

# Lecture # 16 – Valuing Environmental Benefits: Stated Preference Approaches

## I. Stated Preference Valuation

- So far, we have studied revealed preference approaches to valuing environmental amenities. Those approaches look at actual market transactions to infer value.
- Unfortunately, there aren't always market transactions that can serve this purpose. For example, how do we value protection of endangered species?
- In these cases, economists simply ask people for their valuation. These are known as stated preference techniques
- Types of stated preference approaches
  - Contingent valuation
    - Uses surveys to assess WTP
    - Typically a yes/no response to a posted price
    - “Contingent” because asking people what they would do if faced with a market for environmental amenities, since the actual markets do not exist.
      - Uses textual description of scenarios
      - Will respondents identify key features in the text?
  - Choice experiments: conjoint analysis
    - Respondent asked to consider an environmental commodity defined by several attributes
    - Choose among three or more options during each task
    - Reveal information on value of individual attributes, which are not emphasized in CV
      - But short descriptions of attributes may oversimplify
- Contingent valuation versus choice experiments
  - Decision of which to use depends on how respondents tend to perceive the good. Suggest three considerations:
    - Will the change being valued affect specific characteristics of the item or the item as a whole
      - CV works well if estimating fixed set of changes as a whole (e.g. oil spill damages)
      - CE useful if some attributes affected but not others
    - Do respondents think of the change in terms of individual attributes or as a whole?
      - E.g. do they think of an ecosystem holistically?
      - If they think of individual attributes, can the attributes be thought of separately?
        - If so, CE appropriate.
    - How does information presentation affect respondents understanding of item to be valued?
      - If complex, listing individual attributes (CE) may be confusing

- Policy background
  - The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) gave government the right to sue for damages to natural resources for which they are trustees.
    - In 1986, the Department of Interior said nonuse (largely existence) values were recoverable if use values were not measurable.
      - Sanctioned the use of CV
  - In 1989, a federal court of appeals directed the Department of the Interior to redraft its regulations with equal weight on use and nonuse values.
  - After the Exxon Valdez crash in March of 1989, DOI regulations meant that Exxon would have to pay for non-use damages.
    - CV estimate of damages: \$3 billion. Other studies cited in Kling *et al.* range from \$3.8 *million* (recreation demand study) to \$4.9 *billion*
    - Difference is almost entirely that the larger study includes non-use values, rather than just the loss of actual visits to the spill area
      - Thus, CV became a focus of court cases.
    - Settled through a U.S. District Court consent decree in 1991.
      - Paid out \$1 billion in damages and over \$2 billion in immediate responses and restoration efforts
  - In response to the Exxon Valdez, Congress passed the Oil Pollution Act of 1990.
    - In addition to reducing the likelihood of future spills, the act provided for damage recovery.
    - The Department of Commerce was to draw up regulations on damage assessment.
      - The Dept., through the National Oceanic and Atmospheric Administration (NOAA), assembled a blue ribbon panel to evaluate CV.
        - Question to panel: Is the CV method capable of providing estimates of lost nonuse or existence values that are reliable enough to be used in natural resource damage assessments?
  - BP accident in 2010 raised the issue again

- How contingent valuation works:
  1. Ask people their willingness to pay (WTP) to bring about a specific environmental improvement (e.g. improved visibility from air pollution, protection of an endangered species).
  2. A payment mechanism is specified.
    - The respondent must believe that the money paid could lead to the desired effect.
      - E.g.: increase sewer fees to improve water quality.
    - It should not be controversial (e.g. property taxes).
      - Because the survey shouldn't serve as a referendum on the type of payment mechanism chosen.
      - Other alternatives include surcharges on utility bills or generally assessed fees
  3. Information about the respondent is usually gathered.
    - E.g. income, age, education...
    - Allows verification of results, estimation of income elasticities, etc.
- Survey development and implementation – how to describe the scenario
  - Goal of good survey design is to maximize validity and reliability of results
  - Scenario description
    - Both the baseline (status quo) and proposed change must be described in a way that is understood and viewed as credible
      - Description should be specific
        - Want to capture the value of a specific amenity (e.g. cleaning a local lakes) not a general class (e.g. preferences for lakes in general)
        - People may react to news (e.g. oil spills)
          - Do people care about the resource (e.g. Alaskan coast) or preventing the action (e.g. clean up oil spills)?
      - Provide information on the mechanisms that will lead to the change
        - Payment mechanism must be clear and credible (e.g. on tax bill or utility bill)
    - Surveys should be pretested
      - Both qualitative (e.g. focus groups) and quantitative (pilot studies)
      - Do respondents understand the baseline and proposed treatment?
      - Verbal protocol analysis looks at why people answer as they do.
        - Have respondents think out loud.
          - When they do this, they often say irrelevant things: “If we all do our part it won’t cost much.”
          - Pretesting of surveys can avoid these problems.

