AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

333
S45
1999

WESTERN REGIONAL RESEARCH PUBLICATION

W-133
BENEFITS AND COSTS OF RESOURCES POLICIES AFFECTING
PUBLIC AND PRIVATE LAND

12TH INTERIM REPORT
JUNE 1999

Compiled by
W. Douglass Shaw

Department of Applied Economics and Statistics
Mail Stop 204
University of Nevada
Reno, Nevada 89557-0105

INTRODUCTION

This volume contains the proceedings of the 1999 W-133 Western Regional Project Technical Meeting on "Benefits and Costs of Resource Policies Affecting Public and Private Land." Some papers from W-133 members and friends who could not attend the meeting are also included. The meeting took place February 24th - 26th at the Starr Pass Lodge in Tucson, Arizona. Approximately 50 participants attended the 1999 meeting, are listed on the following page, and came from as far away as Oslo, Norway.

The W-133 regional research project was rechartered in October, 1997. The current project objectives encourage members to address problems associated with: 1.) Benefits and Costs of Agro-environmental Policies; 2.) Benefits Transfer for Groundwater Quality Programs; 3.) Valuing Ecosystem Managment of Forests and Watersheds; and 4.) Valuing Changes in Recreational Access.

Experiment station members at most national land-grant academic institutions constitute the official W-133 project participants. North Dakota State, North Carolina State, and the University of Kentucky proposed joining the group at this year's meeting. W-133's list of academic and other "Friends" has grown, and the Universities of New Mexico and Colorado were particularly well represented at the 1999 W-133 Technical Meeting. The meeting also benefitted from the expertise and participation of scientists from many state and federal agencies including California Fish and Game, the U.S. Department of Agriculture's Economic Research and Forest Services, the U.S. Department of Interior's Fish and Wildlife Service, and the Bureau of Reclamation. In addition, a number of representatives from the nation's top environmental and resource consulting firms attended, some presenting papers at this year's meeting.

This volume is organized around the goals and objectives of the project, but organizing the papers is difficult because of overlapping themes. The last section includes papers that are very important to the methodological work done by W-133 participants, but do not exactly fit one of the objectives. -- I apologize for the lack of consistent pagination in this volume.

On A Personal Note... Any meeting or conference is successful (and fun!) only because of its participants, so I would first like to thank all the people who came and participated in 1999 - listed below. I also want to thank Jerry Fletcher for all his help at this meeting and prior to it, and John Loomis who passed on his knowledge of how to get a meeting like this to work, and who continues to have the funniest little comments to lighten the meetings up. I especially thank Paul Jakus, who helped me to organize this conference and have a lot of fun during it and afterward. Finally, I want to thank Nicki Wieseke for all her help in preparing this volume, and Billye French for administrative support on conference matters.

W. Douglass Shaw, Dept. of Applied Economics & Statistics, University of Nevada, Reno.
June, 1999

P.S. P.F. and J.C. - As far as I can tell, that darn scorpion is still dead!

Lessons Learned: A Systematic Look at Validity Issues in Conjoint Analysis

Daniel J. Mullarkey *

Abstract

Contingent Valuation (CV) and Conjoint Analysis (CA) are both stated preference valuation techniques, and have a lot in common. To date, CV has been more widely used in the context of valuing natural resource amenities. Considerable attention has been focused on the validity of CV surveys, and over time we have learned much about how to assess and improve the validity of CV surveys. Many of the lessons learned from the debate over the validity of CV can be applied to the emerging use of CA to value natural resource amenities. This paper uses the content, construct and criterion validity framework to explore a variety of issues related to CA techniques. Among the issues receiving particular attention are: the types of information needed and the problem of information burden; the choice of valuation format (e.g., choice experiments or ranking or rating exercises); the appropriateness of CA for estimating nonuse values; issues associated with the multiple sources of uncertainty that may be present, including whether survey participants view the various alternatives within a CA exercise as equally plausible; and the ability of CA to estimate willingness to accept compensation. The goal of the paper is to use this framework to illuminate techniques to assess validity and to help CA survey designers improve the validity of their surveys. To this end, a variety of validity tests are discussed. The paper concludes with suggestions for future research on CA in the natural resource valuation setting.

