.03] does not reject the ratio, is 1 at the 5% level. Meta-analysis of particular types of goods, such as outdoor recreation [e.g., Rosenberger and Loomis (2000)] suggests a similar conclusion.

done right before the election do well in predicting two candidate races and ballot initiatives, providing some support for this proposition.[168] Carson, Hanemann and Mitchell (1987) did a CV survey modeled on proposed California ballot measure on a water quality bond issue as part of a Field Institute California Poll and found a close correspondence between their survey estimate and the percent in favor in the actual vote.[169] More recently Champ and Brown (1997) found a similar result on a road related referendum in Colorado and Vossler et al. (2003) found a similar result looking at an open-space bond issue in Oregon.[170]

The other approach to looking at whether the theoretical prediction holds in practice is with some type of economic experiment. Here, the evidence is mixed and much more controversial. List and Gallet (2001) survey the experimental literature and find that experiments with "real" economic commitments tend to result in lower estimates than comparable "hypothetical" treatments. As Little and Berrens (2004) note, most of the studies List and Gallet (2001) review are not directly relevant to assessing the performance of binary discrete choice questions cast in a referendum context. Cummings et al. (1997) perform such an experiment that was later extended by Cummings and Taylor (1998). These two papers show that the "hypothetical" treatment had a higher fraction in favor than the treatment where agents were required to pay if their group voted in favor; and that the percent in favor monotonically declined as the probability that the group's vote was binding. The Carson, Groves and Machina (2000) framework shows neoclassical economic theory provides no prediction about how agents faced with a "purely hypothetical," that is an inconsequential survey question should behave. That framework does, however, provide a strong prediction that as long as the probability of influencing the decision is positive, then the vote/survey response should be invariant to the actual probably. Carson et al. (2004b) conduct an experiment looking at this prediction. In the experiment, groups vote (privately) on whether to supply a public good where the probability that the vote would be binding was randomly assigned before the

[168] There are two important caveats here. The first is that the "survey" is taken close to the election. Elections are to a large degree information campaigns where the two sides disseminate their version of the "truth" and the information set held by voters often changes considerably over the course of an election. The second is that the "don't know" responses to a ballot measure question appear to get largely translated into a "not in favor" vote in the election. This is not surprising, since voting for the "not in favor" alternative almost always maintains the *status quo*.

[169] The measure was put on the ballot by the state legislature and did not have any organized group who spent substantial amounts supporting or opposing the measure. Surveys of support for a ballot measure are incentive compatible if respondents believe that there is any type of bandwagon effect such that other voters are more likely to favor a measure the higher the percent in favor when the survey results are released to the public.

[170] In both of these studies, as in the Carson, Hanemann and Mitchell (1987) study it is necessary to disproportionately allocate the don't know respondents to be "not favor" responses. This is consistent with the survey literature on ballot measures. Vossler and Kerkvliet (2003) working with more detailed data from the Oregon land use measure find disproportionate allocation of the don't knows to "not favor" is not needed to get a close correspondence between the actual vote and the survey results, suggesting that the issue may be related to a divergence between the survey population and those who vote.

vote, and find this to be the case.[171] They do find that the inconsequential case does behave differently in terms of providing a higher percent in favor.[172]

As a final note in this section, there is a literature [e.g., Blackburn, Harrison and Rutström (1994)] that attempts to find a calibration factor that would scale CV estimates so they would equal an unbiased measure of "true" willingness to pay. However, it is unlikely that there can exist a single scaling factor that applies to all circumstances. The Carson, Groves and Machina (2000) theoretical framework suggests that in some contexts there should be a divergence and in other cases there should not – a result consistent with the empirical evidence. Further, the inconsequential case which should play no role in CV appears to play a large role in such calibration schemes. Calibration schemes are potentially of more use in circumstances such as marketing where an incentive compatible format is not usually available. In a marketing context, factors such as advertising/informational campaigns can be varied, and there are repeated opportunities to first observe the survey response and then the actual purchase decision over a specific time frame. In such cases, there may be useful empirical regularity that can be exploited for forecasting purposes. Even then, it is important to carefully consider what is being observed and what the researcher would like to observe. For instance, there may well be an empirical regularity between survey indications of likely contributions to a voluntary fund to provide a good and actual contributions toward it, but neither is the correct measure of the benefits of providing the good.

11. Concluding remarks

After over twenty-five years of working on CV from both a theoretical and empirical perspective, we are struck by one key insight – there is nothing particularly unique about CV. Once one realizes that responses to consequential surveys represent economic behavior in the standard sense that economist think about, then much of the previously troubling differences between different ways of eliciting preference information and the differences observed across different types of goods and contexts makes sense. Once one realizes that the predictions of neoclassical economic theory are quite different for the small price changes that characterize most discussions of welfare economics and the imposed quantity changes that typify much of environmental valuation using CV, then

[171] The experimental setup involved participants at a sports memorabilia show who were randomly assigned to treatment groups. Each group voted on whether to provide one ticket stub to a famous baseball game. In the base treatment if a majority voted in favor, all members had to pay the stated price and got one ticket each. Otherwise, all members of the group paid nothing and did not get a ticket stub.

[172] Carson et al. (2004b) develop a treatment in which the vote can influence a second outcome. In this case, the theoretical prediction is that the vote in favor should be monotonically decreasing with the probability that the vote is binding. They further argue that this is the problem in the Cummings and Taylor (1998) experimental setup because a vote in favor might encourage a fundraising effort that could provide the good at no cost.

results that were once characterized as aberrant become consistent with standard theory if not expected.

CV does pose challenges, however. Few economists are experienced at crafting survey instruments. They typically make do with whatever data is available and try to acknowledge its limitations. Economists have little experience in thinking through the nuances of collecting data in an ideal form. Even with ideal data, questions of functional form do not disappear and addressing questions of consistency with theoretical predictions requires auxiliary assumptions.

In part because there is a substantial degree of control of what data is collected, it is easy to demonstrate that respondents violate some of the tenets of the standard narrowly self-interested rational optimizing model that is commonly assumed in modern economics. Often these tenets do not follow directly from neoclassical economic theory, but rather, are assumed because they make the usual analyses much easier to conduct. CV respondents seem to care about many aspects of how programs are implemented and their impacts on other people. This should not be surprising, since this is what actually happens in the real world. Policymakers would like to believe that economic values are independent of details like whether a sales tax or utility bill is used to pay for the good. They would also like to have the value of the public good as if it were provided without any complications. Unfortunately, the willingness to pay estimates must of necessity be for the program to provide the good and not the good itself.[173] CV estimates, like all economic estimates, are context specific. It behooves the analyst and the policymaker to understand the context in which the valuation estimate was obtained.

Where CV has had its largest effect in environmental economics is in opening up the operational possibility of measuring the benefits of a wide range of changes in environmental goods. Obtaining preference about these goods was always technically possible, through having the public directly vote on the issue, but for all practical purposes, infeasible when extended very far. Until the widespread use of CV, economists were often left examining only the cost side of a benefit–cost analysis which gave rise to many of the stereotypical views of economists on environmental issues. Being able to measure the public's willingness to pay for environmental goods is, though, a double edge sword that invites controversy. The value will inevitably be too high or too low for one of the parties with a vested interest. Because the tool involves expensive and time consuming original data collection, CV will tend to be used when there is conflict that needs to be resolved.

While it is always wise to scrutinize specific CV studies and the specific uses that they are put to, the danger of not using CV seems to us much larger. David Starrett (1988, p. 293) makes the same point well with reference to welfare economics in general:

[173] This statement is indeed true of all consumer goods in the marketplace even though it is routinely ignored by economists but not marketers. Reputation matters in the form of brand effects as do measures such as warranties and "money back guarantees." Akerlof's (1970) now classic lemons papers underlines how pervasive perception of quality and likelihood of future service flows are in determining consumer willingness to pay.

> It is very important to reject the view that since welfare measurement is still quite imprecise we may as well leave pubic decision making to the politicians. To do so invites mistakes that are costly on a scale that dwarfs any possible measurement error.[174]

CV can help decision makers to identify the public's interest. It is particularly useful in two cases. One is where the benefits of providing an environmental good are large but diffuse and its provision is opposed by a powerful special interest group. In this case a countervailing interest group pushing for the good's provision is unlikely to spring up. The other is where there is a strong lobby in favor of providing an environmental good, with the public as a whole footing the bill and their aggregate willingness to pay for it being much smaller than its cost. The nature of the political process will often be to supply the good to the determinant of the public's welfare as long as there is not a strong group opposing it. In both cases, an estimate of the public's WTP for the good can help illuminate the nature of the decision at hand.

References

Abala, D.O. (1987). "A theoretical and empirical investigation of the willingness to pay for recreational services: a case study of Nairobi National Park". Eastern Africa Economic Review 3, 111–119.

Acton, J.P. (1973). "Evaluating public progress to save lives: the case of heart attacks". Rand Research Report R-73-02. RAND Corporation, Santa Monica, CA.