- Data collection
  - Survey method
    - E.g. mail vs. electronic
      - Do Internet surveys reach all desired respondents?
    - Choice is often context specific
      - Phone generally worst, as difficult to convey complex information
  - Sample
    - Market studied should be the group of people whose welfare is affected by the change being valued
      - Harder to identify for non-use values
    - Draw sample from known frames consistent with the population
      - Check for non-response bias
  - Choice of sample relates to generalizability
    - To use WTP results to guide policy often requires aggregating results to a population
    - Knowledge of the sample frame is important for establishing generalizability
      - Socioeconomic characteristics of respondents should be documented
      - Test for non-response bias
        - Do the demographics of respondents compare to demographics of the population?
        - If there are differences, demographic data can be used to re-weight the responses
- Include auxiliary questions
  - Can help test validity
    - E.g. does willingness to pay increase with income?
  - Theory should motivate the use of auxiliary questions
    - Be careful to not include endogenous covariates
      - Responses to questions asked after valuation question may be influenced by the valuation
        - E.g. asking about a person's environmental beliefs or the certainty of their answer
  - Can identify demographic information
    - Demographics clearly exogenous

- Value elicitation – a payment mechanism is specified
  - The respondent must believe that the money paid could lead to the desired effect.
    - E.g.: increase sewer fees to improve water quality.
    - Surveys should be designed so participants believe the study's results could affect their well-being (consequentiality)
  - It should not be controversial.
    - Because the survey shouldn't serve as a referendum on the type of payment mechanism chosen.
    - Other alternatives include surcharges on utility bills or generally assessed fees (although Kling *et al.* do consider property taxes an alternative)
  - Use incentive-compatible response formats
    - Incentive compatible means that respondents have incentives to reveal their true value
      - E.g. there aren't strategic reasons to give a higher or lower answer
      - Voluntary contributions to public goods are an example
        - If the respondent believes the survey will be used to decide on the ultimate provision of a public good, may have incentive to report *more* than the true WTP and to contribute less should it become a reality
        - So as to influence provision and have an opportunity to free ride
      - Binding payment vehicles are incentive compatible and avoid free riding
        - Some taxes (e.g. income or sales taxes) can be avoided by consumers, and are thus not good choices
        - The question should involve a firm sense of commitment.
          - "If the program is approved, firms will raise prices by \$X."
    - Must be credible
    - Timing
      - Is it a one-time or annual payment?

- Types of questions:
  - Open-ended – ask respondents for maximum WTP.
    - “What is the highest amount you would be willing to pay each year to ...?”
    - Open-ended questions are not incentive compatible.
      - May think strategically (how much do I need to offer to get the project done), rather than full value of the project.
  - Close-ended – ask respondents whether they are WTP a certain amount. This amount is varied across respondents.
    - “Would you pay \$X for...?”
    - Close-ended questions are more like the everyday decisions that consumers make every day, and are thus easier for respondents to understand.
    - Best practice is to describe the market as a referendum
  - Bidding games – ask respondents if they are WTP a certain amount. If they say yes, raise the bid until they say no.
    - “Would you pay \$5 for...?” “Yes.” “Would you pay \$10?”...
  - While early studies used a range of methods, standard practice now is to use close-ended responses.
    - A single bid offer is best, as iterative bidding can be subject to anchoring (e.g. influenced by where bidding started)
- What value to elicit? Willingness to pay (WTP) or willingness to accept (WTA)?
  - Implied property rights can inform the decision
  - Most studies estimate WTP
    - Framing WTA is more difficult
- Use multiple questions
  - Allows for testing for consistency in answers
  - May help respondents generate a better understanding of the task at hand
  - But be careful that sequencing doesn’t affect answers
    - Questions should be independent of one another
    - Later questions shouldn’t depend on answers to previous questions
    - If that will be a problem, shouldn’t use multiple questions

