JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT **31,** 287–301 (1996)
ARTICLE NO. 0045

Do Contingent Valuation Estimates Pass a ''Scope'' Test? A Meta-analysis*

V. KERRY SMITH†

Department of Economics, Duke University, Durham, North Carolina 27708 and Resources for the Future, Washington, DC 20036

AND

LAURA L. OSBORNE

Department of Economics, Georgia State University, Atlanta, Georgia 30303

Received March 15, 1995; revised January 23, 1996

This paper considers the scope test proposed to judge the internal consistency of contingent valuation estimates. The test is shown to be quite sensitive to the maintained hypotheses required to derive fairly precise expectations for the properties of WTP functions, and, therefore, a different approach may be needed in gauging the reliability of CV. This paper describes an approach that relies on a weight-of-the-evidence criterion and uses meta-analysis to develop a systematic appraisal of what the economic values of changes in amenity resources are. The approach is illustrated for the case of estimating people's willingness to pay for improving (or maintaining) visibility at the national parks.

I. INTRODUCTION

Few aspects of current economic research created as much controversy as the use of contingent valuation to elicit people's values for environmental resources. While contingent valuation (CV) surveys have been in the literature over 20 years, until recently they attracted little attention from most economists.[1] The Exxon-Valdez oil spill changed this state of ''tolerant'' indifference and marked the beginning of a continuing debate over the reliability of CV. A central premise of those arguing that ''. . . CV is a deeply flawed methodology that does not estimate what its proponents claim to be estimating'' [12, p. 62] is that there have been no internal tests of the validity of CV results.[2] For Diamond and Hausman [12], one

* Thanks are due Richard Carson, Laurie Chestnut, Ron Cummings, Peter Diamond, Bob Rowe, and three anonymous referees for most helpful comments on earlier drafts and to Paula Rubio, Brent McLamb, and Carla Skuce for preparing and editing those earlier drafts of this paper. Partial support for this research was provided by the UNC Sea Grant Program under Grant R/MRS-25.

† To whom correspondence should be addressed.

[1] For an overview of the development of contingent valuation, see Hanemann [15, 16].

[2] This interpretation can be found in Diamond and Hausman [11, 12], as well as in Diamond *et al.* [13]. For example, they suggest that ''internal consistency tests (particularly adding-up tests) are required to assess the reliability and validity of such surveys [CV surveys]. When these tests have been done, contingent valuation has come up short. Contingent valuation proponents typically claim that the surveys used for these tests were not done well enough. Yet they have not subjected their own surveys to such tests . . . . There is a history of anomalous results in contingent valuation surveys that seems closely tied to the embedding problem. Although this problem has been recognized in the literature for over a decade, it has not been solved. Thus, we conclude that current contingent valuation methods should not be used for damage assessment or for benefit cost analysis'' (Diamond and Hausman [12], pp. 62–63, bracketed phrase added).

0095-0696/96 $18.00

way to evaluate the CV method is to compare willingness to pay (WTP) functions estimated with CV surveys with the specific, observable properties that economic theory implies WTP should follow. Diamond and Hausman suggest this type of internal validity test is especially important in situations where nonuse values dominate the responses used to construct monetary measures of economic value.

The purpose of this paper is to consider one of the proposed internal-consistency tests for CV-based estimates of WTP. This test requires CV estimates of willingness to pay to be responsive to the amount, or *scope*, of the environmental amenity being offered. To test whether CV estimates of WTP are responsive to scope, we conducted a meta-analysis that pools WTP estimates for visibility improvements at U.S. national parks from five different contingent valuation studies. Our findings support a positive, statistically significant (and robust) relationship between the WTP estimates and the percentage improvement in visible range.

Before the results of our meta-analysis may be considered evidence that WTP estimates based on contingent valuation are consistent with an internal validity test, the requirements of a scope test must be defined. The scope test was initially proposed by a panel of social scientists composed by the General Counsel of the National Oceanic and Atmospheric Administration (NOAA). They were asked to evaluate the reliability of CV for measuring the passive-use values of environmental resources involved in natural-resource damage assessments.[3] The panel's report (Arrow *et al*. [2]) calls for WTP estimates that are ''adequately'' responsive to the scope of the environmental insult. Four of the original six NOAA panelists submitted subsequent comments to clarify the meaning of ''adequate responsiveness.'' They noted a clear distinction between observing a statistically significant difference in the WTP estimates for two different sized changes in an environmental good and ''plausibly responsive'' values, stating that: ''A survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult. This, of course, is a judgment call, and cannot be tested in a context-free manner . . . '' (Arrow *et al*. [1]).