Adamowicz, W.L., Louviere, J.J., Williams, M. (1994). "Combining revealed and stated preference methods for valuing environmental amenities". Journal of Environmental Economics and Management 26, 271–292.

Adamowicz, W.L., Boxall, P.C., Louviere, J.J., Swait, J. (1999). "Stated preference methods for valuing environmental amenities". In: Bateman, I.J., Willis, K.G. (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the USA, EC, and Developing Countries. Oxford University Press, New York.

Akerlof, G. (1970). "The market for lemons: quality uncertainty and the market mechanism". Quarterly Journal of Economics 84, 488–500.

Alberini, A. (1995). "Optimal designs for discrete choice contingent valuation surveys: single-bound, double-bound, and bivariate models". Journal of Environmental Economics and Management 28, 287–306.

Alberini, A., Bjornstad, D., Kahn, J.R. (Eds.) (in press). Handbook of Contingent Valuation. Edward Elgar, Northampton, MA.

Alberini, A., Carson, R.T. (1990). "Choice of thresholds for efficient binary discrete choice estimation". Discussion Paper 90-34. Department of Economics, University of California, San Diego.

Alberini, A., Kanninen, B.J., Carson, R.T. (1997). "Modeling response incentives in dichotomous choice contingent valuation data". Land Economics 73, 309–324.

Arrow, K. (1958). "Utilities, attitudes, choices: a review note". Econometrica 26, 1–23.

Arrow, K., Fisher, A.C. (1974). "Environmental preservation, uncertainty, and irreversibility". Quarterly Journal of Economics 88, 313–319.

[174] It is possible to make a similar statement with respect to "experts" by asking the natural questions "whose experts" and whether there is any reason to believe that their preferences (given the same technical information) correspond to the public's?

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

Atkinson, A.C., Donev, A.N. (1992). Optimum Experimental Designs. Oxford University Press, New York.

Audience Research, Inc. (1958). "A study of outdoor recreational activities and preferences of the population living in the region of the Delaware river basin". Report prepared for the U.S. National Park Service.

Ayer, M., Brunk, H.D., Ewing, G.M., Silverman, E. (1955). "An empirical distribution function for sampling with incomplete information". Annals of Mathematical Statistics 26, 641–647.

Azevedo, C.D., Herriges, J.A., Kling, C.L. (2003). "Combining revealed and stated preferences: consistency tests and their interpretations". American Journal of Agricultural Economics 85, 525–537.

Barrett, L.B., Waddell, T.E. (1973). "Costs of air pollution damage: a status report". Report No. AP-85 to the U.S. Environmental Protection Agency. Center for Economics Research, Research Triangle Institute, Research Triangle Park, NC.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C., Sugden, R. (1997). "Does part-whole bias exist?: an experimental investigation". Economic Journal 107, 322–332.

Bateman, I.J., Langford, I.H., Jones, A.P., Kerr, G.N. (2001). "Bound and path effects in double and triple bounded dichotomous choice contingent valuation". Resource and Energy Economics 23, 191–214.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N.D., Hett, T., Jones-Lee, M.W., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, D.W., Sugden, R., Swanson, J. (2002). Economic Valuation with Stated Preference Techniques. Edward Elgar, Northampton, MA.

Bausell, R.B., Li, Y.-F. (2002). Power Analysis for Experimental Research: A Practical Guide for the Biological, Medical and Social Sciences. Cambridge University Press, New York.

Beardsley, P.L., Kovenock, D.M., Reynolds, W.C. (1974). Measuring Public Opinion on National Priorities: A Report on a Pilot Study. Sage Publications, Beverly Hills.

Beardsley, W.G. (1971). "Economic value of recreation benefits determined by three methods". U.S. Forest Service Research Notes, RM-176, Rocky Mountain Experiment Station, Colorado Springs.

Beggs, S.C., Hausman, J.A. (1981). "Assessing the potential for electric cars". Journal of Econometrics 16, 1–19.

Belsley, D.A., Kuh, E., Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Ben-Akiva, M., Morikawa, T. (1990). "Estimation of mode switching models from revealed preferences and stated intentions". Transportation Research A 24, 485–495.

Berger, M.C., Blomquist, G.C., Kenkel, D., Tolley, G.S. (1987). "Valuing changes in health risks: a comparison of alternative measures". Southern Economic Journal 53, 967–984.

Bergstrom, J.C., Dillman, B.L., Stoll, J.R. (1985). "Public environmental amenity benefits of private land: the case of prime agricultural land". Southern Journal of Agricultural Economics 17, 139–149.

Berry, D. (1974). "Open space values: a household survey of two Philadelphia parks". RSRI Discussion Paper Series No. 76. Regional Science Research Institute, Philadelphia.

Binkley, C.S., Hanemann, W.M. (1978). "The recreation benefits of water quality improvement: analysis of day trips in an urban setting". Report No. EPA-600/5-78-010. U.S. Environmental Protection Agency, Washington, DC.

Bishop, R.C., Heberlein, T.A. (1979). "Measuring values of extra market goods". American Journal of Agricultural Economics 61, 926–930.

Bishop, R.C., Heberlein, T.A. (1980). "Simulated markets, hypothetical markets, and travel cost analysis: alternative methods of estimating outdoor recreation demand". Staff Paper Series No. 187. Department of Agricultural Economics, University of Wisconsin, Madison, WI.

Bishop, R.C., Heberlein, T.A., Kealy, M.J. (1983). "Contingent valuation of environmental assets: comparisons with a simulated market". Natural Resources Journal 23, 619–633.

Blackburn, M., Harrison, G.W., Rutström, E.E. (1994). "Statistical bias functions and informative hypothetical surveys". American Journal of Agricultural Economics 76, 1084–1088.

Blank, F., Brookshire, D.S., Crocker, T.D., d'Arge, R.C., Horst, R., Jr., Rowe, R.D. (1978). "Valuation of aesthetic preferences: a case study of the economic value of visibility". Research Report to EPRI (Electric

Power Research Institute). Resource and Environmental Economics Laboratory, University of Wyoming, Laramie, WY.
Blinder, A.S. (1991). "Why are prices sticky: preliminary results from an interview study". American Economic Review 81, 89–100.
Bohm, P. (1972). "Estimating demand for public goods: an experiment". European Economic Review 3, 111–130.
Bohm, P. (1984). "Revealing demand for an actual public good". Journal of Public Economics 24, 135–151.
Boulier, B.L., Goldfarb, R.S. (1998). "On the use and nonuse of surveys in economics". Journal of Economic Methodology 5, 1–21.
Bowen, H.R. (1943). "The interpretation of voting in the allocation of economic resources". Quarterly Journal of Economics 58, 27–48.
Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. Wiley, New York.
Boyle, K.J., Welsh, M.P., Bishop, R.C. (1988). "Validation of empirical measures of welfare change: comment and extension". Land Economics 64, 94–98.
Bradburn, N., Sudman, S. (1988). Polls and Surveys: Understanding What They Are Telling Us. Jossey-Bass, San Francisco.
Bradburn, N.M., Sudman, S., Wansink, B. (2004). Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires. Jossey-Bass, San Francisco.
Brookshire, D.S., Ives, B.C., Schulze, W.D. (1976). "The valuation of aesthetic preferences". Journal of Environmental Economics and Management 3, 325–346.
Brookshire, D.S., Randall, A., Stoll, J.R. (1980). "Valuing increments and decrements in natural resource service flows". American Journal of Agricultural Economics 62, 478–488.
Brown, G.M., Jr., Charbonneau, J.J., Hay, M.J. (1978). "Estimating values of wildlife: analysis of the 1975 hunting and fishing survey". Working Paper No. 7. Division of Program Plans, U.S. Fish and Wildlife Service, Washington, DC.
Brown, G.M. Jr., Hammack, J. (1972). "A preliminary investigation of the economics of migratory waterfowl". In: Krutilla, J.V. (Ed.), Natural Environments: Studies in Theoretical and Applied Analysis. Johns Hopkins University Press, Baltimore, MD, pp. 171–204.
Brown, T.C., Peterson, G.L. (2003). "Multiple good valuation with focus on the method of paired comparison". In: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation. Kluwer Academic, Norwell, MA.
Brownstone, D., Bunch, D.S., Train, K. (2000). "Joint mixed logit models of stated and revealed references for alternative-fuel vehicles". Transportation Research B 34, 315–338.
Brownstone, D., Small, K. (2005). "Valuing time and reliability: assessing the evidence from road pricing demonstrations". Transportation Research A 39, 279–293.
Burton, A.C., Carson, K.S., Chilton, S.M., Hutchinson, W.G. (2003a). "An experimental investigation of explanations for inconsistencies in responses to second offers in double referenda". Journal of Environmental Economics and Management 46, 472–489.
Burton, A.C., Carson, K.S., Chilton, S.M., Hutchinson, W.G. (2003b). "An analysis of response strategies in double-bounded dichotomous choice contingent valuation". Paper presented at the American Agricultural Economics Association Meeting, Montreal.
Camerer, C.F., Loewenstein, G., Rabin, M. (Eds.) (2003). Advances in Behavioral Economics. Princeton University Press, Princeton, NJ.
Cameron, T.A. (1988). "A new paradigm for valuing nonmarket goods using referendum data: maximum likelihood estimation by censored logistic regression". Journal of Environmental Economics and Management 15, 355–379.
Cameron, T.A. (1992). "Combining contingent valuation and travel cost data for the valuation of nonmarket goods". Land Economics 68, 302–317.
Cameron, T.A., Englin, J. (1997). "Welfare effects of changes in environmental quality under individual uncertainty about use". RAND Journal of Economics 28 (Supplement), 45–70.

