Staff Paper 212   March 1992

RELIABILITY OF CONTINGENT VALUATION ESTIMATES
OF WILLINGNESS-TO-PAY:
THE VALUATION OF TROPICAL RAIN FOREST PRESERVATION

Sharon I. Cowfer
Graduate Research Assistant
Graduate Program in Policy Analysis

Donald J. Epp
Professor of Agricultural Economics

Staff Papers are circulated without formal review
by the Agricultural Economics and Rural Sociology
Department. Content is the responsibility of the
author(s).

Presented at the Eastern Economic Association meeting, New York
City, March 28, 1992

Agricultural Economics and Rural Sociology Department
College of Agricultural Sciences
The Pennsylvania State University
University Park, Pennsylvania 16802

Reliability of Contingent Valuation Estimates of Willingness-To-Pay:

The Valuation of Tropical Rain Forest Preservation

Sharon I. Cowfer and Donald J. Epp

Contingent valuation (CV) is becoming more widely recognized as an important method of resource economists, but it is still a new method. As such, CV is subject to methodological concerns, such as the validity and reliability of the estimates produced. Loomis shows that more studies have examined the validity of CV than have tested the reliability. This fact was also noted by Musser and his colleagues. In order to test the validity of a measure, it is necessary that the measure be reliable (Mitchell and Carson, p. 211). Also, if information from contingent valuation surveys is to be used when making policy decisions, it is essential that the results of the surveys are reliable. Thus, for validity research and for policy applications, the results must be reproducible, or consistent over time.

This study reports a test-retest reliability study of the CV method. We examine the reliability of willingness-to-pay (WTP) responses to a question about preserving the wilderness characteristics of a substantial area of tropical rain forest. This study is interesting for two reasons. First, it expands the relatively small number of reliability studies. Second, it examines the reliability of responses concerning an environmental condition that, while it has considerable policy relevance and occasional news value, is not familiar to most individuals.

Previous Studies

Several studies of CV reliability examined products that were familiar to the respondents. This enhances the ability of respondents to give meaningful answers and may be presumed to give more reliable answers. Examples of reliability studies using familiar commodities or services include Kealy, Montgomery and Dovidio (candy bars), Musser and colleagues (cross-country skiing) and Loomis (sample of visitors to Mono Lake). Others examined situations with which the respondents were familiar, but may not have given much prior thought to the specific characteristic presented in the CV survey. Examples include Jones-Lee, Hammerton and Phillips (highway safety) and Kealy, Montgomery and Dovidio (acid precipitation effects on the Adirondack Park). Only Loomis has tested the reliability of CV responses to questions where a large number of respondents might not know about the topic (protection of Mono Lake for the sample of the entire population of California).

Another important aspect of test-retest reliability studies is the time interval between the initial test and the follow-up. Respondents should use the same process for determining their response to each presentation of the willingness-to-pay (WTP) question. Thus, the interval between tests should be long enough that respondents do not merely repeat their previous response from memory when deciding upon their response to the retest question. This suggests longer rather than a shorter intervals, although the subject being evaluated may influence the memorability of a particular response. Trivial items where a response can easily be determined may not be remembered for more than a few days. A decision that requires thought and

consideration or which is quite important to an individual may be remembered for a longer time. Thus, the interval used should be long enough that it is reasonable to assume that most respondents do not remember their previous responses. The interval should be short enough that the important determinants of a person's bid have not changed between the two tests. Alternatively, if changes in variables that influence the bid are expected during the interval, measures of the determining variables need to be taken with each test.

Kealy, Montgomery and Dovidio used a two week interval in their studies of students' WTP for candy bars and for prevention of additional damage to the Adirondack region from acid precipitation. Loehman and De, also using a student population, used a three week interval in their studies of WTP for improved air quality. Loomis used retest intervals of nine months for the general population sample and a five month interval with the sample of visitors to Mono Lake. Musser and his colleagues used a retest interval of about one year when studying WTP for cross-country skiing opportunities. While there is no hard-and-fast rule about the appropriate time interval, policy makers will want to base policies on values for non-marketed goods or services that can be projected over periods of several years or longer.

The Study

The study consisted of two rounds of mailed CV survey. Each round followed the Total Design Method (Dillman), with the exception that the final mailing was sent by

regular first-class mail. The first round was initiated in June 1990 and the re-test round began in April 1991, about 10 months after the first round. This time interval is believed to permit sufficient time for respondents to forget their previous WTP amount and require that they rethink their WTP. As described below, the survey instrument for each round obtained data about variables that are believed to influence WTP so that the influence of changes in those variables could be considered.

