STATED PREFERENCE: WHAT DO WE KNOW? WHERE DO WE GO?

PROCEEDINGS
SESSION TWO

VALIDITY OF STATED PREFERENCE METHODS

A WORKSHOP SPONSORED BY THE US ENVIRONMENTAL PROTECTION AGENCY'S NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS AND NATIONAL CENTER FOR ENVIRONMENTAL RESEARCH

October 12-13, 2000
Doubletree Hotel, Park Terrace
Washington, DC

Edited by Sylvan Environmental Consultants for the
Environmental Law Institute
1616 P Street NW, Washington, D.C. 20036

ACKNOWLEDGEMENTS

Sections of this report, indicated as "summarizations," were prepared by Sylvan Environmental Consultants for the Environmental Law Institute with funding from the National Center for Environmental Economics. ELI wishes to thank Matthew Clark of EPA's Office of Research and Development and Kelly Brown, Julie Hewitt, Nicole Owens and project officer Alan Carlin of National Center for Environmental Economics.

DISCLAIMER

These proceedings are being distributed in the interest of increasing public understanding and knowledge of the issues discussed at the Workshop and have been prepared independently of the Workshop. Although the proceedings have been funded in part by the United States Environmental Protection Agency under Cooperative Agreement CR-826755-01 to the Environmental Law Institute, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Session II Proceedings
Table of Contents

THE VALUE OF VISIBILITY: A COMPARISON OF STATED PREFERENCE METHODS

Thomas H. Stevens
University of Massachusetts

John M. Halstead
University of New Hampshire

Wendy Harper
Wellesley College

Ina Porras
University of New Hampshire

L. Bruce Hill
Clean Air Task Force

Theresa L. Walker
University of New Hampshire

Cleve Willis
University of Massachusetts[1]

Prepared for presentation at the US EPA National Center for Environmental Economics and National Center for Environmental Research Conference, "Stated Preference: What do we know? Where do we go? Washington, D.C. October 12-13, 2000. Funding from EPA/NSF is gratefully acknowledged. All opinions are those of the authors.

The 1977 Clean Air Act requires the U.S. EPA, the states, and federal land managers to protect and restore visibility in wilderness areas (Harper, 2000). However, despite national

[1] Professor, Department of Resource Economics, University of Massachusetts; Professor, Department of Resource Economics and Development, University of New Hampshire; Visiting Assistant Professor, Wellesley College; Former Research Assistant, University of Massachusetts; Senior Scientist, Clean Air Task Force; Research Assistant, 1 Department of Resource Economics and Development, University of New Hampshire; Professor and Dean, College of Food and Natural Resources, University of Massachusetts.

reductions in sulfur dioxide emissions, visibility in most of the northeastern wilderness has declined substantially since the 1970's. As noted by Hill, et al., (2000), human induced smog conditions have become increasingly worse and average visibility in Class 1 airsheds, such as the Great Gulf Wilderness in New Hampshire's White Mountains is now about one-third of natural conditions. Deregulation of electricity production is likely to result in further degradation as consumers switch to low cost fossil fueled generation, and although EPA regional haze rules attempt to address this problem, many policy makers question whether the value of improved visibility is worth the cost.[2]

This paper focuses on several of the problems associated with the valuation of atmospheric visibility in wilderness areas. One problem is that different forms of the stated preference valuation approach, such as contingent valuation and conjoint or choice analysis, may produce very different results (Stevens, et al., 2000). Results may also differ depending on whether valuation is conducted onsite or offsite, or by mail or in person. Also, little is known about the geographical extent of the "market" for visibility; is it a local, regional, or global public good? Finally, do visibility value estimates adequately exclude the value of joint products like health and ecosystem effects associated with atmospheric pollution?

We begin with a brief review of previous studies. A case study of visibility in the Great Gulf Wilderness of New Hampshire is then presented and discussed.

Background and Previous Studies

Most previous studies of the value of visibility have used the contingent valuation method (CVM). One of the first studies was conducted by Rowe, et al. (1980) who found that non-residents were willing to pay about $4 per day to preserve visual range in southwestern Colorado. Schulze et. al. (1983) reported that residents of Los Angeles, Denver, Albuquerque and Chicago were willing to pay $3.75 to $5.14 per month to preserve visibility in the Grand Canyon. Crocker and Shogren (1991) estimated that residents were willing to pay about $3.00 per day to preserve visibility in the Cascades of Washington State. And, Chestnut and Rowe (1990) found that respondents were willing to pay $4.35 per month to avoid a change in average levels of visibility in the Grand Canyon, Yosemite and Shenandoah National Parks.[3]

With respect to wilderness areas in the northeast, the Appalachian Mountain Club (AMC) administered a survey in the summer of 1996 to ascertain visitor's perceptions of visibility in the White Mountain National Forest. This survey was administered to individuals at three sites: The Pinkham Notch visitors' center at the base of Mt. Washington, the Cardigan lodge at the base of Mt. Cardigan and the Mt Washington Observatory (at the top of Mt. Washington). This survey asked respondents to rate photographs of Mt. Jefferson, a mountain in the Class 1 Presidential Dry River airshed, at various visibility conditions. Each photograph was correlated with a measurement of optical extinction measured by a nephelometer at the site where the photograph was taken. Results of this survey show that individuals were able to consistently perceive different levels of visibility. That is, respondents were clearly able to differentiate between improvements and degradations to visibility (Hill, 2000).

[2] The EPA regional haze rules were recently overturned in Federal Court. However, the EPA has appealed and the current administration plans to take the case to the Supreme Court (Harper, 2000).

[3] Many of these studies were modeled after research and ideas developed or presented at a 1982 conference on visual values (Rowe and Chestnut, 1983).

Although much has been learned, results of previous research suggest that several important questions remain unanswered. The first issue refers to the valuation technique used. As noted by Brookshire, et al.(1982), results should be tested by using valuation techniques other than the traditional CVM. Second, relatively little is known about the relationship between the onsite and offsite value of visibility or about the effects of location (distance) on the value of visibility. The results of location analysis might help to resolve two of the major problems in the valuation of environmental assets: the extent of the market area associated with damage assessments and whether benefit estimates derived from one region can be transferred to other areas. And, very few studies have included analysis of the potential problem of joint products which may be a very important factor in estimating the value of visibility itself.

Theoretical Foundations

As noted, most direct valuation studies of nonmarket goods have used the CVM. The theoretical foundations of generating compensating and equivalent variations have been extensively documented (see, e.g. Mitchell and Carson, 1989; Cummings, Brookshire, and Schulze, 1986; Freeman, 1993) and will not be repeated here. There are several venues for eliciting respondents' stated value for the good(s) in question in the CVM, including open ended valuation questions, iterative bidding, and referendum-style. However, given the primary goal of this study--that is, to directly compare conjoint analysis and contingent valuation--it was necessary to use the dichotomous choice format. Adapting from Freeman (1993), if the individual's indirect utility function can be represented as u(M, v, C) where M is income, v is the level of visibility, and C is a vector of individual characteristics, the individual responds yes if

$$(1)\ u(M - E, v^1, C) - U(M, v^0, c) \geq 0$$

If $w(\beta)$ is the observable component of utility, then

$$(2)\ Pr(yes) = Pr[w(M - E, v^1, C) + \varepsilon_1 > w(M, v^0, C) + \varepsilon_0]$$

Where the ε_i $(i = 0, 1)$ are the random, unobservable components of utility. Willingness to pay for the visibility improvement to v^1 is thus defined implicitly by

$$(3)\ u(M - WTP, v^1) = u(M, v^0)$$

and

$$(4)\ w(M - WTP, v^1) + \varepsilon_1 = w(M, v^0) + \varepsilon_0$$

Conjoint analysis, which has its foundations in hedonic theory, has principally been employed in marketing studies (e.g. Green and Srinivasan, 1978 and 1990; Gineo, 1990; Manalo, 1990), and more recently in recreation studies (Mackenzie, 1990) and community development research (Gruidl and Pulver, 1991). Conjoint analysis is designed to decompose a set of overall responses to "stimulus" factors (such as a list of product attributes) so that the utility of each attribute can be inferred from the respondents' overall evaluations of the stimuli (Green and Wind, 1975; Green and Srinivasan, 1978, 1990). The technique allows disaggregation of the relative importance of each facet of a multi-attribute product. Conjoint measurement has the attractive feature of decomposing the respondent's original evaluations into separate and compatible utility

scales by which the original bundle choices can be reconstituted; a relative value of the importance of each attribute can then be derived. By deriving marginal utility estimates for each attribute, a measure of willingness to pay for changes in the level of these attributes can be estimated, which may then be compared to estimates generated using other valuation methods.

Conjoint analysis, which is essentially a modification of the referendum CVM, asks survey respondents to rate or rank alternative programs. Following MacKenzie (1990), assume an environmental protection program, Q, with M attributes:

(5) $Q = q(Q_i,.., Q_m)$

Where Q_i is the quantity of the i^{th} attribute. In conjoint analysis, product price or cost, P_Q, is treated as just a another attribute so that:

(6) $Q = q(Q_i,.., Q_m, P_Q)$

If the utility function is separable, then:

(7) $U = U(Q_i, - P_Q + M)$,

Where M is all other goods (income).

Assume that respondents are asked to rate two alternative programs, B and C, which differ only in terms of price and attribute Q_i. The utility difference between them is given by:

(8) $U^b - U^c = [U^b(Q^b_i + M - P^b)] - [U^c(Q^i + M - P^c)]$.

If $U^b > U^c$, then program B is rated above C. Rating can, therefore, be expressed empirically as:

(9) $Rating = a(P_Q) + B(Q_i) + e$,

Where (9) represents the empirical utility function (MacKenzie, 1993). Since the two attributes in (9) can be varied while leaving the rating (utility) constant, $-B/a$ is the marginal willingness-to-pay for attribute Q_i (MacKenzie, 1990; Magat et al. 1988). If (9) is expressed in terms of rating differences, Roe et al. (1996) and Johnson et al. (1995) show that CV and conjoint results can then be directly compared, as subsequently argued here.

However, recent studies have explored some of the simplifying assumptions of these earlier approaches and have sought to reconcile CA with the general assumptions of utility theory (e.g. Adamowicz, Louviere, and Williams, 1994; Boxall et al. 1996). Roe et al. (1996) points out that most CA studies utilize an ad hoc functional form, noting that "...respondents' ratings of a single commodity do not provide the information necessary to estimate the welfare gains or losses of moving from one commodity to another" (p. 148). Therefore, this study will use the model developed by Roe et al. wherein it is assumed that the individual utility associated with environmental quality programs is expressed as

(10) $U^i (P_i, q_i, M, z)$

Where P_i is the cost or price of program i borne by this individual, q_i is a vector of program attributes, m is income, and z is a vector of individual characteristics such as age and education. Assuming that utility is related to individual ratings via a transformation function $\Phi(\bullet)$:

$$(11)\ r^i (P_i, q_i, M, z) = \Phi ([U_i (P_i, q_i, M, z)]$$

where $r^i (\bullet)$ is the conjoint rating. A change from the status quo visibility (level 0) to changed visibility (level i) is given by the ratings difference, Δr:

$$(12)\ \Delta r = r^i (P_i, q_i, M, z) - r^0 (P_0, q_o, M, z)$$

Assuming a linear, constant marginal utility of income:

$$(13)\ \Delta r = r^i (q_i, z) + a(M - P_i) - [r^0 (q_0, z) + a (M - P_0)]$$

where a is a constant. From (13):

$$(14)\ \Delta r = r^i (q_i, z) - r^0 (q_0, z) + a (P_i - P_0)$$

A binary response model can also be derived from the conjoint formulation presented in (14). Suppose, for example, that respondents are asked to rate alternatives visibility levels, including the status quo, on a scale of 0 - 10 (with 0 being totally unacceptable and 10 being definitely acceptable) indicating the alternative which the individual would definitely choose. As Roe et al. argue, this formulation follows the standard random utility model:

$$(15)\ \Pr(\text{program I chosen}) = \Pr (U^i (P_i, q_i, M, z) + \varepsilon_i > U^j (P_j, q_j, M, z) + \varepsilon_j)$$

Where ε_i and ε_j are random errors. This binary format should yield the same results as dichotomous choice contingent valuation; thus, equations (15) and (2) should theoretically yield the same results.

To account for possible embedding effects, subsamples of the survey group will be confronted with different levels of the environmental commodity (visibility) where possible. Standard statistical analysis will then be used to test for embedding. It is hypothesized that if embedding effects are not present, value estimates will differ significantly as further degradations in visibility are presented.

Methods

A case study of visibility in the Great Gulf Wilderness in New Hampshire was undertaken during the winter, spring and summer, 1999, and during summer of 2000. Visibility at the study area, which is about one quarter mile northeast of the Mt. Washington summit, is commonly impaired by regional haze that is largely a product of fossil fuel energy production (Hill et al., 2000).

Four surveys were used to measure the value of visibility in the Great Gulf Wilderness region. The first survey was administered *onsite* by a trained interviewer who used a personal

computer (laptop) to present respondents with computer modeled images derived from the WinHaze Visual Air Quality Program. This program allowed us to hold weather conditions constant (cloud cover) while changing visibility only.

The second survey was identical in all respects except that it was administered *offsite* to individuals residing in the Northampton/Amherst area in Western Massachusetts (about a 3 to 4 hour drive from the study site). The third survey which was conducted by mail involved a random sample of 1,000 New England residents. The fourth survey was conducted by mail of a random sample of residents of New Hampshire, Vermont, and Maine.

A split sampling approach was employed throughout. In each of the intercept surveys one half of the respondents received a contingent valuation question that asked for their willingness to accept reduced visibility in exchange for lower electricity bills. The other respondents were asked to rate, on a scale of 1 to 10, the status quo and a scenario with less visibility and lower monthly electricity bills. The first mail survey was modeled after the intercept surveys except that it was possible to confront respondents with multiple scenarios of visibility degradation in eliciting WTA measures via the electric bill vehicle. The second mail survey switched elicitation procedures and sought to estimate respondents' willingness to pay (again using a split sample of conjoint and contingent valuation methods) to *avoid* degradation of visibility.

This sampling strategy allows us to test for differences in economic value estimates due to respondent's place of residence, survey type (mail or personal), type of valuation question (contingent valuation or conjoint/choice), and whether the respondent was contacted onsite or offsite.

The first section of the surveys asked respondents to rate several pictures according to the amount of haze in each. Each picture was a view taken from Camp Dodge, directly across from the Great Gulf Wilderness that had been altered by WinHaze to simulate different levels of atmospheric pollution, all else held constant (cloud cover, etc). Respondents to the personal survey were asked to rate 15 pictures while mail survey respondents rated 4 pictures.

The CVM or ratings (conjoint/choice) question was then presented. Following an introductory statement about electricity deregulation and air quality in the White Mountains (see appendix), each respondent viewed two pictures in this section: picture A represented the status quo visibility and electric bill while picture B represented reduced visibility and a lower electric bill. The CVM and conjoint (choice) questions were asked as follows:

1. WTA

Conjoint Analysis:

- How would you rate the situation in photograph A on a scale of 0 to 10 with 0 being totally unacceptable and 10 indicating that you would definitely be willing to accept this level of visibility along with no change in your monthly electric bill?

- How would you rate the situation in photograph B on a scale of 0 to 10 with 0 being totally unacceptable and 10 indicating that you would definitely be willing to accept this level of visibility along a $x decrease in your monthly electric bill?

Contingent Valuation: Would you be willing to accept this new level of visibility (indicated by picture GB) in the White Mountain National Forest if your monthly electric bill were reduced by $x?

2. WTP

Conjoint Analysis: How would you rate the situation in photograph B on a scale of 1 to 10, with 1 being totally unacceptable and 10 indicating that you would definitely be willing to pay $x per month more for electricity to avoid this new level of visibility?

Contingent Valuation: Would you be willing to pay $x per month more for electricity to avoid this new level of visibility (indicated by picture B) in the White Mountain National Forest?

Picture A, which represented the base scenario, or status quo, describes the average visibility level at the site during the summer months. Picture B represented one of four visual range reductions. The electric bill reduction was 20 percent of the respondent=s total monthly bill in the personal survey and one of 1/4th, 1/3rd, or 2 of the monthly bill for the first mail survey respondents,[4] while respondents to the second survey were confronted with bids ranging from $10 to $50 per month (these values were chosen based on the initial year surveys).

A series of follow up questions were asked to obtain information about each respondent's socio-economic characteristics, motives involved in answering the valuation question, and plans, if any, to visit the wilderness area in the future.

Double wave mailings with postcard follow ups were used in each mail survey. Response rates were approximately 36 percent for the WTA survey and 39 percent for the WTP survey. These response rates are disappointingly low, and raise the issue of non-response bias.

Although previous efforts to obtain willingness to accept (WTA) estimates for environmental commodities have generally been unsuccessful (Hanley, et al., 1997), there are three reasons why a WTA format was initially employed in this study. First, from a theoretical perspective, property rights to a clean environment are often assumed to belong to the public, and consequently environmental losses should be evaluated using a WTA measure (Harper, 2000). If as suggested by Kahneman et al. (1990), individuals value losses more highly than gains, willingness to

[4] Twenty percent is the average savings expected from deregulation.

pay estimates could severely understate value. Second, given deregulation of electricity generation, acceptance of an increase in air pollution in exchange for cheaper electricity is, in our view, a very realistic scenario. Third, few, if any, comparisons of WTA derived from the CVM and conjoint or choice techniques have been conducted.[5] The current year's mail survey was motivated when the results of the original surveys were somewhat ambiguous, leading to use of a willingness to pay format to see if responses differed substantially.

Results

Characteristics of individuals responding to each survey are summarized in Table 1. Respondents to the mail survey tended to be older and have more income as compared to personal survey respondents. One reason for this difference is that personal interviews were conducted on randomly selected individuals who were contacted onsite or offsite at libraries and cafes in the college towns of Amherst and Northampton, MA. On the other hand, the mail survey was sent to a randomly selected list of households in the entire New England region. It is important at this juncture to note that none of the samples are representative of the population as a whole, and therefore the results should not be extrapolated beyond the sample itself.

WTA

Results from each survey in terms of the percentage of respondents accepting reduced visibility in exchange for lower monthly electricity bills is shown in Table 2. For the conjoint responses, three alternative criteria were used to define acceptance; scenario B ranked equal to or greater than A, scenario B ranked greater than A, and scenario B rated a 10 (definitely accept), but not equal to A. Table 2 also shows average electricity bill compensation. It is important to note that relatively few respondents were willing to make a tradeoff between electricity bills and reduced visibility and that willingness to accept was quite sensitive to the criteria of acceptance assumed in the conjoint format.

That relatively few respondents were willing to accept a tradeoff between visibility and electricity cost is not surprising. In this study average electricity bill reductions ranged from only $7.41 to $ 29.14 per month. However, it is important to stress that the scenarios presented are thought to be very realistic given projected conditions for electricity deregulation in New England (Harper, 2000).

To test for the effects of valuation technique, respondent's location, and type of survey (mail or personal), the two logit model set forth in Table 3 were specified. All data derived from the surveys were pooled and dummy variables were included to test for the effect of respondent's residence, whether the survey was a choice or CVM format, whether it was conducted on or offsite, and whether by mail or in person. The dependent variable in the first model is defined as those rating scenario B>A in the conjoint format and yes in the CVM. The dependent variable in the second model takes on a value of one if respondents rated B $\geq$A or answered yes to the CVM question.[6] These two definitions of "yes" responses were used because previous research suggests that value estimates often vary widely depending on how respondent uncertainty is incorporated in

[5] Since the conjoint method avoids pricing the environmental commodity directly, we hypothesize that conjoint or choice analysis might be more reliable in WTA applications.

[6] There were not enough observations to model B = 10 respondents.

the analyses (Elkstrand and Loomis, 1997; Alberini, et al., 1997; and Wang, 1997). One potential advantage of the choice format employed in this study is that as compared to the traditional CVM, respondents can express uncertainty directly. However, the criterion that should be used to define "yes" responses in the choice format has not been determined. One line of argument suggests that from a purely conceptual perspective, responses rating $B \geq A$ are consistent with "yes" CVM responses. On the other hand, the experimental literature shows that the CVM is often subject to the so called "hypothetical bias" problem and that this bias is reduced or eliminated in choice formats that only count $B > A$ or even $B = 10$ as yes responses (See Champ, et al., 1997; Cummings and Taylor, 1999).

In any case, the specifications presented in Table 3 are not rigorously grounded in economic theory. Rather, we view these specifications as similar to Meta Analyses in that we are primarily attempting to examine the influence of location and survey method (choice, personal, mail, onsite, offsite, etc.) on whether or not respondents would accept a reduction in visibility in exchange for cheaper electricity.

As shown in Table 3, WTA reduced visibility is expected to increase with compensation and visibility. We also expect that the probability of accepting a visibility reduction will be less for those who plan future visits to the site and for those interviewed onsite.[7] It is also important to note that about one-half of all respondents were interviewed personally, forty eight percent received a choice survey, 31 percent were interviewed onsite, about 8 percent lived in New Hampshire, and more than two-thirds had plans for future visits. And, only about 15 to 20 percent were willing to trade reduced visibility for cheaper power, depending on definition of a "yes" response in the choice format.

Results obtained from the models are presented in Table 4. WTA reduced visibility increases, as expected, with compensation and visibility.[8] However, residents of New Hampshire were more likely to accept reduced visibility, all else held constant. One possible explanation for this is that individuals who are most familiar with the resource being valued (live relatively nearby in New Hampshire) are simply less concerned about visibility. However, respondents planning future visits to the wilderness area were less likely to accept reduced visibility. It is important to note that whether the survey was conducted in person or onsite was not a statistically significant factor. However, conjoint (choice) respondents were less likely to accept reduced visibility in model 1, but not in model 2.

That the CVM and conjoint models can produce different results should not be too surprising. Although few comparisons of these techniques have been published, most previous empirical comparisons suggest substantial differences (see Stevens, et al., 2000). There are several reasons for this. First, when compared with the CVM, many conjoint questions provide more information about substitutes. Second, from a psychological viewpoint, respondents may react differently when choosing among options then they do when making dollar valuations (Irwin, et al., 1993; Brown, 1984). And, as noted above, Alberini, et al., (1997), Wang (1997), Elkstrand and Loomis (1997) ,Champ, et al. (1997) and others have shown that value estimates can vary widely depending on how respondent uncertainty is included in the analysis.

[7] Those onsite presumably have more at stake.

[8] This suggests that the models pass the so called scope test.

In the case study considered here, the CVM and conjoint questions presented respondents with the same set of substitutes, but conjoint responses were counted as "yes" in two different ways; if B>A or if B≥A. And, this difference seems to be responsible for whether the conjoint results are or are not different from CVM results. In other words, the way in which respondent uncertainty is handled appears to be responsible for the disparity between the CVM and conjoint results in this study.