Cameron, T.A., Huppert, D.D. (1989). "OLS versus ML estimation of nonmarket resource values with payment card interval data". Journal of Environmental Economics and Management 17, 230–246.
Cameron, T.A., Huppert, D.D. (1991). "Referendum contingent valuation estimates: sensitivity to the assignment of offered values". Journal of the American Statistical Association 86 (416), 910–918.
Cameron, T.A., James, M.D. (1987). "Efficient estimation methods for 'closed-ended' contingent valuation surveys". Review of Economics and Statistics 69, 269–276.
Cameron, T.A., Quiggin, J. (1994). "Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire". Journal of Environmental Economics and Management 27, 218–234.
Carson, R.T. (1985). Three essays on contingent valuation. Dissertation, University of California, Berkeley.
Carson, R.T. (1991). "Constructed markets". In: Braden, J.B., Kolstad, C.D. (Eds.), Measuring the Demand for Environmental Commodities. North-Holland, Amsterdam, pp. 121–162.
Carson, R.T. (1997a). "Contingent valuation: theoretical advances since the NOAA panel". American Journal of Agricultural Economics 79, 1501–1507.
Carson, R.T. (1997b). "Contingent valuation and tests of insensitivity to scope". In: Kopp, R.J., Pommerhene, W., Schwartz, N. (Eds.), Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods. Kluwer Academic, Boston.
Carson, R.T. (2000). "Contingent valuation: a user's guide". Environmental Science and Technology 34, 1413–1418.
Carson, R.T. (in press). Contingent Valuation: A Comprehensive Bibliography and History, Edward Elgar, Northampton, MA.
Carson, R.T., Flores, N.E., Hanemann, W.M. (1998). "Sequencing and valuing public goods". Journal of Environmental Economics and Management 36, 314–323.
Carson, R.T., Flores, N.E., Meade, N.F. (2001). "Contingent valuation: controversies and evidence". Environmental and Resource Economics 19, 173–210.
Carson, R.T., Flores, N.E., Mitchell, R.C. (1999). "The theory and measurement of passive use value". In: Bateman, I.J., Willis, K.G. (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries. Oxford University Press, Oxford.
Carson, R.T., Groves, T., Machina, M.J. (2000). "Incentive and informational properties preference questions". Paper presented at the Kobe Conference on Theory and Application of Environmental Valuation, Kobe University, Kobe.
Carson, R.T., Hanemann, W.M., 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., Hanemann, W.M., Steinberg, D. (1990). "A discrete choice contingent valuation estimate of the value of kenai king salmon". Journal of Behavioral Economics 19, 53–68.
Carson, R.T., Jeon, Y. (2000). "On overcoming information deficiencies in estimating willingness to pay distributions". Paper presented at the European Association of Environmental and Resource Economists Annual Conference, University of Crete, Rethymnon.
Carson, R.T., Meade, N.F., Smith, V.K. (1993). "Passive use values and contingent valuation: introducing the issues". Choices 8 (2), 4–8.
Carson, R.T. Mitchell, R.C. (1991). "The value of diamonds and water". Paper presented at the European Association of Environmental and Resource Economists Annual Conference, Stockholm.
Carson, R.T., 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, 2445–2454.
Carson, R.T., 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., Ruud, P.A. (1990). "Valuing air quality improvements: simulating a hedonic equation in the context of a contingent valuation scenario". In: Mathai, C.V. (Ed.), Visibility and Fine Particles. Air & Waste Management Association, Pittsburgh, pp. 639–646.
Carson, R.T., Steinberg, D. (1990). "Experimental design for discrete choice voter preference surveys". In: 1989 Proceeding of the Survey Methodology Section of the American Statistical Association. American Statistical Association, Washington, DC.

Carson, R.T., Wilks, L., Imber, D. (1994). "Valuing the preservation of Australia's kakadu conservation zone". Oxford Economic Papers 46, 727–749.

Carson, R.T., Costanzo, M., Hanemann, W.M., Wegge, T. (1987). "South Central Alaska area sport fishing economic study". Final Report to the Alaska Department of Fish and Game by Jones and Stokes Associates.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S., Ruud, P.A. (1992). A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill. Anchorage, AK.

Carson, R.T., Hanemann, W.M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K. (1994). Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight, vols. 1 & 2. Report to the National Oceanic and Atmospheric Administration.

Carson, R.T., Flores, N.E., Martin, K.M., Wright, J.L. (1996). "Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods". Land Economics 72, 80–99.

Carson, R.T., Hanemann, W.M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K. (1997). "Temporal reliability of estimates from contingent valuation". Land Economics 73, 151–163.

Carson, R.T., Hanemann, W.M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K. (1998). "Referendum design and contingent valuation: the NOAA panel's no-vote recommendation". Review of Economics and Statistics 80, 484–487.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S., Ruud, P.A. (2003). "Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill". Environmental and Resource Economics 25, 257–286.

Carson, R.T., Conaway, M.B., Hanemann, W.M., Krosnick, J.A., Mitchell, R.C., Presser, S. (2004a). Valuing Oil Spill Prevention: A Case Study of California's Central Coast. Kluwer Academic, Boston.

Carson, R.T., Groves, T., List, J., Machina, M. (2004b). "Probabilistic influence and supplemental benefits: a field test of the two key assumptions underlying stated preferences". Paper present at NBER Public Economics Workshop, Palo Alto.

Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.) (2003). A Primer on Non-Market Valuation. Kluwer Academic Press, Boston.

Champ, P.A., Brown, T.C. (1997). "A comparison of contingent and actual voting behavior". Proceedings from W-133 Benefits and Cost Transfer in Natural Resource Planning, 10th Interim Report.

Champ, P.A., Bishop, R.C., Brown, T.C., McCollum, D.W. (1997). "Using donation mechanisms to value nonuse benefits from public goods". Journal of Environmental Economics and Management 33, 151–162.

Chapman, D.J., Hanemann, W.M., Kanninen, B.J. (1993). "Nonmarket valuation using contingent behavior: model specification and consistency tests". Working Paper, Department of Agricultural and Resource Economics, UC Berkeley.

Chapman, R.G., Staelin, R. (1982). "Exploiting rank ordered choice set data within the stochastic utility model". Journal of Marketing Research 19, 281–299.

Chen, H.Z., Randall, A. (1997). "Semi-nonparametric estimation of binary response models with an application to natural resource valuation". Journal of Econometrics 76, 323–340.

Chilton, S.M., Hutchinson, W.G. (1999). "Some further implications of incorporating the warm glow of giving into welfare measures: a comment on the use of donation mechanisms by Champ et al.". Journal of Environmental Economics and Management 37, 202–209.

Cicchetti, C.J., Smith, V.K. (1973). "Congestion, quality deterioration, and optimal use: wilderness recreation in the Spanish peaks primitive area". Social Science Research 2, 15–30.

Cicchetti, C.J., Smith, V.K. (1976a). The Cost of Congestion. Ballinger, Cambridge, MA.

Cicchetti, C.J., Smith, V.K. (1976b). "The measurement of individual congestion costs: an economic application to a wilderness area". In: Lin, S.A. (Ed.), Theory and Measurement of Externalities. Academic Press, New York.

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.

Clawson, M., Knetsch, J. (1966). Economics of Outdoor Recreation. Johns Hopkins University Press, Baltimore.

Clark, J.M. (1915). "The concept of value: a rejoinder". Quarterly Journal of Economics 29, 709–723.

Cocheba, D.J., Langford, W.A. (1978). "Wildlife valuation: the collective good aspect of hunting". Land Economics 54, 490–504.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed. Wiley, New York.

Conrad, J.M., LeBlanc, D. (1979). "The supply of development rights: results from a survey in Hadley, Massachusetts". Land Economics 55, 269–276.

Converse, P.E. (1974). "Comment: the status of nonattitudes". American Political Science Review 68, 650–660.

Cooper, J.C. (1993). "Optimal bid selection for dichotomous choice contingent valuation surveys". Journal of Environmental Economics and Management 24, 25–40.

Cooper, J.C. (2002). "Flexible functional form estimation of willingness to pay using dichotomous choice data". Journal of Environmental Economics and Management 43, 267–279.