- Example: study of the damage from the BP oil spill
  - Part of an effort to determine the value of natural resource damage from the spill
  - Used a nationally representative stated-preference survey
    - Would respondents support investing at least \$17.2 billion to prevent future oil spills of this type in the Gulf of Mexico?
    - Took three years to design and pre-test the survey
    - Surveyed 3656 people
      - Response rate of 48%
  - Tests whether responses are consistent with economic decisions
  - Survey
    - Describes:
      - the state of the Gulf before the accident
      - what caused the accident
      - injuries to the Gulf's natural resources
        - This was randomly varied, describing different sets of injuries, to test for sensitivity to scope of injury.
    - Examples:
      - Smaller set: # of miles of oiled marshes, dead birds, and lost recreation trips
      - Larger set: also included injuries to dolphins, coral, snails, and young fish and sea turtles
    - a proposed program for preventing future accidents
      - Can be seen as insurance against another spill occurring in the next 15 years
      - Involves drilling a second pipe as the same time a well is drilled
        - Because it takes at least three months to drill a pipe, waiting until a leak occurs will not prevent an accident
        - Doubles the cost of drilling, so survey says companies will need government support to pay for it
    - how much extra the household would pay in taxes if the program were implemented
      - Randomly assigned one of five different tax amounts: \$15, \$65, \$135, \$265, or \$435.
      - Votes are used to establish a lower bound for WTP

- *Question:* What did the researchers do to ensure respondents believed their answers were consequential?
  - Respondents received official letter from US Department of Commerce, on letterhead, emphasizing that the survey results would be used for policy-making.
  - Interviewers reviewed content of that letter before administering the survey
  - Questionnaire explained that a decision on the program hadn't been made yet, and if it did move forward, it would require new tax revenue.
- *Question:* How do they show the results are represent consistent decisions?
  - Proportion of people voting for a proposal shouldn't increase as the tax increases
  - For a given tax amount, a program avoiding more injuries should be preferred
  - Table in paper shows both conditions were met.
- Results:
  - WTP for smaller set of injuries: \$136 (s.e. \$6.34)
  - WTP for larger set of injuries: \$153 (s.e. \$6.87)
    - Implies an aggregate benefit of \$17.2 billion
      - Equals  $\$153 \times 112,647,215$  households in the US



- Conjoint analysis example: ([Peng and Oleson, \*Ecological Economics\* 2017](#))
  - Paper uses conjoint analysis to value water quality at Hawaiian beaches
  - Survey
    - Survey given at local beaches in Hawaii
      - 263 interviews conducted
      - 11% refusal rate
    - Socioeconomic information collected
      - Residency, mode of transport, time in transit to beach, primary trip purpose, frequency of visits, # in group, annual income
    - Asked about attitudes towards water quality
      - “I will not enter the water if bacteria is over the safe limit” and “I judge how clean the water is by how clear it is”)
    - Choice experiment
      - Each survey included 8 choice scenarios
        - In each, choose between 3 beaches
        - Baseline beach was the same in all 8 scenarios
          - Baseline beach was similar to current conditions at beaches in the survey area
      - Five attribute survey:
        - Number of days with beach advisory
          - Average beach in Hawaii exceeds bacterial levels for advisory 11 days per year
        - Water quality
        - Coral reef cover
        - Fish diversity
        - Payment vehicle
          - Additional cost to move to a better beach
          - Used because all beaches in Hawaii are free
            - Charging admission could lead to protest 0's
- Results (Table 4)
  - Most respondents agreed that base beach characteristics represented current conditions accurately
  - Bacterial exceedance
    - WTP \$11.43 to reduce days of bacterial exceedance from 11 to 5 per year
    - WTP \$42.15 to reduce days of bacterial exceedance from 11 to 0
  - Water clarity
    - WTP \$35.71 to improve water clarity from 15 to 30 feet
    - WTP \$50.51 to improve water clarity from 15 to 60 feet
  - Coral reef cover
    - WTP \$15.33 to improve coral reef cover from 10% to 25%
    - WTP \$20.22 to improve coral reef cover from 10% to 45%
  - Fish diversity
    - WTP \$7.14 to improve fish density from low to medium
    - WTP \$9.61 to improve fish density from low to high