Estimates of the median economic value of visibility derived from the logit models are presented in Table 5. All median values were calculated by:

$$(16) \quad \text{Pr Accept} = \frac{1}{1+e} - (a + b\,\text{Ln Compensation})$$

where 'a' and b are estimated parameters (see Table 4). The estimated visibility values suggest that the average respondent is not willing to make a tradeoff between energy cost and visibility. The average respondent's monthly electricity bill was approximately $70, substantially less than the median WTA estimates of $924 and $1006 per month derived from models 1 and 2, respectively. And, the median value estimates are very sensitive to whether model 1 or model 2 is used, whether the respondent lives in New Hampshire, or does not plan to visit the site.

Another issue concerns what it is that respondents were valuing. Responses to the follow up questions indicate that many individuals were not just valuing visibility; rather, air quality as a whole was valued. Many respondents linked their WTA response to health problems, now or in the future. Visibility per se did not seem to be the main concern in many cases, regardless of the respondent's geographical location. For example, consider the following quotes from the follow up questions:

> "This 'haze' would in fact be potentially dangerous pollution in the form of air born particulates accompanied by large amounts of invisible sulfur dioxide and some heavy metals. This pollution would be spread and/or funneled by the prevailing winds over a large area. It is the long term effect of these pollutants that is unacceptable. The technology exists to significantly reduce this emission".

> "It will increase sickness and allergies"... "With the increase of haze in the air, more health problems will result. Since I live in Vermont, this will affect my personal health."

> "To me visibility *per se* is cosmetic; what truly concerns me is the contents of that air and its long term effect on human existence..."

Other respondents were more concerned about the effects of the increase in pollution on the ecosystem and wilderness. Context is important here, and respondents felt affected by their environmental "responsibility". For example:

> "This condition is unhealthy for the living things. I am willing to pay a little more to protect the environment"... "Only a small amount of haze can have an enormous impact on the forest ecosystem."... "Endangered species/wild animals that depend on the wild will be likely to migrate or disappear".

> "Clean air and clean water are priceless. I do not think that money is the issue at stake. The health and well-being of humans as well as most other animals and plants is dependent upon the quality of the environment in which we live. To put a price on environmental quality and destroy the resources on which we depend is absurd."

> "Preserve these treasured landmarks"... "Preserving the pristine conditions of National parklands should be a national priority. One that does justify cost to consumers"... "Too much haze for a non-city vacation spot."

Some respondents were also concerned about the effect that visibility might have on the tourism, recreation activities and property values in the White Mountains:

> "As a landowner in the White Mountains I wouldn't accept any increase in air pollution."... "If visibility is poor the usual number of tourist do not come to Maine, New Hampshire or Vermont, there the ripple effect will be seen in less revenues for the states, hotels/motels, restaurants, etc."

Finally, some respondents were totally against energy providers using coal, and advocated the use of alternative forms of electricity that provide the same benefits (reduced costs) without increasing pollution. Some respondents did not believe the assumption that the reduction in visibility would only occur at the White Mountains of New Hampshire and were concerned about the effects of the increased pollution in their own area.

> "Why should the level of visibility be less than it is in the picture A?. There shouldn't be any pollution. Alternative renewable energy sources are available now, which would eliminate pollution and be cheaper than fossil fuels to produce. The use of solar energy and its applications to solar thermal electric and solar photovoltaic electricity, wind energy and hydropower could easily replace fossil fuels and nuclear energy. This would result over a period of just a few years in the elimination of pollution globally and actually reduce the cost of electricity."

Since many respondents valued air pollution in general as opposed to visibility only, the valuation results presented in this study are likely to be biased upward. However, in future studies it may be possible to employ a conjoint format that separates the effects of visibility from the problems of air pollution in general.

WTP

The summer 2000 mail WTP survey is still has only recently been completed, and so much analysis remains. However, Tables 6 and 7 present the results of analysis of the data set in isolation. As can be seen, a number of variables had statistically significant coefficients of the expected sign: gender, income, electric bill, natural log of bid, and natural log of visibility were all significant at the five percent level or above. Unlike the previous mail survey, there were no "state" effects. The model had reasonably good predictive power, as Table 7 illustrates that it predicted about 78 percent of responses correctly. Using equation (16) to calculate median WTP values yields an estimate of $511 monthly. This figure does appear quite high at first blush; however, it appears similar to the

WTA estimates. Clearly, further research aggregating this data set into the overall data pool is necessary and will hopefully shed more light on the issue.

Summary and Conclusions

The findings that emerge from this study can be summarized as follows. First, most respondents were not willing to accept cheaper electricity in exchange for reduced visibility over the range examined in this study. In fact, the estimated economic value of visibility suggests that compensation for improved visibility via lower priced electricity is simply not feasible; the necessary compensation is likely to be greater than the average respondent's actual electricity bill. If respondents are well informed, we might therefore infer that deregulation will not result in a substantial increase in pollution as a result of greater household demand for the cheapest source of electricity.

Second, the effects of location appear to be more complex than previously imagined. Respondents living nearby seem to value visibility less than those living further away, all else held constant. Perhaps absence does make the 'heart grow fonder'. On the other hand, valuation did not differ among those interviewed on or off site, yet those planning future visits were much less likely to accept reduced visibility.

The "market area" for visibility at popular unique sites, such as the Grand Canyon and Yosemite is known to be very large. Although much less is known about the extent of the market for less unique wilderness areas, like the Great Gulf in New Hampshire, this study suggests that its market area may also be quite extensive. On the other hand, conclusions about the effects of location are clouded by the finding that many respondents did not believe that air pollution would be limited to the study site.

Third, the CVM and conjoint models can produce very different results. In this study the difference seems to be a result of the criterion used to define a "yes" response in the conjoint format. Twenty percent of conjoint respondents were WTA the tradeoffs presented in this study when a yes response was defined as $B \geq A$; 9 percent of conjoint respondents were WTA if the criteria is $B > A$; and only about three percent indicated that they would definitely accept ($B = 10$ and $B \neq A$). We therefore believe that future studies should include tests for sensitivity to the valuation question format and to respondent uncertainty.

As noted above, we believe that conjoint (choice) responses rating $B \geq A$ are conceptually consistent with "yes" responses in the traditional CVM. And, our empirical estimates suggest no difference between conjoint and CVM in this case (see Tables 4 and 5). However, the problem of hypothetical bias suggests that "yes" responses should be defined by $B > A$, and when this was done, the resulting value estimates derived from the conjoint format were much different from those derived from the CVM (see Table 5).

Fourth, we did not find differences associated with whether the valuation question was conducted by mail or in person. Perhaps the NOAA guidelines requiring personal interviews should continue to be reevaluated.

Finally, despite survey pre tests and careful wording of the valuation question, many respondents valued air pollution in general. Consequently, the value of visibility is likely to be overestimated. A conjoint analysis that includes several attributes of pollution, including visibility, might clarify this issue, but the problem of sensitivity of this method to the definition of "yes" responses is likely to remain an issue.

Table 1. Socioeconomic Characteristics of Respondents; Sample Means[a]

Survey	Planned Future Visits (%)	Age (years)	Income (thousands)	Residence (%) MA	CT	NH
CJ, Personal Onsite (WTA)	98 (13)	38.8 (13.3)	52.2 (34.7)	20 (40)	07 (25)	12 (32)
CJ, Personal, Offsite (WTA)	60 (49)	36.7 (12.8)	43.8 (36.2)	100 (-)	-	-
CJ, Mail, Offsite (WTA)	71 (45)	48.9 (15.7)	67.2 (39.2)	84 (37)	05 (21)	06 (23)
CVM, Personal, Onsite (WTA)	95 (21)	38.3 (14.6)	38.9 (34.2)	27 (45)	07 (26)	16 (37)
CVM, Mail, Offsite (WTA)	62 (49)	48.0 (13.9)	52.3 (33.9)	16 (37)	39 (49)	11 (31)
CVM, Personal, Offsite (WTA)	54 (50)	31.7 (10.3)	22.4 (20.9)	100 (-)	-	-
CVM, Mail, Offsite (WTP)	67 (47)	51.2 (15.4)	55.1 (29.6)	33	15	52

[a] Standard deviations in parentheses.

Table 2. Summary of Survey Results[a]

Survey	Sample Size	Average Monthly Electric Bill Reduction ($)	Percent Accepting Visibility Reduction		
			B≥A or yes	B>A or yes	B=10[b]
Conjoint, Personal Onsite	60	$13.75 (7.24)	20 (40)	13 (34)	0 (-)
Conjoint, Personal Offsite	60	$12.23 (10.27)	25 (44)	12 (32)	5 (-)
Conjoint, Mail Offsite	105	$25.73 (15.95)	18 (39)	6 (23)	1 (-)
CVM, Personal Onsite	87	$7.41 (2.51)	17 (38)	17 (38)	-
CVM, Mail Offsite	102	$29.14 (25.86)	23 (42)	23 (42)	-
CVM, Personal Offsite	59	$11.35 (6.38)	20 (41)	20 (41)	-

[a] Standard deviation in parentheses.
[b] And B≠A.

Table 3. Logit Model Specification

Variable	Definition	Mean	Standard Deviation	Expected Sign
Model 1 Dependent	Rating B>A or yes to CVM	.15	.36	
Model 2 Dependent	Rating B≥A or yes to CVM	.20	.40	
Ln Compensation	Ln $ monthly electric bill reduction	2.6	.74	+
Ln Visibility	Ln miles	2.78	.56	+
Age	years	41.8	15.1	+/-
Income	thousands	48.3	36.8	+/-
MA	Dummy; Massachusetts resident = 1	.55	.50	+/-
CT	Dummy; Connecticut resident = 1	.12	.32	+/-
NH	Dummy; New Hampshire resident = 1	.08	.27	+/-
PER	Dummy; Personal Interview = 1	.56	.50	+/-
Onsite	Dummy; Onsite Interview = 1	.31	.46	-
FVisit	Dummy; Plans for future visit = 1	.73	.44	-
CJ	Dummy; Conjoint (Choice Survey) = 1	.48	.50	+/-

Table 4. Logit Model Results

Variable	Model 1 (B>A) Parameter Estimates	Standard Error	Model 2 (B≥A) Parameter Estimates	Standard Error
Intercept	-5.01***	1.33	-3.81***	1.09
Ln Compensation	.47*	.25	.35*	.21
Ln Visibility	.67**	.30	.44*	.24
Age	.008	.01	-.0004	.009
Income	.003	.004	.002	.004
MA	.48	.43	.57	.39
CT	.53	.50	.76	.47
NH	1 .10**	.49	1 .30***	.46
PER	.34	.47	.24	.39
Onsite	.54	.51	.23	.43
FVisit	-1.15***	.32	-.87***	.28
CJ	-.99***	.32	.01	.26
N	472		472	
F	38.79***		26.78***	
Percent correct predictions	72.6		65.5	

*** Significant at .01 percent level; ** Significant at .05 percent level; * Significant at .10 percent level.

Table 5. Visibility Value Estimates: Median WTA Per Month[a]

	Model 1 (B>A)	Model 2 (B≥A)
I. Average Respondent	$924	$1006
II. New Hampshire Resident; no visits planned	$36	$17
III. Average resident; No visits planned	$154	$162
IV. Average Respondent Conjoint model	$2790	C[b]
V. Average Respondent CVM Model	$331	C[b]
VI. Average Respondent CVM Model (WTP)	$511	

[a] Values rounded to nearest dollar.
[b] Conjoint dummy variable not different from zero (see Table 4).

Table 6. Coefficient Estimates, WTP to Avoid Degraded Visibility		
Variable	**Coefficient (Standard Error)**	**|P[|Z|>z]**
Constant	-0.9760 (1.6800)	0.5636
Gender	0.9326 (0.4886)	0.0563
Education	0.1408 (0.3118)	0.6516
Age	-0.0044 (0.0172)	0.7965
Income	0.0001 (0.0000)	0.0197
Electric Bill	0.0104 (0.5243)	0.0469
Ln Bid	-1.1742 (0.3882)	0.0025
Ln Visibility	1.3738 (0.4685)	0.0034
Maine	0.2581 (0.5043)	0.6087
Vermont	0.9035 (0.6055)	0.1357
Future Use	0.3035 (0.5337)	0.5696

n = 139

Log likelihood function: -69.87915

Chi-squared: 41.84588

Table 7. Actual vs. Predicted Values, WTP to Avoid Degraded Visibility				
		Predicted		
		0	**1**	**Total**
Actual	0	81	8	89
	1	23	27	50
	Total	104	35	139

References

Alberini, A., Boyle, K., Welsh, M. 1997. "Using Multiple-Bounded Questions to Incorporate Preference Uncertainty in Non-Market Valuation." In: Englin, J. (Compiler), Tenth Interim Report, W-133 Benefits and Costs Transfer in Natural Resource Planning, University of Reno, Reno, NV.

Brookshire, Thayer, Schulze and D=Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," *American Economic Review* Vol 72:165-177.

Brown, T. 1984. "The Concept of Value in Resource Allocation," *Land Economics* 60, 231-246.

Champ, P., Bishop, R., Brown, T., McCollum, D. 1997. "Using Donation Mechanisms to Value Nonuse Benefits From Public Goods". *J. Environ. Econom. Manag.* 32(2), 151-162.

Chestnut, L. and R. Rowe. 1990. "Economic Valuation of Changes in Visibility: A State of the Science Assessment for NAPAP." In Acidic Deposition: State of Science and Technology, Report 27, Washington, D.C. NAPAP.

Crocker, Thomas and Jason Shogren. 1991. "Ex-Ante Valuation of Atmospheric Visibility." *Applied Economics*. Vol. 23, 143-151.

Cummings, R. G. and L. O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review*. Vol. 89(3), 649-65.

Elkstrand, E., Loomis, J. 1997. "Estimated Willingness to Pay For Protecting Critical Habitat For Threatened and Endangered Fish With Respondent Uncertainty." In: Englin, J. (Compiler), Tenth Interim Report, W-133 Benefits and Costs Transfer in Natural Resource Planning, University of Reno, Reno, NV.

Harper, W. 2000. "A Comparison of Direct Methods for Valuing Environmental Amenities: A Case Study of the White Mountain National Forest." (Ph.D. dissertation, University of New Hampshire).

Hill, Bruce, W. Harper, J. Halstead, T. H. Stevens, Ina Porras, and K. D. Kimball. 2000. "Visitor Perceptions and Valuation of Visibility in the Great Gulf Wilderness, New Hampshire." in Cole, et al. Proceedings: Wilderness Science in a Time of Change. Proc., RMRS-P-000. Ogden, VT: U.S.D.A. Forest Service, Rocky Mountain Research Station.

Irwin, J. R., Slovic, P., Licktenstein, S., McClelland, G. 1993. "Preference Reversals and the Measurement of Environmental Values." *J. Risk Uncertain*. 6, 5-18.

Kahneman, D., J. Knetsch and R. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*. Vol. 98, 1325-1347.

Porras, Ina T. 1999. "Off-Site Value of Visibility: New Hampshire's White Mountain Case Study." (M.S. thesis, University of Massachusetts, Amherst).

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

Rowe, Robert, Ralph D'Arge and David Brookshire. 1980. "An Experiment on the Economic Value of Visibility." *Journal of Environmental Economics and Management*. Vol 7, 1-19.

Rowe, Robert and Lauraine G. Chestnut, editors. 1983. Managing Air quality and Scenic Resources at National Parks and Wilderness Areas. Westview Press: Boulder, CO.

Schulze, William D., David Brookshire, et. al. 1983. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest." *Natural Resource Journal*. Vol. 23, 149-165.

Stevens, T. H., R. Belkner, D. Dennis, D. Kittredge, and C. Willis. 2000. "Comparison of Contingent Valuation and Conjoint Analysis in Ecosystem Management." *Ecological Economics*. 32, 63-74.

Wang, H. 1997. "Treatment of Don't Know? Responses in Contingent Valuation Survey: A Random Valuation Model." *J. Environ. Econom. Manag.* 32, 219-232.

APPENDIX

Introduction to Valuation Questions

For the next question, consider the following: Currently, many states are debating the issue of deregulation in the electric utility industry. If deregulation occurs in your state, you may be able to choose your own power provider. Assume for the purposes of this question that cheaper power (that is, less than what you currently pay) is available through a mid-western power company. Further, this power company produces electricity by burning coal. Increased demand for this company's cheaper power will contribute to air pollution and poor visibility in the White Mountains.

Now suppose picture A represents the level of visibility most often experienced in this region during the summer months. Further suppose that you were faced with a situation where the visibility level would change to that in picture B. The purposes of this question assume that visibility would change ONLY in the White Mountain National Forest.

COMPENSATING FOR PUBLIC HARMS: WHY PUBLIC GOODS ARE PREFERRED TO MONEY

Presented by Carol Mansfield, Research Triangle Institute
Co-authored with George Van Houtven, Research Triangle Institute and
Joel Huber, Duke University

Summarization

Dr. Mansfield presented work based on two studies of willingness to accept (WTA) and difficulties in measuring WTA.[1]

Studies attempting to measure WTA often carry indications casting doubt on the results' reliability, such as high protest rates or very high stated WTA amounts. However, being able to measure WTA is important from a practical policy perspective.

In earlier work, Prof. Huber and others had found that surveys involving risk/risk trade-offs produced more stable results than surveys involving risk/dollar trade-offs.[2] That led Dr. Mansfield and her fellow researchers to wonder if surveys offering public goods to offset public bads would yield more valid results than surveys offering dollars to offset public bads.

Some studies in the literature suggest that people may have a negative reaction to cash compensation. Two articles by Frey and co-authors looked at the siting of a low-level nuclear waste dump in Switzerland.[3] The authors found that when residents were asked simply, would you accept the disposal site, 50 percent said yes. When residents were asked, would you accept the site if you received a large cash payment in compensation, only a quarter said yes. The authors concluded that offering compensation crowds out altruistic motivations to support public projects.

Other studies have looked at people who qualified for welfare but refused to accept it.[4] They theorized that there is a social stigma associated with accepting cash compensation.

---

[1] Two papers provide more detail on the results discussed in the presentation: Mansfield, C., G. Van Houtven and J. Huber (2000), "Compensating for Public Harms: Why Public Goods are Preferred to Money," RTI working paper, 2000; and Mansfield, C., G. Van Houtven and J. Huber (forthcoming) "The Efficiency of Political Mechanisms for Siting Nuisance Facilities: Are Opponents More Likely to Participate than Opponents?" Journal of Real Estate Finance and Economics.

[2] Viscusi, W.K., W.A. Magat, and J. Huber (1991), "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-Offs for Chronic Bronchitis," Journal of Environmental Economics and Management, vol. 21, pp. 32-51.

[3] Frey, B.S. and F. Oberholzer-Gee (1997), "The Cost of Price Incentives: An Empirical Analysis of Crowding Out," American Economic Review, vol. 87 (4), pp. 746-755.
Frey, B.S., F. Oberholzer-Gee, and R. Eichenberger (1996), "The Old Lady Visits Your Backyard: A Tale of Morals and Markets," Journal of Political Economy, vol. 104 (6), pp. 1297-1313.

[4] Moffitt, R. (1993), "An Economic Model of the Welfare Stigma," The American Economic Review, December 1993, pp. 1023-1035. Keane, M.P. (1995), "A New Idea for Welfare Reform," Federal Reserve Bank of Minneapolis Quarterly Review, Spring 1995, pp. 2-28.

Possibly too, people find it easier to balance like against like – that is, people find it easier to balance public bad against public good than public bad against private compensation.

Still another study theorized that by accepting compensation, people were also mentally accepting moral responsibility for any future harm from the public bad.[5]

But why would people prefer public goods? Isn't a public good a bribe in the same way as direct cash compensation?

In general, people do not seem to perceive public goods as bribes in the same way as cash. Accepting public goods allows people to feel that they are doing something good for the community. Cognitively, public goods are easier for people to balance against public bads, and public goods may in some way mitigate the harm done by public bads better than cash alone could. For example, a public good could provide a long-term offset to the harm to a neighborhood's reputation caused by a public bad in a way that cash payments never could.

Mansfield and her fellow researchers explored two hypotheses. First, they hypothesized that the existence of a public bad in a neighborhood increases the utility from public goods relative to cash. For example, say you offered respondents the choice between two houses, one with a park nearby and the other with lower taxes. Then you offered a similar choice, except that both houses offered were near a public bad such as a landfill or airport. They hypothesized that with the public bad in the mix, more people would choose the public good (the park) as a counterbalance.

Second, they hypothesized that when you move from a market framework to a WTA or compensation-for-a-public-bad framework, the value of public goods as compensation will increase relative to cash.

In the study, the researchers first asked respondents some neutral market choice questions. They asked respondents to choose between two houses — one enjoying a nearby public good, such as a park, and the other with lower taxes. In some cases researchers also told respondents about a public bad that affected both houses, and in other cases they did not mention the bads.

They also asked some traditional WTA questions. In some they offered cash compensation and in others they offered a public good as compensation.

The researchers tested a variety of public bads and gave respondents a significant amount of detail about the unpleasant aspects of each scenario. In general, the scenarios posed noise, odor, or aesthetic problems, not health or safety risks. Some of the public bads had local effects, involving only the closest houses, and some had broader effects. Similarly, the public goods in the study varied from having local to more general effects.

They collected three data sets. The first, a pre-test, was from a mail-back paper survey offered to church members. The second and third were from self-administered computer surveys, one conducted in a mall in Greensboro, North Carolina, and the other at five different malls in North Carolina, Florida, New Jersey, and New York.

---

[5] Boyce, R.R., T.C. Brown, G.H. McClelland, G.L. Peterson, and W.D. Schulze (1992), "An Experimental Examination of Intrinsic Values as a Cause of the WTA-WTP Disparity," The American Economic Review, vol. 82, pp. 1366-1373.

In terms of demographics, the church respondents were older, wealthier, more likely to be white, and more likely to be retired than the other two samples. The Greensboro mall in the second data set had low income demographics, and for the third data set the researchers deliberately sought out malls with higher income demographics.

The researchers first compared the percentage of respondents who chose the house with the nearby public good over the house with lower taxes in the neutral market choice questions. In all scenarios except one, telling people about the presence of a public bad made them less likely to choose the house with the lower taxes over the house near the public good. Taken individually, not every scenario showed significant differences at the 10 percent level between the public-bad and no-public-bad responses. However, when the researchers pooled the data from all scenarios and ran a random effects logit, the results showed an significant, increased preference for the public good over lower taxes (cash) when the public bad was described in the neutral market choices between two houses. The basic result seemed to be that if you lived in a neighborhood with a public bad, you wanted to live near a public good.

In the same surveys, the researchers also framed questions in terms of compensation, asking what people were willing to accept in return for allowing a public bad to be sited in their neighborhood. In the church-member survey and the multiple-mall survey, more people accepted the public good offered as compensation than accepted cash. In the Greensboro mall survey, more people accepted cash. The sample sizes were small. Pooled together in a fixed-effects logit model, more people accepted the public good.

This alone proves nothing, except perhaps that the researchers picked public goods that people preferred, or that they offered too little cash. The analysis did not speak to what the researchers wanted to know: whether the relative attractiveness of cash dropped when they moved from the neutral market choice to the compensation framework.