* Daniel Mullarkey is a natural resource economist in the Resource Accounting Branch of the Economic Research Service, United States Department of Agriculture. The views expressed in this article do not necessarily represent the policies or views of the U.S. Department of Agriculture.

Introduction

Contingent Valuation (CV) and Conjoint Analysis (CA) are two similar methodologies that can be used to estimate economic values for natural resource amenities. While terminology has yet to be standardized (some authors use the terms stated preference or stated choice), by CA I am referring generally to the use of rating, ranking, or choice experiments that share the following similarities: survey participants are asked about their preferences for alternative bundles of natural resource amenities that are described by a set of attributes; and the levels of the attributes are varied across the alternatives. This allows us to tease out estimates of marginal values for each attribute, which is perhaps the major advantage of CA over CV. The purpose of this paper is not to compare the strengths and weaknesses of the two methods, but rather to adapt what we have learned about validity in CV studies to future research on CA.

The heated debate over the validity of welfare estimates obtained from CV studies has taught us a fair amount about assessing and improving validity, and many of these lessons have analogs for CA. While CA is relatively new to natural resource valuation, it has been widely used in marketing research. Validity research in the marketing literature has focused on predictive validity and test-retest reliability. Natural resource applications present different challenges, however. We are dealing with harder to define goods, which are not bought and sold, for which we often will not have observable behavior to compare predictions to, and for which high-quality surveys will be more expensive. A more useful validity framework for our purposes is the content-construct-criterion validity framework that has been discussed in the CV literature (Mitchell and Carson, 1989; Bishop, Champ and Mullarkey, 1995). This paper uses the 3-C's validity framework to discuss some validity issues that I think require attention as we invest more time and money in using CA for natural resource valuation.

Content Validity

Content validity basically deals with whether the structure of the choice problem and the information provided in the survey are conducive to measuring the economic object of interest. Content validity can be thought of as a necessary but not sufficient condition for the validity of a CA survey. If the survey fails to achieve content validity, there is little point in assessing

construct or criterion validity since the survey is not measuring the object it is intended to. This section discusses several issues related to the content validity of CA studies that I would like to draw attention to.

Information

The most important requirement of any CV or CA survey is to provide participants with the proper information. For either type of survey, the two basic types of information needed are information about the amenity itself and information about the choice framework in which participants will operate. For CA, required information about the amenity includes specifying the status quo level of each attribute (each alternative will provide a target level for each attribute), the source of the change to the amenity, the geographic and temporal extent (i.e., the timing) of the change, and the certainty of the change. Information about baseline levels and any changes to substitutes and complements of the amenity should also be specified where possible.

The key to CA is getting the attributes right. Omission of key attributes will obviously bias the welfare estimates. Focus groups and other preliminary design techniques should be used to develop a survey that includes all the relevant attributes in a clear and useful manner. Experts can and should be used to determine levels of attributes in many applications, but to a large degree the set of attributes should be determined by listening to what lay people say is important to them. It should be noted that the number of attributes included in a survey may need to be restricted in some cases due to cognitive burden (which will be discussed further below). More research is needed into how many attributes participants can effectively deal with in the environmental context.

In addition to selecting the proper attributes, the survey designer will need to select the range of levels and the number of levels to be used across alternatives for each attribute. This is similar to the issue of the number and range of bids to use in a CV survey. In the CA context, Louviere and Timmermans (1990) suggest that in choice experiments, where the individual is asked to choose the preferred alternative from a set of two or more, attributes with a large range of levels may receive more attention that attributes with small ranges. An interesting experiment would be to offer two or more subsamples different ranges for the same attributes. If this results in different marginal values for the same attribute, CA estimates will be open to criticism as being arbitrary.