To understand the theoretical expectations for ''plausibly responsive'' estimates of WTP, Section II outlines the properties WTP responses might have under alternative descriptions of how environmental resources enter individual preferences. We use our meta-analysis to examine whether the existing empirical evidence indicates plausibly responsive estimates of WTP have been elicited in CV surveys for one environmental commodity. Section III describes the studies included in our meta-analysis and presents the results. To include estimates where both use and nonuse considerations are likely to be factors in respondents' choices, we used CV studies that had both on-site surveys and surveys conducted at locations that were substantial distances away from the national parks. Variables that control for features of each study are also included in the WTP models. The last section discusses the implications of this application for broader uses of meta-analysis.

[3] The panel was chaired by Kenneth Arrow and Robert Solow. The other panelists included Edward Leamer, Paul Portney, Roy Radner, and Howard Schuman. See Portney [23] for a discussion of the NOAA Panel's assessment. The full text of the NOAA panel report is in the *Federal Register* **58** (10), 4601–4614 (1993).

II. TESTING ECONOMIC CONSISTENCY OF WTP

Measures of the economic value of any object of choice are constructed from people's decisions. Revealed preference (or indirect) methods used to estimate the value for nonmarketed goods rely on theoretical restrictions to supplement observed decisions involving marketed goods. These restrictions are maintained assumptions that are required to isolate estimates of WTP for nonmarketed environmental resources. As a result, the analyst cannot discern whether inconsistencies in any estimated WTP value indicate an unreliable method or the failure of one or more of the maintained assumptions.[4] Although CV surveys can provide direct estimates of WTP, they are not free of important maintained hypotheses. Indeed, the process of defining tests for hypotheses about the economic consistency of contingent valuation's WTP estimates requires additional maintained hypotheses.

A recent example illustrating the assumptions needed to define a scope test for WTP estimates is Diamond's [10] derivation of the properties of a WTP function. Diamond's analysis demonstrates that if preferences can be described with a quasi-linear, indirect-utility function, and if the object of choice involves saving an amount of the environmental resource from being lost (e.g., avoiding a specified number of birds being killed from a larger population), then a scope test would require the ratio of two WTP estimates for two different numbers of birds saved to be greater than the ratio of the respective numbers of birds saved. For example, if one were to compare saving 1000 versus 100,000 birds, this framework would imply the WTP to save 100,000 birds must be at least 100 times greater than the WTP to save 1000 birds. Unfortunately, this clear-cut expectation for the scope properties of the WTP function does not hold generally. If we assume the amount of a resource saved by the proposal is not a perfect substitute for the original environmental resource, as might be the case if people do not believe the proposal would be completely effective or if the resource is not perceived by individuals on a simple metric such as a count of birds lost, then it is straightforward to show that the only bound on the scope test we can provide is that of the NOAA panel—a larger amount of an object of choice should have a larger WTP than a smaller amount.[5]

To understand fully how an individual's WTP responds to all types of changes in the object of choice (or, equivalently, the economic commodity) would require a

[4] This general point has been carefully illustrated by Harrison [17] in his critique of experiments testing the expected-utility hypothesis.

[5] For instance, contrast the following example describing the implications of assuming perfect substitution between a proposed policy to change a resource and the level of the resource. Assume indirect utility is quasi-linear with a CES sub-function describing the role of the environmental resource (e.g., $V(I, b) = I + \alpha[B^s - b^s]^{1/s}$, where I = income, B = the population of birds, and b = the number of birds lost without policy intervention). In this case, the ratio of WTP to save two different amounts of birds, say 100,000 versus 1,000, can range from about 95.48 when the elasticity of substitution is 100 (which is comparable to the considering the implementation of the policy to affect the birds as a perfect substitute for the initial status of the resource) to only 1.5 for when the elasticity of substitution is 3.55. The rationale for considering alternatives to Diamond's perfect substitution assumption follows from early work by Hirshleifer [18] considering the effects of the production function for public goods on the incentives for individual action contributing to production of a public good. More recently, Cornes [8] has demonstrated that use of a CES function as a social composition function offers one way to generalize the Hirshleifer arguments.