To investigate the possibility that the relative value of cash and the public good changed as respondents moved from a neutral market choice to a compensation or WTA format, the researchers constructed log odds ratios calculated from the two WTA responses (WTA cash and WTA the public good). These odds ratios were used to derive predictions of how people would respond to the neutral market choice question offering a public good or cash in the presence of a public bad. In every case but one, the estimated preference for cash derived from the WTA questions underpredicted the number who actually chose cash in the neutral market choice questions. This suggests that the value of cash declines relative to the marginal value of extra public goods in a compensation framework.

Summing up these findings, Dr. Mansfield observed that the presence of a public bad in a neighborhood increases the value of public goods over monetary incentives, even in the market choice scenarios designed to minimize guilt or other intrinsic motivations for preferring the public good to cash. A compensation framework further increases the value of public goods relative to cash.

The researchers held a focus group between the first and second surveys, and some of the comments of the participants shed light on the thinking involved in these kinds of choices. Regarding cash as compensation, one participant said, “It won’t help if everyone just gets money.” Another said, “The government shouldn’t be able to just pay people to do whatever they want.”

Regarding a public good as compensation, one person said, "I think it would be good for the whole neighborhood." Another said that the public goods would "help make the neighborhood nicer, help address the problem."

Regarding the difference between the compensation framework and the neutral market framework, one participant said, "This [market] choice is easier to make, it's not an ethical dilemma."

Addressing the policy implications, Dr. Mansfield suggested that it might be easier to site nuisance facilities if you offer public goods as compensation. In case studies of actual sitings, public goods are often used as compensation, and Dr. Mansfield could find no examples involving only cash as compensation. Compensation is important to the siting process, and economic insights can help make it more likely to be accepted. People who supply the public bad and bear the burden of compensation should be interested to know that they may be able to save money by supplying a mitigating public good rather than cash.

On a related note, Dr. Mansfield made some observations on the characteristics of people who tend to say "no" to WTA questions. She discussed a study by Hamilton on the expansion of hazardous waste facilities.[6] Hamilton found that in counties with higher voter turnout, existing facilities were less likely to expand. He argued that voter turnout was a proxy for community activism.

In Mansfield's study, the researchers noticed that the demographic characteristics of the people who refused any compensation seemed to match the characteristics of people who tend to vote or engage in collective action. A formal analysis of the data tended to confirm this insight. This fits well with the general observation that it is hard to move sitings of nuisance facilities through the political process in politically engaged communities, even if only a minority of residents oppose the facility.

Dr. Mansfield observed that the percentage of people willing to accept compensation in her studies was rather high compared to other studies. She attributed this to the nature of the public bads presented. People are probably more likely to accept compensation for declines in odor, noise, or aesthetics than in health or safety.

The study highlights the importance of context. Values obtained in one setting may not translate to another. In particular, if a policy choice involves compensation, researchers should frame the issue in a compensation context when determining WTA values. Values of public goods should be seen as a lower bound on WTA.

Future research needs to shed more light on why the presence of the public bad matters. Also, research could further explore how the nature of public goods and public bads affects people's choices. Finally, further research could explore this work's implications in the environmental justice field.

---

[6] Hamilton, James T. (1993). "Politics and Social Costs: Estimating the Impact of Collective Action on Hazardous Waste Facilities," Rand Journal of Economics, vol. 24 (1), pp. 101-125.

PROGRESS TOWARD COMPARING STATED PREFERENCE AND REAL MONEY CONTINGENT VALUES

Stephen K. Swallow
Michael A. Spencer
Laurienne Whinstanley

Department of Environmental
and
Natural Resource Economics

University of Rhode Island
Kingston, RI 02881

Discussion paper prepared for the U.S. EPA Workshop on stated preferences research, Washington, DC., October 12-13, 2000.

This paper discusses work in progress. Any results presented are preliminary, and may change on further analysis. Research is supported by NSF/EPA Science to Achieve Results, Partnership for Environmental Research grant #R825307-01-0 and the University of Rhode Island Agricultural Experiment Station. All opinions and errors are the responsibility of the authors, and do not reflect the opinions of any of the funding agencies. We are deeply indebted to Donald A. Anderson, President of StatDesign, Inc., Evergreen, CO, who provided a fractional factorial design for wetland survey questions discussed in this paper.

Environmental goods affect the general welfare as public goods, by which many individuals can benefit from their conservation without necessarily paying for the costs of conservation. As is well accepted by economic principles, private market transactions, or the free market, will generally under-provide such public goods because the potential suppliers often lack any free-market means to earn compensation for their provision. Thus, wetlands and wildlife habitats have frequently been targets for development or conversion to other uses because markets reward landowners directly for investments made in agricultural, residential, or commercial land uses. For several decades now, economists have been working on methods to identify the value of such public goods in a manner that is comparable to the value of private goods sold in markets.

For both market and non-market (public) goods, the concept of economic value depends upon what an individual is willing to give up, or sacrifice, in order to obtain the good in question. Often this willingness to sacrifice is measured in terms of money, a convention on which economists often focus because individuals, the public, corporations, and politicians are believed able to translate, subjectively, money-measures of value into terms of personal relevance, such as an equivalent value of groceries, entertainment, or commercial value. For market goods, market prices represent the value that the marginal buyer or seller places on a good, so that the value of market goods can be estimated based on observations of behavior by individuals in actual market transactions. For some public goods, individual behavior in markets can also be used to estimate at least some portion of the economic value of the good, such as when home buyers are willing to pay a premium for residential properties located in aesthetically pleasing environments, or such as when outdoor recreationists purchase travel services in order to access environmental resources of wildlife refuges or parks.

However, individuals do benefit from public environmental goods in ways that are independent of their decisions in existing markets. In order to estimate the economic value, in monetary terms, of these services from public (environmental) goods, economists have proposed the contingent valuation method (CVM). CVM is thoroughly described in Mitchell and Carson's (1989) book. For the present purpose, it is a valuation methodology by which economists use a survey of the public to ask individual respondents whether or not they would be willing to pay a specified dollar amount to obtain, maintain, or conserve a particular environmental good, such as an undeveloped wetland acre or a particular level of water quality in a local river.

While many environmental economists have accepted the CV methodology, others remain skeptical concerning the validity of money measures of value derived from hypothetical statements, or from the stated preferences of survey respondents. This skepticism has bred controversy as the CVM has been applied to successively larger policy questions or to determine liability for large-scale damages to public trust resources caused by commercial accidents. In response, economists are providing research to identify whether estimates of public good values under CVM match estimates that would be obtained if survey respondents had faced the actual, rather than simply the hypothetical, consequences of their stated preferences, stated choices, or stated willingness to pay. In particular, researchers are concerned with identifying whether CVM-estimated willingness to pay departs from an individual's actual willingness to pay for a public good under conditions that provide her with incentives to be truthful (and realistic) about her true values (incentive-compatible valuation).

This paper provides an overview of some economic experiments or experimental surveys designed to look for and to examine any departures between willingness to pay values based on stated preference (CVM) and values based on surveys in which respondents were required to provide cash or a check if they stated they were actually willing to pay a particular amount for a public good. Two main groups of experiments are discussed, and partial results from some of these are provided.

One group of experiments follows from Spencer et al. (1998) who focused on estimating the value of a monitoring water quality at different ponds in Rhode Island. The new experiments discussed here examine features of the payment mechanism in relation to willingness to pay for adding a pond-site to the statewide water quality monitoring program. These experiments were intended to help identify the most appropriate way to present the payment mechanism in surveys involving real money.

The second group of experiments involves the valuation of wetlands in northwestern Rhode Island. In particular, these experiments focus on estimating the value of purchasing the development rights for wetlands with various attributes, such as the wetland's location relative to developed and undeveloped lands, an indicator of the potential role of the wetland in wildlife or biodiversity conservation, the availability of public access, and acreage. Of course, in Rhode Island, as in many states, wetland conservation laws are fairly strict, often severely restricting development. However, state law does provide a process by which landowners may apply for permits to allow the modification of wetlands, most often as a part of accessing or developing adjacent upland parcels considered suitable for development. Survey respondents were asked about their willingness to pay for a binding "conservation contract" or "easement" on different parcels of wetlands, where this agreement would be placed as a deed restriction on a real wetland parcel, but where the deed restriction would expire after 10 years. This group of experiments is intended to mimic actual land conservation programs wherein public or private conservation agencies purchase the development rights from landowners who continue to hold the title to the land. In these experiments, the survey question was formatted as a choice experiment or contingent choice (Opaluch et al. 1993; Adamowicz et al. 1998), wherein respondents viewed descriptions of two wetland parcels and stated their preference to pay and preserve one parcel or to forgo conservation on both parcels. The real money surveys also requested actual payment.

I. Background Literature and Implications

Previous research illustrates the potential that contingent valuation may produce willingness to pay (WTP) estimates that differ from WTP estimates produced when respondents must actually make a payment. In order to mitigate concerns about incentive-compatibility, a series of experiments applied contingent valuation to estimating the willingness to pay for traditional market (non-public) goods, such as chocolate bars and coffee mugs (e.g., Cummings et al. 1995; Cummings et al. 1997; Fox et al. 1995; Neill et al. 1994). The results generally indicate that hypothetical valuation exceeds valuation based on questions requiring real payments. These results have led to work in designing various methods to adjust or calibrate hypothetical values to real values, using, for example, experimental economics or statistical functions as complements to traditional CVM (Blackburn et al. 1994; Champ et al. 1997; Shogren 1993; Swallow 1994). Other research involves

developing hypothetical surveys that produce value estimates closer to estimates based on real payments (e.g., Cummings and Taylor 1997).[1]

However, there remain questions involving the comparability of results based on market (private or non-public) goods to those based on public goods, because, for example, with market goods an individual's stated value may reflect her prior expectation of market price but she may not actually be in the market for the particular good. Findings based on public (environmental) goods have also found departures between stated-preference values and values based on actual payments for public goods (Brown et al. 1996; Seip and Strand 1992).[2] Yet these results are not entirely conclusive because of the difficulty in eliminating incentive-incompatibility in the respondents' answer to real-money questions about public goods. Indeed, it was the concern about incentive-compatibility that led Brown et al. (1996) to focus on donations in their evaluation of question formats, explicitly conditioning their contribution to contingent valuation research on the assumption that, if the donation vehicle for payment induces a bias in real payments, any real-money bias would be similar to the bias with hypothetical payments.

Unfortunately, it appears that characteristics of the payment vehicle can induce differential effects on valuation of a particular public good or environmental attribute. Johnston et al. (1999) use a hypothetical survey to demonstrate that the relative value of environmental attributes can be conditional on attributes of the payment vehicle, or the means of collecting money, even though respondents may treat a marginal dollar equally under different payment vehicles. Their result is consistent with the substitutability of environmental policy attributes, as elucidated by Hoehn and Randall (1989) and Hoehn (1991), so that there may be a neoclassical explanation for differences between stated-preference (hypothetical) and real-money WTP estimates since, by definition, these payment vehicles (hypothetical and real) have different attributes. Moreover, Johnston et al. (1999) show that their result may not obtain for all environmental attributes, even within the same policy context (watershed management). Therefore, we expect the implied challenge to studies of the validity of CVM to be unpredictable.

Nonetheless, research into the validity of CVM is an important topic, due to the dependence of large-scale public policies on CVM-estimated WTP. Experimental economists, such as Shogren (1993), suggest using laboratory markets to develop improved contingent valuation approaches (see also, Blackburn et al. 1994; Shogren et al. 1994).

For example, experimental economists evaluate alternative mechanisms to reduce incentives for individuals to "free ride" on others' financial contributions to the provision of public goods (Marks and Croson 1998). These mechanisms include establishing a threshold of aggregate payments that are necessary to provide the particular public good at all, establishing a money back guarantee for contributors in the event the public good is not provided, and defining an means to dispose of contributions in excess of the threshold required. Excess contributions may be disposed of through rebates to the original donors or through provision of additional units of a public good, which creates "extended benefits" to contributors or to all individuals. Within these mechanisms, Marks and Croson (1998) hypothesize that the threshold provides an incentive to contribute to the

[1] Greg Poe and colleagues have evaluated a survey method wherein respondents indicate how certain they are that they would pay a specific amount, and using a multiple- bounded method of valuation.

[2] Of course, Bishop and Heberlein's (1979) classic study in the *American Journal of Agricultural Economics* lies in this literature as well.

public good, the money-back guarantee eliminates the risk of incurring costs when others fail to contribute to the threshold, and the disposition of excess contributions reduces the cost to an individual who over contributes. Rondeau et al. (1999) assess these mechanisms in a laboratory experiment to estimate real-WTP when researchers control the value of the public good to each participant, finding that the estimated WTP of the average experimental participant is a satisfactory predictor of the average value of the public good. Spencer et al. (1998) used a money back guarantee and a threshold for aggregate contributions (a provision point) in estimating the value of adding a site to a statewide set of water quality monitoring sites in Rhode Island, finding that stated-preference estimates of WTP exceeded real-money estimates by a ratio of 4.7 to one, although the difference was not statistically significant.

This literature contributed several considerations to the experiments described below. First, following the laboratory results, we designed a set of experiments to evaluate the implications of alternative mechanisms to reduce the tendency of individuals to free-ride on the contributions of others. In this case, we anticipate that a money-back guarantee coupled with a threshold provision point for an environmental good and with a rebate of excess contributions will reduce the divergence between hypothetical-money and real-money contingent valuation estimates.

Second, we attempt to control for the effects of changes in the attributes of the payment mechanism. In most cases, this implies that the stated preference survey and the real-money survey instruments are quite similar, except for the presence or absence of a requirement to pay actual dollars. That is, in most cases the presentation of a stated-preference (hypothetical) survey includes a description of the money-back guarantee, the provision point, and the rebate or other distribution of excess contributions.

Third, the wetlands experiments are designed to allow an evaluation of effects on the estimated WTP for several attributes of the wetland parcels. This design attempts to anticipate that attributes of the payment vehicle, possibly including the mere definition of hypothetical and real payment mechanisms, may induce a different degree of divergence between hypothetical and real-money WTP estimates for different attributes of the wetlands parcel, the public good of interest.

II. Experiments for Mechanisms to Reduce Incentives to Free Ride: Valuing Water Quality Monitoring

A set of experiments was designed around an "adopt-a-pond," water quality monitoring program similar to that reported in Spencer et al. (1998). This application, or environmental good, was defined in collaboration with the University of Rhode Island Watershed Watch program. As described below, the experimental surveys asked student college students about their willingness to pay to have a water quality monitoring schedule established at a pond that was not previously part of the URI Watershed Watch program. This experiment was designed to examine the effects of the method of payment on a respondent's willingness to pay in a discrete choice (yes-no) format, for either hypothetical or real payments to support water quality monitoring at a pond described in the survey.

Payment Formats

Three payment formats were used, as motivated by the experimental economics literature. In all three formats, respondents were informed that the pond would be monitored if the group of

respondents they were part of (50 respondents per group) reached an aggregate level of contributions of at least $350. This aggregate contributions level represents a "provision point" that is constant across each sub-sample of respondents. Each sub-sample, or group, of respondents then faced one format for payments supporting water quality monitoring.

The base format stated to respondents that if their group failed to reach the provision point, their personal contribution would be refunded to them. This format establishes a money-back-guarantee (MBG) but fails to address the issue of how excess contributions above the provision point would be distributed. This base format is called the No Rebate (NR) format.

The second format included the money-back-guarantee along with a statement that excess contributions would be rebated to contributors on a proportionate basis. In this case, if a group of respondents contributed 10% more than the provision point (e.g., contributing a total of $385 rather than just $350), then each contributor would receive a rebate of 10% of their personal contribution. This payment format is called the Uniform Proportional Rebate (UPR) format.

The third format again included the money-back-guarantee, but rather than a rebate, the format included a statement that contributions in excess of the provision point would be retained by URI Watershed Watch and used for the purpose of covering administrative costs. This format is called the Extended Benefits (EB) format.

The Experiment

The experiment was conducted at the University of Rhode Island with 435 students. The students-participants were obtained by prior arrangement with the instructors of one introductory class in sociology and one introductory class in psychology, and by direct solicitation in the university's cafeteria. The experiment took place during regular class hours in the sociology and psychology classes, and none of the students were forced or pressured to participate in the experiment. Classroom students were given prior notice that the experiment would not be a class assignment, but that they would be paid to make real decisions regarding a real environmental resource. The students recruited in the university's cafeteria were recruited on the spot during lunchtime to complete a survey involving environmental decisions. All survey respondents were paid to participate—respondents in the sociology class were paid $10, while respondents recruited from the psychology class and the cafeteria were paid $2.

The experiment required respondents to fill-out a survey, which elicited their attitudes and values for an environmental resource. Standard socio-economic questions were also elicited. In a split-sample style design, respondents were given either a hypothetical-money or a real-money survey treatment. Parallel language was used in both survey formats, except for slight modifications where respondents were asked to donate money to support a program involving the environmental resource. In the hypothetical-money survey treatment respondents faced a decision involving a hypothetical-money cost, while in the real-money survey treatment respondents faced decisions with real-money consequences. Of the 435 respondents 150 participated in the real-money treatment; 50 each under the three payment rules. The rest participated in some version of the hypothetical-money treatment.

The environmental good used in the experiment was an "adopt-a-pond" water-quality monitoring program. Over 90% of Rhode Island's water quality data on ponds and lakes comes

from volunteer monitors coordinated through a non-profit program called URI Watershed Watch. Watershed Watch relies on voluntary donations to operate, and each pond costs $500 per year to monitor [see Spencer, Swallow, and Miller (1998) for an overview of pond monitoring in Rhode Island]. If Watershed Watch collects enough funds to monitor a pond then the pond will be monitored; otherwise, it will not be monitored. Through Watershed Watch we were able to offer, for sale, a water-quality monitoring program for a specific pond that was not currently being monitored. Thus, those respondents who participated in the real-money survey treatments had a real opportunity to fund water-quality monitoring on the pond. Making a contribution involved a one-time payment to support water-quality monitoring on the pond for the upcoming monitoring season.

In both the hypothetical and real-money survey treatments, the WTP question was preceded by information regarding water-quality monitoring and the conditions under which the respondent was asked to donate money. Each respondent's survey contained information pertaining to either a no rebates, uniform proportional rebates, or extended benefits rule. The WTP question involved a "yes" or "no" response of the form depicted in Figure 1. Across all survey treatments, individuals faced a cost of $6, $8, $12, or $15.

In the real-money survey treatments, each respondent was assigned an arbitrary identification number. Real-money respondents were required to present this number in order to receive any money-back or rebates at the conclusion of the experiment. Experimental monitors collected all surveys and summed the contributions collected under each rebate rule. All cash transactions were handled on the same day before each respondent left the classroom.

Data Analysis

Data were analyzed using Hanemann's (1984) standard, utility-theoretic framework for estimating willingness to pay based on yes-no answers to discrete choice questions (such as in Figure 1). Within this random utility framework, the probability that individual respondent *i* will choose "yes," to monetarily support water quality monitoring on the pond, is modeled as follows:

$$\Pr(\text{Yes}) = \Pr(U_{i,Yes} > U_{i,N0}) ,$$

where

$$U_{i,Yes} = V(Q_{Yes}, \mathbf{S_i}, \mathbf{R}, M_i - C_i) + e_{i,Yes},$$

$$U_{i,N0} = V(Q_{N0}, \mathbf{S_i}, \mathbf{R}, M_i) + e_{i,N0},$$

Q_{Yes} represents the state of the world with monitoring on the pond, Q_{N0} represents the state of the world without monitoring on the pond, $\mathbf{S_i}$ is a vector of characteristics which describe individual *i*, $\mathbf{R}$ is a vector of variables indicating which payment rule a respondent faced, $V(\cdot)$ represents the deterministic component of utility that is econometrically measurable by the researcher, and e is the random or unobservable (to the researcher) component of utility. After rearranging terms, one has

$$\Pr(\text{Yes}) = \Pr(e_{i,N0} - e_{i,Yes} < V_{i,Yes} - V_{i,N0}).$$

If one assumes the *e*'s are independent and identically Type I Extreme Value distributed in standard form, then Pr(Yes) can be modeled as the following logit model (McFadden 1974):

$$\Pr(\text{Yes}) = \frac{1}{1+e^{-\Delta V}} ,$$

where $\Delta V = V_{i,Yes} - V_{i,No}$.

In estimating Pr(Yes), we hypothesized that the probability that a respondent would answer yes will be higher for both a uniform proportional rebate (UPR) and extended benefits (EB) payment rule than for a no rebate (NR) rule; will be lower for respondents facing a higher cost of monitoring the pond; will be higher for respondents who have declared an environmental major in college. Since environmental majors have expressed a preference to address environmental issues as professionals, it is reasonable to expect this factor to increase their willingness to pay (see also Spencer et al. 1998).

Finally, if one assumes V is linear in its parameters,

$$V_{i,Yes} = \beta'(\boldsymbol{Z_{i,Yes}}) + \beta_C C_{i,Yes},$$

then one can derive a utility-theoretic measure of WTP as follows (Hanemann 1984)

$$WTP = -(\beta'/\beta_C)\cdot(\boldsymbol{Z_{i,Yes}} - \boldsymbol{Z_{i,No}}) , \qquad \text{[equation (1)]}$$

where $\boldsymbol{Z_{i,Yes}} = z(Q_{Yes}, \mathbf{S_i}, \mathbf{R})$, β' is a vector of parameters, and β_C is the coefficient on individual cost for pond monitoring.[3] Estimates of β' and β_C can be obtained through maximum likelihood estimation.

Results and Discussion for Water Quality Monitoring

Table 1 describes the explanatory variables used in the behavioral model of pond monitoring, and Table 2 presents the logit estimation results for three specifications of the model. Specification 1 allows for varying intercepts and slopes across rebate rules (i.e., treatments), while specifications 2 and 3 represent reduced forms of specification 1 that only allow for varying slopes and intercepts, respectively, across rebate rules. For each specification, the χ^2-statistic for a likelihood ratio test of model significance is highly significant ($P< 0.001$). Across all specifications, the signs of the parameter estimates are as expected. Both *Hypothetical Cost* and *Real Cost* have negative signs that are highly significant. This indicates that the survey respondents are less likely to respond "yes" as the individual cost of pond monitoring increases. Dummy variables indicating the payment rule faced by a respondent, *UPR* and *EB*, both have positive signs, which indicate respondents are more likely to support pond monitoring under a uniform proportional rebates (*UPR*) or extended benefits (*EB*) rule than under a no rebates (*NR*) rule.[4] This result follows expectations. However, the effect of the *UPR* rule unexpectedly appears stronger than the *EB* rule.

[3] If cost is interacted with elements of R, then we mean hear that β_C includes the adjustment for this interaction.
[4] The dummy variable for NR is omitted to prevent colinearity with the intercept.

In specification 3, *UPR* is highly significant (*one*-tailed $P < 0.01$) while *EB* is only weakly significant (*one*-tailed $P < 0.09$).

Neither specification 2 ($\chi^2_{1vs.2}$-statistic = 1.0398, $P > 0.59$) nor specification 3 ($\chi^2_{1vs.3}$-statistic = 3.2758, $P > 0.51$) are statistically different from specification 1, suggesting that both specifications 2 and 3 are preferred to specification 1 based on a likelihood ratio test. However, since specification 2, unlike specification 3, allows one to obtain different hypothetical WTP to real WTP ratios across rebate rules, this discussion only reports WTP estimates for specification 2.