It is also important to get the economic setting specified accurately. Features of payment mechanism that have been identified as important in the CV literature include the payment vehicle (e.g., income taxes, user fees, etc.), the decision-making unit, the timing of the payment(s), and the prices of substitutes and complements. Three aspects of the context of valuation need to be addressed. The survey should specify all parties that will pay for the change as well as who is expected to be affected by the change. Second, the survey designer needs to decide whether to measure willingness to pay (WTP) or willingness to accept (WTA), or both. Third, the value elicitation device needs to be selected. In CV, this refers to the choice between open-ended, payment card, or referendum questions. In CA, the researcher needs to choose between ranking, ratings, pairwise or three-way choice experiments, or some combination of these formats.

Ranking experiments ask participants to rank a set of alternatives according to their preferences. Ratings experiments typically ask for a numerical rating, often using a 10-point scale, for each alternative. Choice experiments ask the participant to select a preferred alternative from a set of two (pairwise) or more (three-way, four-way, etc.) alternatives. Each format has its strengths and weaknesses. I do not go into much detail here (see Louviere and Timmermans (1990), McKenzie (1993), and Roe, Boyle and Teisl (1996)), but I would like to argue that we should explore the differences carefully. Choice experiments utilize the random utility model that we are used to working with, and therefore offer some advantages over ordinal ratings or rankings, which the researcher must then translate into welfare estimates. However, choice experiments may suffer from biases related to the task of making comparisons, such as greater attention being paid to attributes with high variances. Research on decision heuristics for each form of CA would be very useful in helping us understand what survey participants are actually doing.

Nonuse Values

I want to touch briefly on how nonuse values are handled in CA studies. When participants are evaluating any particular alternative, their nonuse value for the amenity in question will likely influence their response, so CA analysis does incorporate nonuse values. We should avoid the temptation to try to measure nonuse values directly, however, since it is not very tenable to tie nonuse values to any particular attribute. One issue I do see for CA

researchers to deal with is how nonuse values affect the choice of functional form specified for the indirect utility function. The recent paper by Rollins and Lyke (1998) formalized the idea that nonuse values are likely to exhibit diminishing returns to scale. This is inconsistent with the common CA practice of specifying an indirect utility function that is linear and additive in the attributes. With nonuse values present, we would not expect a doubling of the change in attributes to result in a doubling of WTP. Thus other functional forms should be considered when nonuse values are expected to be nontrivial. To the degree that this alters the design of the choice sets, the functional form issue needs to be addressed prior to survey design rather than after the data are collected.

Uniform Plausibility

The different structures of the choice tasks in CV and CA studies raises an additional issue for content validity. There is an implicit assumption in CA that participants find each alternative equally plausible. This assumption makes sense for marketing applications, where there is no uncertainty regarding the well-defined attributes of the product. However, anecdotal evidence from a number of CV surveys supports the notion that people are skeptical that humans can successfully implement large scale improvements in environmental amenities. Specifically, some people are likely to find a small-scale improvement much more likely to be achieved than a large-scale change. People do not believe humankind can perfectly replicate natural processes, thus they will not believe, for example, that manmade wetlands will work as well as natural wetlands. In CV, there is no such issue since while people may be uncertain about the change being offered, there is only one possible change, not many, each with different subjective probability of success. In CA, attribute differences are assumed to be the only differences between alternatives. Hence, the presence of plausibility differences between alternatives will bias estimates of marginal attribute values. In applications with relatively little uncertainty, this may not be much of an issue, but many environmental applications do involve considerable uncertainty. Thus it seems incumbent upon the CA researcher to invest effort in the design phase aimed at determining whether people perceive plausibility differences across scenarios, and to minimize them to the extent possible.