rich CV design in which the issues influencing the perceived substitutes for the resource and other aspects of a proposed policy could be varied as part of the survey instrument. By controlling for these other influences it would be possible to evaluate the partial effect of scope. A more modest alternative is available if we are willing to use past CV research. That is, by using existing CV estimates for specific environmental resources, it is possible to partially mimic the survey design required to evaluate these issues. Evaluating WTP estimates in this way is one particular use of meta-analysis for testing the theoretical properties of an underlying behavioral relationship. In what follows, we use measures of the economic value of visibility improvements at parks in the U.S. to illustrate our proposed approach for evaluating CV estimates of WTP.

III. META-ANALYSIS ESTIMATES OF WTP FUNCTIONS FOR VISIBILITY CHANGES AT NATIONAL PARKS

A scope test meeting the recommendations of the NOAA Panel must demonstrate that WTP is responsive to the amount of the economic commodity offered. As Arrow *et al*. [1] note, this demonstration is not simply isolating statistically significant differences; it also requires that the differences be consistent with economic intuition. Unfortunately, as they acknowledge, these prior expectations will be context dependent. Ideally, we would have some prior knowledge about how people conceptualize what is offered to them in a CV choice. Although we know people can accurately relate differences in photos of different visibility conditions to changes in the visible range, this does not mean these differences fully characterize how each individual interprets a CV question offering a policy or plan intended to provide some visibility change.[6] To address this issue, we consider the sensitivity of our results to both the models used and the sample composition.

The data for our analysis of households' valuation of improvements in visibility were assembled from a review of thirteen contingent valuation studies presenting air-pollution changes and their influences on visibility and other local conditions. We selected five of these studies for the meta-analysis because they used comparable methods. These studies focused on air quality as a key element in visibility and associated vistas that are present at specific national parks.[7] In addition, these studies present the change in air quality in a way that permits computation of the change in visible range. Table I identifies the five studies and indicates the number of observations drawn from each study, the author(s) of each study, the park at which visibility is valued, a few summary measures of the survey design, WTP estimates, and the visibility changes considered in each study. Each observation in our meta-analysis includes, as the dependent variable, the estimate of mean

[6] See Chestnut and Rowe [5] and Balson *et al*. [3] for a summary of this evidence. Rowe and Chestnut [25], Stewart *et al*. [27], and Stewart *et al*. [28] provide support for the linkage between perceptible visible range and visible range measured in miles.

[7] Several of the original studies considered for this meta-analysis valued visibility changes in urban areas. These studies are not considered here because they sought to evaluate multiple interrelated components of air pollution including visibility changes and health effects. Synthesizing these studies requires a consistent treatment of the health effects, which requires more consistency of the health-effect measures than appears to be available from the studies. For this reason, and because the health effects are separate from the scope question considered here, these studies were not included in the final analysis.

TABLE I
Summary of CV Studies for Visibility at National Parks[a]

Authors	Summary statistics[b]		Location	Type of survey
Row *et al*. [24]	O	6	Navaho Recreation Area	In-person interviews administered to to households in area
	$\bar{x}$	\$9.27		
	$\tilde{x}$	\$8.64		
	q	\$6.83–\$10.82		
	$\bar{v}$	0.50		
MacFarland *et al*. [22]	O	8	Grand Canyon and Mesa Verde National Parks	In-person interviews administered to visitors to the area
	$\bar{x}$	\$2.75		
	$\tilde{x}$	\$2.68		
	q	\$1.69–\$3.73		
	$\bar{v}$	1.18		
Schulze *et al*. [26]	O	20	Grand Canyon, Mesa Verde, and Zion National Parks	In-person interviews administered to households in Albuquerque, Los Angeles, Denver, and Chicago
	$\bar{x}$	\$8.50		
	$\tilde{x}$	\$7.00		
	q	\$4.42–11.67		
	$\bar{v}$	0.79		
Chestnut and Rowe [5]	O	72	Grand Canyon, Yosemite, and Shenandoah National Parks	Mail with telephone follow-up survey administered to households in Arizona, Virginia, California, New York, and Missouri
	$\bar{x}$	\$4.35		
	$\tilde{x}$	\$4.20		
	q	\$3.15–\$5.48		
	$\bar{v}$	0.62		
Balson *et al*. [3]	O	10	Grand Canyon National Park	In-person interviews administered through a pilot study conducted in St. Louis and San Diego Counties
	$\bar{x}$	\$0.46		
	$\tilde{x}$	\$0.12		
	q	\$0.007–\$0.97		
	$\bar{v}$	0.955		