Cooper, J.C., Hanemann, W.M., Signorello, G. (2002). "One-and-one-half-bound dichotomous choice contingent valuation". Review of Economics and Statistics 84, 742–750.

Copas, J.B. (1983). "Plotting p against x". Applied Statistics 32, 25–31.

Cornes, R., Sandler, T. (1996). The Theory of Externalities, Public Goods, and Club Goods. Cambridge University Press, New York.

Corso, P.S., Hammitt, J.K., 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.

Creel, M.D., Loomis, J.B. (1997). "Semi-nonparametric distribution-free dichotomous choice contingent valuation". Journal of Environmental Economics and Management 32, 341–358.

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

Cummings, R.G., Harrison, G.W. (1992). "Identifying and measuring nonuse values for natural and environmental resources: a critical review of the state of the art". Report to American Petroleum Institute.

Cummings, R.G., Harrison, G.W. (1995). "The measurement and decomposition of nonuse values: a critical review". Environmental and Resource Economics 5, 225–247.

Cummings, R.G., Harrison, G.W., Rutström, E.E. (1995). "Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?". American Economic Review 85, 260–266.

Cummings, R.G., Taylor, L.O. (1998). "Does realism matter in contingent valuation surveys". Land Economics 74, 203–215.

Cummings, R., Taylor, L.O. (1999). "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method". American Economic Review 89, 649–665.

Cummings, R.G., Elliot, S., Harrison, G.W., Murphy, J. (1997). "Are hypothetical referenda incentive compatible?". Journal of Political Economy 105, 609–621.

Currie, J.M., Murphy, J.A., Schmitz, A. (1971). "The concept of economic surplus". Economic Journal 81, 741–799.

Darling, A.H. (1973). "Measuring benefits generated by urban water parks". Land Economics 49, 22–34.

Davidson, J.D. (1973). "Forecasting traffic on STOL". Operations Research Quarterly 24, 561–569.

Davis, R.K. (1963a). The value of outdoor recreation: an economic study of the maine woods. Dissertation, Harvard University.

Davis, R.K. (1963b). "Recreation planning as an economic problem". Natural Resources Journal 3, 239–249.

Deaton, A.S. (1974). "The analysis of consumer demand in the United Kingdom". Econometrica 42, 341–367.

Debreu, G. (1959). Theory of Value. Wiley, New York.

DeShazo, J.R., Fermo, G. (2002). "Designing choice sets for stated preference methods: the effects of complexity on choice consistency". Journal of Environmental Economics and Management 43, 360–385.

Devine, D.G., Marion, B.W. (1979). "The influence of consumer price information on retail pricing and consumer behavior". American Journal of Agricultural Economics 61, 228–237.

Desvousges, W.H., Smith, V.K., Fisher, A. (1987). "Option price estimates for water quality improvements: a contingent valuation study for the monongahela river". Journal of Environmental Economics and Management 14, 248–267.

Desvousges, W.H., Gable, A.R., Dunford, R.W., Hudson, S.P. (1993a). "Contingent valuation: the wrong tool for damage assessment". Choices 8 (2), 9–11.

Desvousges, W.H., Johnson, F.R., Dunford, R.W., Boyle, K.J., Hudson, S.P., Wilson, K.N. (1993b). "Measuring natural resource damages with contingent valuation: tests of validity and reliability". In: Hausman, J.A. (Ed.), Contingent Valuation: A Critical Assessment. North-Holland, Amsterdam, pp. 91–164.

Diamond, P.A. (1996). "Testing the internal consistency of contingent valuation surveys". Journal of Environmental Economics and Management 30, 337–347.

Diamond, P.A., Hausman, J.A. (1994). "Contingent valuation: is some number better than no number?". Journal of Economic Perspectives 8 (4), 45–64.

Dillman, D.A. (in press). Mail and Internet Surveys, 3rd ed. Wiley, New York.

Dubourg, W.R., Jones-Lee, M.W., Loomes, G. (1997). "Imprecise preferences and survey design in contingent valuation". Economica 64, 681–702.

Ducci, J. (1988). "Metodolodia de cuantificacion de beneficios: Uruguay, sanemiento ambiental de Montevideo". Unpublished Background Paper. Technical Annex, Inter-American Development Bank, Washington, DC.

Duffield, J.W., Patterson, D.A. (1991a). "Inference and optimal design for a welfare measure in dichotomous choice contingent valuation". Land Economics 67, 225–239.

Duffield, J.W., Patterson, D.A. (1991b). "Field testing existence values: an instream flow trust fund for Montana rivers". Paper presented at the American Economic Association Annual Conference.

Eastaugh, S.R. (1991). "Valuation of the benefits of risk-free blood: willingness to pay for hemoglobin solutions". International Journal of Technology Assessment in Health Care 7, 51–57.

Eastman, C., Randall, A., Hoffer, P.L. (1974). "How much to abate pollution?". Public Opinion Quarterly 38, 574–584.

Eckstein, O. (1958). Water Resource Development. Harvard University Press, Cambridge.

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

Fiedler, J.A. (1972). "Condominium design and pricing: a case study in consumer trade-off analysis". In: Venkatesan, M. (Ed.), Proceedings of the Third Annual Conference of the Association for Consumer Research, pp. 279–293.

Fischhoff, B., Furby, L. (1988). "Measuring values: a conceptual framework for interpreting transactions with special reference to contingent valuation of visibility". Journal of Risk and Uncertainty 1, 147–184.

Fischhoff, B., Slovic, P., Lichtenstein, S. (1980). "Knowing what you want: measuring labile values". In: Wallsten, T.S. (Ed.), Cognitive Processes in Choice and Decision Behavior. Lawrence Erlbaum Associates, Hillsdale, NJ.

Flores, N.E. (1995). "The effects of rationing and virtual price elasticities". Discussion Paper 95-20. Department of Economics, University of California, San Diego.

Flores, N.E. (1999). "The importance of agenda and willingness to pay". Working Paper No. 99-30. Department of Economics, University of Colorado.

Flores, N.E., Carson, R.T. (1997). "The relationship between the income elasticities of demand and willingness to pay". Journal of Environmental Economics and Management 33, 287–295.

Flores, N.E., Thatcher, J. (2002). "Money, who needs it? Natural resource Damage Assessment". Contemporary Economic Policy 20, 171–178.

Fowler, F.J., Mangione, T.W. (1989). Standardized Survey Interviewing: Minimizing Interviewer-Related Error. Sage, Beverly Hills, CA.

Freeman III, A.M. (1979). The Benefits of Environmental Improvement: In Theory and Practice. Johns Hopkins University Press, Baltimore MD.

Freeman, A.M. (1993). The Measurement of Environmental and Resource Values: Theory and Methods. Resources for the Future, Washington.

Ganiats, T.G., Carson, R.T., Hamm, R.M., Cantor, S.B. (2000). "Population-based time preferences for future health outcomes". Medical Decision Making 20, 263–270.

Garbacz, C., Thayer, M.A. (1983). "An experiment in valuing senior companion program services". Journal of Human Resources 18, 147–152.
Georgiou, S., Whittington, D., Pearce, D.W., Moran, D. (1997). Economic Values and the Environment in the Developing World. Edward Elgar, Northampton, MA.
Gerking, S., de Haan, M., Schulze, W.D. (1988). "The marginal value of job safety: a contingent valuation study". Journal of Risk and Uncertainty 1, 185–199.
Gibbard, A. (1973). "Manipulation of voting schemes: a general result". Econometrica 41, 587–601.
Gluck, R.J. (1974). An economic evaluation of the rakai fishery as a recreational resource. Master's Thesis, University of Canterbury.
Graham, D.A. (1981). "Cost benefit analysis under uncertainty". American Economic Review 71, 715–725.
Graham, D.A. (1984). "Cost benefit analysis under uncertainty: reply". American Economic Review 74, 1100–1102.
Gramlich, F.W. (1977). "The demand for clean water: the case of the Charles river". National Tax Journal 30, 183–194.
Grandstaff, S., Dixon, J.A. (1986). "Evaluation of Lumpinee park in Bangkok, Thailand". In: Dixon, J.A., Hufschmidt, M.M. (Eds.), Economic Valuation Techniques for the Environment: A Case Study Workbook. Johns Hopkins University Press, Baltimore.
Green, J.R., Laffont, J.J. (1978). "A sampling approach to the free rider problem". In: Sandmo, A. (Ed.), Essays in Public Economics. Lexington Books, Lexington, MA.
Greene, W.H. (2002). Econometric Analysis, 5th ed. Prentice-Hall, Englewood Cliffs, NJ.
Greenley, D.A., Walsh, R.G., Young, R.A. (1981). "Option value: empirical evidence from a case study of recreation and water quality". Quarterly Journal of Economics 96, 657–672.
Groothuis, P.A., Whitehead, J.C. (2002). "Does don't know mean no? Analysis of 'don't know' responses in dichotomous choice contingent valuation questions". Applied Economics 34, 1935–1940.
Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (2001). Survey Nonresponse. Wiley, New York.
Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (2004). Survey Methodology. Wiley, New York.
Groves, T. (1973). "Incentives in teams". Econometrica 41, 617–631.
Haab, T.C. (1999). "Nonparticipation or misspecification? The impacts of nonparticipation on dichotomous choice contingent valuation". Environmental and Resource Economics 14, 443–461.
Haab, T.C., McConnell, K.E. (1997). "Referendum models and negative willingness to pay: alternative solutions". Journal of Environmental Economics and Management 32, 251–270.
Haab, T.C., McConnell, K.E. (1998). "Referendum models and economic values: theoretical, intuitive, and practical bounds on willingness to pay". Land Economics 74, 216–229.
Haab, T.C., McConnell, K.E. (2002). Valuing Environmental and Natural Resources. Edward Elgar, Northampton.
Hammack, J., Brown, G.M. Jr. (1974). Waterfowl and Wetlands: Toward Bioeconomic Analysis. Johns Hopkins University Press, Baltimore, MD.
Hanemann, W.M. (1978). A methodological and empirical study of the recreation benefits from water quality improvement. Dissertation, Harvard University.
Hanemann, W.M. (1984a). "Discrete/continuous models of consumer demand". Econometrica 52, 541–561.
Hanemann, W.M. (1984b). "Welfare evaluations in contingent valuation experiments with discrete responses". American Journal of Agricultural Economics 66, 332–341.
Hanemann, W.M. (1985). "Some issues in continuous- and discrete-response contingent valuation studies". Northeastern Journal of Agricultural Economics 14, 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. (1992). "Preface". In: Navrud, S. (Ed.), Pricing the European Environment. Oxford University Press, New York, pp. 9–35.
Hanemann, W.M. (1994). "Valuing the environment through contingent valuation". Journal of Economic Perspectives 8 (4), 19–43.