Information Burden

In terms of accuracy of welfare estimates, I think that CA sacrifices precision in favor of flexibility when compared to CV. If we look at it from the perspective of the information burden placed on the survey participant, CV generally asks people to consider two situations – a reference or baseline situation and a target situation. CA, on the other hand, often asks people to evaluate between six and sixteen or more alternatives. If you accept that participants have a fixed burden capacity for either CV or CA, after which point more information is either ignored or the participant starts forgetting or confusing information, then CA must either provide less information per alternative or risk lowering the quality of responses due to information overload. Clearly some information will be constant across conjoint alternatives, so there are some economies of scale. However, some information will not be constant across alternatives. For example, the conjoint surveys I have seen have not specified differences in the time required to implement various alternatives, but it is quite conceivable that larger scale changes will take longer to implement. This type of information may be omitted in an effort to limit the number of attributes included in the survey, and may not matter to some participants, but it is information that can be more easily included in CV, and increases the precision of the definition of the economic construct being considered. My intent here is not to condemn CA since some of this information may not be very important to people, but I do wish to suggest that due to information burden, CA seems to offer less precision than CV, all else equal.

Assessing Content Validity

Assessing content validity is largely a subjective endeavor. As with CV, it will be very important for CA researchers to heavily invest in the design phase. Input and feedback from lay people will be critical elements in developing a successful CA survey. That being said, there are two techniques that I and others have found helpful for developing some quantitative information on content validity.

First, True-False quizzes preceding the actual valuation or choice questions can be used to achieve several goals. This type of exercise causes many participants to review the information provided. This translates into better understanding and absorption of information, and therefore better informed responses. It also provides the researcher with evidence of how effectively the survey communicated important information. CA appears to offer less

opportunity to fully employ True-False questions than CV since key attribute information varies by alternative. However, information that could be included in True-False questions is the status quo level of each attribute, and facts that are assumed constant across alternatives (perhaps the timing of the change and the source of the change, and questions about the payment vehicle).

The second type of quantitative information that can be obtained on content validity comes from follow-up questions that explore whether participants accepted key elements of the scenario. If large numbers of participants reject key elements of the scenario, the survey clearly suffers from a lack of content validity. Potentially important issues include whether the participant believed the payment vehicle, such as whether they really believed the tax would be a one-time only payment; whether they believed the alternatives would actually cost the amount specified; whether certain changes that were presented as certain to occur were perceived as certain or uncertain; and whether the scenarios were equally plausible. Questions about plausibility would need to be carefully worded. Simply asking if they found all scenarios equally plausible would probably yield lots of "yes" responses in an effort to please the researcher. A better way to ask the question might be, "Were there one or more alternatives that seemed less likely to occur than the others?" This could be followed up questions asking for identification of these alternatives, and asking how much it affected their responses.

Construct Validity

Construct validity centers on the relationship between the measure of interest and other economic variables (referred to as theoretical construct validity) or other measures of the value of the amenity (convergent construct validity). Much of the CV validity literature has focused on testing construct validity. With a little forethought, CA surveys can be structured to gather the types of information that will allow for a variety of tests of construct validity.

Theoretical Construct Validity

Theoretical construct validity is concerned with how well the relationship between the measure and other economic variables conforms to economic theory. To assess theoretical construct validity, the survey needs to be designed to collect the appropriate information. This can involve both gathering socio-demographic information and selecting a set of questions that

allows for testing of various hypotheses. Below I discuss how some of the tests that have been proposed in the CV literature can be applied to CA studies, as well as some additional tests that are possible due to the multiple-question format of CA.