[a] A detailed description of these studies is available from the authors upon request.
[b] O = number of observations taken from the study, $\bar{x}$ = mean of the mean willingness to pay per month used for the meta-analysis in 1990 dollars, $\tilde{x}$ = median value of willingness to pay, q = interquartile range of mean willingness to pay, $\bar{v}$ = mean proportionate change in visibility.

willingness to pay (MWTP) for a visibility change as originally reported in each study. This mean is typically computed over some subset of the total study sample (i.e., subsamples based on specific changes in visibility or survey design). Each MWTP estimate has been elicited as (or converted to) a monthly payment expressed in 1990 dollars using the Consumer Price Index.

The proportionate change in distance of visible range is the representation for the amenity being valued. Each CV study used photos depicting the change in visibility offered at each national park to describe the commodity to survey respondents. Because there is substantial evidence supporting people's ability to connect the differences in visible range presented with photos displaying different visibility conditions to the proportionate change in the distance, this proportionate

change is used in our statistical summaries of people's responses to the CV questions associated with these proposed changes. As might be expected, the photos used in each study depicted different scenes (even when they were studying the same parks). Moreover, the CV questions were intended to ask respondents to abstract from the specific location of the pictures and to use each photo as a means of portraying these conditions. Thus, it is important to acknowledge that the role of visibility as a complementary dimension of a vista depends on what is actually presented as part of the CV question in relation to the intended object of choice—the visibility conditions at a specified park.

The text of the CV questions requires a description of the frequency (or probability) that visibility conditions would resemble those in the photographs. For example, the survey used in the Chestnut-Rowe [5] study described a range of visibility conditions and identified the average conditions as part of the description. In their CV questions, respondents were asked about changes in those average conditions. By contrast, the Balson *et al*. [3] study presented and asked valuation questions about the changes for a specific number of days during a season. Although there is reasonably wide variation in the proportionate changes in visibility conditions across all five studies, each study does not treat other characteristics that might also define the commodity, such as the number of days when the changes take place, in a systematic way. Therefore, in addition to the absolute value of the proportionate change in the visible range, we include variables describing the commodity, the complementary resources, or any descriptive information that we hypothesized would influence how the commodity is interpreted.

The first commodity-related variable included in the data is a qualitative variable that distinguishes surveys for eastern versus western parks. We separated the visibility effects for these two locations because the vistas affected by changes in visible range are quite different in western parks than in eastern sites.[8] A second issue related to the commodity arises because in some cases the survey questions ask respondents about their WTP to avoid a decline in visibility. This framing corresponds to an equivalent surplus definition rather than the compensating surplus format of the WTP questions. Because we have treated all visibility changes as positive values (i.e., the proportionate change in visible range that is provided or avoided), a qualitative variable indicating whether or not the WTP is for an increase or decrease in visible range is also included. The last commodity-related variable included concerns the extent of the change—whether visibility changes are to occur at a single park or for the southwest region as a whole. In the abstract, we might regard this variable as expanding the ''amount'' of improvement in visibility. However, it is not clear how respondents would interpret questions that focus on local verses regional changes. Because visibility changes are generally explained to respondents as arising from changes in air pollution, the physical nature of the process that ''delivers'' them to a single park affects, to some degree, all parks in the same region. To the extent respondents understand this mechanism, they may well include some perception of effects throughout the area in their responses to a question about visibility at a specific park, even when they are not explicitly identified.

[8] This format (measuring the commodity as the proportionate change in visibility, and distinguishing between eastern and western sites) is also used in the first statistical summary of CV evidence reported by Chestnut and Rowe [6].