Table 3 reports both hypothetical and real WTP estimates for the average respondent based on the parameter estimates for specification 2 in Table 2. [5] For hypothetical WTP, the observed pattern suggests that $WTP_{NR} > WTP_{UPR} \approx WTP_{EB}$, where WTP_r is defined by equation one for the NR, UPR, or EB rule as indicated by subscript r. As expected, the UPR and EB rules yield higher hypothetical WTP estimates than the NR rule—in fact, both UPR and EB yield hypothetical WTP estimates that are nearly double those of the WTP estimates under the NR rule. The observation that $WTP_{UPR} \approx WTP_{EB}$, for the hypothetical (stated preference) respondents, is, however, unexpected. This result suggests that stated-preference willingness to pay may be robust with respect to the means by which excess contributions would be distributed if money were actually collected. Given these point estimates of hypothetical willingness to pay, it appears that leaving the issue of how excess contributions would be distributed may cause respondents to express a lower WTP, as illustrated by comparing the NR rule to the UPR and EB rules in table 3. Of course, the experimental design does not allow a comparison between these estimates of hypothetical WTP and the estimated WTP that might arise if a stated-preference survey omitted most of the details about how money could be collected and used. We leave this issue for future research.

For real-money WTP, the observed pattern suggests that $WTP_{NR} \approx WTP_{EB} < WTP_{UPR}$. As expected, $WTP_{UPR} > WTP_{NR}$, but unexpectedly WTP_{EB} is not greater than WTP with either the NR or UPR rules. Although the latter result seems counter-intuitive, based on experimental economics literature (e.g., Marks and Croson 1998) it may be that respondents did not value the extended benefits (EB) offered. In this experiment, any excess contributions were designated to help cover the general operating costs of the pond monitoring organization. Perhaps other forms of extended benefits, such as support for monitoring additional ponds, would have stimulated higher WTP estimates under the extended benefits rule. In contrast to the hypothetical WTP estimates, it appears that real WTP is not robust with respect to the disposition of excess contributions.

The last column of Table 3 reports the ratio of hypothetical WTP to real-money WTP for each rebate rule. All the ratios fall well within the range of ratios reported across the few other discrete-choice studies that compare hypothetical and real-money WTP for a public good (Foster, Bateman, and Harley, 1997). The NR and UPR rules have approximately the same hypothetical-to-real ratio, about 2, while the EB rule produces a ratio that is nearly twice that of the NR and UPR rules. Compared to Spencer, Swallow, and Miller (1998), who report a ratio of 4.67 for a water quality monitoring program, a ratio of 2 is a substantial improvement in the observed gap between hypothetical and real-money WTP.

[5] An interesting note is that specification 3 yields negative WTP estimates for the average respondent under the no rebate rule. This result could indicate that most non-environmental majors were not in the market for pond monitoring. However, willingness to pay differences from specification 3 remain consistent with those discussed for specification 2.

A tentative conclusion from this experiment is that the UPR rule appears preferable for use in real-money, contingent surveys. Experimental economics literature suggests that such rebate rules can result in valuation estimates that approximate well aggregate demand for an environmental good. Moreover, UPR produces a favorable ratio of hypothetical WTP versus real-money WTP, while the NR rule may be suspect in the case of a hypothetical survey that is worded in a fashion parallel to the corresponding real money survey. For these reasons, the UPR rule was adopted in experimental surveys concerning the valuation of wetlands development rights.

III. Experiments on the Valuation of Wetlands

The final set of experiments concerns the value of protecting wooded wetlands from development. These experiments constitute an attempt to compare stated preference (hypothetical-mone) values and real-money willingness to pay for wetlands in an actual "field experiment" involving randomly drawn citizens and a mail-survey format similar to formats commonly used in CVM. The wetland application is sufficient to raise challenges that may be associated with valuation of many environmental goods.

Wetland Focus Groups

Prior to and during survey development, we conducted extensive focus groups with citizens. These focus groups elicited citizens' views on what features of wooded wetlands would affect their willingness to pay for wetland protection. Focus group observations led to the following list of attributes considered most relevant to participants: size or acreage of the parcel, character of the surrounding land, availability of public access, role in expanding or connecting existing conservation areas or in establishing a new conservation area, relative diversity of wildlife, relative sustainability of wildlife habitat qualities, location of the parcel relative to roads, and the cost of protection.

In several focus groups, and in survey pretesting, a primary subject was the method by which to present a real-money survey question to potential respondents. Commonly, focus group or pretest participants expressed surprise after completing several preliminary parts of the survey and then finding a survey question that required that they contribute real money if their answer indicated a willingness to pay for protecting a wetland. These observations are difficult to quantify, but a general impression is worth noting for future researchers.

First, we attempted to convey, through cover letters and through instructions within the survey, that the survey was <u>not</u> a simple fundraising device, despite its request for money. This approach was soundly rejected by focus group participants who accepted that the survey was about valuing environmental goods, but failed to accept the idea that it was <u>not</u> a fundraising gimmick. In various revisions of the survey materials, we attempted to identify a timing (or location within the materials) that would mitigate these concerns, but we judged that the fundamental approach was ineffective.

We then altered the approach by dropping all claims that the survey was not about fundraising. Our cover letter included a statement that wetlands may be important to some citizens, as are the services of development, and that in some instances it is important for towns or state agencies to compare the value of wetlands to the value of development "in economic terms." The cover letter did not explicitly state that the survey would request real payments of money, but it did explicitly state that the questions involved money. Within the survey, we developed an introduction

to the real-money questions that reminded respondents that our purpose was to ask about the value of wooded wetlands to the respondent's household. At this point, we stated that we would be asking about their willingness to pay a particular amount to protect one of two wetland parcels. This information, within the survey, emphasized that whatever choice the respondent made, it was important to provide us with an answer.

The survey then provided additional details. These details described the method of protection for the wetland and indicated that protection for a particular wetland would occur ". . . *only if* enough respondents decide to financially support the wetland parcel you chose." This latter phrase is indicative of a provision point for the public good, but it does not specify a specific dollar amount for the provision of the wetland protection program for a parcel.[6] Prior to mailing the survey, we decided that the threshold for provision of the wetland protection contract would be equal to one year's payment on a 10 year conservation agreement with the respective landowners (amounting to $300 and $400 for the parcels located in the towns of Scituate and North Smithfield, respectively).

Next the survey details indicated that any excess money collected would be returned (rebated) to respondents who contributed, while any funds retained by the university would be used only for establishing a conservation contract for the wetland owners (while a research grant would cover all salaries and other costs at the university). The survey details also stated that, upon completion of the survey, contributors would either receive a letter indicating that a wetland conservation contract was executed and returning "your share of any left over funds" or a letter indicating that the contract was not executed and a refund of their entire contribution. While we did not explicitly state that we would give rebates to all contributors in proportion to the excess collected, focus group participants discussed their understanding of this refund mechanism in a manner consistent with the Uniform Proportional Rebate mechanism examined in the water-quality monitoring experiments above.

The Wetland Survey Experiment

The wetland survey experiment involves two landowners who have contracted with the university in a Land Conservation Contract that can be converted to a deed restriction that will legally prohibit development on a specified wetland parcel. One parcel is located in the Town of Scituate, Rhode Island, and is 29 acres, with medium wildlife habitat quality (as judged by a university biologist), located on local roads, an representing an opportunity to establish a new conservation area isolated from other legally protected lands. The second parcel is located in the Town of North Smithfield, Rhode Island, containing about 50 acres of wetlands on a 70 acre wooded tract. Given the locations of these parcels, the survey experiments focused on samples of residents in these towns.

Six survey-versions were designed and administered to a randomly selected sample of residents obtained from a commercial mail-marketing company. Table 4 summarizes the survey versions and samples, as well as response rates and collections of real money. All stated-preference

[6] We did not quantify the provision point because we had contracted with two different landowners for potential protection for different parcels of land at different costs and we decided not to examine whether differences in a quantitatively defined provision point would affect willingness to pay results in this field experiment. This decision leaves an open question for future research.

questions in all versions were presented in a contingent choice or choice experiment format (Opaluch et al. 1993; Adamowicz et al. 1998), wherein the respondent evaluates descriptions of two wetland parcels and then decides to state a preference to protect one of these parcels or to choose "neither" parcel and thereby forego protection on both parcels while saving any (hypothetical or real) monetary costs to their household.

Survey version 1 presented respondents with a single, real-money contingent choice question. Version 2 involved a hypothetical (stated-preference) contingent choice question, followed by a different, real-money question; prior to answering the hypothetical question, respondents were informed that a real-money question would follow. These two survey versions allow an evaluation of whether respondents altered their willingness to pay in switching between hypothetical and real-money questions within the same survey (version 2) and whether the opportunity to view a set of hypothetical substitutes would alter real-money willingness to pay in a split-sample design (version 1 versus version 2).

Versions 3, 4, and 5 all involved two hypothetical contingent choice questions. Version 3 was designed from comparability to real-money version 1. The main design feature here is that a single contingent choice question was presented in version 3, much as in version 1. However, after completion of the same demographic questions as presented in version 1 (and all versions), version 3 asked a second, hypothetical contingent choice question. This second question was included in version 3 in order to augment the sample of hypothetical responses.

Version 4 was motivated by concerns in the contingent valuation literature that respondents to hypothetical surveys may be motivated by a desire to show support for the environment. This motivation could be described as "yea-saying." In version 4, respondents were given three ways in which they could choose not to pay for conservation on either parcel (i.e., three versions of the "neither" choice): (a) choice to save the respondent's money rather than pay for conservation (a statement that was identical to the "neither" choice in other survey versions); (b) choice stating a general unwillingness to pay for wetland protection; and (c) a choice stating that protection of wetlands is valuable to the respondent's household, but that the respondent is unwilling to pay the specified costs at the time of response.

Version 5 is a straightforward format presenting two hypothetical contingent choice questions followed by the demographic questions. This version was intended to allow a comparison to the two questions (one hypothetical and one involving real-money) from version 2.

Finally, version 6 presented respondents with twelve hypothetical questions, primarily in an effort to gain additional data from hypothetical respondents. This motivation arises from results in Spencer et al. (1998) suggesting that the precision of willingness to pay estimation is poorer with stated-preference surveys than it is with real money surveys.

All versions of the survey were mailed following principles of the Dillman Total Design method, including an initial mailing of the survey, a reminder post card, a second mailing of the survey to non-respondents, a second reminder post card, and a follow-up letter to non-respondents with a brief, post-card survey collecting demographic information and a statement concerning the individual's choice to not respond. Inspection of table 4 shows that response rates (based on deliverable surveys) were generally in the range of 35% for real-money surveys (versions 1 and 2),

while response rates for surveys containing only hypothetical contingent choice questions (versions 3 through 6) were generally in the range of 45%.

Analysis and Results

In each contingent choice questions, wetland parcels were described according to the variables listed in Table 5. For hypothetical contingent choice questions, contingent choice questions were designed to cover the range of variables and combinations of their levels as described in the table. This design led to 36 different parcel-choice pairs in the hypothetical surveys. For real-money contingent choice questions, parcel descriptions were constrained to apply to all or at least a portion of the real wetland parcels that were contracted for this project. This constraint led to 9 parcel-choice pairs in the real-money surveys for Scituate, and 18 parcel-choice pairs in the real-money surveys for North Smithfield. The parcel-choice pairs were designed by a fractional factorial procedure provided by StatDesign, Inc.

At the time of this writing, the available data is limited to survey versions 2, 4, and 5 for both towns, and responses to version 1 from Scituate. These data are considered reliable, but are as yet unverified for coding accuracy. For the purpose of this preliminary analysis, the data were pooled for analysis using a multinomial logit version of the random utility model described above (see also, Adamowicz et al. 1998). This analysis assumes respondents stated a choice for a parcel, or for the neither-option, only if the chosen alternative would maximize their utility if implemented.

Three specifications of the choice model are presented in Table 6. All three models are statistically significant predictors of respondents' choices at a high level ($P < 0.01$). Specifications 1 and 3 include all of the attributes describing a parcel. Specification 2 is a model nested within Specification 1, omitting those parcel attributes that were found to be statistically insignificant in Specification 1 (omitting public access variables, variables for location on different types of roads, variables identifying whether the parcel connects to existing undeveloped land, the dummy variable for medium sustainability of wildlife habitat, and the dummy variable indicating whether the surrounding landscape is farmland).

The likelihood ratio test of Specification 2 versus the unrestricted Specification 1 fails to reject the null hypothesis that the coefficients on the variables dropped in specification 2 are zero (chi-square = 3.268, 8 df, $P>0.90$). Based on Specification 2, it appears that wetland parcels in residential landscapes were less preferred for conservation than were parcels in farmland or undeveloped-wooded landscapes, and it appears that parcels with low or medium diversity of wildlife or low sustainability of wildlife habitat were significantly less preferred for conservation as compared to parcels with high diversity.

Specification 1 allows the coefficient on cost to differ between real and hypothetical surveys, and for the version 4 format for the neither choice to lead to a willingness to pay estimate that is different than other versions of the hypothetical survey. Specification 3 restricts these cost coefficients such that there is a single coefficient on cost across all forms of the choice questions. A log-likelihood ratio test of this restriction, compared to unrestricted model 1, rejects the null hypotheses (chi-square = 20.539, 2 df, $P<0.005$). This preliminary analysis indicates that respondents treated money costs differently between the real-money surveys and the hypothetical-money surveys.

A comparison of the cost coefficient from real-money surveys to the cost coefficient from hypothetical surveys (other than version 4) indicates that respondents were less likely to state a willingness to pay to protect a parcel if the survey involved real money.[7] Based on these results, however, it is pre-mature to discuss whether this difference is significant. The coefficient on hypothetical costs is only significantly different from zero at a level of only about 11% for a one-tail test. Availability of the additional hypothetical responses (unavailable at this writing) may permit a more definitive comparison of these survey versions.

Finally, it is worth pointing out that respondents who contributed real money provided an average of $18 from North Smithfield, and an average of $16 from Scituate. Respondents from both towns contributed about $1000 to protect the wetland associated with their town. These contributions exceeded the pre-identified provision point of a one-year payment on the ten-year contract with landowners. Excess contributions are being rebated to these respondents.

[7] This conclusion is consistent with a few other preliminary tests, even though the variance on the coefficient of H_COST is rather high.

References

Adamowicz, W., P. 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(February):64-75.

Blackburn, M., G.W. Harrison, and E.E. Rutström. 1994. Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics* 76(December):1084-88.

Brown, T.C., P.A. Champ, R.C. Bishop, and D.W. McCollum. 1996. Which response format reveals the truth about donations to a public good? *Land Economics* 72(2):152-66.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1997. Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management* 33(2):151-62.

Cummings, R.G., S. Elliott, G.W. Harrison, and J. Murphy. 1997. Are hypothetical referenda incentive compatible? *Journal of Political Economy* 105(3):609-21.

Cummings, R.G., G.W. Harrison, and E.E. Rustrom. 1995. Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review* 85(1):260-66.

Cummings, R.G. and L.O. Taylor. 1997. Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. Typescript, Department of Economics, School of Policy Studies, Georgia State University. August.

Foster, V., I.J. Bateman, and D. Harley. 1997. "Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-Experimental Comparison." *Journal of Agricultural Economics* 48:123-138.

Fox, J.A., J.F. Shogren, D.J. Hayes, and J.B. Kliebenstein. 1995. CVM-X: Calibrating contingent values with experimental auction markets. Kansas Agricultural Experiment Station Contribution no. 95-470-J. Manhattan: Kansas State University. November.

Hanemann, W.M. 1984. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 66(August):332-341.

Hoehn, J.P. 1991. Valuing the multidimensional impacts of environmental policy: Theory and methods. *American Journal of Agricultural Economics* 73:289-299.

Hoehn, J.P., and A. Randall. 1989. Too many proposals pass the benefit cost test. *American Economic Review* 79:544-551.

Johnston, R.J., S.K. Swallow, and T.F. Weaver. 1999. Estimating willingness to pay and resource tradeoffs with different payment mechanisms: An evaluation of a funding guarantee for watershed management. *Journal of Environmental Economics and Management* 38:97-120.

Marks, M., and R. Croson. 1998. Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics* 67:195-220

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. P. Zarembka, pp. 105-142. New York: Academic Press.

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

Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison, and T. McGuckin. 1994. Hypothetical surveys and real economic commitments. *Land Economics* 70(2):145-54.

Opaluch, J.J., S.K. Swallow, T.F. Weaver, C.R> Wessells, and D. Wichelns. 1993. Evaluating impacts from noxious facilities: Including public preferences in current siting mechanisms. *Journal of Environmental Economics and Management* 24(January):41-59.

Rondeau, D., W.D. Schulze, and G.L. Poe. 1999. Voluntary revelation of the demand for public goods using a provision point mechanism. *Journal of Public Economics* 72:455-470.

Seip, K., and J. Strand. 1992. Willingness to pay for environmental goods in Norway: A contingent valuation study with real payment. *Environmental and Resource Economics* 2(1):91-106.

Shogren, , J.F. 1993. Experimental markets and environmental policy. *Agricultural and Resource Economics Review* 22(October):117-29.

Shogren, J.F., J.A. Fox, D.J. Hayes, and J.B. Kliebenstein. 1994. Bid sensitivity and the structure of the Vickrey auction. *American Journal of Agricultural Economics* 76(December).

Spencer, M.A., S.K. Swallow, and C.J. Miller. 1998. Valuing water quality monitoring: A contingent valuation experiment involving hypothetical and real payments. *Agricultural and Resource Economics Review* 27 (April):28-42.

Swallow, S.K. 1994. Value elicitation in laboratory markets: Discussion and applicability to contingent valuation. *American Journal of Agricultural Economics* 76(December):1096-1100.

<u>Review the pond described below and then check</u> the appropriate box (in the middle of the page) indicating whether you do or do not prefer to have U.R.I. Watershed Watch monitor the pond. <u>Note</u>: we ask that you please make an <u>actual</u> monetary payment, here and now, if you prefer to have the pond monitored.

DESCRIPTION OF POND

R 46 acres,
Suspected to have average to poor water quality,
Located inland,
2 Surrounded by wooded area,
\+ Is accessible to public,
Has low to moderate fishing usage, and
Monitoring will help find source(s) of current problem.

<u>Cost to you</u>:
We ask that you pay **$8**, here and now, to help fund monitoring for next season.

⇓

Please check only ***<u>one</u>*** box below:

❑ **<u>Yes</u>**, I prefer to pay the $8 and have the pond monitored.		❑ **<u>No</u>**, I prefer to pay nothing and <u>not</u> have the pond monitored.
⇓		⇓
If you choose **"Yes"** to having the pond monitored, ***please turn-in an envelope <u>filled with</u> $8 cash or a check for $8 (payable to U.R.I. Watershed Watch)*** with your completed survey. If you do not include such funds, we will assume that, at this time, you prefer not to monitor the pond.		If you choose **"No"** to having the pond monitored, ***please turn-in an <u>empty</u> envelope with your completed survey***.

Figure 1. Example of WTP Question

Table 1. Description of Explanatory Variables

Variable	Description
Hypothetical Cost	The required contribution (or cost) for an individual in the *hypothetical-money* survey treatment to support water quality monitoring on the pond. [a]
Real Cost	The required contribution (or cost) for an individual in the *real-money* survey treatment to support water quality monitoring on the pond.
UPR [b]	Indicator variable = 1 if *uniform proportional rebate* rule; = 0 otherwise.
EB	Indicator variable = 1 if *extended benefits* rule; = 0 otherwise.
UPR × *Hyp. Cost*	Interaction term between the *uniform proportional rebate* rule and the required *hypothetical-money cost* to support pond monitoring.
EB × *Hyp. Cost*	Interaction term between the *extended benefits* rule and the required *hypothetical-money cost* to support pond monitoring.
UPR × *Real Cost*	Interaction term between the *uniform proportional rebate* rule and the required *real-money cost* to support pond monitoring.
EB × *Real Cost*	Interaction term between the *extended benefits* rule and the required *real-money cost* to support pond monitoring.
Environ. Major	Indicator variable = 1 if environmental major; = 0 otherwise. [c]

[a] Notice that the no rebate (*NR*) rule is used as the base level.

[b] For both the hypothetical-money and real-money survey treatments, individuals faced a cost of $6, $8, $12, or $15.

[c] Environmental majors included animal science and technology, aquaculture and fishery technology, biological sciences, biology, clinical laboratory science, environmental science, marine biology, microbiology, plant science, water and soil science, wildlife biology, and zoology.

Table 2. Estimation Results for Various Specifications of the Logit Model

Variable	Specification 1		Specification 2		Specification 3	
	Parameter Estimate	Pr > |*Z*| (*P*-value)[a]	Parameter Estimate	Pr > |*Z*| (*P*-value)	Parameter Estimate	Pr > |*Z*| (*P*-value)
CONSTANT	–0.4796	0.4381	0.0463	0.8902	–0.3511	0.3506
	(–0.775)		(0.138)		(–0.933)	
Hypothetical Cost	–0.0571	0.3362	–0.1043	0.0061	–0.0700	0.0306
	(–0.962)		(–2.741)		(–2.162)	
REAL COST	–0.1595	0.0336	–0.2123	0.0002	–0.1754	< 0.001
	(–2.124)		(–3.715)		(–4.345)	
UPR	0.7553	0.3568			0.6549	0.0171
	(0.921)				(2.385)	
EB	0.7416	0.3898			0.3913	0.1654
	(0.860)				(1.387)	
UPR × *Hyp. Cost*	–0.0214	0.7862	0.0462	0.1256		
	(–0.271)		(1.532)			
EB × *Hyp. Cost*	–0.0211	0.7985	0.0453	0.1384		
	(–0.255)		(1.482)			
UPR × *Real Cost*	0.0193	0.8399	0.0933	0.0926		
	(0.202)		(1.682)			
EB × *Real Cost*	–0.0977	0.3831	–0.0218	0.7494		
	(–0.872)		(–0.319)			
Environ. Major	0.9304	0.0036	0.9386	0.0033	0.9528	0.0028
	(2.910)		(2.934)		(2.990)	
Total Observations	427		427		427	
Log-likelihood	–238.476		–238.996		–240.114	
χ^2-statistic	42.849	< 0.001[b]	41.809	< 0.001	39.573	< 0.001
Degrees of Freedom	9		7		5	

Note: numbers in parentheses represent *Z*-statistics.

[a] The *P*-values reported in this table correspond to a *two*-tailed test of H_0: $\beta = 0$ versus H_A: $\beta \neq 0$.

[b] The level of significance (i.e., *P*-value) for the χ^2-statistic.

Table 3. Willingness-to-Pay Estimates for the Average Respondent

Rebate Rule	Hypothetical WTP	Real-Money WTP	Ratio ($WTP_{Hyp.}/WTP_{Real}$)
No Rebates	\$1.58	\$0.78	2.03
Uniform Prop. Rebates	\$2.84	\$1.39	2.04
Extended Benefits	\$2.80	\$0.71	3.94

Note: the WTP estimates reported here are based on the parameter estimates to specification 2 in Table 2 and the average respondent, identified by the sample mean of the variable *Environ. Major*, which is 0.1265.