Scope Tests

Theory predicts that up until the point of satiation, larger quantities of a normal good should engender larger WTP estimates. Critics of CV contend that CV estimates are not sensitive to the scope of the amenity (Kahneman 1986, Diamond et al. 1993, Desvousges et al. 1993, Schkade and Payne, 1994). A number of CV studies have included scope tests by having one subsample value a larger change in one or more attributes than a second subsample. The vast majority of these studies show that CV estimates can in fact be sensitive to the scope of the construct being measured (e.g., Boyle, Welsh and Bishop 1993, Carson, Wilks and Imber 1994, Mullarkey 1997). With conjoint studies, scope tests are essentially built into the empirical analysis. If the coefficients of the attributes are significantly different from zero and of the proper sign, then the welfare measure will be affected by changes in attributes as predicted by theory. If the coefficient on an attribute is not statistically significant, but the attribute was identified by focus groups as being important, something is amiss.

Demographic Variables

Relative to CV, it is harder for CA studies to demonstrate that welfare estimates are related to demographic variables in the expected manner. With CV, demographic variables can be included as regressors in bid equations to help explain variation in WTP. Unfortunately, this will not work for CA since these variables are constant across an individual's choices. However, the data can be split into demographic groups (kids vs. no kids, urban vs. rural resident, income above vs. below some amount, etc.), and likelihood ratio tests can be used to check for differences in coefficients.[1]

---

[1] Likelihood ratio tests determine whether the estimated coefficients are equal across samples by comparing value of the log likelihood function (LL) for each the two samples to the value of the log likelihood function for a pooled sample. The likelihood ratio test statistic (LR) is calculated as LR = -2 * [LL(group A) + LL(group B) - LL(pooled)], and compared to the chi-square distribution with degrees of freedom equal to the number of estimated parameters (Judge et al.. 1988).

Adding-up Tests

Economic theory implies that WTP for good A, plus WTP for good B conditional on already having purchased good A, should equal WTP for A and B together. A test of this hypothesis is called an adding-up test. With CV, adding-up tests can be done using split samples. The major difficulty lies in accounting for income and substitution effects. It will be more difficult to conduct adding-up tests with CA. The prospect of asking an individual to evaluate two separate amenities, and then to evaluate them together as part of a CA study, is not promising. Aside from the usual within-sample issue of independence, this task would double the information burden placed on the individual and could be somewhat confusing.[2] Therefore, three subsamples would be needed. Group 1 could be asked about good A, group 2 about good B, and group 3 about A and B combined. Group 3 would obviously face a larger information burden, but it is hard to see how this could be avoided. Given the complexity of the task, amenities that can be described by relatively few attributes would be preferred candidates for this type of test. As with CV, income and substitution effects will need to be accounted for. An adding-up test of this sort would considerably increase the expense of a CA study (and some funding sources may not be interested in paying for this type of methodological experiment), but could be very informative.

Transitivity

With choice experiments, one can test the transitivity assumption. The survey can be designed to ask an individual to choose between alternatives A and B, then between B and C, and finally between A and C. For those that preferred A to B and B to C, transitivity requires them to prefer A to C. The converse is also true; however, if A and C are both preferred to B, then theory offers no prediction on the relationship between A and C. This type of test would be fairly easy to include in a CA study. Inevitably some participants will violate the transitivity assumption (they should be removed from the sample). The larger the percentage of participants that fail this test, the more questionable would be the theoretical construct validity of the survey.

---

[2] Within product adding-up tests are not appropriate since functional form will dictate the results. For example, WTP to improve the first three attributes plus WTP to improve the second three attributes will always equal WTP to improve all six attributes if the common linear additive utility function is specified. Similarly, other specifications will drive the welfare estimates.