The first round survey instrument included a brief description of a way to protect a high proportion of the remaining area of tropical rain forest including suggestions that non-governmental organizations might organize a specific protection scheme. Respondents were then asked the following question.

> Thinking about your current monthly expenditures for food, clothing, charities, bills, etc., what is the maximum, one-time amount your household would be willing to pay to promote tropical rain forest protection?

This question was followed by a question examining several possible reasons for a $0.00 answer, including several possible protest and non-protest answers and a final opportunity to list "other" reasons. The first round instrument also included questions to measure the respondent's level of agreement with the New Environmental Paradigm (Dunlap and Van Liere), three different ways to measure respondent's knowledge of tropical rain forests (Griffith), level of schooling attained by the respondent, household income, knowledge of household use of tropical rain forest products, previous or intended visits to tropical rain forests, and sources of information or news about tropical rain forests.

The questionnaire was sent to 1000 individuals drawn randomly from all Pennsylvania residents listed in current telephone directories. The sample was drawn by a professional survey research company. Questionnaires were returned by 424 households. After adjustment for incomplete or duplicate returns, 394 useable returns were analyzed, a response rate of about 39 percent.

The second round questionnaire was mailed to all respondents to the first round and to 200 additional households not surveyed in the first round. The added households permit testing for any influence on responses due to prior exposure to the first round questionnaire. Statistical analysis showed that WTP responses from those participating in both rounds were not significantly different from those receiving only the second round. For this study of reliability, only the respondents participating in both rounds are analyzed.

Two hundred eighteen usable responses were received from the 394 respondents to the first round, giving a second round response rate of about 55 percent. The questionnaire for the second round was modified slightly in light of results from the first round questionnaire. First, a section of true / false questions about tropical rain forests was eliminated. Analysis of the first round data indicated that other ways of determining respondent knowledge were more significant in explaining WTP. Second, three demographic questions were added including the number of people in household, the number of people who work outside the home, and the number of people in high school and in college. The format of the income question was changed to increase the number of responses. The first round used an

open-ended income question; the second round presented 20 income categories and asked the respondent to indicate the category that included their household income for the previous year. The second round questionnaire also included a few sentences explaining why the income data was important to the interpretation of the survey data.

A fourth change incorporated Dillman's suggestion to have the first question be neutral, easy, and applicable-to-everyone. The added questions asked respondents to describe where they live (large city, town, farm, etc.) and to describe family participation in outdoor recreational activities by checking activities that apply from a list. The fifth change made the final question in the section about attitude toward the environment more neutral. The last question of that section in the first survey asked for level of agreement or disagreement with the statement, "Humans are severely abusing the environment". Ending the section with such a strong statement led several respondents to comment on a perceived bias in the questionnaire. We added another statement so that the final question is, "As we learn more about the environment, people are managing it more wisely." (Responses to each round were scored only on the 12 questions testing agreement with the New Environmental Paradigm.)

The final changes modified the WTP scenario by suggesting a more specific form of organization to provide tropical rain forest protection and added a question about willingness to donate time (in addition to money) to a sponsorship organization for tropical rain forest protection. All of the changes were believed to correct minor

deficiencies in the initial questionnaire, yet leave the task presented to the respondents unchanged.

Results

A measure is reliable if the same results occur on repeated trials of an experiment, test, or any measuring procedure. Four methods were used in this study to assess the reliability of WTP bids: stability in the type of bid, the correlation between two parallel measures, common significant variables appearing in the "best fit" regression equation for each period, and similarity of coefficients in a "best fit" regression equation applied to each period.

Change in Type of Bid

Reliability of WTP bids can be assessed by checking whether the respondents who answered both surveys changed the type of their bid. Three types of bids could be given: a positive bid, a zero non-protest bid, and a zero protest bid. Table 1 is divided into three sections--those who gave the same type of bid, those who gave a similar type of bid (a positive bid and then a zero non-protest bid or vice versa), and those who had a complete change of bid type (a zero non-protest bid to a zero protest bid or vice versa). Nearly two-thirds of the respondents kept the same bid and 78% gave the same or similar types of bids.