Table 4. Summary of Survey Treatments in Wetlands Survey

Town of North Smithfield	**Town of Scituate**
1NS: Real-Money Survey • Sample size = 700 • Deliverables = 629 • Response rate = 34% • 40 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **North Smithfield** was $725.00	1S: Real-Money Survey • Sample size = 500 • Deliverables = 468 • Response rate = 39% • 40 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **Scituate** was $640.00
2NS: Hypothetical-Real Money Survey • Sample size = 200 • Deliverables = 171 • Response rate = 30% 13 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **North Smithfield** was $255.00	2S: Hypothetical-Real Money Survey • Sample size = 200 • Deliverables = 185 • Response rate = 35% 16 respondents were willing to pay for a parcel. Total amount sent in to preserve parcel in **Scituate** was $260.00
3NS: Two question Hypothetical Survey with Demographic Questions • Sample size = 129 • Deliverables = 115 • Response rate = 46%	3S: Two question Hypothetical Survey with Demographic Questions • Sample size = 271 • Deliverables = 256 • Response rate = 39%
4NS: Two question Hypothetical Survey with no responses allowing "support for environment" • Sample size = 150 • Deliverables = 136 • Response rate = 45%	4S: Two question Hypothetical Survey with no responses allowing "support for environment" • Sample size = 250 • Deliverables = 240 • Response rate = 50%
5NS. Two question Hypothetical Survey • Sample size = 100 • Deliverables = 93 • Response rate = 43%	5S. Two question Hypothetical Survey • Sample size = 100 • Deliverables = 94 • Response rate = 54%
6NS: Twelve-question Hypothetical Survey • Sample size = 100 • Deliverables = 95 • Response rate = 42%	6S: Twelve-Question Hypothetical Survey • Sample size = 100 • Deliverables = 95 • Response rate = 44%

Table 5. Description of Explanatory Variable for Wetlands Survey

Variable	Description
Wildlife diversity	Wildlife diversity of the parcel. Can be listed as one of the three: WILD_DVL-- Low wildlife diversity WILD_DVM-- Medium wildlife diversity WILD_DVH --High wildlife diversity
Access	The level of access the respondent would have to the parcel No_ACC—No access LIM_ACC—Limited access, permission required from owner FULL_ACC—Open access to users
Road	Type of road bordering the parcel LCL_RD—Locally traveled road CMN_RD—Commonly traveled road HVY_RD—Heavily traveled road
Sustainability	Indicates how much it was able to sustain habitat quality LOW_SUS—Low sustainability MED_SUS—Medium sustainability HI_SUS—High sustainability
Conservation	Indicates the role of the parcel as a conservation area CNSRVTNI—Parcel is isolated from other protected areas CNSRVTNE— Parcel is connected to an unprotected, undeveloped area CNSRVTNC—Parcel connects two protected areas
Surrounding land	Indicates the type of land surrounding the parcel SUR_LNDR—Parcel is surrounded by rural residential land SUR_LNDF—Parcel is surrounded by farmland SUR_LNDW—Parcel is surrounded by woodland
Size	Acreage of parcel (29, 45, 60)
Cost	The required contribution for a respondent to support protection of the parcel ($5, $10, $15, $20, $25, $30) R_COST—Dummy variable for real money interacted with the cost variable H_COST—Dummy variable for hypothetical money interacted with the cost variable LEGITCST—Dummy variable for legitimizing no's interacted with the cost variable

Table 6. Estimation Results for Various Specifications for Wetland Survey Data

Variable	Specification 1		Specification 2		Specification 3	
	Parameter Estimate	Pr>|Z| (P-value)	Parameter Estimate	Pr>|Z| (P-value)	Parameter Estimate	Pr>|Z| (P-value)
DMY_N (neither dummy)	-.1261 (-.393)	.6940	.1069 (.423)	.6725	-.0885 (-.272)	.7853
WILD_DVL (wildlife low)	-.5549 (-3.969)	.0001	-.5246 (-3.878)	.0001	-.5651 (-3.996)	.0001
WILD_DVM (wildlife medium)	-.4186 (-3.154)	.0016	-4137 (-3.184)	.0015	-.5189 (-3.875)	.0001
NO_ACC (no public access)	-.02988 (-.212)	.8319			.02094 (.148)	.8827
LIM_ACC (limited public access)	-.1288 (-1.015)	.3101			-.1251 (-1.011)	.3120
LCL_RD (local road)	.01205 (.078)	.9376			-.1106 (-.719)	.4720
CMN_RD (road commonly traveled)	.01897 (.120)	.9042			.0281 (.177)	.8595
LOW_SUS (low sustainable habitat)	-.5297 (-3.773)	.0002	-.4669 (-3.797)	.0001	-.5289 (-3.727)	.0002
MED_SUS (medium sustainable habitat)	-.1020 (-.784)	.4333			-.2215 (-1.704)	.0884
CNSRVTNI (isolated area)	-.1054 (-.726)	.4678			-.2474 (-1.714)	.0865
CNSRVTNE (expands conservation area)	-.04735 (-.309)	.7570			-.0354 (-.230)	.8184
SUR_LNDR (residential lands)	-.5328 (-3.895)	.0001	-.4564 (-3.734)	.0002	-.5820 (-4.273)	.0000
SUR_LNDF (farmland)	-.1364 (-1.020)	.3075			-.2175 (-1.644)	.1002
SIZE (acres)	.01762 (3.538)	.0004	.01741 (3.556)	.0004	.0236 (4.816)	.0000
COST					-.01357 (-2.368)	.0179
R_COST (COST x real dummy)	-.05226 (-4.717)	.0000	-.05714 (-5.560)	.0000		
H_COST (COST x hypothet. Dummy)	-.009651 (-1.295)	.1954	-.009015 (-1.239)	.2152		
LEGITCST (COST x version 4 dummy)	.008194 (1.049)	.2943	.008066 (1.039)	.2990		
Total Observations	847		847		847	
Log-Likelihood	-858.5940		-860.2280		-868.8636	
χ^2- statistic	124.633	<.005	121.365	<.005	104.094	<.005
Degrees of freedom	17		9		15	

Note: numbers in parentheses represent Z-statistics

ASK A HYPOTHETICAL QUESTION, GET A VALUABLE ANSWER?

by

Christopher D. Azevedo
Joseph A. Herriges
Catherine L. Kling*

Department of Economics

Iowa State University

June 12, 2000

This paper models the recreation demand for Iowa wetlands, combining survey data on both actual usage patterns (i.e., revealed preferences) and anticipated changes to those patterns under hypothetical increases in trip costs (i.e., stated preferences). We formulate and test specific hypotheses concerning potential sources of bias in each data type. We consistently reject consistency between the two data sources, both in terms of implied wetland values and underlying preference parameters. Careful attention is paid to the interpretations of the test results, noting particularly how the interpretation of the same results can vary with the "school of thought" of the reader.

* The authors would like to thank Dan Phaneuf, Kerry Smith, and Wally Thurman for helpful comments on earlier drafts of this paper. We would also like to thank seminar participants at North Carolina State University, the University of California-Davis, and the University of Kansas. This research was supported in part by the U.S. Environmental Protection Agency and by the Western Regional Research Project W-133. Although the research described in this article has been funded in part by the United States Environmental Protection Agency through R825310-01-0 to the authors, it has not been subject to the Agency's required peer review policy and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

"Economists have inherited from the physical sciences the myth that scientific inference is objective, and free of personal prejudice. This is utter nonsense." ¾ Leamer [1983, p. 36]

I. Introduction

Economists began investigating the use of surveys to elicit consumers' willingness-to-pay for public goods over three decades ago. The vast majority of the early applications concerned environmental goods, although numerous other public goods have been the subjects of valuation surveys. In the last decade, the use of surveys to elicit welfare values has come under debate by the profession at large. The now famous NOAA blue ribbon panel (Arrow, *et al.* [1993]) was assembled to consider the validity of these survey methods (also called contingent valuation or stated preference methods) and the *Journal of Economic Perspectives* ran a symposium with competing views on the topic (Diamond and Hausman [1994]; Hanemann [1994]; and Portney [1994]).

The fundamental question in this debate might be stated: can carefully designed survey methods provide informative data on consumers' willingness to pay for public goods, or does the hypothetical nature of these instruments render them irrelevant, regardless of how much attention is given to truth-revealing mechanisms in their construction? The initial reaction of most economists to this question is that hypothetical questions will generate hypothetical answers that likely bear little relation to the "true" answers. Critics of stated preference (SP) methods point to numerous potential sources of bias. For example, it has been argued that survey respondents may ignore or downplay their budget constraint in answering hypothetical questions (e.g., Arrow, *et al.* [1983]; Loomis, Gonzales-Caban, and Gregory [1994], and Kemp and Maxwell [1993]). Additional criticisms include concerns that SP-based willingness-to-pay estimates fail to vary sufficiently with the scope of the resource being valued, the so-called "embedding effect" (e.g., Desvousges, *et al.* [1993] and Kahneman and Knetsch [1992]), and that they are inordinately sensitive to the elicitation format used (e.g., McFadden [1994] and Diamond and Hausman [1994]).

Yet, there is also substantial evidence that answers to carefully designed surveys contain at least some information content. At the most basic level, Mitchell and Carson [1989, pp. 206-207] note that valuation estimates based upon SP methods are typically correlated in the expected direction with those independent variables that theory predicts should influence consumer preferences. Thus, the demand for an environmental improvement is usually found to decrease with the cost of its provision and increase with the income level of the respondent. Of course, internal validity checks are not sufficient to justify a reliance on SP-based valuations. To further investigate their validity, researchers have turned to comparisons to valuation estimates based upon revealed preference (RP) data, including actual market transactions for related goods (e.g., Cameron [1992a], and Adamowicz, Louviere, and Williams [1994]) or simulated market transactions obtained in a laboratory setting (e.g., Cummings *et al.* [1997], and Cummings and Taylor [1999]). The shear number of such comparisons has grown dramatically over the past two decades. In summarizing 83 studies with 616 such comparisons, Carson *et al.* [1996] find that the ratio of SP to RP valuations for essentially to same good ranges from 0.005 to 10.269. However, the ratio generally lies near 1.00 and the RP and SP estimates are highly correlated (with a rank correlation between 0.78 and 0.92). For many, including the now infamous NOAA panel, these results hold out the hope that useful information can be gleaned from properly designed SP surveys. The question remains as to how best

to use that information: in isolation, calibrated to similar revealed preference results, or combined with revealed preference data.

In this paper, we address this issue in the context of valuing recreational wetland usage in Iowa. The empirical analysis draws on a 1997 survey of 6000 Iowa residents, which elicited both revealed and stated preferences regarding the use of wetland regions in the state. Like previous efforts to combine RP and SP data (e.g., Cameron *et al.* [1999]; McConnell, Weninger, and Strand [1999]; and Huang, Haab, and Whitehead [1997]), we develop a joint model of individual responses to both the RP and SP questions and use the resulting framework to test for consistency between the two sources of information. In doing so, however, we also emphasize testing specific hypotheses concerning sources of bias in *both* revealed and stated preference data. In carrying out these sets of hypothesis tests, we carefully investigate alternative interpretations of the results, noting particularly how the interpretation of the same results can vary with the "school of thought" of the reader. These arguments are in the vein of Leamer's [1983] famous discussion of alternative interpretations of the same data by those with different initial views of capital punishment.

The remainder of the paper is divided into six sections. Section II begins by reviewing the various schools of thought that have emerged in the literature in interpreting RP and SP data. The underlying behavioral and econometric models are detailed in Section III, along with hypothesis tests designed to test for specific sources of bias in each data source. The Wetlands database is then described in Section IV, with the empirical specification of the model detailed in Section V. Section VI provides the results of the estimation and hypothesis testing and Section VI concludes the paper.

II. Schools of Thought

Nearly two decades ago, Leamer warned of the potential fragility of econometric analysis to hidden "whimsical assumptions" and the delusion of the "…goal of objective inference" [1983, p. 37], arguing that the initial beliefs of a researcher too often drive the conclusions gleaned from a database. These concerns are no less relevant today and nowhere more apparent than in the debate over the relative merits of stated and revealed preference methods in valuing nonmarket amenities. With literally hundreds of comparisons between RP and SP valuations appearing in the literature, the prior beliefs of the analyst often color the hypotheses tested and inferences made from conflicts between the two sources of data. In this section, we describe three basic schools of thought that seem to have emerged.

The first school of thought, which we will refer to as the "RP-lovers" sect, views SP preference data with skepticism. In its extreme form, the group denounces stated preference methods, characterizing its use in the policy arena as misguided. Proponents of this view argue that "…contingent valuation surveys do not measure the preferences they attempt to measure…" and that "…changes in survey methods are not likely to change this conclusion" (Diamond and Hausman [1994, p. 46]). The less extreme portion of this sect holds out some hope for stated preferences to inform policy makers, while holding onto the sanctity of revealed preference data. McConnell, Weninger and Strand [1999, p. 201] echo this sentiment when they consider a joint model of RP and SP data.

"In other words, the preference structure that governs the RP decision may be completely independent from the preference structure that governs the SP decision. If so, the information

gleaned from SP methods is arbitrary and can lead to unreliable estimates of trip value. On the other hand, when the underlying preference structure is identical, both the RP and SP decisions provide useful information for valuing environmental amenities."

Thus, SP methods are only of value if they are consistent with their RP counterparts. Indeed, the notion of "calibrating" welfare estimates obtained from SP data is essentially a tenant derived from this school of thought, as it implies a calibration to revealed preference estimates. A smaller, though no less devoted, sect is the "SP-Lovers" school of thought. Much like its counterpart, "SP-Lovers" view with skepticism results based on commonly used revealed preference techniques, such as hedonic pricing and travel cost methods, noting their many sources of bias and limited applicability to the valuation of environmental amenities. Travel cost methods suffer, among other things, from "…value-allocation assumptions related to multi-purpose 'visits'; dependence of costs on assumptions concerning fixed/variable direct travel costs, costs (benefit?) of time in travel and on-site; and problems in obtaining values which are appropriately 'marginal' vis-à-vis the site/activity in questions." (Cummings, Brookshire, and Schulze [1986, p. 95]). As Randall [1994, p. 88] argues, the trip quantities employed in a travel cost model may be "revealed," but the prices necessary to derive trip demand functions, and hence welfare estimates, are not. Consequently, the travel cost method "…yields only ordinally measurable welfare estimates…" and "…cannot serve as a stand-alone technique for estimating recreation benefits." Similarly, the hedonic pricing methods is plagued by "…persistent collinearity between 'important' variables and extraordinarily low explanatory power in regression equations" (Cummings, Brookshire, and Schulze [1986, p. 96]). At the very least, this school of thought would argue that SP methods deserve equal footing in the pantheon of valuation techniques. At its extreme, it would hold that RP preference methods are not only fraught with limitations, but are simply not capable of capturing the range of use and nonuse values associated with environmental changes, leaving SP methods as the only game in town.

In recent years, a third school of thought has emerged, arguing for a reliance on both revealed and stated preferences (e.g., Cameron [1992a], Adamowicz, Louviere, and Williams [1994], and Kling [1997]). These "agnostics" suggest a shift in focus away from viewing RP and SP as competing sources of values and towards seeing them as complementary sources of information. In this view, both data sources illuminate, though imperfectly, consumer preferences for environmental amenities of interest. Revealed preference methods bring the "discipline of the market" to stated preference valuations, whereas stated preference methods can shed light on consumer preferences for price and quality attribute levels that are not currently observed in the market (Cameron, [1992b]). Discrepancies between the individual parameter estimates obtained using RP and SP estimates are not necessarily indicative of a failure of either method, but instead suggestive that the two sources are working in correcting the limitations inherent in each method. This ecumenical view would not fall back on either data source as "correct", but rely instead on their combination as providing the best source of welfare estimates.

III. Behavioral and Econometric Models of Preferences

In developing behavioral models underlying revealed and stated preference data sources, we focus our attention on a single recreation good (q), though obvious extensions exist to the multiple good situations. Individual i's "true" preferences are assumed to result in an ordinary (Marshallian) demand equation given by:

$$q_i = f\left(p_i, y_i; b\right) + e_i \qquad (1)$$

where q_i denotes the quantity consumed by individual *i*, p_i denotes the associated price, y_i is the individual's income, and b is a vector of unknown parameters. The additive stochastic term is used to capture heterogeneity in individual preferences within the population. Ideally, the analyst would have available data on the actual price and quantity for environmental good in question, along with quality attributes for that good and individual socio-demographic characteristics. Unfortunately, for both revealed and stated preference models, environmental valuation efforts often rely upon imperfect and indirect measures for these variables.

Revealed Preferences

Revealed preference models, such as travel cost models, rely upon surveys to elicit information on the number of trips taken to the site in question (q_i^R) and to construct proxies for the price of traveling to the site (p_i^R), including the cost of expended time in traveling to the site. The demand relationship modeled using this data is assumed to take the form:

$$q_i^R = f^R\left(p_i^R, y_i^R; b^R\right) + e_i^R. \qquad (2)$$

As is often the case in the literature, the error term is assumed to be drawn from a normal distribution, with $e_i^R \sim N\left(0, s_R^2\right)$. Standard econometric procedures can then be employed to obtain consistent estimates of the parameters of the function accounting for the censoring from the left at zero. Specifically, maximum likelihood estimation would rely upon the log-likelihood function

$$LL^R = \sum_{i=1}^{n}\left(D_i^R \ln\left\{ s_R^{-1} f\left[\frac{q_i^R - f^R\left(p_i^R, y_i^R; b^R\right)}{s_R} \right] \right\} + \left(1 - D_i^R\right) \ln\left\{ \Phi\left[\frac{-f^R\left(p_i^R, y_i^R; b^R\right)}{s_R} \right] \right\} \right) \qquad (3)$$

where $\Phi(\cdot)$ and $f(\cdot)$ are the standard normal cdf and pdf, respectively, and $D_i^R = 1$ if $q_i^R > 0; = 0$ otherwise.

There are, of course, numerous reasons for discrepancies between the demand relationships in equations (1) and (2). First of all, trip data are typically quantities based upon survey questions asking the individual to recall their number of trips over the past year. Recall errors and rounding by survey respondent contribute to differences between both the quantities and the error distributions in the two equations. Second, and perhaps of greater concern, are the discrepancies between the actual visitation costs (p_i) and those constructed by the analyst (p_i^R). It is standard practice to compute travel costs as the sum of the out-of-pocket cost of visiting the site in question (typically using a fixed cost per mile times the round-trip distance to the site) and a proxy for the value of the time used in getting to the site. Time is most often valued as a fixed fraction of the individual's wage rate (Cesario [1976]).[1] The reliance on this, or similar, conventions, drives a wedge between the individual's true demand equation in (1) and the estimated relationship in (2). In turn, this has led a number of critics to argue that, because "…visitation costs are inherently subjective…any particular welfare estimate is in part an artifact of the particular conventions selected for imposition" and hence suspect (Randall [1994, pp. 88-89]).

[1] There have been a number of efforts to better capture the value of travel time by estimating the appropriate fraction of the individual's wage rate (McConnell and Strand [1981] and Bockstael, Strand, and Hanemann [1987]) or the monetary value of leisure time (Feather and Shaw [1999]). However, the use a fixed wage fraction (between twenty-five and fifty percent) continues to dominate the applied literature.

Stated preference data can take a variety of forms, including direct value elicitations (i.e., contingent valuations) or contingent behavior data (e.g., providing information on how individual trip quantities might change as the price per trip or site attributes changed). We will focus on the latter type of SP data. In particular, suppose that in the process of gathering RP data, the survey respondents are asked: "How many recreation trips would you have taken to the site if the cost per trip increased by $B?" In this case, the resulting database looks identical to the RP data set, providing both price (p_i^S) and quantity (q_i^S) data for each individual. The preferences revealed through these questions can be represented by the demand equation

$$q_i^S = f^S\left(p_i^S, y_i^S; b^S\right) + e_i^S . \tag{4}$$

where $p_i^S = p_i^R + B$ and $e_i^S \sim N\left(0, s_S^2\right)$. Estimation of the demand relationship in (4) alone would proceed much like in the RP setting, with a log-likelihood function identical to that in equation (3), except that *R* would be replaced with *S* in all of the superscripts and subscripts. Efficiency gains can be achieved by recognizing the likely correlation between RP and SP responses (i.e., between e_i^R and e_i^S). The appropriate log-likelihood function for the joint estimation is given by:[2,3]

$$LL = \sum_{i=1}^{n}\left\{ D_i^R\left[\ln f\left(\frac{q_i^R - f_i^R}{s_R}\right) - \ln\left(s_R\right)\right] + D_i^R D_i^S\left[\ln f\left(\frac{\left(q_i^S - f_i^S\right) - q^S\left(q_i^R - f_i^R\right)}{s_S\sqrt{1-r^2}}\right) - \ln\left(s_S\sqrt{1-r^2}\right)\right]\right.$$
$$\left. + D_i^R\left(1 - D_i^S\right)\left[\ln\Phi\left(\frac{-f_i^S - q^S\left(q_i^R - f_i^R\right)}{s_S\sqrt{1-r^2}}\right)\right] + \left(1 - D_i^R\right)\left(1 - D_i^S\right)\ln\int_{-\infty}^{\frac{-f_i^R}{s_R}}\int_{-\infty}^{\frac{-f_i^S}{s_S}} f_2\left(h_1, h_2; r\right)dh_1 dh_2\right\} \tag{5}$$

where $Corr\left(e_i^R, e_i^S\right) = r$, $q \equiv r s_S / s_R$, $f_i^k = f^k\left(p_i^k, y_i^k; b^k\right)$ (k = *R*, *S*), and $f_2\left(\cdot,\cdot; r\right)$ denotes the standard bivariate normal pdf. This model can be used to test a variety of hypotheses concerning the consistency of the RP and SP data.

As with the RP data, there are numerous reasons one might expect preferences revealed by SP questioning, and modeled in equation (4), to differ from the individual's true preferences. As noted above, one of the most common concerns is that individuals fail to adequately account for their budget constraint and the availability of substitute commodities in responding to hypothetical scenarios presented to them in a survey. Thus, the income they perceive as being available for allocation towards the commodity in question may be inflated (i.e., $y_i^S > y_i$). In addition, individuals may have greater uncertainty regarding their preferences under unfamiliar price and quality settings; i.e., one might expect that $e_i^S = e_i + h_i^S$, where h_i^S captures the additional uncertainty or variability in preferences under the hypothetical conditions.

---

[2] The derivation of this log-likelihood function is available from the authors upon request.

[3] Louviere [1996] refers to this integration of RP and SP data sets having the same form as *pooling* models. A second category of models, *combining* models, integrates information from RP and SP data sets with different underlying forms. For example, Cameron [1992] combines RP fishing trip demand data with discrete choice SP data on whether individuals would take any fishing trips if a lump-sum access fee was imposed. Similar efforts have been undertaken by Larson [1990] and Huang, Haab, and Whitehead [1997].