Our sample is composed of a limited number of studies and includes studies using practices that are not part of current CV methodology. To attempt to account for this limitation, variables describing several aspects of the survey design are also included. These qualitative variables identify (1) the elicitation format—whether the WTP was elicited with iterative bidding, a direct question, or a format that allowed respondents to select their desired response from a list of values presented to them (payment card or checklist); (2) whether the interview took place at the park in question (or with households at other locations); and (3) whether households live in the state where the park is located. Some surveys asked respondents their WTP to achieve the specified visibility change at the respondents' next visit to the park. These scenarios are identified with a qualitative variable because they introduce some uncertainty (i.e., whether the respondent would have a future use) and because they link the change in visible range to future use.[9]

Before discussing our results, several reasons for caution in interpreting them need to be highlighted. In some cases, the qualitative variables defined to represent specific effects may capture multiple influences because the effects are present in only one study. For example, Balson *et al*. [3] was the only study to use a direct question to elicit WTP. As we noted earlier, it also focused on short durations of visibility change. Similarly, Rowe *et al*. [24] was the only study to use iterative bidding to elicit WTP, and Chestnut and Rowe [5] was the only study that involved a mailed survey. These examples of study-specific factors highlight how important it is to acknowledge that a statistical summary involving a limited number of studies can confound specific study attributes with a fixed effect that could be included to identify each study. Thus, a summary model using the data across studies can include either attributes that are unique to studies or a fixed effect factor, but not both. Whichever approach to distinguishing features of the studies is chosen, the estimated parameters will reflect elements of the other omitted factor.

Table II provides eight model specifications for visibility improvements at national parks. The variable names used in Table II describe the hypothesized influences, economic and methodological, for estimated WTP. Each of these effects is a qualitative variable (other than the proportionate change in visibility) that has a value of one when the effect is present. The qualitative variable identifying eastern sites is entered in the models as an interaction term with the proportionate change in visibility. The default category for qualitative descriptions of the question format is check-off or payment-card format. Because our estimates are based on a pooled sample using the mean WTP as the dependent variable and are drawn from different sample sizes, it is reasonable to assume there will be violations in the ideal conditions usually maintained with ordinary least squares (OLS). To take account of these influences, we used a Huber [19]–White [29] consistent covariance estimator in computing the standard errors for our hypothesis tests. This approach treats each study as the equivalent of a sample cluster with

[9] In each study, the payment mechanism was either an increased entrance fee at the park in question or a monthly addition to the household's electricity bill. A qualitative variable identifying the payment mechanism was not included in the final analysis because it was identified as being highly collinear with the qualitative variable identifying whether or not respondents were asked the WTP for a visibility change that occurs on their next visit to the park.

TABLE II

Summary of WTP Estimates for Improved Visibility at National Parks: Semi-log Models[a]

Independent variables	Mean values	Model							
		(1a)	(1b)	(1c)	(2)	(3)	(4)	(5)	(6)
Intercept		1.326	1.911	1.326	0.916	1.091	0.919	1.102	1.131
		(5.67)	(471)	(3.35)	(4.31)	(13.44)	(4.41)	(9.65)	(20.55)
Proportionate change in visibility[b]	0.709	1.127	1.213	0.806	1.127	0.575	1.112	0.555	0.543
		(2.41)	(6.11)	(6.72)	(2.39)	(21.824)	(2.30)	(11.20)	(18.36)
Eastern park* proportionate change in visibility	0.156[c]	−0.814	−0.892	−0.583	−0.843	−0.420			
		(−2.52)	(−4.46)	(−3.06)	(−2.57)	(−14.10)			
Decline in visibility (= 1)	0.304	0.408	0.607	0.482	0.496	0.416	0.478	0.448	0.447
		(1.91)	(3.40)	(2.26)	(2.18)	(1.33)	(1.53)	(1.23)	(1.23)
Asked region (= 1)	0.661	−0.468	−0.744	−0.306		0.026		0.028	
		(−1.86)	(−1.96)	(−0.84)		(0.49)		(0.424)	
Iterative bidding (= 1)	0.052	0.116	0.321	0.116		0.770	0.571	0.732	0.716
		(0.76)	(0.67)	(0.23)		(2.79)	(1.85)	(2.378)	(2.14)
Direct question (= 1)	0.078	−4.616	−4.252	−4.330	−4.206		−4.196		
		(−14.42)	(−9.06)	(−9.34)	(−20.37)		(−19.92)		
Park interview (= 1)	0.096	−0.621	−0.964	−0.661		−0.775	−0.693	−0.764	0.776
		(−4.88)	(−3.31)	(−2.38)		(−7.41)	(−9.08)	(−7.303)	(−8.01)
Sample composed of residents (= 1)	0.261	0.148	0.200	0.146					
		(1.00)	(1.25)	(0.79)					
Question framed for visibility change at revisit to site (= 1)	0.174	−1.041	−1.212	−0.674	−0.849		−0.558		
		(−2.14)	(−2.65)	(−1.43)	(−1.87)		(−1.64)		
n:	115	115	115	115	115	106	97	88	88
R^2		0.725	0.711[d]	0.738[d]	0.703	0.408	0.725	0.428	0.428
Box–Cox (λ)			0.291						