Hanemann, W.M. (1995). "Contingent valuation and economics". In: Willis, K.G., Corkindale, J.T. (Eds.), Environmental Valuation: New Perspectives. CAB International, Wallingford, UK, pp. 79–117.
Hanemann, W.M. (1999). "Neo-classical economic theory and contingent valuation". In: Bateman, I.J., Willis, K.G. (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries. Oxford University Press, New York, pp. 42–96.
Hanemann, W.M., Felipe Vasquez, F. (2005). "Strictly for the birds: A scope test revisited". Working Paper, UC Berkeley Department of Agricultural & Resource Economics, October.
Hanemann, W.M., Kanninen, B.J. (1999). "The statistical analysis of discrete-response CV data". In: Bateman, I.J., Willis, K.G. (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries. Oxford University Press, Oxford, UK, pp. 302–441.
Hanemann, W.M., Kriström, B. (1995). "Preference uncertainty, optimal designs and spikes". In: Johansson, P.-O., Kriström, B., Mäler, K.-G. (Eds.), Current Issues in Environmental Economics. Manchester University Press, New York, pp. 58–77.
Hanemann, W.M., Loomis, J.B., Kanninen, B.J. (1991). "Statistical efficiency of double-bounded dichotomous choice contingent valuation". American Journal of Agricultural Economics 73, 1255–1263.
Hanley, N.D., Mourato, S., Wright, R.E. (2001). "Choice modelling approaches: a superior alternative for environmental valuation?". Journal of Economic Surveys 15, 435–462.
Hanley, N.D., Wright, R.E., Adamowicz, W.L. (1998). "Using choice experiments to value the environment-design issues, current experience and future prospects". Environmental and Resource Economics 11, 413–428.
Hardie, I., Strand, I.E. (1979). "Measurement of economic benefits for potential public goods". American Journal of Agricultural Economics 61, 311–317.
Harrison, G., List, J.A. (2004). "Field experiments". Journal of Economic Literature 42, 1009–1055.
Hausman, J.A. (Ed.) (1993). Contingent Valuation: A Critical Assessment. North-Holland, Amsterdam.
Hausman, J.A., Ruud, P.A. (1987). "Specifying and testing econometric models for rank-ordered data". Journal of Econometrics 34, 83–104.
Heckman, J. (1979). "Sample selection bias as a specification error". Econometrica 47, 153–161.
Hensher, D.A. (1994). "Stated preference analysis of travel choices: the state of practice". Transportation 21, 107–133.
Hensher, D.A., Louviere, J.J., Swait, J. (1999). "Combining sources of preference data". Journal of Econometrics 89, 197–221.
Hensher, D.A., Rose, J.M., Greene, W.H. (2005). Applied Choice Analysis: A Primer. Cambridge University Press, New York.
Hensher, D.A., Truong, P. (1985). "Valuation of travel time savings: a direct experimental approach". Journal of Transport Economics and Policy 19, 237–261.
Herriges, J.A., Kling, C.L. (Eds.) (1999). Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice. Edward Elgar, Northampton, MA.
Herriges, J.A., Shogren, J.F. (1996). "Starting point bias in dichotomous choice valuation with follow-up questioning". Journal of Environmental Economics and Management 30, 112–131.
Hicks, J.R. (1943). "The four consumer surpluses". Review of Economic Studies 11, 31–41.
Hirschleifer, J. (1965). "Investment decision under uncertainty: choice theoretic approaches". Quarterly Journal of Economics 79, 509–536.
Hirschleifer, J. (1966). "Investment decision under uncertainty: applications of the stated preference approach". Quarterly Journal of Economics 80, 252–277.
Hines, L.G. (1951). "Wilderness areas: an extra-market problem in resource allocation". Land Economics 27, 306–313.
Hoehn, J.P., Randall, A. (1987). "A satisfactory benefit cost indicator from contingent valuation". Journal of Environmental Economics and Management 14, 226–247.
Hoehn, J.P., Randall, A. (1989). "Too many proposals past the benefit cost test". American Economics Review 79, 544–551.

Hoinville, G., Berthoud, R. (1970). Identifying and Evaluating Trade-Off Preferences: An Analysis of Environmental Accessibility Priorities. Social and Planning Research, London. Publication P-117.

Holmes, T.P., Adamowicz, W.L. (2003). "Attribute base method". In: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation. Kluwer Academic, Norwell, MA, pp. 171–219.

Horowitz, J.K., McConnell, K.E. (2002). "A review of WTA/WTP studies". Journal of Environmental Economics and Management 44, 426–447.

Infosino, W.J. (1986). "Forecasting new product sales for likelihood of purchase ratings". Marketing Science 5, 372–390.

Jansen, H.M.A., Opschoor, J.B. (1973). "Waardering van de invloed van het vliegtuiglawaai op woongebied rond de potentiele locaties van de Tweede nationale luchthaven (Valuation of the effects of aircraft noise on residential areas around the potential locations of the second Dutch national airport)". Working Paper Series A, Nos. 4 and 5. Institute for Environmental Studies, Vrije University, Amsterdam.

Johannesson, M., Fagerberg, B. (1992). "A health economic comparison of diet and drug treatment in obese men with mild hypertension". Journal of Hypertension 10, 1063–1070.

Johansson, P.-O. (1995). Evaluating Health Risks: An Economic Approach. Cambridge University Press, New York.

Johnson, F.R., Desvousges, W.H. (1997). "Estimating stated preferences with rated-pair data: environmental, health, and employment effects of energy programs". Journal of Environmental Economics and Management 34, 79–99.

Johnson, R.M. (1974). "Trade-off analysis of consumer values". Journal of Marketing Research 11, 121–127.

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.

Jones-Lee, M.W., Hammerton, M., Philips, P.R. (1985). "The value of safety: results of a national sample survey". Economic Journal 95, 49–72.

Just, R.E., Hueth, D.L., Schmitz, A. (1982). Applied Welfare Economics and Public Policy. Prentice-Hall, Englewood Cliffs, NJ.

Kahneman, D., Knetsch, J.L. (1992). "Valuing public goods: the purchase of moral satisfaction". Journal of Environmental Economics and Management 22, 57–70.

Kahneman, D., Tversky, A. (1979). "Prospect theory: an analysis of decision under risk". Econometrica 47, 263–291.

Kanninen, B.J. (1993a). "Optimal experimental design for double-bounded dichotomous choice contingent valuation". Land Economics 69, 138–146.

Kanninen, B.J. (1993b). "Design of sequential experiments for contingent valuation studies". Journal of Environmental Economics and Management 25, S1–S11.

Keeney, R., Raiffa, H. (1976). Decisions with Multiple Objectives. Wiley, New York.

Kish, L.F. (1967). Survey Sampling. Wiley, New York.

Kling, C.L., List, J.A., Zhao, J. (2003). "WTP/WTA disparity: have we been observing dynamic values but interpreting them as static?". CARD Working Paper #03-WP 333, Iowa State University.

Knetsch, J.L. (1990). "Environmental policy implications of disparities between willingness to pay and compensation demanded measures of values". Journal of Environmental Economics and Management 18, 227–238.

