Journal of Agricultural and Resource Economics 22(1):44–60

Contingent Valuation of Rural Tourism Development with Tests of Scope and Mode Stability

Kreg Lindberg, Rebecca L. Johnson, and Robert P. Berrens

Contingent valuation is used to measure the social impacts of tourism in rural Oregon communities. Impacts are substantial, for example, annual household willingness to pay (*WTP*) to reduce traffic congestion is $186. Study features included tests of sensitivity to a change in scope, tests of stability across survey mode, and a thorough system of "no"-vote follow-up questions in a referendum format. While there is no evidence of scope effects (at the 0.05 level), results indicate that conclusions regarding sensitivity to scope may be dependent on the test used. *WTP* estimates are substantially less with the mail versus telephone survey mode.

Key words: contingent valuation, mode stability, scope, social impacts, tourism

Introduction

Oregon coast communities, and many other rural American communities, are changing as employment in traditional natural resource industries declines while tourism development increases. Because tourism is consumed at the place of production, it tends to generate more social impacts than other industries. Although these nonmarket impacts have been recognized and evaluated (e.g., Lankford and Howard), they have not been assessed using a money measure. Economists have evaluated the environmental impacts of development (Freeman), and there is a parallel need to evaluate the social impacts (Portney). However, the literature lacks examples of the valuation of social impacts associated with specific industries. This article presents results from a contingent valuation study of selected social impacts, including increased traffic congestion and minor crime, associated with tourism development on the Oregon coast.

The survey-based contingent valuation (CV) method asks respondents for statements of willingness to pay (WTP) or willingness to accept (WTA) compensation for changes in nonmarket goods. Use of CV has increased rapidly, motivated in part by evolving natural resource damage law. A recent "blue ribbon panel" provided a qualified endorsement of CV (Arrow et al.), but some economists remain critical of the method. This article addresses some of those criticisms.

Critics assert that CV will not provide valid measures of economic value, in part because survey responses may reflect scenario features rather than the good itself (e.g., Diamond and Hausman; Green, Kahneman, and Kunreuther). For example, CV scenarios

Lindberg is a lecturer in the Johnstone Centre for Parks, Recreation and Heritage at Charles Sturt University; Johnson is an associate professor in the Department of Forest Resources at Oregon State University; and Berrens is an assistant professor in the Department of Economics at the University of New Mexico. This research was funded by the Oregon Sea Grant (project No. R/PPA-40, grant No. NA36RG0451) and Coastal Oregon Productivity Enhancement programs. The authors thank three anonymous reviewers for their comments.

generally include the following features that might affect responses: payment vehicle, payment distribution, implementing agency, implementation method, and implementation rule. In dichotomous choice referendum formats, if respondents object to one of the scenario features, they may vote "no" even if their value is greater than the required payment amount (or "bid"). Likewise, recent research suggests that, in addition to valuing the good itself, motivations for "yes" votes might include the desire to contribute a "fair share," and concerns about a broader issue or "good cause" (e.g., Kahneman and Knetsch; Loomis, Lockwood, and DeLacy; Schkade and Payne; Stevens, More, and Glass). CV supporters counter that the CV decision-making process is difficult, complex, and context dependent, but so too are many other consumer decisions that are not rejected by economists (e.g., Hanemann).

To some extent, critics and supporters are discussing different conceptual models, such that disagreement may be reduced by tailoring the models and, hence, the assumptions and data analyses, to the focus of the valuation exercise. For example, decision makers may want to know whether WTP for a program to reduce traffic congestion is greater than the cost of the program. The most analogous market in this case is a bond measure, where "yes" votes are assumed to represent a WTP greater than expected cost, regardless of whether that WTP is affected by program features or motivations other than personal benefits stemming from reduced congestion. The interest is in whether people would vote for the program or policy, given its features and cost. This case is the basis for the "policy" models described below. Individuals are valuing the proposed policy change in its entirety.

Alternatively, decision makers may desire an estimate of the (negative) value of increased traffic congestion that would arise from future increases in tourism-related traffic. This estimate could be compared with value estimates for economic and environmental impacts to determine whether, and how, to pursue tourism development. In this case, the desired value estimate is for the decrement in quality *independent* of scenario features; the scenario is simply a necessary device for deriving the estimates. The issue becomes *why* respondents would or would not pay, and analysts typically use follow-up questions to "no" votes to determine whether the vote resulted from objection to a scenario feature. These protest "no" votes generally are excluded when calculating WTP, thereby increasing estimated WTP. This case is the basis for the "commodity" models described below.

Follow-up questions have been recommended as standard practice in CV studies (e.g., Arrow et al.). However, there is a need to further refine these approaches to more accurately isolate WTP for the good itself. Respondent votes may be based on multiple considerations, and there is no standard for distinguishing between "valid" and "invalid" considerations, or somehow allocating between them for the same vote. Moreover, the simple follow-up questions used in most studies may generate misleading information. This study is characterized by an especially thorough system of "no"-vote follow-up questions.

Another criticism of CV is that responses are not adequately sensitive to changes in the scope of the good (Boyle et al.; Desvousges et al.).[1] For example, Boyle et al. found no significant difference in WTP to prevent three different levels of waterfowl deaths: 2,000; 20,000; and 200,000 (which represent from much less than 1% to about 2% of the specific population). Though Boyle et al. present five possible explanations for this

[1] As used here, scope sensitivity is equivalent to Carson and Mitchell's "component sensitivity."

finding, critics tend to focus on the explanation that CV cannot measure the difference in values and to conclude that CV is therefore unable to generate valid estimates of WTP. However, the studies used by CV critics have themselves been strongly criticized (cf., Carson and Mitchell; Hanemann; Smith). Moreover, numerous CV studies have identified significant scope effects (Hanemann). It would, therefore, be premature to reject CV on the basis of failure to pick up scope effects. This study evaluates WTP for different levels of congestion mitigation and the potential sensitivity of conclusions to the type of scope test used.