TABLE 1.
COMPARING TYPES OF WTP BIDS

1st --> 2nd BID	# IN BOTH SURVEYS	% OF TOTAL
KEPT THE SAME BID	**141**	**64.68%**
BID --> BID	35	16.06%
NP --> NP	82	37.61%
P --> P	24	11.01%
SAME TYPE OF BID	**29**	**13.30%**
BID --> NP	17	7.80%
NP --> BID	12	5.50%
COMPLETE CHANGE	**48**	**22.02%**
BID --> P	5	2.29%
NP --> P	17	7.80%
P --> BID	8	3.67%
P --> NP	18	8.26%
TOTAL	**218**	**100%**

BID: The respondent gave a positive WTP bid.
NP : The respondent gave a zero non-protest bid.
P : The respondent gave a zero protest bid.

If the bids from those respondents who received each survey are aggregated into the three general types of bids, then a similar story is told. Table 2 illustrates the number of bids, zero non-protest bids, and zero protest bids for each survey. Clearly

the number of bids in each type remained very stable over the time period, even on a

subject the respondents knew little about.

TABLE 2.
AGGREGATE TABLE OF BID TYPES

TYPE OF BID	1st SURVEY	2nd SURVEY
POSITIVE BID	57	55
NON-PROTEST BID	111	117
PROTEST BID	50	46

Correlation Between Two Parallel Measures

The standard measure of reliability is the correlation between parallel scores on

the same test at two different points in time. This method can be used with a test-

retest survey. The correlation coefficient between the two WTP bids of the 146

respondents who provided non-protest bids on each round was 0.63, indicating a fairly

high level of reliability.

Common Significant Variables

If a measure of WTP is reliable, the independent variables that explain variation

in the responses from each round should be the same. The responses for each

round were used to estimate a regression equation explaining the variation in

expressed WTP. Similar independent variables were considered in developing the

models for each round, although the exact form of some variables differed between the two questionnaires. The second round added a question about the respondent's willingness to donate time in addition to the monetary bid given.

The best equation for each round is presented in Table 3. Independent variables used are INCOME, representing income categories ($5,000 intervals); KNOWLEDGE, a dichotomous variable where 1=high knowledge (scores of 5 or 6 on responses to an open-ended question where scores could range from -3 to 6); INCOME*KNOWLEDGE, an interaction of two dichotomous variables where income=1 for income categories greater than $65,000 and knowledge was scored as above; SCHOOLING, representing categories of schooling completed ranging from 1= "no formal education" to 10= "a graduate degree;" ENVIRONMENTAL ATTITUDE, the score on the test of agreement with the new environmental paradigm (range=12-60, higher scores indicate stronger agreement); DONATE TIME, a dichotomous variable where 1=willing to donate time (only asked on the second round), AREA, a dichotomous variable where 1=an urban residential location; LEARNING, a dichotomous variable where 1=active forms of learning about tropical rain forests, such as, attending meetings or talking with family or friends (passive learning included watching TV news programs or specials), and USE VALUE, a dichotomous variable where 1=an affirmative response to any of three questions about visits or intended visits to tropical rain forests or the household use of products from tropical rain forests.

TABLE 3.
COMPARISON OF TOBIT REGRESSION VARIABLES

VARIABLES	1st Survey Coefficients	1st Survey Significance	2nd Survey Coefficients	2nd Survey Significance
INTERCEPT	159.80	0.0000	144.50	0.0000
INCOME	5.43	0.1825		
KNOWLEDGE (1=HIGH)	115.95	0.0084		
INCOME* KNOWLEDGE (1=EACH HI)			308.13	0.0000
SCHOOLING	18.98	0.0829	18.07	0.0134
ENVIRON. ATTITUDE	-5.65	0.0001	-4.34	0.0002
DONATE TIME (1=YES)	NA	NA	85.70	0.0122
AREA (1=URBAN)			-55.25	0.0932
LEARNING (1=ACTIVE)			149.44	0.0322
USE VALUE (1=YES)	87.05	0.0239		
N	123		130	
ADJ R^2	.2016		.4268	
LOG LIKELIHOOD	-358.18		-337.35	

The independent variables included in the best fit equations all have coefficient

estimates significantly different from zero at the usual levels of significance except

income in the 1st survey equation. The significance level ($\alpha = 0.18$) is greater than usually accepted in hypothesis testing, but the variable is retained in the equation because of its theoretic significance and the fact that other variables considered but rejected required much higher significance levels to be considered significantly different from zero. With that *caveat*, we observe that each equation includes several of the same variables with statistically significant coefficients. Income and knowledge of tropical rain forests, either directly or as an interaction, schooling, and environmental attitude are included in each equation. These variables account for four of the five variables included in the 1st round equation and three of the six variables in the 2nd round equation. While this test shows a weaker indication of reliability than the previous ones examined, it indicates a stability of underlying relationships over time.