Knetsch, J.L., Davis, R.K. (1966). "Comparisons of methods for resource evaluation". In: Kneese, A.V., Smith, S.C. (Eds.), Water Research. Johns Hopkins University Press, Baltimore, MD, pp. 125–142.

Knetsch, J.L., Thaler, R.H., Kahneman, D. (1990). "Experimental tests of the endowment effect and the coase theorem". Journal of Political Economy 98, 1325–1349.

Kolstad, C.D., 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.

Kopp, R.J., Smith, V.K. (1997). "Constructing measures of economic value". In: Kopp, R.J., Pommerehne, W.W., Schwarz, N. (Eds.), Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods. Kluwer Academic, Boston.

Kriström, B. (1990). "A nonparametric approach to the estimation of welfare measures in discrete response valuation studies". Land Economics 66, 135–139.
Kriström, B. (1997). "Spike models in contingent valuation". American Journal of Agricultural Economics 79, 1013–1023.
Krosnick, J.A., Holbrook, A., Berent, M.K., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Ruud, P.A., Smith, V.K., Moody, W.K., Green, M.C., Conaway, M.B. (2002). "The impact of 'no opinion' response options on data quality: non-attitude reduction or an invitation to satisfy?". Public Opinion Quarterly 66, 371–403.
Krutilla, J.V. (1967). "Conservation reconsidered". American Economic Review 57, 777–786.
Krutilla, J.V., Eckstein, O. (1958). Multiple Purpose River Development. Johns Hopkins University Press, Baltimore, MD.
Lansing, J.B., Morgan, J.N. (1971). Economic Survey Methods. Survey Research Center of the Institute for Social Research, University of Michigan, Ann Arbor.
LaPage, W.F. (1968). "The role of fees in camper's decisions". USDA Forest Service Research Paper No. 188. Northeastern Forest Experiment Station, Upper Darby, PA.
Lavrakas, P.J. (1993). Telephone Survey Methods: Sampling, Selection, and Supervision. Sage, Beverly Hills, CA.
Lave, L.B. (1984). "Controlling contradictions among regulations". American Economic Review 74, 471–476.
Layton, D.F. (2000). "Random coefficient models for stated preference surveys". Journal of Environmental Economics and Management 40, 21–36.
Lee, L.F., Pitt, M. (1986). "Microeconomic demand systems with binding nonnegativity constraints: the dual approach". Econometrica 54, 1237–1242.
León, C.J., Vazquez-Polo, F.J. (1998). "A Bayesian approach to double bounded contingent valuation". Environmental and Resource Economics 11, 197–215.
Lester, R.A. (1946). "Shortcomings of marginal analysis for wage-unemployment problems". American Economic Review 36, 63–82.
Levy, P., Lemeshow, S. (1999). Sampling of Populations: Methods and Applications. Wiley, New York.
Li, C.-Z., Mattsson, L. (1995). "Discrete choice under preference uncertainty: an improved structural model for contingent valuation". Journal of Environmental Economics and Management 28, 256–269.
Lichtenstein, S., Slovic, P. (1971). "Reversals of preferences between bids and choices in gambling decisions". Journal of Experimental Psychology 89, 46–55.
List, J. (2000). "The incredible disappearing act: the effect of market experience on the WTP/WTA disparity". Paper presented at the European Association of Environmental and Resource Economists Meeting, Rythymnon, Greece.
List, J., Gallet, C. (2001). "What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis". Environmental and Resource Economics 20, 241–254.
Little, I.M.D. (1957). A Critique of Welfare Economics. Oxford University Press, Oxford, UK.
Little, J., Berrens, R. (2004). "Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis". Economic Bulletin 3 (6), 1–13.
Little, R.J.A., Rubin, D.B. (2002). Statistical Analysis with Missing Data, 2nd ed. Wiley, New York.
Loehman, E.T., De, V.H. (1982). "Application of stochastic choice modeling to policy analysis of public goods: a case study of air quality improvements". Review of Economics and Statistics 64, 474–480.
Louviere, J.J. (1974). "Predicting the evaluation of real stimulus objects from an abstract evaluation of their attributes: the case of trout streams". Journal of Applied Psychology 59, 572–577.
Louviere, J.J. (1994). "Conjoint analysis". In: Bagozzi, R. (Ed.), Advanced Methods in Marketing Research. Blackwell, Oxford, UK.
Louviere, J.J., 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.D. (2000). Stated Choice Methods: Analysis and Application. Cambridge University Press, New York.

Louviere, J.J., Woodworth, G. (1983). "Design and analysis of stimulated consumer choice experiments or allocation experiments: an approach based on aggregate data". Journal of Marketing Research 20, 350–367.

Louviere, J., Ainslie, J.A., Cameron, T., Carson, R.T., DeShazo, J.R., Hensher, D., Kohn, R., Marley, T., Street, D. (2002). "Dissecting the random component". Marketing Letters 13, 177–193.

Machlup, F. (1946). "Marginal analysis and empirical research". American Economic Review 36, 519–554.

Mackenzie, J. (1993). "A comparison of contingent preference models". American Journal of Agricultural Economics 75, 593–603.

Madden, P. (1991). "A generalization of Hicksian substitutes and complements with application to demand rationing". Econometrica 59, 1497–1508.

Madow, W.G., Olkin, I., Rubin, D.B. (Eds.) (1983). Incomplete Data in Sample Surveys. Academic Press, San Diego.

Magat, W.A., Viscusi, W.K., Huber, J. (1988). "Paired comparison and contingent valuation approaches to morbidity risk valuation". Journal of Environmental Economics and Management 15, 395–411.

Mäler, K.-G. (1974). Environmental Economics: A Theoretical Inquiry. Johns Hopkins University Press, Baltimore, MD.

Manski, C. (1977). "The structure of random utility models". Theory and Decision 8, 229–254.

Manski, C.F., Straub, J.D. (2000). "Worker perceptions of job insecurity in the mid-1990s: evidence from the survey of economic expectations". Journal of Human Resources 35, 447–479.

Marley, A.A.J., Louviere, J.J. (in press). "Some probabilistic models of best, worse, and best–worst choices". Journal of Mathematical Psychology.

Mathews, S.B., Brown, G.M. (1970). "Economic evaluation of the 1967 sport salmon fisheries of Washington". Technical Report No. 2. Department of Fisheries, Olympia, Washington, DC.

Melinek, S.J. (1974). "A method of evaluating human life for economic purposes". Accident Analysis and Prevention 6, 103–114.

McConnell, K.E. (1977). "Congestion and willingness to pay: a study of beach use". Land Economics 53, 186–195.

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., Ducci, J.H. (1989). "Valuing environmental quality in developing countries: two case studies". Paper presented at the Association of Environmental and Resource Economists Annual Conference.

McFadden, D.L. (1974). "Conditional logit analysis of qualitative choice behavior". In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

McFadden, D.L. (1986). "The choice theory approach to market research". Marketing Science 5, 275–297.

McFadden, D.L. (1994). "Contingent valuation and social choice". American Journal of Agricultural Economics 76, 689–708.

McFadden, D.L. (1999a). "Rationality for Economists?". Journal of Risk and Uncertainty 19, 73–105.

McFadden, D.L. (1999b). "Computing willingness-to-pay in random utility models". In: Moore, J., Riezman, R., Melvin, J. (Eds.), Trade, Theory, and Econometrics: Essays in Honor of John S. Chipman. Routledge, London.

McFadden, D.L., Leonard, G.K. (1993). "Issues in the contingent valuation of environmental goods: methodologies for data collection and analysis". In: Hausman, J.A. (Ed.), Contingent Valuation: A Critical Assessment. North-Holland, Amsterdam, pp. 165–216.

Meyer, P.A. (1974). "Recreation and preservation values associated with salmon of the Fraser river". Information Report Series No. PAC/IN-74-1. Environmental Canada Fisheries and Marine Service, Southern Operations Branch, Vancouver.

Meyer, R.J., Levin, I.P., Louviere, J.J. (1978). "Functional analysis of mode choice". In: Transportation Forecasting and Travel Behavior. Transportation Research Records Report #673. Transportation Research Board, Washington, DC, 1–7.