The final issue is whether WTP estimates are stable across CV survey modes (in person, telephone, or mail). The strengths and weaknesses of alternative survey modes have been evaluated by various reviewers (e.g., Arrow et al.; Mitchell and Carson), who generally discourage mail surveys due to the potential for unacceptable nonresponse bias. However, there has been surprisingly little empirical evaluation of the stability of WTP estimates across modes.

In a study of boater WTP for wetlands, Mannesto and Loomis found that mean WTP from mail surveys was less than that from in-person surveys (cf., Loomis and King). Due to the much lower response rate for the mail survey (24%) than for the in-person survey (97%), the difference in mean WTP likely was conservative (insofar as motivation to complete and return the survey is positively related to WTP [Mitchell and Carson]). Given the realized sample, the significance of the difference in means depended entirely on the experience of the interviewer. Mannesto and Loomis conclude that this difference may result from interviewer bias associated with the more enthusiastic experienced interviewer. In addition, Mannesto and Loomis postulate that respondents are more likely to vote "no" on mail surveys than during in-person surveys because of the greater time available to evaluate one's budget constraint. Results from Whittington et al. support this relationship between time and stated WTP.

In simplified terms, respondent behavior can be placed into one of three categories. First, regardless of time available, the respondent fully evaluates her preferences and budget constraint before voting. Second, regardless of time available, the respondent does not fully evaluate her preferences and budget constraint, but rather reverts to an alternative basis for responding to the scenario. One of these bases is to provide an answer that is perceived as desired by the interviewer and/or the sponsoring agency. The likelihood of reverting to an alternative basis may be affected by interviewer behavior. Third, when provided sufficient time, the respondent fully evaluates her preferences and budget constraint, but she reverts to an alternative basis when provided insufficient time. Insofar as the first case generates true WTP, mail surveys can generate more accurate WTP estimates than nonmail surveys when either the second or third case occurs. In the second case, interviewer bias is avoided because mail surveys do not involve interviewers. In the third case, mail surveys provide the time needed to fully evaluate preferences and the budget constraint. This study investigates mode stability with particular attention to evaluation of the budget constraint.

Theoretical Basis

The solution to the consumer's expenditure minimization problem, subject to a utility constraint, can be represented by the expenditure function:

(1) $$e(p, q, U) = Y,$$

where p is a vector of prices, q is a vector of fixed public goods, and Y is the minimum income needed to maintain utility level U (this presentation follows Mitchell and Carson closely). Given initial levels of p_0, q_0, U_0, and Y_0 and subsequent levels q_1 and Y_1, the Hicksian compensating surplus (CS) can be represented by

(2) $$CS = [e(p_0, q_0, U_0) = Y_0] - [e(p_0, q_1, U_0) = Y_1] = Y_0 - Y_1.$$

The difference $Y_0 - Y_1$ is the focus of CV; survey respondents are asked for the income adjustment that, when combined with a specified change in q, leaves their utility unchanged. When the move from q_0 to q_1 represents an increment in the public good, CS is positive and can be interpreted as the maximum willingness to pay (WTP) for the increment. Using the congestion scenario described below as an example, q_0 represents the current levels of public goods, including traffic congestion, while q_1 represents the specified reduction in traffic congestion and current levels of all other public goods. The difference $Y_0 - Y_1$ represents the maximum annual household WTP to achieve this reduction. When the move from q_0 to q_1 represents a decrement in the public good, CS is negative and can be interpreted as the minimum willingness to accept compensation for the decrement.

In dichotomous choice referendum CV, a statement of maximum WTP (hereinafter referred to simply as WTP) is not obtained directly, but rather inferred through a discrete indicator variable I (Cameron 1988, 1991). Given a bid t_i presented to respondent i,

(3) $$I_i = \begin{cases} 1 & \text{if } WTP_i \geq t_i \\ 0 & \text{otherwise.} \end{cases}$$

That is, the respondent will vote "yes" to the CV referendum scenario if her willingness to pay is equal to or greater than the payment amount (bid) presented. The probability of a "yes" response ($I_i = 1$) is commonly modeled using logit or related approaches. In the logit approach, the probability of a "yes" response, P_i, is given by

(4) $$P_i = (1 + e^{-Z})^{-1},$$

where $Z = (t_i\alpha + x_i'\gamma + u_i)$, and t_i is the bid, x_i is a vector of observations on independent variables for respondent i, and u_i is the error term. Following Cameron (1988), the equation for predicted WTP is then derived by dividing $x_i'\gamma$ by the negative of α:

(5) $$WTP_i = x_i'\gamma/-\alpha = x_i'\beta.$$

Cameron (1991) also provides a way to obtain confidence intervals around WTP estimates.

Background on Study Site and Survey Administration

The economies of Oregon coast communities historically have depended on extractive natural resource industries (forestry, fishing, and agriculture). While these industries remain important, the forestry and fishing sectors in particular have declined due to harvest restrictions. Conversely, tourism and retiree in-migration have become increasingly important. Although tourism generates jobs and personal income, it also can generate negative social and sociophysical impacts, including traffic congestion, noise and minor crime (e.g., disorderly conduct), and crowding in stores, bayfronts, and other areas. Based on responses to the survey pretest, as well as discussion with community leaders, CV scenarios were created for programs that would (*a*) reduce traffic congestion on Highway 101 by 25% or 50% during busy periods (each respondent was presented either the 25% or the 50% reduction scenario), (*b*) reduce noise and minor crime by 30% during summer and holiday periods, or (*c*) provide low-income housing for all qualifying families in the community (wording for the congestion scenario is provided in the appendix). These programs were designed to address problems associated with tourism.