[a] Numbers in parentheses for models (1a) through (6) are ratios of coefficients to Huber–White consistent variance estimates. With models (1b) and (1c) they correspond to the ratio of the estimated coefficient to the relevant standard error in each case.

[b] This variable is scaled by 100 for these estimates so it is the proportionate change in the miles of visibility in absolute magnitude: [revised − initial]/initial.

[c] Fraction of observations for visibility changes at eastern sites.

[d] The R^2 statistics computed for these models are not comparable to OLS because they are not measured as the same dependent variables.

the potential for heteroscedasticity (i.e., differences in error variance across clusters).

Each model, labeled (1a), (1c), and (2) through (6), was estimated using the log of mean WTP (for specified subsets of the full sample in each study) specified as a linear function of variables describing the type of elicitation procedure and location of the interview. The models labeled (1b) and (1c) correspond to the Box–Cox [4] and feasible generalized least squares (FGLS) estimates for the most complete specification, respectively. The estimated Box–Cox parameter, λ, is

significantly different from zero suggesting the model may have less curvature than what is implied by the semi-log form.[10] The Box–Cox parameter is also significantly different from one, implying that the form lies between linear and semi-log by this criterion. While the coefficients of the Box–Cox form are not directly comparable to the other estimates (i.e., they are not measures of the percentage change in WTP with a change in each independent variable, as is the case of the semi-log), the sign and statistical significance of the visibility-change measures and all other variables are consistent with the semi-log results using either OLS with the Huber estimates of coefficients' standard errors or the FGLS results. The FGLS estimates use the number of observations in each sub-sample, providing the estimate of the mean WTP to correct for heteroscedasticity.

All of the models in Table II implicitly assume that the proportionate change in visible range is an adequate measure of the economic commodity. This formulation implies that respondents interpret the policy as providing improvements highly substitutable for the original conditions.[11] The remaining models in Table II are distinguished by the variables included as potential determinants of mean WTP responses and the studies composing the sample. Model (1a) is the most inclusive, using the full sample and each of the explanatory variables described above. Model (2) also uses the full sample, but it includes a limited number of explanatory variables. Model (3) is our preferred specification for the WTP function when the Balson *et al.* [3] study of the Grand Canyon is excluded from the data. Because this study emphasizes the small number of days that would be involved in the visibility changes, one might hypothesize that this commodity may be considered differently by respondents. Model (4) is our preferred specification for the WTP function when eastern sites are not included. Again, we hypothesize that changes in visibility conditions at eastern sites might be considered a different commodity and as such arguably should not be considered in the same meta-analysis as visibility changes at western parks. Models (5) and (6) exclude both the Balson *et al.* [3] study and eastern sites, and model (5) includes more explanatory variables than model (6). These models illustrate that our conclusions about the link between WTP and changes in the visible range are not sensitive to the specified determinants, whether or not the most restricted sample is used.

Regardless of the sample composition or model specification, there is a statistically significant, positive relationship between willingness to pay and proportionate improvement in the visible range. The effect of a change in visible range on WTP is smaller in magnitude when the Balson *et al.* [3] study is excluded from the data. This result is consistent with the framing of the visibility change used in Balson *et al.* It was the only study in our sample to describe the number of days when the visibility change would take place. Because it is not possible to capture this feature,

[10] The general form for the Box–Cox transformation is $(y^{\lambda} - 1)/\lambda$, with $\lambda = 0$ implying the dependent variable is equivalent to $\ln y$ and $\lambda = 1$ implying the dependent variable is y. The likelihood ratio tests using the most detailed model are given as follows: H_0: $\lambda = 0$, $\chi_{(1)}^2 = 87.62$, p-value = 0.000; H_0: $\lambda = 1$, $\chi_{(1)}^2 = 120.72$, p-value = 0.000.