Mierheim, H. (1974). Nutzen–Kosten Analysen Öffentlicher Grunanlagen im Innerstädtischen Bereich. Eine Untersuchung über die Anwendbarkeit am Beispiel Berlin-West (Benefit–cost analysis of public parks in

inner city areas: an evaluation of its applicability concerning the example of Berlin-West). Dissertation, Technical University of Berlin.
Miles, G.A. (1967). "Water based recreation in Nevada". Report B-14. College of Agriculture, University of Nevada, Reno.
Mishan, E.J. (1971). "Evaluation of life and limb: a theoretical approach". Journal of Political Economy 90, 827–853.
Mishan, E.J. (1976). Cost–Benefit Analysis, 2nd ed. Praeger, New York.
Mitchell, R.C. (1982). "On the use of the contingent valuation approach to value public services in developing nations". Report prepared for the Country Policy Department, World Bank, Washington, DC.
Mitchell, R.C. (2002). "On designing constructed markets in valuation surveys". Environmental and Resource Economics 22, 297–321.
Mitchell, R.C., 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., 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., Carson, R.T. (1989). Using Surveys to Value Public Goods: The Contingent Valuation Method. Johns Hopkins University Press, Baltimore, MD.
Moeltner, K., Shonkwiler, J.S. (2005). "Correcting for on-site sampling in random utility models". American Journal of Agricultural Economics 87, 327–339.
Morey, E.R., Thatcher, J., Breffle, W.S. (in press). "Using angler characteristics and attitudinal data to identify environmental preferences classes: a latent class model". Environmental and Resource Economics.
Moser, D.A., Dunning, C.M. (1986.) A Guide for Using the Contingent Valuation Method in Recreation Studies, National Economic Development Procedures Manual – Recreation, vol. 2. IWR Report 86-R-T. Institute for Water Resources, U.S. Army Corp of Engineers, Fort Belvoir, VA.
Mueller, D.C. (1989). Public Choice II. Cambridge University Press, Cambridge.
Mulligan, P.J. (1978). "Willingness to pay for decreased risk from nuclear plant accidents". Working Paper No. 43. Center for the Study of Environmental Policy, Pennsylvania State University.
Navrud, S. (1992). "Willingness to pay for preservation of species: an experiment with actual payments". In: Navrud, S. (Ed.), Pricing the European Environment. Oxford University Press, New York, pp. 231–246.
Neill, H.R., Cummings, R.G., Ganderton, P.T., Harrison, G.W., McGuckin, T. (1994). "Hypothetical surveys and real economic commitments". Land Economics 70, 145–154.
O'Hanlon, P.W., Sinden, J.A. (1978). "Scope for valuation of environmental goods: comment". Land Economics 54, 381–387.
Oster, S. (1977). "Survey results on the benefits of water pollution abatement in the Merrimack river basin". Water Resources Research 13, 882–884.
Parzen, E. (1960). Modern Probability Theory and its Applications. Wiley, New York.
Palmatier, M. (1969). "Willingness to pay for health services: a sampling of consumer preferences". Unpublished Paper. Department of Economics, University of Southern California, Los Angeles.
Pattison, W.S. (1970). Economics of moose hunting activities. Master's Thesis, Department of Rural Economy, University of Alberta.
Pawitan, Y. (2001). In All Likelihood: Modeling and Inference Using the Likelihood. Oxford University Press, New York.
Payne, S. (1951). The Art of Asking Questions. Princeton University Press, Princeton, NJ.
Pearce, D.W., Markandy, A. (1989). Environmental Policy Benefits: Monetary Evaluation. Organization for Economic Cooperation and Development (OECD), Paris.
Pendse, D., Wyckoff, J.B. (1972). "Determining preference for environmental goods". Paper presented at the Meeting of the Pacific Division of the American Academy for the Advancement of Science, Eugene, OR.
Peterson, G.L., Worrall, R.D. (1970). "An analysis of individual preferences for accessibility to selected neighborhood services". In: Highway Research Record, vol. 305. Highway Research Board, Washington, DC.

Plowden, S.P.C. (1970). The Cost of Noise. Metra Consulting Group Ltd., London.
Plowden, S.P.C., Sinnott, P.R.J. (1977). "Evaluation of noise nuisance: a study of willingness to receive payment for noise introduced into the home". Supplementary Report 261. Transport and Road Research Laboratory, Crowthorne, UK.
Poe, G.L., Giraud, K.L., Loomis, J.B. (2005). "Computational methods for measuring the difference of empirical distributions". American Journal of Agricultural Economics 87, 353–365.
Portney, P.R. (1994). "The contingent valuation debate: why economists should care". Journal of Economic Perspectives 8 (4), 3–18.
Propper, C. (1990). "Contingent valuation of time spent on NHS lists". Economic Journal 100, 193–199.
Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Singer, E. (2004). Methods for Testing and Evaluating Survey Questionnaires. Wiley, New York.
Rae, D.A. (1982). "Benefits of visual air quality in Cincinnati". Report to the Electric Power Research Institute, Charles River Associates, Boston.
Randall, A. (1973). "Frontiers of nonmarket valuation". In: Stoll, J.R., Shulstad, R.N., Smathers, W.M., Jr. (Eds.), Nonmarket Valuation: Current Status and Future Directions, Proceedings of a Regional Workshop. SNREC Publication No. 18. Southern Rural Development Center/Farm Foundation.
Randall, A. (1993). "Passive-use values and contingent valuation: valid for damage assessment". Choices 8 (2), 12–15.
Randall, A. (1998). "Beyond the crucial experiment: mapping the performance characteristics of contingent valuation". Resource and Energy Economics 20, 197–206.
Randall, A., Grunewald, O., Pagoulatos, A., Ausness, R., Johnson, S. (1978). "Reclaiming coal surface mines in central Appalachia: a case study of the benefits and costs". Land Economics 54, 427–489.
Randall, A., Hoehn, J.P. (1996). "Embedding in market demand systems". Journal of Environmental Economics and Management 30, 369–380.
Randall, A., Hoehn, J.P., Tolley, G.S. (1981). "The structure of contingent markets: some experimental results". Paper presented at American Economic Association Meeting, Washington, DC.
Randall, A., Ives, B.C., Eastman, C. (1974). "Bidding games for the valuation of aesthetic environmental improvements". Journal of Environmental Economics and Management 1, 132–149.
Randall, A., Stoll, J.R. (1980). "Consumer's surplus in commodity space". American Economic Review 70, 449–455.
Ready, R.C., Hu, D. (1995). "Statistical approaches to the fat tail problem for dichotomous choice contingent valuation". Land Economics 71, 491–499.
Ready, R.C., Navrud, S., Dubourg, W.R. (2001). "How do respondents with uncertain willingness to pay answer contingent valuation questions?". Land Economics 77, 315–326.
Ready, R.C., Whitehead, J.C., Blomquist, G.C. (1995). "Contingent valuation when respondents are ambivalent". Journal of Environmental Economics and Management 29, 181–196.
Reardon, G., Pathak, D.S. (1988). "Monetary valuation of benefits of pharmacy services". Topics in Hospital Pharmacy Management 8 (2), 66–75.
Reizenstein, R.C., Hills, G.E., Philpot, J.W. (1974). "Willingness to pay for control of air pollution: a demographic analysis". In: Curhan, R.C. (Ed.), New Marketing for Social and Economic Progress and Marketing's Contributions to the Firm and to Society. American Marketing Association, Chicago.
Richer, J.R. (1995). "Willingness to pay for desert protection". Contemporary Economic Policy 13, 93–104.
Ridker, R.G. (1967). Economic Cost of Air Pollution. Praeger, New York.
Ridker, R.G., Henning, J.A. (1967). "The determinants of residential property values with special reference to air pollution". Review of Economics and Statistics 49, 246–257.
Roberts, K., Thompson, M.E., Pawlyk, P.W. (1985). "Contingent valuation of recreational diving at petroleum rigs, gulf of Mexico". Transactions of the American Fisheries Society 114, 214–219.
Rollins, K., Lyke, A. (1998). "The case for diminishing marginal existence values". Journal of Environmental Economics and Management 36, 324–344.
Rosenberger, R.S., Loomis, J.B. (2000). "Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database". Water Resources Research 36, 1097–1107.