The dichotomous choice referendum method was used to elicit valuation responses, where each respondent was given a scenario with a specified payment amount and asked whether they would vote "yes" or "no" for the program. Each survey contained all three scenarios (the order was randomized). In each scenario, respondents were presented with a bid that was randomly selected from a group of 16 values in the range of $5 to $1,000 per household per year. The CV questions were part of a larger survey of attitudes toward tourism, and economic development generally, administered to residents in eight geographically and economically diverse communities during November and early December 1993.[2] In each community, a random sample of households was contacted by telephone using random digit dialing. One member from each household was chosen at random, based on date of birth, to complete the telephone survey, which lasted an average of 15 minutes. Residents who completed the telephone survey were asked to complete a mail survey. Half of those accepting the mail survey were sent a tourism version while the other half were sent a version focused on more general issues. The principles of Dillman's "total design method" were followed in survey preparation, pretest, and administration. A professional survey research firm conducted the telephone surveys on behalf of Oregon State University. The university conducted the mail surveys.

A large number (873) of contacted households refused to participate in the telephone survey before hearing any details. High refusal rates are common in telephone surveys, and specific factors increased the refusal rate for this survey. For example, many of the residential telephones in the communities are located in second homes and vacation rentals. Potential respondents contacted in such locations did not consider themselves to be residents and therefore declined to participate in the survey. High response rates were achieved once residents were engaged in the survey. Of the 962 residents who initiated the telephone survey, 17 (1.8%) terminated midway, leaving 945 (98%) completed surveys. Of those completing the telephone survey, 793 (84%) accepted the follow-up mail survey. Of these, 571 (72%) completed and returned the mail survey. Response rates are similar to other CV surveys (Mitchell and Carson).

Results and Discussion

Policy and Commodity Models

Two models were developed for each scenario. The first is the policy model, which values the mitigation program (i.e., the reduction in congestion *and* the method for achieving

[2] The communities were not randomly selected. In addition, random samples were taken from individual communities rather than from the group of all communities combined. However, the general insignificance of community dummy variables suggests that the sample is representative of the communities taken as a group.

Table 1. Treatment of "No" Votes, Congestion Commodity Model

Reason for "No" Vote (first follow-up)	Number Retained as "No"	Number Converted to "Yes"	Number Excluded[a]
Congestion is not a problem	57	0	0
A problem, but not worth cost	22	0	0
Can't afford it	58	0	0
Opposed to taxes/new taxes	4	0	28
Opposed to government	6	2	2
Should not have to pay/not my responsibility	5	1	14
Would not work/would cause more problems than solve	3	6	23
Other	3	0	0
Don't know	0	0	6
Total	158	9	73
Total with typical treatment (only one follow-up question)	137	0	103

[a] The large number of excluded "no" votes resulted from the tendency of interviewers to classify the first follow-up responses into the "other" category. The verbatim responses were recorded and later use to reclassify most of the observations into one of the preceding categories. However, the second follow-up question was not asked in such cases; without additional information, "no" votes were excluded.

it). Each program necessarily includes provisions for payment and implementation, so the valuation of each program includes valuation of these components. Therefore, all "no" votes are retained in the policy model. Because respondents are valuing the scenario components, these components should be specific and realistic. However, some level of generality is necessary because actual mitigation programs will vary across communities. For example, the scenario presents a generic payment vehicle: payment by each household into an independent fund. Because the actual payment vehicle (e.g., a property tax) may negatively affect *WTP*, the gain in generality from using a generic payment vehicle is achieved at the possible expense of upward bias in value estimates.

The second model is for the commodity, which reflects the value of the good (e.g., a reduction in congestion) independent of the method for achieving it. This model uses the traditional method of excluding protest "no" votes. However, the system of follow-up questions used in this study is unusually thorough. For example, all respondents who voted "no" for the congestion scenario were asked a first follow-up question: "And why would you vote against the measure?" Seven responses to this question were classified by interviewers as reflecting beliefs that residents should not be responsible for paying for the program (table 1).[3] Each of the seven respondents was then asked a second follow-up question: "If you knew that funding by local residents was the only way to solve this problem, would you vote for the measure?" For the five respondents saying "no," the original "no" votes were retained. For the one respondent saying "yes," the original "no" vote was converted to a "yes" vote. For the one respondent saying "don't

[3] When uncertain about classification, interviewers transcribed responses verbatim. These responses were later classified by a member of the research team. In the congestion scenario, 13 such responses were classified as "should not have to pay." This is the cause of the discrepancy between the figure of 7 cited in the text and the figure of 20 (sum across "should not have to pay" row) in table 1.

know," the original "no" vote was excluded from the sample. CV surveys typically involve only one follow-up question. As shown at the bottom of table 1, the information gained from the second follow-up question led to substantially fewer exclusions and more classifications as "no" and "yes" votes. This process reduced the loss of data from exclusion and, because the information gained in the second follow-up affected vote classification, provided more accurate estimation of *WTP* for the commodity than would have been the case with a single follow-up.