## II. Evaluating CV

- Validity assessment
  - Note that bias does not preclude validity
    - Need to consider how much bias is acceptable
    - Virtually all empirical studies have some bias, not just stated preference.
  - Economists consider four types of evaluating validity when assessing contingent valuation. I describe each below.
- Content validity: Do surveys lead to true values? Is best practice being followed?
  - Are the procedures used to design and implement a survey appropriate?
  - Can be evaluated using the guidelines from this article
    - Each study must be evaluated separately
  - Kling et al. point out that best practices are now well known. Three important ones are:
    - The environmental good needs to be defined with a high level of specificity
      - Both status quo and changed level need to be explained in a way accessible to laypeople
    - Constructed market should represent a realistic mechanism for bringing about the proposed change
      - Size of change needs to be physically plausible
      - Context matters: estimated values should not be independent of context
    - Understanding how to encourage and/or test for rationality and truthfulness has evolved
      - Not as simple as using a referendum format
      - Voluntary vs. coercive payment matters
      - Framing the survey to be consequential matters
      - Criteria to evaluate construct validity are case specific and depend on context

- Lessons from behavioral economics
  - Behavioral economics suggests that people are not always rational
  - Behavioral anomalies matter
    - Individual preferences may not be well-behaved in neoclassical sense
    - Individuals do not always optimize when making choices
  - If behavioral anomalies are observed in stated preference outcomes, is it a failure of stated preference methods or of the neoclassical paradigm?
    - Are anomalies because respondents aren't revealing true preferences or because they are simply inconsistent with the expectations of the researcher?
    - Note that true behavioral anomalies are not unique to stated preference. Can occur in revealed preference as well.
  - Types of anomalies
    - Individuals apply simplified decision rules to reduce cognitive burden (e.g. because the good is unfamiliar)
    - Scenario adjustment or rejection
      - E.g. value something other than intended by the researcher
      - Needs to be checked for during survey pre-testing
    - Anomalies due to survey design
      - Anchoring
        - Whether ask for value of a seal and then a whale, or a whale and then a seal, the first value given is around \$100.
        - Therefore, WTP depends on order of questions.
      - Therefore, sequencing matters
    - Endowment effect
      - People require more compensation to part with something they already have than they would give up to acquire it
      - Explains divergence in WTP and WTA, which is often cited as evidence of CV's failings
    - Warm glow
      - Now understood as one of many reasons for pro-social behavior such as contributing to public goods
      - Social norms and other-regarding preferences such as altruism and reciprocity also lead individuals to value an environmental good more than its private benefits

- Reasons for departures from optimizing behavior:
  - People may make “mistakes” due to bounded rationality
    - E.g. mental accounting implies that it is not one’s overall budget but a specific expense category that matters
    - Evidence: Li *et al* (2005) found that respondents had lower WTP for reduction of global warming when they received reminders about their discretionary income and its use for environmental causes, compared to only receiving reminders about their household budget
  - Rationality may be the results of repeated participation in markets
    - Mistakes are costly, so people learn
    - Therefore, complex or unfamiliar decision environmental may lead to mistakes resulting from “rule of thumb” behaviors
      - CV studies may be particularly vulnerable to this
- These lessons from behavioral economics help explain anomalies, but do they also raise concerns?
  - If anomalies exist, comparing estimates of surveys and market data may not be valid
    - Must choose a paradigm first (e.g. behavioral or neo-classical) and then design a study to test accuracy of stated preference based on that paradigm
    - If stated preference fails the test, is it because stated preference is flawed or the paradigm is flawed?
  - But if so, how do we make use of the numbers from CV in policy?
- Construct validity: are stated preference estimates consistent with theoretical predictions?
  - Most common way of testing validity
  - Note that some constructs, such as adding-up test, are also failed by marketing studies
    - Thus, shouldn’t invalidate the use of stated preference
  - Are the results similar and consistent with theory?
    - Values increase when income increases.
    - Environmentalists WTP more.
      - Hausman and Diamond argue that would happen despite flaws in survey.
      - For example, if measuring general taste for environmental quality.