Roskill, The Honorable Mr. Justice Chairman (1970). Commission on the Third London Airport: Papers and Proceedings. HMSO, London.
Rowe, R.D., d'Arge, R.C., Brookshire, D.S. (1980). "An experiment on the economic value of visibility". Journal of Environmental Economics and Management 7, 1–19.
Rowe, R.D., Schulze, W.D., Breffle, W.S. (1996). "A test for payment card biases". Journal of Environmental Economics and Management 31, 178–185.
Rushton, G. (1969). "Analysis of spatial behavior by revealed space preference". Annals of the Association of American Geographers 59, 391–400.
Ruud, P.A. (1986). "Contingent ranking surveys: their application and design in estimating the value of visibility". In: Bhardwaja, P.S. (Ed.), Visibility Protection: Research and Policy Aspects. Air Pollution Control Association, Pittsburgh, pp. 206–217.
Sagoff, M. (1994). "Should preferences count". Land Economics 70, 127–144.
Samples, K.C., Dixon, J.A., Gower, M.M. (1985). "Information disclosure and endangered species valuation". Paper presented at the Annual Meeting of the American Agricultural Economics Association, Ames, IA.
Samuelson, P. (1954). "The pure theory of public expenditures". Review of Economics and Statistics 36, 387–389.
Sarja, M. (1969). Lappajärven ja Venetjoen tekojärven virkistyskäyttötutkimus ja virkistsarvon määrittäminen (A study on recreation use and recreation value of lake Lappjärvi and artificial lake in Venetjoki). Master's Thesis. Department of Geography, Helsinki University.
Satterthwaite, M. (1975). "Strategy-proofness and arrow conditions: existence and correspondence theorems for voting procedures and welfare functions". Journal of Economic Theory 10, 187–217.
Schelling, T. (1968). "The life you save may be your own". In: Chase, S. (Ed.), Problems in Public Expenditure Analysis. Brookings Institute, Washington, DC.
Schuman, H., Presser, S. (1981). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Academic Press, New York.
Seip, K., Strand, J. (1992). "Willingness to pay for environmental goods in Norway: a CV study with real payment". Environmental and Resource Economics 2, 91–106.
Shechter, M., Enis, R., Baron, M. (1974). "Mt. Carmel National Park: the demand for outdoor recreation". Unpublished Paper. Center for Urban & Regional Studies. Technion (Israel Institute of Technology), Haifa.
Sinden, J.A. (1974). "A utility approach to the valuation of recreational and aesthetic experiences". American Journal of Agricultural Economics 56, 61–72.
Sinden, J.A., Worrell, A.C. (1979). Unpriced Values: Decisions without Market Prices. Wiley, New York.
Smith, V.K. (1976). "The treatment of intangibles in benefit–cost analysis". Environmental Control Newsletters 4 (5).
Smith, V.K. (1984). Environmental Policy Under Reagan's Executive Order: The Role of Benefit–Cost Analysis. University of North Carolina Press, Chapel Hill, NC.
Smith, V.K., Desvousges, W.H. (1986). Measuring Water Quality Benefits. Kluwer Academic, Boston.
Smith, V.K., Desvousges, W.H., Freeman, A.M. III (1985). Valuing Changes in Hazardous Waste Risks: A Contingent Valuation Analysis, vol. I. Draft Report to the Environmental Protection Agency no. RTI Project No. 41U-2699. Research Triangle Institute, Research Triangle Park, NC.
Smith, V.K., Mansfield, C. (1998). "Buying time: real and hypothetical offers". Journal of Environmental Economics and Management 36, 209–224.
Smith, V.K., Osborne, L. (1996). "Do contingent valuation estimates pass a scope test? A meta analysis". Journal of Environmental Economics and Management 31, 287–301.
Starrett, D.A. (1988). Foundations of Public Economics. Cambridge University Press, New York.
Stavig, G.R., Gibbons, J.D. (1977). "Comparing the mean and median as measures of centrality". International Statistical Review 45, 63–70.
Stewart, B.E. (1961). Recreational use of private land in a portion of Eastern Maine. Master's Thesis, Department of Agricultural and Resource Economics, University of Maine.
Stigler, G.J., Becker, G.S. (1977). "De gustibus non est disputandum". American Economic Review 67, 76–90.

Stigler, S.M. (1977). "Do robust estimators work with real data?". Annals of Statistics 5, 1055–1098.
Strauss, R.P., Hughes, G.D. (1976). "A new approach to the demand for public goods". Journal of Public Economics 6, 191–204.
Strazzera, E., Scarpa, R., Calia, P., Garrod, G.D., Willis, K.G. (2003). "Modelling zero values and protest responses in contingent valuation surveys". Applied Economics 35, 133–138.
Sudman, S., Bradburn, N.M. (1982). Asking Questions: A Practical Guide to Questionnaire Design. Jossey-Bass, San Francisco.
Sugden, R. (1999). "Alternatives to the neo-classical theory of choice". In: Bateman, I.J., Willis, K.G. (Eds.), Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the USA, EC, and Developing Countries. Oxford University Press, New York.
Swait, J., Louviere, J.J. (1993). "The role of the scale parameter in the estimation and use of multinomial logit models". Journal of Marketing Research 30, 305–314.
Thaler, R.H. (1985). "Mental accounts and consumer choice". Marketing Science 4, 199–214.
Thaler, R.H., Rosen, S. (1976). "The value of saving a life". In: Terleckyj, N.E. (Ed.), Household Production and Consumption. National Bureau of Economic Research, New York.
Thayer, M.A. (1981). "Contingent valuation techniques for assessing environmental impacts: further evidence". Journal of Environmental Economics and Management 8, 27–44.
Throsby, C.D. (1984). "The measurement of willingness to pay for mixed goods". Oxford Bulletin of Economics and Statistics 46, 279–289.
Thurstone, L. (1927). "A law of comparative judgment". Psychological Review 34, 273–286.
Tolley, G.S., Randall, A., Blomquist, G.C., Fabian, R.G., Fishelson, G., Frankel, A., Hoehn, J.P., Krumm, R., Mensah, E. (1986). Establishing and valuing the effects of improved visibility in the Eastern United States. Report to the Environmental Protection Agency No. Cooperative Agreement 807768-01-0, University of Chicago, Chicago.
Tourangeau, R., Rips, L.J., Rasinski, K. (2000). The Psychology of Survey Response. Cambridge University Press, New York.
Turnbull, B. (1976). "The empirical distribution function with arbitrarily grouped, censored and truncated data". Journal of the Royal Statistical Society B 38, 290–295.
Tversky, A., Kahneman, D. (1986). "Rational choice and the framing of decisions". Journal of Business 59, S251–S278.
Ullman, E.L., Volk, D.J. (1961). "An operational model for predicting reservoir attendance and benefits: implications of a location approach to water recreation". Washington University, St. Louis, MO.
U.S. Department of the Interior. United States (1986). "Final rule for natural resources damage assessments under the comprehensive environmental response compensation and liability act of (CERCLA)". Federal Register 51, No. 1148, 27674–27753.
U.S. Senate Committee on Public Works (1957). "Evaluation of recreation benefits". 85th Congress 1st Session, Washington, DC.
U.S. Water Resources Council (1979). "Procedures for evaluation of national economic development (NED) benefits and costs in water resources planning (level C), final rule." Federal Register 44, No. 242, 72892–72976.
Viscusi, W.K. (1993). "The value of risks to life and health". Journal of Economics Literature 31, 1912–1946.
Vossler, C.A., Kerkvliet, J. (2003). "A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum". Journal of Environmental Economics and Management 45, 631–649.
Vossler, C.A., Kerkvliet, J., Polasky, S., Gainutdinova, O. (2003). "Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon". Journal of Economic Behavior and Organization 51, 261–277.
Walsh, R.G., Loomis, J.B., Gillman, R.A. (1984). "Valuing option, existence and bequest demands for wilderness". Land Economics 60, 14–29.
Walsh, R.G., Miller, N.P., Gilliam, L.O. (1983). "Congestion and willingness to pay for expansion of skiing capacity". Land Economics 59, 195–210.

Walsh, R.G., Sander, L.D., Loomis, J.B. (1985). Wild and scenic river economics: recreation use and preservation values. Report to the American Wilderness Alliance, Department of Agriculture and Natural Resource Economics, Colorado State University.

Water Demand Research Team. World Bank (1993). "The demand for water in rural areas: determinants and policy implications". World Bank Research Observer 8, 47–70.

Weisbrod, B.A. (1964). "Collective consumption services of individual consumption goods". Quarterly Journal of Economics 78, 471–477.

Werner, M.J. (1999). "Allowing for zeros in dichotomous-choice contingent-valuation models". Journal of Business and Economic Statistics 17, 479–486.

Werner, M.J., Groves, T. (1993). "A practical procedure for public policy decisions: or 'contingent valuation and demand revelation – without apology'." Discussion Paper 93-51. Department of Economics. University of California, San Diego.

Whittington, D. (1988). Guidelines for conducting willingness to pay studies for improved water services in developing countries. USAID Water and Sanitation for Health Project No. Washington Field Report No. 306. U.S. Agency for International Development, Washington, DC.

Whittington, D. (1998). "Administering contingent valuation surveys in developing countries". World Development 26, 21–30.

Whittington, D. (2002). "Improving the performance of contingent valuation studies in developing countries". Environmental and Resource Economics 22, 323–367.

Whittington, D., Mujwahuzi, M., McMahon, G., Choe, K.-A. (1988). Willingness to pay for water in Newala district, Tanzania: strategies for cost recovery. USAID Water and Sanitation for Health Project Field Report No. 246. U.S. Agency for International Development, Washington, DC.

Willig, R. (1976). "Consumer's surplus without apology". American Economic Review 66, 589–597.

World Bank (1987). "Willingness to pay for water in rural areas". Research Proposal. Water and Urban Development Department, World Bank, Washington, DC.

Wu, C.F.J., Hamada, M. (2000). Experiments: Planning, Analysis, and Parameter Design Optimization. Wiley, New York.

Zeckhauser, R., Shepard, D. (1976). "Where now for saving lives?". Law and Contemporary Problems 40, 5–45.

Zhao, J., Kling, C.L. (2004). "Willingness to pay, compensating variation, and the cost of commitment". Economic Inquiry 42, 503–517.