Despite this process, limitations remain. For example respondents often vote "no" for a combination of reasons. In some cases, these multiple reasons were identified during the survey and "no" votes were allocated in a conservative manner (i.e., a manner favoring allocation as valid "no" votes). However, it simply is not possible to fully explore the reasons for "no" votes in the course of a telephone survey. In addition, the categories developed for interviewers are, by necessity, aggregated and may lead to inaccurate classification.[4]

The bids presented to respondents were random and reflected a wide range. As a result, some low-income households were presented with large bids. In a few cases, respondents voted "yes" even though they would be unlikely to pay such amounts. Previous studies have converted such "yes" votes to "no" votes or to missing values (e.g., Duffield and Neher; Mitchell and Carson). This process increases the proportion of "no" votes, thereby generating a conservative estimate of *WTP*. For this analysis, "yes" votes were converted to "no" votes when the bid was greater than approximately 1% of reported annual household income (the precise percentage varied slightly because income categories were used in the survey). Of the 1,160 total "yes" votes on bids for the three different programs, 46 (4%) were converted to "no" votes. Analysis of unconstrained *WTP* models indicates that this conversion reduced estimated mean *WTP* by approximately 20%.[5]

The linear specification of the logit model was chosen after exploring the fit of alternative models, including probit and logarithmic specifications. The initial set of independent variables was selected based on economic theory (e.g., the bid and household income variables) and social psychological theory (Eagly and Chaiken). Variable selection and model evaluation also followed previous CV analyses (Loomis, Gonzalez-Caban, and Gregory; Whittington et al.). Model variables are based on responses to items in both the telephone and mail surveys. For example, the CV scenarios, and thus the "yes" or "no" votes, were contained in the telephone survey, while demographic and several attitudinal variables were contained in the mail survey.

Several categorical variables, such as education, can be modeled either as interval variables or as sets of dummy variables. Models were evaluated with each alternative, including logarithmic and exponential transformations of the interval form. In order to achieve parsimonious specification, insignificant variables (at the 0.10 level) were

[4] This concern is true even for the second round of follow-up questions. For example, an increase in traffic not only would increase travel time, but would increase business opportunities in the community. Portney argues that respondents may hold existence values for the jobs associated with these business opportunities, just as they hold existence values for environmental goods. In fact, interviewer transcriptions indicate that some of the "no" votes for the congestion scenario stem from respondent concern that the program would reduce business opportunities for others within the community.

[5] The small percentage of votes converted had a relatively large impact on mean *WTP* because these votes were in response to large bids. The results described below in the *Mode Stability* section support the use of this income constraint; 32% of respondents whose congestion scenario bid was greater than 1% of reported annual household income changed their "yes" vote to "no" when provided the opportunity to do so in the follow-up mail survey. The conversion of all "yes" votes in this category is consistent with the recommendation of Arrow et al. to be conservative in CV design and analysis (i.e., where decisions may lead to biased *WTP* estimates, to decide in favor of downward bias).

Table 2. Description of Variables

Bid	The bid amount presented to respondent.
HH income	8-category variable for total annual household income before taxes.
High income	Variable allowing piecewise regression on income, with break at annual household income >$40,000.
Property value	8-category variable for assessed value of home. Set to zero for nonhomeowners.
High prop value	Dummy variable for respondents with assessed home value of $200,000 or more.
Rent/own	Dummy variable for home ownership. Renters = 0, owners = 1.
Important	Response to question about importance of scenario-related issues: congestion, low-income housing, noise/minor crime. Not important = 1, somewhat important = 2, very important = 3.
Govt. resp.	Response to statement "Local government works hard to address the concerns of local residents" using five-point Likert scale. Strongly disagree = 1, strongly agree = 5.
Govt. active	Response to statement "Local government should take an active role in controlling negative aspects of tourism and other development." Same scale as *Govt. resp.*
Move	Response to statement "If I had the opportunity, I would move away from this community." Same scale as *Govt. resp.*
Sacrifice	Response to statement "Residents sometimes need to make personal sacrifices for the good of the community." Same scale as *Govt. resp.*
Scope dummy	Respondents were presented one of two congestion scenarios: 25% reduction or 50% reduction in traffic on Highway 101 during busy periods. This variable is a dummy with a value of 0 for the 25% reduction and 1 for the 50% reduction.
Base categories for each set of dummies:	
Age	18–29 years.
Growth	Decrease in number of people living in community in the next five years.
Communities	Combined set of the small, adjacent communities of Gleneden Beach, Depoe Bay, and Lincoln Beach.
Order	The order in which the scenarios were presented was randomized across respondents. For the base, congestion was presented first. For *Order1,* congestion was presented last. For *Order2,* congestion was presented second.

dropped (variable exclusion did not substantially affect *WTP* estimates). The two exceptions to this rule were the *scope dummy* variable (representing level of congestion reduced) and individual dummy variables contained within a set that showed significance using a likelihood ratio test. Insignificant variables excluded from final models include employment status (whether employed and whether in tourism or retail sector), desired level of growth in tourism, length of residence, education, and gender. Table 2 describes each of the included variables. Results for the final logit models are shown in table 3. Goodness-of-fit measures for these models are relatively high for CV analyses.

The logit models were converted to *WTP* equations (Cameron 1988), here using the model for the noise commodity as an example:

$$(6)\qquad WTP(\$) = -302.88 + 57.16*HH\ income - 72.38*High\ income - 9.70*Property\ value + 102.46*High\ prop.\ value + 50.77*Important + 46.06*Govt.\ active.$$