Monotonicity or Dominance

A similar test for violations of the monotonicity hypothesis can be incorporated into the design of the alternatives. If one alternative offers equal or greater levels of each (good) attribute at the same cost as a second alternative, the first alternative is said to dominate the second, and should always be preferred by the participant. With ranking or rating exercises, testing this is straightforward. With choice experiments, there are several options available. The simplest way would be to ask the participant to choose between a dominating alternative and the alternative that it dominates. Since this might render the choice too obvious, a less direct test can be devised. Consider three alternatives, where F strictly dominates D, and E is any alternative that neither dominates nor is dominated by D. Ask participants to choose between D and E, and between E and F (the order of questions should not matter). Monotonicity requires those participants who prefer D to E to also prefer F to E (the test is indeterminant for those that prefer E to D). One possible explanation for violations of the monotonicity assumption is that the participant does not place equal probabilities of success on each alternative. If the participant does not believe that the dominant alternative is politically or physically feasible at the cost specified, or feels that it is less likely to be realized than the dominated strategy, she may feel less inclined to choose the dominant strategy.

Ranking Attributes

One simple piece of evidence that can be collected in every CA survey is a non-monetary ranking of attributes. Participants could be asked before the central questions to simply rank the importance of the various attributes. This ranking should match or be very close to the ranking of attributes revealed by the marginal attribute values. A lack of consistency between the two rankings could result from a number of factors, ranging from a lack of content validity to inappropriate econometric analysis. Unfortunately, the small sample sizes found in focus groups may not provide an early warning of this problem, but perhaps the larger sample sizes of pilot studies would allow identification of a problem before the final survey instrument is fielded. Subsequent focus groups could then be conducted to uncover the root of the problem.

Convergent Construct Validity

Convergent construct validity is the degree to which the estimate in question resembles other measures of the same construct. CA estimates can be compared to either CV estimates or revealed preference (RP) estimates. Comparisons of CA to CV include Hanley et al (1998), Boxall et al (1996), Adamowicz et al (1998). A few studies comparing dichotomous-choice CV to CE show that the CV estimates tend to be larger than the CA estimates. Comparisons to RP estimates may be possible when the construct has little or no nonuse value. In fact, TCM and CE share the RUM framework, and can be combined into one data set, as Adamowicz, Louviere and Williams (1994) did. These types of comparisons are helpful and should be encouraged where applicable.

Another test of convergent construct validity would be to use multiple conjoint-type formats to estimate values for the same amenities. Thus a study that compared welfare estimates of an amenity derived from various forms of ratings, rankings, and choice experiments would be one type of convergent validity test. There are a few studies that do this to some degree, including McKenzie (1993) and Roe, Boyle and Teisl (1996). Coupled with research on the types of decision heuristics participants use in each format, this would be a most informative study.

Criterion Validity

Criterion validity considers the relationship between the measure and an alternative measure that is closer to the underlying construct. In order to assess the criterion validity of a CA study, one would therefore need to have an external welfare measure that is unequivocally closer to the "true" total value than the CA estimate.[3] As with CV, it will be quite difficult to assess the criterion validity of CA studies. For the applications we are concerned with, there will seldom be an observable criterion that offers a measure that is unequivocally closer to true total value. This is particularly true for applications dealing with nonuse values.

Some authors may argue that CA studies of market goods can be compared to actual market behavior or to simulated markets as a test of criterion validity. While this is appropriate

---

[3] The existence of a measure that is considered closer to total value does not preclude the use of CA in cases where CA studies would be less expensive. If estimates from a less expensive technique consistently approach the criterion measure of a more expensive technique, it may prove cost effective to adopt the less expensive measure.

for market goods, it is not clear that these comparisons shed any light on the criterion validity of CA applications that deal with nonmarket and nonuse amenities. The contexts are quite different. There is likely to be supply-side uncertainty with environmental amenities, and the alternatives being considered may involve irreversible impacts on the amenities. Respondents may have less experience with the amenity or similar amenities than in market-goods contexts, and they may have different incentives to carefully search their preferences. The penalty, in terms of welfare loss, of making a poor decision for an inexpensive private good is likely to be small, potentially reversible, and the loss is borne solely by the respondent or the respondent's household. With public goods, on the other hand, not only might a poor decision be irreversible, but it potentially affects millions of other lives, human and nonhuman. This is an added responsibility that some civic-minded people take very seriously. Thus while it can be argued that if CA does not work well for private goods it would be unlikely to work well for public goods, it can also be argued that respondents have greater incentives to carefully search their preferences in studies of public goods. Therefore I am not too comfortable with using private goods to learn about how people value public goods, and even less so when the public goods have nonuse value components.