Hypothesis Testing

There are two primary reasons for jointly estimating the RP and SP demand equations. The first and least ambitious is to garner improved efficiency in the parameter estimates by recognizing the likely correlation between the two data sets. It is hard to argue with this objective regardless of the school of thought from which one starts.

The second reason for combining the two models is to allow for hypothesis testing. Unfortunately, both the tests conducted and the interpretation of the results often depends upon the perspective brought to exercise. For example, the most common hypothesis tested is one that imposes complete consistency between the two models; i.e.,

$$H_0^1 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S . \qquad (6)$$

Traditionally, rejection of this hypothesis has been widely viewed as a repudiation of the SP data and methodology, though this is clearly an "RP Lovers" perspective. An equally valid viewpoint is that the discrepancies arise due to problems with the RP data; i.e., the "SP Lovers" perspective. These viewpoints also influence which model one falls back on for welfare analysis in the event that H_0^1 is rejected. Obviously, "RP Lovers" and "SP Lovers" would fall back on their respective preferred models. The "agnostics", on the other hand, might argue that discrepancies between the two models should in fact be expected. The RP data is bringing the discipline of the market to bear on the SP data and the SP data is examining regions of consumer preferences outside of range of consumers' experiences in the marketplace. From this perspective, welfare analysis is best served by relying upon the combined model using both sources of information.[4]

A modification to the overall consistency test allows for differences in the underlying variation in the two models, while constraining the demand parameters themselves to be equal; i.e.,

$$H_0^2 : \beta^R = \beta^S . \qquad (7)$$

Allowing for differences in the distribution of preferences in the revealed and stated preferences models has been suggested by a number of authors, including Haab, Huang, and Whitehead [1999] and Cameron *et al.* [1999].

In addition to testing for overall consistency between the two models, one can, as suggested by Cameron [1992b], focus on specific weakness of each model. For example, one might generalize the SP demand equation as follows:

$$q_i^S = f^S\left(p_i^S, \tau_y y_i^R; \beta^S\right) + \varepsilon_i^S . \qquad (8)$$

This replaces y_i^S in equation (4) with $\tau_y y_i^R$, where τ_y is a parameter to be estimated. The argument that respondents to SP questions tend to downplay their budget constraints is then reflected by the expectation that $\tau_y > 1$. The hypothesis test in this case would correspond to restricting (2) and (8) such that

[4] A weaker form of "agnosticism" might argue for presenting the welfare estimates obtained from the unconstrained RP and SP models as bounding the true values. An alternative ("atheistic") perspective is, of course, to view the rejection of H_0^1 as a rejection of both data sources, suggesting that neither model be used for welfare analysis.

$$H_0^3 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S, \text{ while } \tau_y \text{ remains unrestricted.} \quad (9)$$

This hypothesis test essentially treats the RP data as the base against which the SP data are tested.

An alternative approach is to reverses the roles of the RP and SP data. If the analyst believes the SP data are correct, but the RP data are subject to error, then the SP data can be used as the basis for a validity test of the RP data. As noted above, Randall [1994] has argued forcibly that the price term in RP data is poorly measured and is likely the cause of significant bias. An external validity test of the RP data can then be performed by substituting p_i^R in the RP models with $\tau_p p_i^R$, where $\tau_p > 0$, and generalizing equation (2) by using

$$q_i^R = f^R\left(\tau_p p_i^R, y_i^R; \beta^R\right) + \varepsilon_i^R. \quad (10)$$

Again, if τ_p is not estimated to be significantly different from one, external validity for the RP data would not be rejected. The corresponding hypothesis test would test the following restriction on equations (4) and (10):

$$H_0^4 : \beta^R = \beta^S \text{ and } \sigma_R = \sigma_S, \text{ while } \tau_p \text{ remains unrestricted.} \quad (11)$$

All four of the hypothesis tests identified above can be viewed as restrictions on the generalized system of demand equations in (8) and (10), with:

- $\tilde{H}_0^1 : \beta^R = \beta^S$, τ_y=1, τ_p=1, and $\sigma_R = \sigma_S$; i.e., complete consistency.
- $\tilde{H}_0^2 : \beta^R = \beta^S$, τ_y=1, and τ_p=1; i.e., consistency in demand parameters but not in terms of error variances.
- $\tilde{H}_0^3 : \beta^R = \beta^S$ and τ_p=1; i.e., when respondents answer the stated preference questions, in addition to having a different error variance, they also ignore their budget constraint. Consistency holds in all other respects.
- $\tilde{H}_0^4 : \beta^R = \beta^S$ and τ_y=1; when respondents answer the revealed preference question, in addition to having a different error variance, they also do not treat the computed travel cost term (*p*) as the cost of accessing the recreation site (analysts have calculated the incorrect price). Consistency holds in all other respects.

IV. Data

The data used to illustrate the above pooling model are drawn from a 1997 survey of 6,000 Iowa residents concerning their use of the state's wetland resources. The goal of the survey was to elicit information from respondents about their visits to, knowledge about, and attitude towards both the existing wetlands in Iowa, as well as efforts to preserve and expand these resources. Of the deliverable surveys, there was a fifty-nine percent response rate (with 3,143 surveys returned).[5]

Our analysis draws on the first section of the survey, which focused on visits to wetlands during the past year. After carefully defining what was meant by a "wetland," the survey asked respondents to recall the number of trips they had taken during the past year to wetlands in each of

[5] 594 surveys were returned by the post office as undeliverable. Details of the survey design and administration can be found in Azevedo [1999].

fifteen possible zones (see Figure 1). These zones, defined along county lines for convenience of the survey respondent, were designed to reflect major types of wetlands within the state.[6] Responses to this first question provide the basis for the revealed preference variable q_i^R. Individuals were then asked how their pattern of usage would change if the cost of visiting wetland zones near their residence (X, Y, and Z) increased.[7] In particular, they were asked to "…[c]onsider all of the recreation trips you made to wetlands in zones X, Y, and Z in 1997. Suppose that the **total cost per trip of each of your trips** to these areas had been $B more (for example, suppose that landowners charged a fee of this amount to use their land or that public areas charged this amount as an access fee)."[8] They were then asked to detail how their behavior would have changed with the increased cost, both in terms of reductions in visitations to areas X, Y, and Z, and changes in visits to other zones within the state. These questions provide the basis for constructing the SP quantity variable q_i^S.

While the surveys provide direct information on the trip quantities, the travel costs themselves must be constructed. The first step in the process was to establish travel time (t_i^z) and travel distance (d_i^z) for visits to the wetland zones (z = 1,…,15). One of the survey questions asked the respondent to place an X on a map, similar to Figure 1, locating their most recent wetland visit. The longitude/latitude coordinates for the visitation points in each of the 99 counties were averaged to find the mean visitation point in that county. Travel time and distance, for each respondent from their residence to each of the 99 mean visitation points, was then constructed using PC Miler, a software package designed for use in the transportation and logistics industry. This gave us a data set with 99 travel times and distances for each respondent in the database. Finally, zonal travel times (t_i^z) and distances (d_i^z) were calculated as a weighted average of their respective county travel county level values.[9]

For the purposes of this paper, we have focused our attention on a subset of the survey sample; i.e., those households in the Prairie Pothole Region of north central Iowa (zones 4, 5, and 8). We consider the aggregate number of trips to this region, with $q_i^k = q_i^4 + q_i^5 + q_i^8$ (k = R, S). Travel times and distances were formed as weighted averages of the zone specific values, where the weights used for individual i were the average percentage of trips to each zone among individuals in i's zone of residence. On average, for the 278 households with completed surveys in the Prairie

[6] For example, zones 4, 5, and 8 represent the prairie pothole region of north central Iowa. Pothole wetlands are the result of glacial activity and characterized by depressions in the land, most of which are less than two feet deep and filled with water for at least part of the year. In contrast, riverine wetlands dominate regions 1 through 3 and 13 through 15, and are associate with marshy land near rivers and streams.

[7] The wetland zones (X, Y, and Z) were assigned to each individual based upon the region of the state in which they lived. Specifically, the fifteen zones in **Figure 1** were grouped into five "megazones," reflecting regional wetland areas. The megazones were defined as the Missouri River Region (1,2,3), the Prairie Pothole Region (4,5,8), the Iowa River Corridor Region (9,10,11), the Mississippi Region (13,14,15), and the remainder of the state (6, 7, 12). The zones (X, Y, Z) for a given survey respondent consisted of those zones defining the megazone in which the respondent resided.

[8] The value of $B varied across the individuals surveyed, with bid values $5, $10, and $15 each randomly assigned to 20% of the sample and bid values of $20, $30, $40, and $50 each assigned 10% of the sample.

[9] Each county's weight was determined by the percentage of trips within that county's zone taken to that county. For example, zone three is made up of Pottawattamie, Mills, and Fremont counties. There were 92 trips taken to zone three. 41 trips were taken to Pottawattamie County, 24 trips were taken to Mills County, and 27 trips were taken to Fremont County. Therefore, the weight for Pottawattamie County was 0.45, the weight for Mills County was 0.26, and the weight for Fremont County was 0.29.

Pothole region, 8.2 trips (q_i^R) were actually taken. Respondents indicated that they would average only 2.7 trips (q_i^S) with the average hypothetical price increase (\$B) of \$27 per trip.

V. Model Specification

The parametric specifications for the demand functions in (8) and (10) were chosen to be of a simple linear form, with

$$q_i^R = b_0^R + b_p^R t_p p_i^R + b_y^R y_i^R + e_i^R \tag{12}$$

and

$$q_i^S = b_0^S + b_p^S p_i^S + b_y^S t_y y_i^S + e_i^S . \tag{13}$$

Clearly, given the linear form in equation (12), the parameters b_p^R and t_p are not separately identified. b_y^S and t_y are likewise not separately identified in equation (13). To deal with this issue, and to simplify the exposition below, we reparameterize the model as follows:

$$q_i^R = \tilde{b}_0^R + \tilde{b}_p^R p_i^R + \tilde{b}_y^R y_i^R + e_i^R \tag{14}$$

and

$$q_i^S = k_0 \tilde{b}_0^R + k_p \tilde{b}_p^R p_i^S + k_y \tilde{b}_y^R y_i^S + e_i^S . \tag{15}$$

In this case, the k_j's measure the discrepancies between the RP and SP demand equation parameters, with $k_j = 1 \; \forall j$ and $s_R = s_S$ corresponding to complete consistency between the two models.

Finally, the trip prices p_i^R and p_i^S must be specified. As suggested by McConnell and Strand [1981], we parameterize these prices as:

$$p_i^k = c_i + l^k L_i \quad k = R, S \tag{16}$$

where l^k is a parameter to be estimated, $c_i \equiv 0.21 d_i$ denotes the travel cost (computed as roundtrip travel distance times 21¢ per mile), and $L_i = w_i t_i$ denotes full time costs (computed as the individual's wage rate times the roundtrip travel time). As noted above, it is standard practice in the recreation demand literature to fix l^k at some value, typically between 0.25 and 0.5, rather than estimating its value.

Substituting (16) into equations (14) and (15) results in the following system of demand equations:

$$q_i^R = \tilde{b}_0^R + \tilde{b}_p^R \left(c_i + l^R L_i \right) + \tilde{b}_y^R y_i^R + e_i^R \tag{17}$$

and

$$q_i^S = k_0 \tilde{b}_0^R + k_p \tilde{b}_p^R \left(c_i + k_l l^R L_i \right) + k_y \tilde{b}_y^R y_i^S + e_i^S . \tag{18}$$

where $k_l \equiv l^S / l^R$. The hypothesis tests corresponding to those outlined above become:

- Complete Consistency. $\tilde{H}_0^1 : k_j = 1 \; \forall j$ and $s_R = s_S$.
- Heteroskedasticity Hypothesis. $\tilde{H}_0^2 : k_j = 1 \; \forall j$.
- Price Hypothesis. $\tilde{H}_0^3 : k_0 = 1, \; k_l = 1, \text{ and } k_y = 1$.

- Income Hypothesis. $\tilde{H}_0^4 : k_0 = 1,\ k_p = 1$ and $k_I = 1$.

VI. Results

The system of demand equations in (17) and (18) was estimated using standard maximum likelihood procedures, with the likelihood function in (5) correcting for the potential censoring at zero for either or both RP and SP trip data and accounting for possible correlation between the two error terms. The resulting parameter estimates are provided in Table I for the unconstrained model and for each of the constrained specifications outlined above. For the stated preference parameters, we provide both the estimated values for the k_j's and the implied levels for the stated preference parameters $\tilde{b}_j^S \equiv k_j \tilde{b}_j^R$ in square brackets.

Focussing first on the unconstrained model, we note that all of the parameters have the expected signs. The price coefficient for the revealed preference data is just above negative one and statistically significant. While the income coefficient is small and statistically insignificant, it is positive and with a small standard error. Interestingly, the coefficient on the wage rate, which is traditionally fixed at level between 0.25 and 0.50, is also close to zero. Given these parameter estimates, the probability that l^R exceeds 0.25 is less than 0.1%. Indeed, our estimate of l^R is not statistically different from zero at any reasonable confidence level.

The stated preference parameters differ from their revealed preference counterparts in a number of areas. In particular, both the intercept and price terms are significantly different, with both k_0 and k_p significantly less than one. This indicates, for example, that the price response in the SP data set is much smaller than in the RP counterpart. The remaining k_j's are all greater than one, though not significantly so. The estimate of k_I suggests that greater weight is placed on the wage rate in the stated preference choices than in their revealed preference counterpart, with the point estimate for $l^S = 0.88$. The underlying variability of preferences revealed by the two data sets is similar, with $k_s \equiv s_S / s_R$ estimated to be close to one and precisely measured. This result is in contrast to other efforts integrating RP and SP data sets that found the source of inconsistency between the two data sets to be in the degree of preference dispersion (e.g., Cameron *et al.* [1999] and Haab, Huang, and Whitehead [2000]). Finally, we note that there is substantial correlation between the two data sources, with $r = 0.72$ and significantly different from zero.

While parameters of the demand equations are important, they are typically used in the policy arena as an intermediate step to the computation of welfare impacts. As one might expect, the substantial differences in price coefficients between the RP and SP models lead to substantial differences in the associated welfare measures. In Table I, we provide estimates of the average consumer surplus associated with wetland visitations. The surplus derived from the SP model (CS_S) is over two and half times that obtained using the revealed preference model (CS_R).

Turning to the constrained models, we find that each of the model restrictions is rejected at any reasonable significance level. Consistency between RP and SP data is not borne out by the data, either in its strict form or when we allow for different levels of variances for two data sources. Indeed, even in its unconstrained form, the demand systems suggest $s_R = s_S$ (with $k_s = 1.04$). Less

restrictive forms of consistency are no more successful. The price hypothesis, which an SP-Lover might argue for, suggests that the source of the inconsistency between the two demand systems lies in the formation of the trip price. Yet this hypothesis is soundly rejected. Similarly, the income hypothesis, which our RP-Lover might suggest, argues that the source of the inconsistency lies in the lack of attention paid to income by stated preference survey respondents, but this hypothesis is also rejected.

The problem, of course, is where do we go from here? We have two sources of data with conflicting conclusions regarding the value of wetland use. One can always hope that the two valuations lead to the same policy conclusions. If they do not, one's "school of thought" regarding the RP and SP controversy will play a key role. Moderate RP- and SP-Lovers are likely to stop at the unconstrained model, taking solace in the improved efficiency achieved by jointly estimating the two demand systems, and relying on their respective valuations (CS_R and CS_S). Indeed, they may never even estimate the various constrained models. The zealots within each group, however, will rely on tests as a source of vindication for their respective positions. For the RP-Lover, the various consistency tests only serve to solidify their suspicion of the stated preference approach. However, these same results can be employed by the SP-Lover to argue against revealed preference models by simply reversing the presumption as to which is the correct model!

In contrast, the agnostic's approach to the results in Table I is likely to be substantively different. Rather than falling back on the unconstrained model in the face of the various rejections of consistency, they are more like to embrace the specification under $\tilde{H}_0^1$ or $\tilde{H}_0^2$. According to the agnostic, the rejection is expected. Indeed, the agnostic might even be disappointed if the constrained parameter estimates were not significantly different from the unconstrained ones, as that would indicate that the gains sought by combining data sources did not occur! That is, the SP data was not carefully enough designed to "fill in" missing pieces from the RP data and/or the RP data had sufficient measurement error or other problems such that it could not impose market discipline on the SP data. The agnostic's preferred strategy, would be to jointly estimate the models, imposing consistency across all parameters, or possibly allowing for difference variances (the complete consistency or heteroskedasticity hypothesis in Table 1). By so doing, all information available on consumers' preferences is best utilized.

VII. Summary and Conclusions

This paper has investigated the combining of revealed and stated preference data to jointly estimate the parameters of consumer preferences for a public good. We have presented a simple model for combining two different data sources and demonstrated how several hypotheses concerning these data sources can be formalized and tested in such a model. In the wetlands application discussed here, consistency between the RP and SP data was consistently rejected. In interpreting the lack of consistency found between the estimated parameters, we identified and discussed several "schools of thought." Analysts who identify with RP-Lovers will be inclined to take parameter and welfare estimates from RP data as the most accurate representation of the underlying true parameters, interpreting consistency tests with SP data as evidence for or against the validity of SP as a method. In contrast, SP-Lovers will view welfare and parameter estimates from SP data as the best available estimates and interpret consistency tests with RP data as referendums on RP's validity. In contrast, those identifying with an "agnostic" approach will be inclined to see

consistency tests as relatively unimportant, but instead prefer to combine data sources to estimate a single set of parameters and welfare estimates, thus enjoying the strengths of each method. Clearly in each of these three approaches, the role and interpretation of consistency tests in nonmarket valuation is closely related to the prior point of view of the analyst.

The fundamental debate concerning the legitimacy of SP methods relative to RP discussed in the introduction will continue to be much discussed. However, we are not optimistic that consistency tests between RP and SP data such as those described here will play a key role in settling the disagreement between those who see SP data as inherently flawed and those who see it as the last, best hope for nonmarket valuation. It is simply too easy to interpret the results of these tests to be consistent with one's prior views.

On the other hand, combining data sources holds considerable promise for those who are willing to credit each method with strengths and weaknesses. By designing questionnaires to elicit RP and SP data to take advantage of their respective strengths and then combining the data sources in estimation, analysts may be able to provide policy makers with more efficient and accurate estimates of the value of public goods. This research agenda is only begun and will require significant effort on the part of environmental economists to identify the parameters under which improved welfare estimates can be obtained through such methods.

References

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

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register,* 58(10), 1993, 4601-4614.

Azevedo, C. D., "Linking Revealed and Stated Preference Data in Recreation Demand Modeling." Ph.D. dissertation, Iowa State University, 1999.

Bockstael, N., I. Strand, and M. Hanemann. "Time and the Recreation Demand Model." *American Journal of Agricultural Economics*, May 1987.

Cameron, T. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics*, August 1992a, 302-317.

Cameron, T. "Nonuser Resource Values." *American Journal of Agricultural Economics*, December 1992b, 1133-37.

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze. "Alternative Nonmarket Value-Elicitation Methods: Are Revealed and State Preferences the Same?" Working Paper, UCLA, December, 1999.

Carson, R. T., N. E. Flores, K. E. Martin, and J. L. Wright. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics*, February 1996, 80-99.

Cesario, F. J. "Value of Time in Recreation Benefit Studies." *Land Economics*, February 1976.

Cummings, R. G., D. S. Brookshire, and W. D. Schulze. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Totowa, N.J.: Rowman & Allandeld, 1986.

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

Cummings, R. G., and L. O. Taylor. "Unbiased Value Estimates for Environmental Goods: Cheap Talk Design for Contingent Valuation Method." *The American Economic Review*, June 1999, 648-665.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, S. P. Hudson, K. N. Wilson, and K. J. Boyle. "Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy," in *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, New York: North Holland, 1993, 91-164.

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

Feather, P., and W. D. Shaw. "Estimating the Cost of Leisure Time for Recreation Demand Models." *Journal of Environmental Economics and Management*, July 1999, 49-65.

Haab, T. C., J. C. Huang, and J. C. Whitehead. "Are Hypothetical Referenda Incentive Compatible? A Comment." *Journal of Political Economy*, February 2000, 186-196.

Hanemann, W. M. "Valuing the Environment through Contingent Valuation." *Journal of Economic Perspectives*, Fall 1994, 19-43.

Huang, J., T. Haab, and J. Whitehead. "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data be Combined?" *Journal of Environmental Economics and Management,* November 1997, 240-255.

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

Kemp, M. A., and C. Maxwell, "Exploring a Budget Context for Contingent Valuation." in *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, New York: North Holland, 217-70.

Kling, C. L. "The Gains from Combining Travel Cost and Contingent Valuation Data to Value NonMarket Goods." *Land Economics*, August 1997, pp. 428-39.

Larson, D. "Testing Consistency of Direct and Indirect Methods for Valuing Nonmarket Goods," University of California, Davis, draft manuscript, December 1990.

Leamer, E. E. "Let's Take the Con out of Econometrics." *The American Economic Review*, March 1983, pp. 31-43.

Loomis, J., A. Gonzalez-Caban, and R. Gregory. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics*, November 1994, 499-506.

Louviere, J. "Combining Revealed and Stated Preference Data: The Rescaling Revolution," paper presented at the AERE conference, June 1996.

McConnell, K., and I. Strand. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sportfishing." *American Journal of Agricultural Economics*, February 1981, 153-156.

McConnell, K., Q. Weninger, and I. Strand. "Joint Estimation of Contingent Valuation and Truncated Recreational Demands," Chapter 7 in *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, edited by J. A. Herriges and C. L. Kling, Cheltenham, U.K.: Edward Elgar Press, 1999, 199-216.

McFadden, D. "Contingent Valuation and Social Choice" *American Journal of Agricultural Economics*, November 1994, 689-708.

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

Portney, P "The Contingent Valuation Debate: Why Economists Should Care." *Journal of Economic Perspectives*, Fall 1994, 3-17.

Randall, A. "A Difficulty with the Travel Cost Method." *Land Economics*, February 1994, 88-96.