Similarity of Coefficients

In this test of reliability we applied the best fit equation for the second round survey to data from the first round and compared the coefficients in the two equations (Table 4). TIME, the willingness to contribute time to efforts to protect tropical rain forests, was not available for the first round, but the remaining variables from the second round equation were used. The signs of the coefficients for the common variables are the same in each equation. The alpha-level for three of the variables is significantly higher with data from the first survey. The effects of AREA and of LEARN are not significant, while the significance of INCOME*KNOWLEDGE is doubtful. The

low adjusted R^2 in the equation using first round data also suggests that the best fit equation for the second round does not fit the first round data very well. (It should be noted that the best fit equation for the first round also had a low adjusted R^2.) Still, this measure provides little support for the hypothesis of CV reliability.

TABLE 4.
COMPARISON OF TOBIT REGRESSION COEFFICIENTS

VARIABLES	1st Survey Coefficients	1st Survey Significance	2nd Survey Coefficients	2nd Survey Significance
INTERCEPT	170.73	0.0000	144.50	0.0000
ENVIRON. ATTITUDE	-6.12	0.0001	-.4.34	0.0002
SCHOOLING	36.02	0.0006	18.07	0.0134
INCOME* KNOWLEDGE (1=EACH HI)	108.39	0.1692	308.13	0.0000
AREA (1=URBAN)	-19.41	0.6157	-55.25	0.0932
LEARN (1=ACTIVE)	49.15	0.3934	149.44	0.0322
DONATE TIME (1=YES)	---	---	85.70	0.0122
N	122		130	
ADJ R^2	.131		.4268	
LOG-LIKELIHOOD	-369.47		-337.35	

Conclusions

This study examines the reliability of CV estimates of a general population's willingness to pay for the protection of tropical rain forests, a subject unfamiliar to most respondents. Of the four measures of reliability examined, two gave indications of reliability. First, bids given nearly a year apart were of the same type for most respondents. That is, there was little switching among positive value bids, zero bids, and protest bids. Second, there was a rather high correlation between the bids given in the two rounds of the study. The third measure of reliability, common significant variables in best fit regression equations for each round, showed that the underlying explanatory relationships are rather stable over the period of the study, although some differences appear. The fourth measure, similarity of coefficients in equations estimated with the same variables using data from each round was not strong in supporting similarity and thus, gives little support to a hypothesis of reliability. Overall, we conclude that respondents seem to have given similar answers to each round of the study and there is evidence, albeit not overwhelming, that the CV method is reliable even when applied to problems not familiar to most respondents.

References

Dillman, Don A. Mail and Telephone Surveys--The Total Design Method. Wiley, New York. 1978.

Dunlap, Riley E. and Kent D. van Liere. "The 'New Environmental Paradigm': A Proposed Instrument and Preliminary Results." J. of Environ. Educ. 9(1978):10-19.

Griffith, Alan Brent. A Contingent Valuation Analysis of the Value of Tropical Rain Forest Protection by Pennsylvania Citizens. unpublished M.S. thesis. The Pennsylvania State University. 1991.

Kealey, Mary Jo, Mark Montgomery and John F. Dovidio. "Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?" J. Environ. Econ. and Mgmt. 19(1990):244-263.

Loehman, Edna T. and Vo Hu De. "Application of Stochastic Choice Modeling to Policy Analysis of Public Goods: A Case Study of Air Quality Improvements." Rev. Econ. and Statist. 64(1982):474-480.

Loomis, John B. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." Am. J. Agr. Econ. 71(1989):76-84.

Mitchell, Robert C. and Richard T. Carson, Using Surveys to Value Public Goods: The Contingent Valuation Method. Washington: Resources For The Future, 1989.

Musser, Wesley N., Keith A. Lampi, Lynn M. Musser and Frederick W. Obermiller. "Test-Retest Reliability of Contingent Valuation Methods." Paper presented at the Northeastern Agricultural and Resource Economics Association meetings, Orono, ME, 1988.