A second avenue for criterion validity assessments is the use of simulated market experiments that deal with nonmarket amenities. Simulated markets do produce observable behavior, which may be considered to lead to better welfare estimates as long as the simulated market is appropriately structured (simulated markets need to meet the same content validity standards as CV or CA studies).[4] However, it is exceedingly difficult for most researchers to have sufficient control over natural resources and payment collection mechanisms to actually construct and utilize an appropriately structured simulated market for a nonmarket amenity (particularly for those that evoke nonuse values). It is also necessary to have a large enough sample to fund the project in order for participants to find the survey realistic.

Simulated markets for CA studies will be slightly different than those used in conjunction with CV surveys. In each case, split samples are used, with the simulated market survey instrument differing from the hypothetical CV or CA instrument only in that participants are told

---

[4] It should be noted that unless the simulated market can be shown to have completely optimal incentive structure (e.g., no incentive to free-ride), it should be viewed as a test of convergent construct validity rather than criterion validity.

that they will actually have to pay for the project being evaluated. Given the structure of CA studies, the data will have to be analyzed before payments are made in order to identify the alternative that provides the largest welfare gain (it does not make much sense to first collect for each individual's preferred alternative and then make refunds or require additional payments from those participants that did not initially select the best alternative). Thus the simulated market participants are told that they will be required to pay for whichever alternative is chosen as the best by the group. Assuming the enforcement mechanism is effective[5], several questions can be asked of the data. Do the CA and simulated market instruments identify the same alternative as the best? Do they provide identical rankings of the marginal attribute values? And are these marginal values statistically different between treatments?

Laboratory experiments, typically using college students, are another form of simulated market. These experiments are useful as a relatively inexpensive method for exploring survey design issues, such as the range of levels or the number of levels to use per attribute. However, they have several features that make them suspect for measuring welfare changes. First, students are not a representative sample of the target population for most environmental projects. Second, a fairly large sample would typically be needed to actually pay for an environmental project of any magnitude. A sample size that is too small (or too large) for the project weakens the plausibility and validity of a survey. Third, it is unlikely that managers of natural resources will allow a handful of students to determine the availability of the type of public amenity that we are interested in.

Conclusions and Future Research

Conjoint Analysis has the potential to be quite useful for economic analyses of natural resource and environmental amenities. The ability to measure marginal values for individual attributes is useful both for determining the design of public projects and for use in benefits transfer studies. CA studies face many of the same challenges as CV studies, as well as some that specific to the structure of CA. To help realize its potential, CA could benefit from the same focus on validity that CV has.

---

[5] If payments are not collected from each individual, the validity of the simulated market results is questionable.

There is much to be explored as we apply CA to the natural resources – environmental context. Uncertainty, irreversibility, nonuse values and the public-goods nature of natural amenities are issues not generally dealt with in the marketing literature. The essence of CA is the tradeoffs between attributes. We need to know more about the decision heuristics participants use to make these tradeoffs. Decision heuristics may well vary between different forms of CA. For example, differences in the variance of levels between the attributes may cause some attributes to receive too much or too little attention in choice experiments, but this may not be a problem for ranking exercises. We also need to know whether participants consider each alternative as equally plausible, and if not, how that affects their decision-making process.

In general, issues related to information burden need to be explored. Given that the information burden can be large, we should explore such design considerations as how many attributes are participants willing to consider, and how many levels for each attribute should be used. It may be that the number of attributes presents a greater cognitive burden than the number of levels per attribute. Increasing the number of levels may therefore increase statistical power without substantially increasing burden, and may also help avoid the potential variance bias discussed above. These types of considerations are critical to developing CA surveys with strong content validity. Content validity needs to receive serious attention, perhaps even more so than with CV since it has been argued that due to greater information burden, CA offers improved flexibility at the potential expense of precision.