Figure 1. Wetland Zones

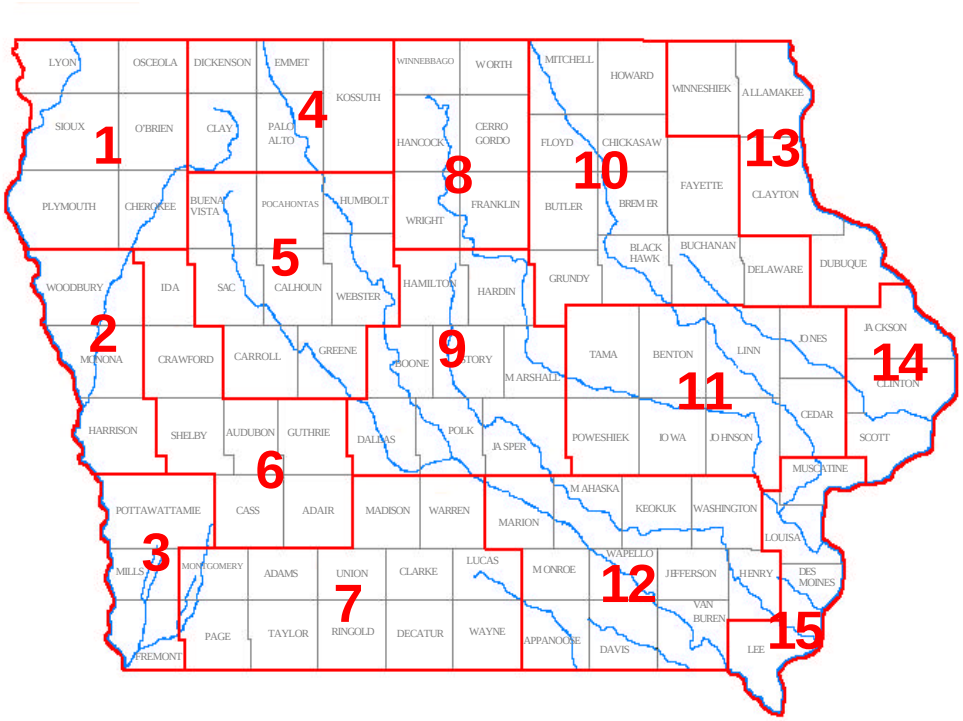

Table I. Parameter Estimates (Standard Errors in Parentheses)

Parameter	Unconstrained	Complete Consistency $\tilde{H}_0^1$	Heteroskedasticity Hypothesis $\tilde{H}_0^2$	Price Hypothesis $\tilde{H}_0^3$	Income Hypothesis $\tilde{H}_0^4$
b_0^R	23.38** (2.91)	14.52** (1.94)	14.22** (2.08)	13.47** (2.40)	13.77** (2.26)
b_p^R	-0.94** (-0.13)	-0.47** (-0.04)	-0.46** (-0.06)	-0.42** (-0.09)	-0.44** (-0.06)
l^R	0.06 (0.06)	0.44** (0.11)	0.47** (0.14)	0.53** (0.19)	0.50** (0.15)
b_y^R	0.08 (0.05)	0.21** (0.05)	0.22** (0.05)	0.22** (0.05)	0.22** (0.05)
s_R	13.84** (0.73)	14.34** (0.78)	14.48** (0.80)	14.47** (0.84)	14.47** (0.85)
k_0	0.48[a] [11.22] (0.14)	1.00	1.00	1.00	1.00
k_p	0.38[a] [-0.36] (0.07)	1.00	1.00	1.04 [-0.44] (0.08)	1.00
k_l	14.69 [0.88] (15.55)	1.00	1.00	1.00	1.00
k_y	3.33 [0.27] (2.18)	1.00	1.00	1.00	0.93 [0.21] (0.12)
k_s	1.04 [14.39] (0.09)	1.00	0.98 [14.21] (0.08)	0.99 [14.40] (0.08)	0.98 [14.24] (0.07)
r	0.72** (0.04)	0.72** (0.04)	0.72** (0.04)	0.73** (0.04)	0.73** (0.04)
-log L	1114.37	1127.56	1127.50	1127.32	1127.35
P-values		<.001	<.001	<.001	<.001
CS_R	99.61	197.39	203.50	222.95	212.25
CS_S	264.65	197.39	203.50	213.52	212.25

** and * denote significance at 99% and 95% confidence levels respectively, a and b denote significant departures from 1.00 at 99% and 95% confidence levels.

Discussion of Session II Papers

by Kelly Brown, U.S. EPA, National Center for Environmental Economics

Introduction

The title of this session is "Validity of Stated Preference Methods." I will begin my comments by providing a brief overview of how I see this session contributing to the overall theme of the workshop. Then I will discuss how each of the papers in this session contributes to this theme, being sensitive to the fact that the presenters did not necessarily choose this particular session and therefore may see their research fitting into the picture in a slightly different manner. My task is to comment on this research from a policy perspective and therefore these comments will focus on the usefulness of the research to environmental policy decision making.

Validity refers to the degree to which a method produces the desired results. For our purposes, validity is the degree to which stated preference methods produce a measure of the economic value associated with an environmental good or service of interest. One of the reasons for devoting an entire session to this topic is because research has suggested that perhaps the current state of the art does not provide a valid measure of economic values associated with non-market goods; stated preference results may not reflect individuals' true willingness to pay for environmental goods or services. This is a serious issue for policy makers attempting to make decisions over devoting resources to clean water or clean air, for example, or trying to convey the total benefits of a particular environmental regulation to the public or fellow government agencies charged with the task of collecting such information.

As we are all well aware, the hypothetical nature of stated preference methods does not necessarily provide us with a result that has been proven to consistently represent true willingness to pay. The question then becomes if, and how, we can use stated preference methods to obtain meaningful measures of economic value or how to modify the methods so as to achieve this result. Both approaches are seen in the research presented in this session. Swallow, et. al. conduct real and hypothetical experiments using different payment mechanisms to determine if the disparity between treatments is affected by the payment mechanism. Herriges and Kling, as well as Stevens, et. al. (presented by John Halstead) compare stated preference results to results from surveys using other valuation techniques to determine if, and how, values differ according to the valuation mechanism. Finally, Mansfield, et. al. analyze the use of public goods as an alternative to monetary compensation for public bads.

Each of these papers brings us closer to making better use of stated preference results, without actually supporting or debunking the method as a whole. I will now turn to a discussion of each of the papers in this session.

Real versus Hypothetical Treatments

Swallow, et. al. examine how validity might be achieved by modifying the payment mechanism used in the survey. To briefly recap, they conduct two sets of experiments. The first experiment is an in-person survey of University of Rhode Island students regarding their values associated with an adopt-a-pond program. Swallow, et. al. use a provision point mechanism with a variety of rebate plans to elicit values. Students are randomly assigned to either a real or hypothetical treatment and within each treatment students are randomly assigned to one of three

rebate plans. The rebate options allow them to assess if, and how, various payment programs affect the disparity between real and hypothetical results. It is this hypothesis that provides the most useful contribution to assessing the validity of stated preference methods.

Unfortunately, despite the very preliminary nature of their results, it does not appear as though the introduction of the various rebate plans "corrects for" the hypothetical bias. This, of course, would be a lofty goal, at best. The degree to which respondents find the payment mechanism realistic is an important aspect of survey design. However, the fact that the results do not appear to move stated preference results closer to a "true" willingness to pay indicates that it is not the payment mechanism that is causing hypothetical and real results to diverge (a familiar result), but that there is some other feature of the method that is at the root of the differences. That is, if Swallow, et. al. were to have found no significant difference between real and hypothetical responses under the provision point mechanism, then there would be evidence that the hypothetical bias might be driven by the payment mechanism, as opposed to say, the hypothetical nature of the survey. Unfortunately, for the stated preference practitioners, this was not the case.

Swallow, et. al. also conduct a second experiment to examine willingness to pay values for a new wetland conservation area in two different Rhode Island towns. Individuals are randomly assigned to one of six treatments distinguished by their real and hypothetical combinations. Again, this experimental design does not introduce any new aspects to the survey methodology. The most useful aspect of this study design is the treatment of 'no' responses in which individuals are asked to further delineate their views in terms of not wanting to pay for the wetland at the specified costs, not valuing wetland protection, and a desire to save their money. Analysis of 'no' responses is merited as researchers are not always sure what a no, or yes, response means. The results of this research are too preliminary to analyze, however, the technique looks promising for helping researchers better understand the variance underlying responses.

Stated Preference versus Other Valuation Methods

Next, I turn to research that attempts to extract meaningful information from stated preference results (as opposed to altering the mechanism itself, as seen above). First, I will discuss Herriges and Kling's research. They address validity by comparing stated preference results to results from a revealed preference survey. The idea behind this approach to assessing validity is to compare real market data with hypothetical market data in hopes of gleaning some information about the hypothetical market. Any discrepancies between the two indicate, perhaps, some flaws, in one, or either method.

Herriges and Kling develop a recreational demand survey that includes a hypothetical question about how behavior would change in response to a change in the price of travel to a state wetland. In other words, they collect real data (how many times did you go to the swamp) and then they have a measure of willingness to pay in the event that a parameter changed (cost). By estimating a joint model they are able to test the consistency between the revealed preference and stated preference data, where the revealed preference data is reflected by questions about previous behavior and the stated preference data is reflected by responses to hypothetical changes in behavior.

Their results reject a hypothesis of consistency between the values revealed by the revealed and stated preference questions. That is, the data do not provide similar information on values.

They also test several other hypotheses related to constrained models and find that these are also rejected. Contrary to the disappointment with this result expressed by Joe Herriges during his presentation, I do not find these results surprising. Revealed preferences provide information on behavior that has taken place in a previous time period, where as stated preferences indicate how an individual anticipates behaving in the future, under different conditions. These are two very different notions and methods trying to compare the two should not necessarily expect to similar results. The researcher does not know if the individual is anticipating a large raise, better health, or merely wants to get out more in the future. Likewise, the researcher does not know if the individual had anticipated participating in more recreational pursuits in the past, but did not do so because of a family emergency or a flat tire. It is not at all unreasonable to observe individuals' thinking about the future in a different light than the past, hence anticipating different behavior than that which has been exhibited.

What does this mean for stated preference methods? Research of this approach addresses the best (or worst) of both worlds. That is, the revealed preference questions "discipline" the stated preference questions by providing a real market context, while at the same time allowing the individual to incorporate other factors into their valuation decision, such as anticipated future changes in behavior or non-use values. The within-sample approach to valuation provides a nice range of values that should not necessarily be rejected because they differ. From a policy perspective I think the revealed preference/stated preference option is a practical alternative that provides a range of values. Researchers may even expect stated preference results to provide an upper bound on values because they incorporate the non-use valuation components.

Stevens, et. al. (presented by John Halstead) also examine the validity of stated preference methods by comparing results from traditional contingent valuation surveys to results from a conjoint analysis. This study employs both a conjoint question and a contingent valuation question regarding willingness-to-accept (and willingness-to-pay in one treatment) for reduced visibility in exchange for lower electric bills. The authors find that contingent valuation (CV) results and conjoint analysis (CA) results yield different results. I focus my comments on the differences as a result of the treatment (as opposed to some of their other hypotheses regarding survey mode and the like). First, they find that CV-CA depending on a somewhat arbitrary measure of indifference. That is, when they impose the restriction that a YES vote is equivalent to situations where people rated B>A they find no difference between the results from the two methods. This clearly needs some more work because such arbitrary distinctions are difficult to adopt on a universal basis. One suggestion I have is as follows. CA provides information on how individuals feel about both the status quo and the change, whereas CV provides information on the change, but not the status quo, assuming that the status quo is acceptable. I propose asking individuals if they are willing to accept the status quo and then use this information to reveal something about their indifference. In other words, consider the following, rough, experimental design.

CV:	status quo	change	
	a. YES	YES	=>1 or 2
	b. YES	NO	=> 3
	c. NO	YES	=> 1
	d. NO	NO	=> 3 or 2
CA:	1. A<B		=> YES
	2. A=B		=> indifferent
	3. A>B		=> NO

In a CV survey there are four possible outcomes. An individual can have positive or negative preferences for the status quo (as indicated by YES and NO under the status quo column above), but only states a positive or negative preference for the proposed change. Likewise, in the CA there are three possible outcomes: B is strictly preferred to A (option 1 under CA, above), A is equivalent to B (option 2), or option A is strictly preferred to B (option 3 above). There is no one-to-one mapping between the CV and CA, even though the analysis attempts to make such a comparison. In the CV questions, researchers only observe a YES or NO to the change. Indifference is typically not allowed. Whereas, in the CA questions, indifference is allowed, however, it is arbitrarily considered a YES or a NO.

By asking individuals in the CV context, about the status quo, researchers are gaining more information about underlying preferences and can map the indifferent individuals (those who answer YES or NO to both the status quo and the change) to those who are indifferent to the CA choices.[1] In addition, I would argue against arbitrarily lumping indifferent individuals in the CA analysis into the YES (or NO) category. This design helps to expand the available information and may provide more insight into the differences, if any, between CV and CA results.

The Stevens, et. al. study also provides some useful information on the choices individuals are given when valuing visibility benefits. The relatively few people who actually agreed to accept the compensation does indicate that actual values may be higher than provided by the change in electric bills and other forms of compensation should be considered. Also, it is clear from the follow-up questions that people have a difficult time untangling visibility from the other issues associated with poor air quality and this should not be ignored.

Alternative Compensation

Finally, turning to the Mansfield, et. al. paper, these authors offer a unique alternative money in compensating for public bads. That is, they introduce the idea of offering a public good as opposed to monetary compensation for public bads to restore utility to individuals. Their results indicate that people are more willing to accept public goods than cash in the presence of public bads and this could serve to be useful in order to gain support for politically unpopular projects, such as siting of landfills. I have a few comments on the survey design itself. The paper does not indicate if individuals were told the total cash award that would be made, as well as their individual award. If given the choice between $1000 and a park, individuals might think that a park is worth more than $1000, not actually accounting for the fact that each individual would receive $1000, or considering whether their own personal use is valued at $1000.

[1] Obviously, individuals who answer YES to both CV questions are expressing very different preferences from those who answer NO to both CV questions. This would need to be addressed in the analysis.

In addition, this method does not provide information on the actual values associated with the good (or bad) in question. It is important to determine the value of the public bad and then determine a reasonable valued alternative to compensate, which is more burdensome than coming up with a single value in the first place. However, it is perhaps more equitable to use public goods as opposed to cash for such compensation because use then is not determined by some ability to pay.

Summary

Where do we stand on the issue of validity? None of these studies provides any overwhelming support or disapproval of stated preference methods. These studies do, however, move the state of the art forward in terms of the ways in which questions are asked to derive values, be it with alternative payment mechanisms or tradeoffs with other public goods. In addition, revealed preference/stated preference combinations will help us move closer to identify ranges of useful values. This research also suggests that it is useful to continue moving the state of the art forward both in terms of tailoring the method itself and expanding alternative ways to conduct stated preference research.

Discussion of Session II Papers

by Anna Alberini, University of Maryland

I. Introduction

The papers presented in the session "Validity of Stated Preference Methods" investigate various aspects of the validity of stated preference data and methods. Specifically, they examine:

(i) Willingness to Pay (WTP) v. Willingness to Accept (WTA) compensation measures (Stevens et al paper, Mansfield et al paper);

(ii) The effects of different survey format, different elicitation approaches (conjoint analysis v. dichotomous choice contingent valuation), populations at different distance from the amenity to be valued (Stevens et al paper);

(iii) Stated preference data v. revealed preference data (Azevedo et al paper);

(iv) Hypothetical v. actual WTP under settings with various degrees of incentive compatibility (Swallow et al paper).

Regarding (i) and (iv), a large body of theoretical literature has been developed that explains the empirical results, and so the ideal output of the research is to discriminate between theories, provide support for or against each theory, and/or identify situations where the existing theories are insufficient to explain the observed behaviors.

For example, we expect to see that WTA for a degradation in environmental quality is greater than WTP to avoid the degradation or obtain a similar improvement. This is the finding of much earlier empirical work, and is consistent with the theory of prospects (Tversky, 1979) and the work by Hanemann (1991) showing that the difference between WTA and WTP depends on the elasticity of substitution between a public good and its private substitutes. In empirical work presented in this session, one would therefore expect to see results compatible with these findings. One would also expect to see that if some of the factors advocated by theory (e.g., elasticity of substitution between the public good and private substitutes) are removed or attenuated (when valuation concerns only private good, or when private substitutes are available), the difference between WTP and WTA should become smaller.

Regarding (ii) and (iii), one would expect to see that, provided that survey instruments are carefully developed and pretested, written, self-administered questionnaires give the same WTP results as computer-assisted in-person surveys. Similarly, one would hope to see that stated preference data and revealed preference (e.g., trips to wetlands, contingent on a certain cost per trip, and actual trips) subsume the same underlying utility functions.

If these are the expectations, is the research presented in this session confirming our expectations? Because most research projects are still underway, it is still too early to say. If the expectations are not borne out in the data, it will be important to understand the reasons why.

The remainder of this discussion is organized as follows. First, I briefly discussed each of the individual papers in sections II, III, IV and V. Finally, I summarized what I have learned from the research projects here presented and offer some lessons for future research.

II. Comments on the paper "Compensating for Public Harms: Why Public Goods Are Preferred to Money" by Carol Mansfield, George Van Houtven and Joel Huber

The purpose of this paper is to determine whether the siting of facilities imposing environmental disamenities can be made more acceptable to a community by offering public goods in lieu of cash compensation.

The paper begins with a nice summary of theories and empirical reasons why willingness to accept to forgo the siting of noxious facilities has been exceedingly high, and why in some studies the rate of acceptance of the proposed siting did not increase—and in fact, decreased—with cash offers. Reasons include the facts that:

cash offers may be perceived as bribes, and as such they carry a stigma;

cash offers crowd out people's willingness to accept the siting of a noxious facility for the good of the community;

accepting a cash offer may be seen as accepting personal responsibility for the negative consequences of an action, while accepting a public good may be seen as distributing such a responsibility over a much larger group of people.

Accepting a public good may be seen as prioritizing other people's well-being more highly than one's own, which is a desirable feature.

All of these explanations, of course, build nicely on the theory of prospects (Tversky, 1979), which suggests that people value differently losses from gains, and on Hanemann's theoretical model which implies that the difference between WTP and WTA depends crucially on the elasticity of substitution between a public good and its private substitutes. They are also compatible with the discussion offered in Mitchell and Carson (1989), who point out that people may report very high WTA values when they are described scenarios that contradict their perception of who has the property right over environmental quality.

In addition, the scenarios constructed by the authors for this research suggest additional reasons why people may hold very high WTA values, or refuse cash payments altogether:

(a) Fear that the agreement whereby cash is paid to individuals may be not be binding for the other party: if the latter does not wish to honor its commitment, an individual has little recourse, but a community who has been denied the public good promised to them will find it easier to put pressure over that party, either informally or through formal channels;

(b) Fear that acceptance of a cash payment opens the door to having to make more concessions to the party seeking to build the noxious facility, an issue potentially related to (c) below;

(c) Uncertainty over what one is really committing to, especially if the sponsors of the survey are not known.

The Experimental design devised by the authors consists of hypothetical scenarios describing noxious facilities being proposed for siting near the respondent's home, in exchange for either (a) cash

compensation or (b) another public good. For each scenario, neither the cash amount nor the public good were varied to the respondents.

In addition, the authors construct a neutral market choice, whereby the respondent must choose between houses located in neighborhood A or B, the houses and neighborhoods being identical in all respects, except for the property tax and the level of a public good. The public bad is the same at both locations, or absent at both locations.

Finding 1. The authors' prior implies that one should expect that the percentage willing to accept public good, among those respondents who were offered the public good as a compensation for accepting siting, is greater than the percentage of respondents willing to accept cash. In practice, depending on the sample and survey format, the former is in some cases greater than and in other cases less than the latter, implying that prior expectations cannot be confirmed or refuted.

It is difficult to draw conclusions from the percentages reported in table X for 3 reasons. First, it is unclear that respondents necessarily understood the scenarios as intended by the researchers.[1] Second, there are many differences in the demographics in the various areas where the study was conducted, and the mall intercept recruiting frame loses the researcher's ability to control for the population. Third, the cash amounts offered as compensation is not varied to the respondents. I believe that more conclusive findings would have been possible, had the cash amounts been varied to the respondents (holding the scenario the same). Doing so would allow the researchers to show if the percentage of respondents willing to accept the dollar amount increases with the dollar amount, or remains virtually unchanged over a wide range of dollar figures.

It is also difficult to say what aspects of a public good prompt the respondents to accept it in exchange for allowing the siting of the facility in their neighborhood. This point and the previous one suggest that perhaps future research might want to conceive this study in terms of choice experiments, where people choose between A, a scenario that offers cash as compensation, B, a scenario that offers a public good as compensation, and a third option allowing them to reject the facility. Respondents would be asked to evaluate several triplets, with varying cash amounts and attributes of the public good. This approach allows the researchers to find out what tradeoffs respondents are making, and would improve the efficiency of the estimates of the coefficients of the underlying utility function. It would allow to compute marginal rates of substitution between different attributes of the public good, and would allow the researcher to estimate the WTA associated with each proposed scenario.

Finding 2. The neutral market choice is in my opinion the most promising part of the experimental design. It avoids the "status quo bias" likely to affect responses in the hypothetical siting scenarios described to the respondents. Future research could attempt to estimate WTA separately from the hypothetical scenario and neutral market choice. Provided that this is done carefully, holding the other aspects of the hypothetical siting scenario and neutral market choice the same (including the payment vehicle, which should be described in terms of savings on property tax in both), any differences in WTA would be ascribed to the formation of a "reference point" and the related "status quo bias."

[1] To elaborate on this point, in the Superfund cleanup scenario people may have been thinking of increased health risks posed by the cheaper barrier method, as opposed to complete cleanup. In the scenario where the neighbors of a farm are offered a recreational facility in exchange for allowing the farm to become a livestock operation, a recreational facility might be an unrealistically expensive option to appease neighbors for the odors associated with the livestock operation. I was also surprised to read about a scenario describing the burning of waste yard in Greensboro—shouldn't waste yard be composted, or taken to landfills, or incinerated at a facility which captures the smoke, rather than just burned creating smoke that bothers residents?

Once again, this goal could be accomplished only if the dollar amounts are varied to the individuals, holding all else the same. Varying the dollar amounts across individuals also helps rule out another possible cause for the key finding reported by the researchers--that the percentage of people that in the neutral market choice choose cash over the public good *is* influenced by the presence or absence of a public bad. The researchers interpret this finding as evidence that the marginal utility of a public good is altered by the presence of a public bad.

To illustrate, the authors ask respondents to choose between identical houses at two locations with different level of public goods and the same level of public bad (either the specified level, or none at all).

The utility from choosing one of the two locations is:

(1) $U = a + b \cdot PG + g \cdot PB + d\,(t - \text{tax saving}) + e$

with α the utility from the house *per se*, which remains identical across the two locations; PG is a dummy for the presence of the specified public good; PB is a dummy for the presence of the specified public bad, t is the total property tax liability, and the tax savings (e.g., $500 a year) is specified in the survey, and applies only to one of the two locations. Here, γ and δ are negative, while β is positive.

When comparing the choice between house A and house B, where both have the public bad, but only the former has the public good, the difference in the underlying utility levels is:

(2) $\Delta U = b + d \cdot \text{tax saving} + e_A - e_B$

Notice that (2) is the same whether or not the public bad is there. Assuming that the error term $(e_A - e_B)$ has the same variance whether or not the public bad is present, the probability of choosing A over B is the same in the situations with and without the public bad, and is equal to:

$$(2)\ \Pr(A) = \Pr(\Delta U \geq 0) = \Pr\left(\frac{e}{s} \geq \frac{b}{s} + \frac{d}{s} \cdot \text{tax saving}\right) = \Pr\left(\frac{e}{s} \leq -\frac{b}{s} - \frac{d}{s} \cdot \text{tax saving}\right),$$

where $e = e_A - e_B$, and σ is the standard deviation of ε.