Table 3. Variables and Estimated Coefficients of the CV Logit Models

Variable	Noise/Crime		Congestion	
	Commodity	Policy	Commodity	Policy
Constant	−2.87[c]	−3.33[c]	−1.44	−1.65[a]
Bid (each $\times 10^{-2}$)	−0.95[c]	−1.03[c]	−0.67[c]	−0.77[c]
HH income	0.54[c]	0.40[c]	0.27[c]	0.18[c]
High income	−0.69[c]	−0.53[b]		
Property value	−0.092[a]		−0.20[b]	
High property value	0.97[a]		1.77[c]	0.84[a]
Rent/own			1.33[b]	
Important	0.48[c]	0.43[c]	0.69[c]	0.54[c]
Dummies for age				
30–39			−0.92[a]	−0.86[a]
40–49			−0.96[a]	−1.15[b]
50–59			−1.00[a]	−1.03[b]
60+			−0.83	−0.74
Dummies for desired growth				
Stay as now			−1.25[b]	−0.96[a]
Grow a little			−1.15[b]	−0.82
Grow a lot			−1.50[b]	−0.99[a]
Dummies for communities				
Seaside				0.42
Cannon Beach				−0.023
Newport				0.67
Coos Bay				0.17
Bandon				−0.36
Dummies for order of CV scenarios				
Order 1			0.63[b]	0.48[a]
Order 2			0.45	0.42
Govt. response		0.18[b]		
Govt. active	0.44[c]	0.40[c]	0.26[b]	0.34[c]
Scope dummy			0.24	0.30
Maddala R^2	0.29	0.27	0.26	0.26
McFadden R^2	0.25	0.23	0.23	0.22
Adjusted for df	0.23	0.22	0.19	0.18
Percent correct predict	72	72	73	73
Mean *WTP*	144	105	194	109
95% CI for mean *WTP*	118–171	85–126	152–236	82–136
Percent correct predict	72	72	73	73
Median *WTP*	148	108	188	109
Population (adjusted) mean *WTP*	130	95	186	110
Number of observations	443	481	412	497

Note: Sample of eight rural Oregon communities as described in text. Commodity models exclude protest "no" votes in order to value the good independent of scenario features. Policy models retain all "no" votes in order to value the good *and* the scenario features necessary for providing it.

[a] Significant at the $p = 0.10$ level or better.

[b] Significant at the $p = 0.05$ level or better.

[c] Significant at the $p = 0.01$ level or better.

The logit models and resulting *WTP* equations generally are consistent with factors thought to affect WTP. For example, *WTP* should increase with increases in ability to pay and increases in the importance of the problem and, thus, benefit from mitigation. The results in table 3 show large *WTP,* which indicates that tourism development has generated significant social costs.

The potential for bias due to item (individual question) and unit (entire survey) nonresponse recently has received significant attention in the CV literature (e.g., Dalecki, Whitehead, and Blomquist). Because there was relatively little item nonresponse for the independent variables, imputation methods were not used. However, the large number of telephone refusals and the modest number of refusals to accept or return the mail survey (unit nonresponse) may lead to sample selection bias insofar as these refusals are associated with demographic or attitudinal variables that affect *WTP.* Population means were used to correct for both item and unit nonresponse. The distributions of the *HH income, High income,* as well as education and age variables for the samples of observations included in the models were compared with the 1990 U.S. census data distributions for the group of communities sampled, weighted by community size. There were modest, but noticeable, differences between the population distributions and the sample distributions, with the samples being on average somewhat older, better educated, and wealthier than the population. Therefore, population means for these variables were substituted into the *WTP* equations to generate adjusted mean *WTP* for each model. The adjusted means are shown at the bottom of table 3. For the commodity models, the adjustment also accounts for any differences in these demographic characteristics between those included in the sample and those excluded because of protest "no" votes.

Scope

When evaluating whether CV is sensitive to scope (e.g., to the difference between a 25% or a 50% reduction in traffic congestion), analysts typically evaluate differences in estimated mean *WTP* (e.g., Boyle et al.), though differences in the distribution of "yes" and "no" votes also have been evaluated (e.g., Desvousges et al.). The conclusions from scope evaluations may depend on the method used insofar as power and assumptions vary across tests. For example, tests of means involve more assumption than tests of raw "yes" and "no" votes, including the assumption that the model is properly specified.

In the present study, one half of the respondents were presented with a scenario that would reduce traffic congestion by 25% during busy periods (low congestion) while the other half were presented a 50% reduction (high congestion) scenario. Table 4 shows *p*-values for various tests of the null hypothesis that votes, and *WTP,* are independent of the level of reduction in mitigation; that is, that CV is insensitive to changes in the scope of the good. None of the tests indicates sensitivity at the 0.05 level, though the Cochran-Mantel-Haenszel (CMH) test on "raw data, final" indicates sensitivity at the 0.10 level (two-tailed tests were used for the combined and separate models to ensure comparability with the χ^2 and CMH tests; one-tailed tests would indicate sensitivity at the 0.10 level).[6]

[6] CMH is a nonparametric test that evaluates association between two variables (in this case, the level of reduction and the vote) while allowing stratification based on one or more additional variables (in this case, the bid) (Landis, Heyman, and Koch; SAS Institute). CMH provides an alternative to χ^2 in cases, like the present, when the high number of bid categories relative to sample size generates small expected frequencies at high and low bid levels.

Table 4. *p*-Values for Various Scope Tests (25% versus 50% Reduction in Congestion)

	Model Type	
Test Basis	Policy[a]	Commodity[b]
Raw data, original (χ^2)[c]	0.17	0.21
Raw data, original (CMH)[d]	0.11	0.33
Raw data, final (CMH)[e]	0.06	0.25
Combined model (dummy)[f]	0.16	0.30
Separate models (means)[g]	0.12	$WTP_{25\%} > WTP_{50\%}$

Note: All tests incorporate the effects of constraining votes based on income. Tests on unconstrained votes generate similar results (the constraint affects both scenarios).
[a] Protest "no" votes retained.
[b] Protest "no" votes excluded or converted as described in text.
[c] χ^2 on votes of all respondents receiving mail survey.
[d] Cochran-Mantel-Haenszel (CMH) test on votes of all respondents receiving mail survey.
[e] CMH on votes of respondents retained in estimation of separate models (excludes respondents with missing values for independent variables in those models).
[f] Significance of coefficient for dummy variable included in combined model (*scope dummy* in table 3).
[g] Significance of *t*-test for differences in mean *WTP* from separate (low and high) congestion models.