One area that is very intriguing is the possibility that CA may produce reasonable estimates of WTA compensation for environmental decrements. Surveys in which some alternatives offer improvements and some offer deteriorations, thereby allowing the participants to evaluate both "goods" and "bads", may reduce the reluctance people have shown in CV studies to explicitly trade the environment for money. Studies that combine WTP and WTA questions should test for status quo effects, as in the Adamowicz et al. (1998) study. As for analysis of the data, be it WTP or WTA or both, sensitivity analysis of the choice of functional form for the indirect utility function would be useful. Most studies to date have relied on linear functions, but the presence of nonuse values makes it unlikely that marginal attribute values are constant.

CA studies can and should be designed to allow for multiple validity tests. The basis of CA is to observe tradeoffs, and thus including both a transitivity test and a monotonicity test may be overkill. But half the sample could receive a transitivity test and the other half a monotonicity test without compromising statistical power. One test that can easily be included in any study is to ask participants to rank the attributes in order of importance. The aggregate ranking and the ranking based on marginal attribute values should be consistent. Where appropriate, comparisons to either CV or RP studies can be useful for establishing convergent construct validity. However, as with CV studies that include nonuse values, there is no good way to test CA studies for criterion validity when nonuse values are present. Hopefully, as more attention is focused on assessing and improving validity, additional validity tests will be identified that are appropriate to the structure of CA.

References

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

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

Bishop, R.C., P.A. Champ and D.J. Mullarkey. 1995. "Contingent Valuation." In *Handbook of Environmental Economics*. Ed. Daniel W. Bromley. Oxford: Blackwell Publishers, Inc. 629-654.

Boxall, P.C., W.L. Adamowicz, J. Swait, M. Williams and J. Louviere. 1996. "A Comparison of Stated Preference Methods for Environmental Valuation." *Ecological Economics*, 18:243-253.

Boyle, K.J., M.P. Welsh and R.C. Bishop. 1993. "The Role of Question Order and Respondent Experience in Contingent-Valuation Studies." *Journal of Environmental Economics and Management*, 25:S80-S99.

Carson, R.T, L. Wilks and D. Imber. 1994. "Valuing the Preservation of Australia's Kakadu Conservation Zone." *Oxford Economic Papers*, 46:727-749.

Desvousges, W.H., F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson and K.N. Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In "*Contingent Valuation: A Critical Assessment*", J.A. Hausman ed. North-Holland, Amsterdam.

Diamond, P.A., J.A. Hausman, F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson, and K.N. Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In "*Contingent Valuation: A Critical Assessment*", J.A. Hausman ed. North-Holland, Amsterdam.

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

Judge, G.C., R.C. Hill, W.E. Griffiths, H. Lutkepohl and T.C Lee. 1988. *Introduction to the Theory and Practice of Econometrics*. John Wiley & Sons.

Kahneman, D. and J.L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management*, 22(1):57-70.

Louviere, J. and H. Timmermans. 1990. "Stated Preference and Choice Models Applied to Recreation Research: A Review." *Leisure Sciences*, 12:9-32.

Mackenzie, J. 1993. "A Comparison of Contingent Preference Models." *American Journal of Agricultural Economics*, 75:593-603.

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

Mullarkey, D.J. 1997. "Contingent Valuation of Wetlands: Testing Sensitivity to Scope." Ph.D. dissertation, University of Wisconsin, Madison.

Roe, B., K.J. Boyle and M.F. Teisl. 1996. "Using Conjoint Analysis to Derive Estimates of Compensating Variation." *Journal of Environmental Economics and Management*, 31:145-159.

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

Schkade, D.A. and J.W. Payne. 1994. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation." *Journal of Environmental Economics and Management*, 26:88-109.