If the marginal utility of the public good is affected by the public bad (as hypothesized by the authors), one can write:

(3) $U = a + b \cdot PG + b' \cdot PG \cdot PB + g \cdot PB + d\,(t - \text{tax saving}) + e$

in which case the difference in utility levels is $\Delta U = b + d \cdot \text{tax saving} + e_A - e_B$ for the situation with no public bads and $\Delta U = (b + b') + d \cdot \text{tax saving} + e_A - e_B$.

Clearly, the fraction of the sample who chooses one A differs across the treatment with and without public bad if b' is different from zero. But b' could be zero, and the percentages still be different if the variance of $(e_A - e_B)$ is different when the public bad is present. Here, for instance, choosing A (the house with the public good and no tax credit) is more likely in the situation with the

public bad if the variance of the error term, σ, is lower than that for the situation without public good. Is this reasonable to believe? Once again, this could be determined if one were able to estimate separately the β, γ and δ from σ, which can be done only if the tax credit is varied to the individuals.

Finally, this research has attempted to document the existence of some effects, but has not conclusively pinpointed to the reasons why we see them happen. It is likely that they are the result of several concurrent causes, and that it will be extremely difficult to devise an experiment where such causes can be disentangled. The paper needs to acknowledge that.

III. Comments on the paper "Ask a Hypothetical Question, Get a Valuable Answer?" by Christopher Azevedo, Joe Herriges, and Cathy Kling

This paper begins with a discussion where valuation economists are placed into one of three categories: Revealed-Preference (RP) lovers, Stated Preference (SP) lovers and "agnostics" who believe that stated-preference data and revealed-preference data should be used for joint estimation of the parameters influencing utility and of welfare effects, without necessarily considering one type better than the other.

An example is then provided where RP data documenting individual trips to wetland locations in Iowa are combined with the same individuals' statements as to the number of trips that they would take if the price of a trip increased by $X. Joint estimation is carried out that allows for correlation between the SP and RP data.

The authors specify 4 hypotheses explaining differences between stated and revealed preferences: namely, that:

(a) all coefficients are identical;

(b) the variances of the error terms are different;

(c) the coefficient of disposal income is different;

(d) the coefficient of price is different.

The model is first estimated in a completely unrestricted fashion, and then each of the above restrictions is imposed, and rejected. The conclusion I draw from these results is that the statistically significant differences between the parameters of the SP and RP demands for trips offer little hope that SP and RP data can be reconciled, at least in this example.

In general, the paper is well written, but the discussion at the beginning and the end of the paper should be a little more balanced. I believe that most valuation researchers are aware of the limitations and advantages of using stated v. revealed preference data, and do not endorse blindly one approach over the other, although because of their training, background and data availability they sometimes elect to work with only one of these types of data.

For instance, I believe that many of the valuation researchers that have relied primarily on RP data are fully aware of the empirical and theoretical difficulties associated with valuation using RP data: consider for instance multiple-day trips, multiple-site trips, the difficulties in specifying the choice set (when using models of site choice), the participation problem, the definition of the appropriate market, etc.

In addition, studies based on RP data sometimes rely on untested assumptions about individuals' perceive environmental quality or risk and react to it—consider for instance compensating wage studies, where the individual is routinely assumed to know and react to *objective* workplace risks measured at the *industry* or *occupation* levels. RP studies also pose econometric problems—for instance, in compensating wage studies it may be very difficult to disentangle other factors influencing wages from the workplace risk, and this may have huge effects on the estimate of value of a statistical life. Last but not least, actual data may preclude consideration of certain changes in environmental quality at one location. These are very real and valid concerns and limitations, and many researchers known for their seminal work in RP modeling have turned to supplementing RP data with questions about what respondents would do under hypothetical circumstances.

The most important finding of the paper is that RP and SP seem to be driven by completely different set of parameters—the differences are not just limited to specific slope coefficients (such as that of price or disposable income) or the variances of the error terms. This is an interesting result, and not necessarily one that invalidates the use of stated preference data, especially if one recognizes—as nicely pointed out by the previous discussant, Dr. Brown—that in this study we have actual trips taken in the past and trips that *would be taken in the future* under specified conditions. Before we draw firm conclusions from this study, I would like to have more information about the following:

(a) The likelihood function appears to be that of a bivariate tobit model. Were restrictions imposed to reflect the fact that most respondents presumably stated a lower number of trips at the higher price than the number of trips at the current price? In other words, the stated-preference trips are truncated from above at the actual trips, but I believe the model does not impose such truncation.[2]

(b) It is interesting that when one relaxes one of the restrictions of the fully restricted model (in moving from col. (B) to col. (C), (D) and (E)) the final estimates of the coefficients for which the constraint was relaxed are effectively equal to the constrained value (one).

(c) It almost appears like the estimation algorithm used the fully constrained parameter values as the starting point, and failed to move away from that. Did you try different estimation algorithms (e.g., Newton, BFGS, etc)? Did you try different starting points?

(d) If the average number of trips taken by the individual is 8.2 and 2.7 for an increase of $27 in the price of a trip, is a bivariate tobit appropriate? Would the results be robust to a bivariate Poisson or negative binomial model?

(e) Are your results robust to other subsets of your 6000 person survey? Other wetland areas?

(f) Might the results change when the models include other individual characteristics that influence the demand for trip, the variances of the error term, and a person's ability to respond to the hypothetical questions?

---

[2] To illustrate, consider individuals with non-zero actual and stated trips. Their actual trips would be written as $R_i = \mathbf{x}_i b + e_i$, where R is the number of actual trips, **x** is a vector of variables influencing the demand for trips (price per trip, income, other individual characteristics), β is a vector of unknown parameters and ε is an error term. Stated-preference trips would be written as $A_i = \mathbf{z}_i g + h_i$, where the vector z reflects the different price per trip, γ is a vector of coefficients, and η is an error term potentially correlated with ε. If $A_i < R_i$, then $h_i < (\mathbf{z}_i g - \mathbf{x}_i b) + e_i$. This implies specifying a truncated distribution for η.

(g) Is there anything from your other questions that could help us understand why the differences between actual and hypothetical trips? If it is not the measurement of price and income, what else could it be that was not explicitly hypothesized in this study?

IV. Comments on the paper "The Value of Visibility: A Comparison of Stated Preference Methods" by Thomas Stevens, John Halstead et al.

This paper presents the results of a study based on split samples which attempts to answer the following questions:

(a) holding all else the same, do values from on-site stated-preference valuation studies differ from off-site studies?

(b) Holding all else the same, does WTA differ from WTP?

(c) Holding all else the same, do computer-assisted, in-person surveys results in different values than mail surveys based on written questionnaires?

(d) Holding all else the same, does conjoint analysis (CJ) produce the same values as dichotomous-choice contingent valuation (CV)?

Clearly, answering all of these questions results in a very complex study design. As shown in the chart, samples are split over a variety a criteria, and must include sufficient respondents to be able to administer different levels of visibility loss, so that WTP can be tested for scope. In practice, at this stage of the research, the "cells" spanned by the experimental design contain too few respondents for proper comparisons to be meaningful. The problem is compounded by the fact that effectively estimation of WTA or WTP is by dichotomous-choice observations—however desirable these are from a number of respects, their informational content about WTA/WTP is necessarily limited.

(a) *Holding all else the same, do values from on-site stated-preference valuation studies differ from off-site studies?*

The regression model for WTA run by the authors pools all data, controls for survey mode and location, and finds no evidence of a statistically significant difference.

Here, however, the appropriate comparison would be between groups A and C, and between groups B and D. Unfortunately, the paper does not present sufficient details so that conclusions can be drawn from such a comparison. In addition, it is unclear how respondents were recruited to participate in the off-site personal interviews. There is some concern that participants in the off-site and on-site computer-assisted interviews differ from age and income, but the differences are in different directions, depending on whether one looks at the CJ or CV study.

(b) *Holding all else the same, does WTA differ from WTP?*

Table 5 suggests that, if anything, WTP is greater than WTA ($511 v. 331). But these figures are dollars per month, and additional analyses are necessary before they can be deemed as conclusive.

(c) *Holding all else the same, do computer-assisted, in-person surveys results in different values than mail surveys based on written questionnaires?*

The problem with this aspect of the study is that the authors are not simply testing for difference in values resulting from computer-assisted in-person surveys and the use of self-administered surveys based on written materials. What they are actually examining is how laptop-assisted, in-person surveys compare with self-administered-surveys *compounded with the self-selection* typically expected of mail surveys.

The authors should first test for the presence of such effects, at least to the extent that it can be captured in observable characteristics of the respondents, such as their incomes, ages, education levels, etc. Cameron et al (forthcoming) demonstrate how one can check for the presence of self-selection biases using zip-code level averages for income, age etc., which should be available for those persons who were sent the questionnaire, but did not return, and for those who did.

Even if the completing and returning of the mail questionnaires is not associated with any particular characteristic of the recipient, it is important than in comparing WTA across the authors keep in mind that their on-site and off-site samples are different in terms of demographics. The two groups' mean or median WTA may differ just because of the different ages and incomes. For a proper comparison, the authors should fit a regression relating WTA to age, income etc. for each of the two groups, and then statistically compare the coefficients of these variables across groups, or predict WTA using the coefficient from one group, but the individual characteristics typical of the other group.

(d) *Holding all else the same, does conjoint analysis produce the same values as dichotomous-choice contingent valuation?*

Most recent high-quality applications of CJ analysis for valuation purposes have tended to be so-called "choice experiments." In other words, respondents are to choose between A and B, where A and B are environmental amenities described by a set of attributes, including cost to the respondent. Respondents are to choose between A and B, and, in what I consider the best applications, also between A, B, and a "do nothing, pay nothing" option. This allows the researcher to estimate WTA or WTP for each commodity. Another advantage is that the same respondent can be asked to evaluate several pairs of commodities, which increases the number of observations, holding the number of respondents the same (of course, one should control for possible correlation between the responses provided by a respondent).

Here, instead, the researchers ask people to rate A and B, where A is always the status quo and B is a reduced-visibility situation accompanied by a reduction in the electricity bill paid by the respondent. The rating is between 0 (absolutely unacceptable) and 10 (perfectly acceptable).

In practice, however, the authors end up using the ratings only in an ordinal sense: If A rated higher than B, it is taken to imply that the respondent answered his WTA question as a "no."

But is this correct? If A, the status quo, is rated 10 and B, the proposed change, is rated 0, then I would agree with their interpretation that B is rejected at the stated price. But what does it mean when A is rated 4 and B is rated 3? The authors treat these ratings in the same fashion as in the previous answer, although here it would appear that option A is not liked very much at all in the first place, and B is liked less than A, but not very much less. (Were there any respondents that rated the status quo so poorly? Does it even make sense to ask people to rate the status quo?)

At a minimum, the authors should try to experiment with ordinal logit models, or some other models that account for "liking A a lot better than B." To conclude, I do not believe that asking people to *rate* visibility levels (accompanied by reduced electricity bills) is the right way to do the comparison between CJ and CV, especially if the ratings are recoded as dichotomous choice observations, and if interpretation of the ratings themselves could be criticized as arbitrary.

In practice, the regression model run by the authors cannot provide firm conclusions about whether the CJ and CV approach give different results. This is because, depending on how the "ties" are treated, the coefficient of the CJ dummy is negative, large and significant in one specification, and virtually zero (and insignificant) in the other.

The authors also note the estimated median WTA is very different, depending on whether a "tie" (i.e., A and B are rated the same) is treated as an acceptance of the proposed change or whether only when the rating of B is strictly greater than that of A is this response treated as an acceptance. The model should be re-estimated to account for the fact that a "tie" indicates indifference—effectively, this produces a continuous observation on WTA.[3]

The authors should also make sure that median WTA is calculated separately for each of the levels of visibility loss examined in this study.

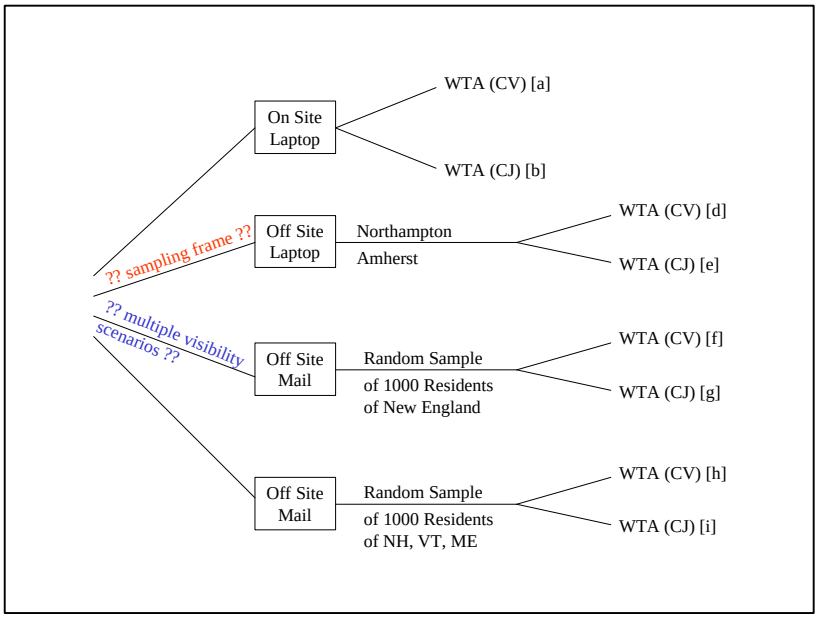

Figure 1. Experimental design in the Stevens et al study.

Finally, the authors state that "That the CVM and conjoint models can produce different results should not be too surprising. Although few comparisons of these techniques have been published, most previous empirical comparisons suggest substantial differences (see Stevens et al., 2000). There are several reasons for this. First, when compared with the CVM, many conjoint questions provide more information

[3] The authors seem to be aware of the problems associated in modeling their ratings. However, they should be careful not to confuse ratings of the alternative scenarios with measures of the *uncertainty* associated with responses to CV questions, as they discuss on page 16 of the paper. What is being measured here is not uncertainty about what one likes, but how much one likes or dislikes a given scenario.

about substitutes. Second, from a psychological viewpoint, respondents may react more differently when choosing among options than they do when making dollar evaluations (Irwin et al, 1993; Brown, 1984)."

It is unclear that this reasoning applies to the present study, since substitutes are not mentioned to the respondent, and we do not know whether respondents were spontaneously considering substitutes when answering either the CV or the CJ questions.

V. Comments on the paper "Toward Comparing Stated Preference and Real Money Contingent Values: Wetlands Valuation" by Stephen Swallow, Michael Spencer and Laurienne Whinstanley

Using two different valuation approaches (standard dichotomous choice and "choice experiments," a variant of conjoint analysis), and two different populations (students recruited at a University campus; state residents) and survey methods (questionnaire taken in the classroom, and mail surveys), this paper examines the relationship between values from hypothetical choices and values from choices involving actual payments. The setting is a donation mechanism ("adopt a pond" for water quality monitoring purposes), and three experimental treatments are devised: one where excess funds are not rebated back to the donors, one where excess funds are rebated back to the donors using a uniform proportional rule, and one where excess funds are retained and used to cover administrative costs of the program (the "extended benefit" mechanism).

Unsurprisingly, WTP is highest for the uniform proportional rebate scheme. Respondents involved in the real money experiment appear to dislike the "extended benefit" mechanism, perhaps because administrative costs are often associated with wasteful use of funds. While willingness to pay for the "adopt a pond" program declines sharply with the real dollar amount requested of the respondent, participants in the "hypothetical" experimental treatment are relatively insensitive to the dollar amount. The authors show that hypothetical WTP exceeds actual WTP by a factor of two or three, depending on the particular experimental treatment.

In general, I remain unconvinced that studies based on the donation mechanism can shed much light on the relationship between actual and hypothetical WTP, because of the lack of incentive compatibility. Perhaps the authors believe that if the donation mechanism encourages free riding behaviors, there is no reason why these should differ across hypothetical and actual donations. In my opinion, there is no particular reason why they should be equal, either.

I also do not find that incorporating various assurances about the use of the excess funds truly helps understand differences between hypothetical and actual WTP. It is important, in my opinion, that experiments be designed to avoid the use of the donation mechanism and to implement the same degree of incentive-compatibility that can be expected of dichotomous-choice, referendum contingent valuation.

VI. What are the lessons that I have personally learned from the papers I have read and seen presented at this conference?

In brief, keep your study design and experimental treatments as "clean" as possible. This means…

1. In designing experimental treatments intended to prove or disprove effects, it is important to keep the design as simple as possible. Unless the available budget and sample sizes are very large, a complicated design may end up with too few observations in each cell to be able to draw firm conclusions.

2. When testing for an effect using split samples, one needs to make sure all of the other aspects of the scenario (payment mechanism, etc.) and of the population are as close as possible across the two subsamples.

3. It is important to avoid drawing samples of convenience from different populations—it is difficult to say what the combined effect of the different population (residents of New England) and the different sampling frame (mall intercept, return of mail questionnaire) will be.

4. Until we know more about them, it is probably better to avoid using elicitation formats that are not clearly interpreted (e.g., ratings) and that can be fed into statistical models of WTP/WTA responses only by making arbitrary assumptions.

References

Cameron, Trudy A., Douglass W. Shaw, and Shannon Ragland (forthcoming), "Non-response Bias in Mail Survey Data: Salience v. Endogenous Survey Complexity," in Joseph Herriges and Catherine Kling, eds., *Valuing the Environment Using Recreation Demand Models*, Edward Elgar Publishing, Ltd.

Hanemann, W. Michael (1991), "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review*, 81(3), 635-647.

Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk" Econometrica, 263-291.

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

Question and Answer Period for Session II

Barbara Kanninen, University of Minnesota, addressed a comment to Stephen Swallow. She suggested that estimating a universal calibration ratio may ultimately be the wrong goal. The Office of Management and Budget (OMB) may have done economists a disservice by giving them the idea that 2 could be the natural calibration ratio. Perhaps economists should search for a technique or calibration model, which would yield different ratios based upon the biases and other aspects of each specific situation.

Swallow agreed in principle. He remarked that his work had come up with a calibration ratio in the wetlands situation of about 6, and between 2 and 3.5 for ponds. His experimental design used real wetlands parcels, presenting the parcels' attributes, such as location and size, to the respondents. However, the experimenters could describe a single parcel in different ways — for example, offering the opportunity to help preserve 30 acres of a 70 acre wetland or offering an opportunity of helping preserve the whole 70 acres. The design allowed the experimenters to come up with real or hypothetical values for these different attributes from the same survey. Others have been working on how to model that difference, to attribute it to education or other factors. That is a worthwhile aim, but beyond the scope of his present work.

Richard Carson, University of California, San Diego, remarked that he saw a serious problem with hypothetical bias. For example, in a survey, when incentives are reduced to hypothetical fund-raising, "free riding" will bias the numbers.

Swallow agreed. His numbers reflect some free riding and other departures. But that doesn't mean stated preference (SP) is wrong, just that it is different. One of the weaknesses in his existing data is that much of the wording in the SP work was parallel to wording in the revealed preference (RP) study. Maybe we should be comparing RP numbers with SP numbers generated in somewhat different ways.

V. Kerry Smith, North Carolina State University, raised two points. First, regarding the Iowa wetlands study, if the price of visiting wetlands increases, people can compensate by visiting closer wetlands. Did the study questions require people to assume their trips would be to the same area, or did the study lump all possible wetlands trips into the same demand function?

Second, in Carol Mansfield's work renters may weigh the possibility of public bads reducing rents while homeowners think about falling property values. That ought to be detectable in comparing the responses of renters and owners.

Joseph Herriges responded to the first point saying his analysis used a simple aggregate model for the prairie pothole region. The next step in analysis is to break it down to three separate zones.

Carol Mansfield said when they regress willingness to accept, the home ownership coefficient is positive and significant. Homeowners were more likely to say yes to both scenarios, an unexpected result. Perhaps it had to do with the nature of the public bads used in the study. The point needs further study.

Anna Alberini observed that some of the results could be explained by variances being different when respondents were offered the public good with the public bad and when respondents heard no mention of the public bad whatsoever.

F. Reed Johnson, Triangle Economic Research, observed to Joe Herriges that the design of the Iowa study necessarily embodied some prior assumptions about preferences. For example, should you weight one real trip as equivalent to ten repetitions of one conjoint? Also, the data in the RP model are going to be riddled with errors and variables. Might these errors and variables swamp underlying similarities between the RP and SP data?

Joe Herriges agreed that what you get from the data depends in part on what you bring to them. Both the approaches of the RP lover and the SP lover will have flaws. The weight you give depends on your prior, and the correct answer is hard to pin down.

Michael Hanemann, University of California, Berkeley, said that based on experience gained in a survey of sport fishing, the number of trips taken could be insensitive to price, but strongly affected by time constraints and other factors. Choice of destination was much more sensitive to price than number of trips was.

This leads to the issue of functional form. He did not believe that demand was linear and in his study used a cubic function to model demand. Also, SP and RP may differ in how they are non-linear.

Further, he pointed out that the number of trips is always an integer, but we are plugging it into a continuous model. If the price goes up, the model does not pick up the integral nature of the variable. But it is possible that SP responses can reflect more subtle changes in demand.

One study where willingness to accept (WTA) gave useful results involved electricity pricing. The trade-off was lower electricity rates if the utility discontinued assistance with energy conservation. In contrast, using WTA in cases dealing with purely public goods raises issues of property rights, asking someone to give up something they do not entirely own.

Finally, he observed that market choice can be worrisome in contexts where the choice offered leaves much unspecified. People want to know details. If you offer them a choice of homes as in the Mansfield study, they will want to know details about the neighborhoods, such as the quality of the schools, which the study does not offer. If you don't tell them, they may make their own assumptions about missing details.

Carol Mansfield replied that they told their respondents that the two choices were identical except for the factors described. However people may take negative factors as signaling other problems with the neighborhood. The issue needs a closer look.

Richard Carson remarked that he had just finished a study of successes in siting hazardous waste facilities. One case involved a pure reduction in property tax, but the others were mostly mixtures of public goods and some type of reduction in tax. It may be good to incorporate these mixed options in future studies. Without money in the picture, you can't answer the question, have you made the public whole?

Kerry Smith observed that EPA needs an inexpensive way to get information about public preferences. It is not necessary to match SP and PR responses from the same individuals. It would be acceptable in principle to stack a small SP sample on, say, the large recreational surveys that EPA does, as long as there were opportunities for restrictions across parameters. Similar sampling is done in other fields.

Julie Hewitt, U.S. Environmental Protection Agency, addressed a question to John Halstead. People have some control over the amount of electricity they use. When people said they were willing to accept higher prices, do you think they were weighing that they could offset part of the rate increase by using less?

John Halstead said the study implied that people would have the same consumption but in hindsight should have stated it explicitly.

Hewitt asked if Halstead could check for variances by income.

Halstead replied that it would be more illuminating to check by state, since New Hampshire has much higher power rates than Vermont or Maine.