Unlike Desvousges et al. and Boyle et al., the two scenarios reflect substantially different levels, in percentage terms, of the good being valued.

Of interest here are the different *p*-values generated by the alternate tests. The lower three rows provide the most appropriate comparison.[7] As indicated by the *p*-values in the policy column, there is substantial variability across tests, which may affect conclusions regarding sensitivity to scope. Although no test indicates sensitivity at the 0.05 level, there is an almost three-fold difference between the smallest and largest *p*-values in the policy column. In other data sets, this difference may span the chosen significance level and thus affect conclusions. For example, for a given data set, a χ^2 test might indicate a *p*-value of 0.09 while a CMH test might indicate a *p*-value of 0.04. If only the former were used, one would conclude that CV is not sensitive to scope; if only the latter were used, the opposite conclusion would be drawn.

Moreover, there is even greater variability in *p*-values across model types. The scope tests based on commodity models indicate far less sensitivity than those based on policy models; in fact, the mean *WTP* for low congestion is greater than the mean *WTP* for high congestion in the "separate models" case. Examination of the data indicates that the *WTP* similarity of low congestion and high congestion commodity models results from a relatively higher proportion of "no" votes excluded as protests for the low congestion scenario. This apparently random effect narrowed the *WTP* difference between the scenarios for the commodity models.

[7] The "separate models" and "raw data, final" are based on the same data. There are small differences between this data and that used for "combined model" as a result of missing observations for included variables in the respective models.

Mode Stability

Mode stability is evaluated based on whether people respond the same to telephone and mail survey questions. Because the mail survey followed the telephone survey, some differences may be due to the passage of time. However, previous research has found stable values over several months, or even several years (e.g., Stevens, More, and Glass). Given that respondents generally completed the mail survey within one to three weeks of the telephone survey, any effects are likely to result from mode, rather then temporal, differences.

During the telephone survey, respondents voted "yes" or "no" for a congestion scenario. During the mail survey, respondents were reminded of the congestion scenario, the bid, and the votes they gave during the telephone survey. They were then asked whether they would like to vote differently (wording is provided in the appendix). This direct questioning approach was chosen over split-sample (e.g., Loomis and King) and test-retest (e.g., Stevens, More, and Glass) approaches for several reasons. The split-sample approach, which uses independent samples, was rejected because it may lead to differences in response rates (and potential nonresponse bias), does not provide information on individual respondent behavior, and was inconsistent with other research needs for the survey. The test-retest approach, which resurveys the same sample, typically presents the same scenario with an introductory statement that willingness to pay may have changed since the original administration. Because of the short time between the phone and mail surveys, respondents likely would reject such a statement and would not respond seriously to the question.

To minimize any possible stigma against vote changes, respondents were reminded that people do change their minds and why this might happen.[8] Of the 571 returned mail surveys, 74 (13%) contained "don't know" or missing responses to this question. Of the remaining respondents, 55 (11%) changed their vote, with the majority (43) of changes being from "yes" to "no." Although there is no consistent relationship between the size of the bid and the likelihood of changing a "yes" vote, there is a consistently positive relationship between the size of the bid as a percentage of income and the likelihood of changing a "yes" vote (table 5).[9] However, a χ^2 test does not indicate that this relationship is statistically significant ($\chi^2_4 = 5.77$, $p = .22$). Thus, the results in table 5 suggest, but do not strongly indicate, that respondents use the additional time offered by mail surveys to more fully evaluate their budget constraint.

Table 6 shows the final logit models and estimated *WTP* for telephone versus mail modes (as with the models in table 3, those in table 6 initially were estimated with the same full set of variables, with insignificant variables dropped). The telephone mode is based on votes given during the telephone survey. The mail mode incorporates vote changes indicated by mail survey responses.[10] To avoid nonresponse bias, observations with "don't know" or missing response to the vote change question were omitted during

---

[8] Unlike in test-retest surveys, this reminder did not suggest that values may have changed. Rather, the reminder stated that "sometimes people change their mind about how they would vote, perhaps because there isn't much time to think about it on the telephone."

[9] Table 5 excludes observations with missing values for the household income question. This exclusion reduced the number of "yes" votes changed.

[10] Only policy models are shown here. It was not possible to reestimate the commodity model because no information was available concerning reasons for "no" votes created by responses to the mail survey. As before, "yes" votes were constrained based on income.

Table 5. Vote Changes by Size of Bid Relative to Income (Congestion Scenario)

	Size of Bid as Percent of Annual Household Income					
	0.0–0.1	0.1–0.2	0.2–0.4	0.4–1.0	>1.0	Total
Telephone Survey						
Combined frequency offered	99	82	90	98	92	461
Number of "yes" votes	73	55	45	37	22	232
	(74)[a]	(67)	(50)	(38)	(24)	(50)
Number of "no" votes	26	27	45	61	70	229
	(26)[b]	(33)	(50)	(62)	(76)	(50)
Mail Survey						
Number of "yes" votes changed	8	8	8	7	7	38
	(11)[a]	(15)	(18)	(19)	(32)	(16)

[a] Numbers in parentheses are percentages of "yes" votes offered.
[b] Numbers in parentheses are percentages of "no" votes offered.

estimation of both models.[11] As shown in table 6, mean *WTP* with the mail mode ($83) is substantially lower than mean *WTP* with the telephone mode ($115). A paired *t*-test indicates this difference is statistically significant (t=5.5, p<0.01).[12]

Conclusions

This article presents CV-based estimates of *WTP* for mitigation of selected negative social impacts from tourism development. Such information can complement values for mitigating negative environmental impacts in evaluating tourism development proposals. Although no previous analyses are directly comparable, the few similar studies have produced generally similar results. For example, Ahearn found a mean annual household *WTP* of $51 ($74 in 1993 dollars) for a 33% reduction in the risk of burglary in Oregon communities. Navarro and Carson used an election returns method to infer that the average San Diego household is willing to pay $138 per year to increase jail and court capacity in an effort to reduce crime.