- Kling et al. note four testable predictions:
  - Number of people willing to contribute to an environmental good in a survey should increase when requested payment falls
  - People should be willing to pay more to have a higher quantity of the good (scope effect)
    - While scope effects missing in some early studies, recent meta-analyses show scope effects present in well-designed studies
  - Income elasticity of WTP should be greater than one
    - Assumes environmental quality is a luxury good
  - WTP and WTA should not be substantially different
    - Observed divergence appears to be because of behavioral anomalies, not design of hypothetical studies
- The first generally holds, but the other three were often violated for stated preference data, particularly in early surveys
- Other problems noted by critics of CV
  - Embedding
    - Suppose 1 group asked about X
    - A second group is asked about Y
    - A third group is asked about X + Y
    - WTP of (1) + (2) should equal (3)
    - However, (1) + (2) often > 3
    - Critics see embedding as evidence of a warm glow.
      - It's not just that people value X or Y, but that people feel good about supporting the environment.
      - When you ask the values individually, the warm glow is added up twice.
      - Thus, adding up results of individual studies leads to double counting of benefits, making aggregation of results for policy difficult.
  - Defenses of embedding
    - Diminishing returns: protecting one lake is valuable. Protecting a second lake isn't as important.
    - Income effects could be a factor.
      - However, WTP is a small percentage of income.
    - Could be large substitution effects between X and Y

- Kling *et al.* note that more recent work shows that these three predictions are sensitive to two common features of environmental goods:
  - Fixed quantities
    - Demand will be downward sloping
    - But relationship to income depends on:
      - Implied income elasticity of demand
      - Substitutability among all quantity-constrained goods
      - Share of augmented income allocated to market goods
        - Implies income elasticity of WTP may be less than one even if income elasticity of demand for the fixed quantity is greater than one
    - Limited substitutability with other consumption goods
- Kling *et al.* caution that as new stated preference techniques are developed, construct validity must be kept in mind
- Convergent validity
  - Are the results similar to results from revealed preference approaches?
    - A 1996 meta-analysis of studies finds results are usually consistent
  - Examples
    - Sellar *et al.* (1985) used CV and travel cost at 4 Texas lakes.
      - At 3 of the 4, the 95% confidence intervals overlapped.
      - However, means did vary (TC \$102.09 at one lake, CV \$35.21, but within 95% interval).
    - Brookshire *et al.* (1982) compare hedonics and CV on air quality.
      - Results are similar and consistent with theory
  - Not possible for non-use values, but can be used in other cases (e.g. recreation)

- Criterion validity: do stated preference estimates match real payments?
  - Compare prediction from stated preference to a standard thought to be a suitable proxy for true measurement
    - Examples include experimental studies and voting studies
  - Two types of lab experiments
    - Induced value experiment: Participants are assigned a value for the experimental good as part of the research design
      - Advantage: can focus on value elicitation
      - Used to examine accuracy of hypothetical referendum vehicles versus binding real payments votes
      - Generally find that distribution of values from hypothetical votes matches induced-value
    - Homegrown value experiments: participants' actual values for a real commodity used as the criterion
      - E.g. consider a referendum where all participants must pay a given amount if the majority votes in favor
      - Results compared to real payment mechanism as a test of validity
      - Generally find that stated values are higher
        - This is known as hypothetical bias
        - Critique (Kling et al. p. 15): Not all of these studies satisfy the incentive compatibility and consequentiality requirements
          - Recent work finds hypothetical bias goes away when consequentiality holds
  - Why might results differ?
    - Ongoing work attempting to answer this
      - E.g. asking people if they are certain and recoding uncertain "yes" as "no" improves results
    - Understanding why would make it possible to calibrate answers