The fact that CV responses are affected by scenario features and motivations beyond consuming the good itself complicates estimation of *WTP* for commodity-type models. Researchers have tried to isolate values for the good itself (apart from values based on these other considerations) through the use of follow-up questions to "no" votes and, increasingly, "yes" votes. However, more research is needed to determine the appropriate method for handling responses reflecting multiple considerations. As illustrated here, there is also a need to carefully probe these considerations in order to treat them consistently with the good being valued.

As Schkade and Payne note, follow-up questions used to probe previous valuation

[11] This omission did not substantially affect mean *WTP*, which is $109 for the original telephone model (table 3) and $115 for the reestimated telephone model (table 6). The difference in sample size between telephone and mail models reflects missing values for variables that were significant, and thus included, in the telephone model.

[12] To test whether differences in included variables or associated sample size affect results, each model was reevaluated using the included variables from the other. Though the difference in means narrowed (telephone mean = $112, mail mean = $94), it remained significant (paired *t*-test yields t = 3.6, p < 0.01).

Table 6. Mode Stability of Congestion Policy Models

Variable	Telephone Mode	Mail Mode
Constant	−2.72[b]	−3.87[c]
Bid (each $\times 10^{-2}$)	−0.87[c]	−0.53[c]
HH income	0.27[c]	0.19[c]
Property value	−0.24[b]	
High property value	1.24[b]	
Rent/own	1.31[b]	
Important	0.56[c]	0.78[c]
Dummies for age		
30–39	−1.02[a]	
40–49	−0.90	
50–59	−0.79	
60+	−0.66	
Dummies for desired growth		
Stay as now	−1.54[b]	−1.56[c]
Grow a little	−1.10[a]	−1.20[b]
Grow a lot	−1.54[b]	−1.15[a]
Dummies for communities		
Seaside	0.61	
Cannon Beach	−0.11	
Newport	0.80[a]	
Coos Bay	0.19	
Bandon	−0.56	
Dummies for order of CV scenarios		
Order 1		0.57[b]
Order 2		0.19
Govt. active	0.36[c]	0.30[b]
Move	0.21[b]	
Sacrifice	0.22[a]	0.28[b]
Scope dummy	0.39	0.47[b]
Maddala R^2	0.31	0.22
McFadden R^2	0.26	0.18
Adjusted for df	0.22	0.16
Percent correct predict.	75	72
Mean *WTP*	115	83
95% CI for mean *WTP*	87–143	42–125
Median *WTP*	115	85
Number of observations	399	430

[a] Variable is significant at the $p = 0.10$ level or better.
[b] Variable is significant at the $p = 0.05$ level or better.
[c] Variable is significant at the $p = 0.01$ level or better.

responses are a form of "retrospective protocol." As such, they may be more susceptible to bias and unreliability than verbal protocol procedures used to probe respondent concerns and motivations leading up to a specific valuation response. Although verbal protocol likely will remain impractical for most applications, it is an appropriate pretest tool for developing follow-up questions.

Sensitivity in valuation responses to the change in the scope of congestion mitigation was not found at the 0.05 level. However, this result must be viewed within the context that many CV studies have found scope effects and that insensitivity is only one explanation for the present lack of effect. Further, the fact that alternative statistical tests may lead to different conclusions suggests that scope evaluations should continue to use multiple tests (e.g., Carson and Mitchell). Finally, tests for differences in voting patterns and estimated *WTP* across congestion, noise, and housing scenarios indicate that respondents were sensitive to the specific good presented.

Lastly, the results from the test of survey mode stability are suggestive that mail surveys provide the time necessary for thorough evaluation of budget constraints. In addition, mail surveys avoid potential interviewer bias. However, these advantages are likely outweighed by the disadvantages of relatively high nonresponse and the difficulty of probing valuation responses. While the issue of nonresponse bias may be overstated insofar as telephone refusal rates also can be quite high, the importance of follow-up questions is sufficiently great to warrant favoring telephone over mail surveys. Nonetheless, the limited available evidence concerning mode stability indicates the importance of (*a*) minimizing interviewer bias through careful selection and training of interviewers and (*b*) providing adequate time for respondents to evaluate preferences and budget constraints.

[*Received August 1995; final version received February 1997.*]

References

Ahearn, M. C. "Analysis of Contingent Valuation Applied to Air Quality and Public Safety from Crime." Unpub. Ph.D. diss., Oregon State University, 1984.

Arrow, K., R. Solow, P. Portney, E. Learner, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(1993):4602–614.

Boyle, K. J., W. H. Desvousges, F. R. Johnson, R. W. Dunford, and S. P. Hudson. "An Investigation of Part-Whole Biases in Contingent-Valuation Studies." *J. Environ. Econ. and Manage.* 27(1994):64–83.

Cameron, T. A. "A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *J. Environ. Econ and Manage.* 15(1988):355–79.

———. "Interval Estimates of Non-market Resource Values from Referendum Contingent Valuation Surveys." *Land Econ.* 67(1991):413–21.

Carson, R. T., and R. C. Mitchell. "Sequencing and Nesting in Contingent Valuation Surveys." *J. Environ. Econ. and Manage.* 28(1995):155–73.

Dalecki, M. G., J. C. Whitehead, and G. C. Blomquist. "Sample Non-response Bias and Aggregate Benefits in Contingent Valuation: An Examination of Early, Late and Non-respondents." *J. Environ. Manage.* 38(1993):133–43.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In *Contingent Valuation: A Critical Assessment,* ed., J. A. Hausman, pp. 91–164. Amsterdam: North Holland, 1993.

Diamond, P. A., and J. A. Hausman. "On Contingent Valuation Measurement of Nonuse Values." In *Contingent Valuation: A Critical Assessment,* ed., J.A. Hausman, pp. 3–38. Amsterdam: North Holland, 1993.

Dillman, D. *Mail and Telephone Surveys.* New York: John Wiley and Sons, 1978.

Duffield, J., and C. Neher. "Montana Bioeconomics Study—A Contingent Valuation Assessment of Montana Waterfowl Hunting: Hunter Attitudes and Economic Benefits." Montana Department of Fish, Wildlife, and Parks, Helena MT, 1991.

Eagly, A. H., and S. Chaiken. *The Psychology of Attitudes.* Fort Worth: Harcourt Brace Jovanovich, 1993.

Freeman, A. M. *The Measurement of Environmental and Resource Values.* Washington DC: Resources for the Future, 1993.

Green, D. P., D. Kahneman, and H. Kunreuther. "How the Scope and Method of Public Funding Affect Willingness to Pay for Public Goods." *Public Opinion Quarterly* 58(1994):49–67.

Hanemann, W. M. "Valuing the Environment through Contingent Valuation." *J. Econ. Perspectives* 8(1994): 19–43.

Kahneman, D., and J. L. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *J. Environ. Econ. and Manage.* 22(1992):57–70.

Landis, R. J., E. R. Heyman, and G. G. Koch. "Average Partial Association in Three-Way Contingency Tables: A Review and Discussion of Alternative Tests." *Internat. Statis. Rev.* 46(1978):237–54.

Lankford, S. V., and D. R. Howard. "Developing a Tourism Impact Attitude Scale." *Annals Tourism Res.* 21(1994):121–39.

Loomis, J., A. Gonzalez-Caban, and R. Gregory. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Econ.* 70(1994):499–506.

Loomis, J., and M. King. "Comparison of Mail and Telephone-Mail Contingent Valuation Surveys." *J. Environ. Manage.* 41(1994):309–24.

Loomis, J., M. Lockwood, and T. DeLacy. "Some Empirical Evidence on Embedding Effects in Contingent Valuation of Forest Protection." *J. Environ. Econ. and Manage.* 24(1993):45–55.

Mannesto, G., and J. B. Loomis. "Evaluation of Mail and In-Person Contingent Value Surveys: Results of a Study of Recreational Boaters." *J. Environ. Manage.* 32(1991):177–90.

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

Navarro, P., and R. Carson. "Growth Controls: Policy Analysis for the Second Generation." *Policy Sciences* 24(1991):127–52.

Portney, P. R. "The Contingent Valuation Debate: Why Economists Should Care." *J. Econ. Perspectives* 8(1994):3–17.

SAS Institute, Inc. *SAS/STAT User's Guide, Version 6,* 4th ed. (Vol. 1). Cary NC: SAS Institute, Inc., 1990.

Schkade, D. A., and J. W. Payne. "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation." *J. Environ. Econ. and Manage.* 26(1994):88–109.

Smith, V. K. "Comment: Arbitrary Values, Good Causes, and Premature Verdicts." *J. Environ. Econ. and Manage.* 22(1992):71–89.

Stevens, T. H., T. A. More, and R. J. Glass. "Interpretation and Temporal Stability of CV Bids for Wildlife Existence: A Panel Study." *Land Econ.* 70(1994):355–63.

Whittington, D., V. K. Smith, A. Okorafor, A. Okore, J. L. Liu, and A. McPhail. "Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application." *J. Environ. Econ. and Manage.* 22(1992):205–25.

Appendix: Sample Survey Wording

Wording for the CV introduction and congestion scenario is as follows:

> In this next section, I would like to ask you about programs that would deal with issues that are problems in some coastal communities. These programs cost money. One way of paying for them is for your community to set up an independent fund paid for by all local households. Fund revenues would be used only for the program described—they will not go to the government.
>
> These programs are hypothetical. However, your responses may be used to guide future policies so please answer the questions as carefully as possible.

> The Oregon Department of Transportation (ODOT) is currently developing options for reducing traffic congestion along Highway 101 by, for example, adding turning or passing lanes. Some of the cost of these options may have to be paid by local communities.
>
> We estimate that one option would reduce traffic congestion on Highway 101 by 25% during busy periods. This would mean there would be as little traffic congestion on 101 during August as there currently is during May.
>
> If you had a chance to vote on a ballot measure that would reduce congestion on Highway 101 by this amount, but would require your household to pay $ *[X]* each year, would you vote for or against it? As with all ballot measures, at least half of the voters would have to support the measure for it to pass.
>
> __For the measure
> __Against the measure
> __Don't know

One half of the sample was presented with a 25% reduction and the other half a 50% reduction. The reference months (August and May) varied across levels of reduction and across communities. Wording for the "vote change" question is as follows:

> During the telephone interview we asked if you would vote for or against a program that would reduce traffic congestion on Highway 101 by 25% during busy periods (so that traffic during August would be about the same as it currently is in May). This program would cost your household $*[X]* each year. You said that you would vote__this program.
>
> Sometimes people change their mind about how they would vote, perhaps because there isn't much time to think about it on the telephone. Have you changed your mind about this program? Please circle one of the following answers:
>
> No, I would still vote__this program.
> Yes, I have changed my mind and would vote__this program.

The percentage reduction and reference months were customized based on the telephone scenario.