[11] To investigate this issue, we used the CES index function following the arguments outlined in footnote 5 to characterize the proposed change in visibility for each observation in our sample. The incremental improvement in visibility was added to the initial visibility conditions if the situation was described as WTP for an improvement and subtracted if the situation was paying to avoid a decline. A simple grid search using the complete model specification and full sample with selected values indicated that values with the highest R^2 imply large elasticities of substitution (e.g., $\sigma = 12.5$ and $\sigma = 33$).

TABLE III
Analysis of the Specification Sensitivity Analysis for WTP Models

Independent variables	Model: Semi-log, no intercept	Model: Linear with intercept	Model: Linear, no intercept
Intercept	—	5.359	—
		(2.68)	
Proportionate change in visibility	1.056	2.946	3.490
	(5.09)	(4.16)	(4.25)
Eastern park* proportionate change in visibility	−0.771	−2.307	−2.499
	(−4.62)	(−3.07)	(−3.65)
Decline (= 1)	0.671	2.364	3.474
	(1.47)	(1.13)	(1.28)
Asked region (= 1)	0.763	−2.376	2.260
	(8.90)	(−1.07)	(7.02)
Iterative bidding (= 1)	1.155	1.568	5.331
	(3.16)	(4.29)	(2.29)
Direct question (= 1)	−3.345	−7.716	−2.877
	(−16.11)	(−4.59)	(−3.67)
Park interview (= 1)	−0.579	−3.354	−3.230
	(−4.81)	(−7.92)	(−8.86)
Sample composed of residents (= 1)	0.226	0.358	0.662
	(2.42)	(0.42)	(0.92)
Question framed for visibility change at revisit to site (= 1)	0.334	−2.780	1.848
	(1.25)	(−1.42)	(2.09)
n	115	116	116
R^2	0.821	0.425	0.780

as distinct from other differences in this study's implementation of CV, the observed response in WTP to proposed visibility change includes another dimension when this study is included in the sample. Table III repeats the analysis with the full sample, using a linear form with an intercept as well as both a linear and semi-log specification without an intercept.[12] Overall, the results remain unchanged with these models, although the sign changes for some qualitative variables because the ''default state'' for these variables has changed.

The estimated WTP function embodies a scope response that is due to the specification selected for the function and the measure used to quantify the commodity offered to respondents. Any comparison of our results with Diamond's [10] scope multiplier (i.e. the ratio of the WTP for one quantity change in comparison to the WTP for another smaller quantity change of the same commodity ''should'' exceed the ratio of the quantities offered) is conditional on these maintained hypotheses. Nevertheless, if we simply compare different levels for the proportionate change in visible range, models with estimated coefficients for this variable exceeding about 0.5 will satisfy the Diamond lower bound. This is consistent with our estimates of the implied substitution between the change offered and natural conditions (see footnote 11). It would not be satisfied for visibility changes

[12] Based on the argument that the WTP function should initiate at zero when there is no change in visibility, Chestnut and Rowe [6] used a model without an intercept in their analysis of a smaller set of the CV studies. Thus, we include a similar analysis for comparison purposes.

at the eastern sites. However, this is not a violation of the scope condition because it can as easily reflect the nature of visibility changes in that context.

Before considering a second measure of the plausibility of our estimated responsiveness of WTP to scope, we consider the effects that the factors related to the commodity have on the WTP estimates. In an earlier metasummary of CV studies for visibility, Chestnut and Rowe [6] indicated that eastern sites have lower values than the western locations. This distinction is also supported by our models, suggesting a different type of complementarity between visible range and the scenic vista for eastern sites. The WTP to avoid a decline is larger than WTP to acquire an improvement over what are described as current conditions. This effect is somewhat sensitive to sample composition, and the coefficient estimates are only statistically significant (at the 10% level) for models that include the Balson *et al*. data.

When the questions are linked to future uses, through the description of how the change would be available (at the time of the next visit to the park), the WTP responses are significantly lower. Questions about visibility conditions for the whole region lead to lower WTP responses as well, but these findings are sensitive to the sample composition and to the estimator (e.g., model (1b)). Moreover, when we consider models excluding an intercept (see Table III), this effect becomes positive and highly statistically significant.[13] If this variable were treated as a measure of the ''amount'' of improvement, then this result could be interpreted as being incompatible with one dimension of a scope test. However, respondents may not interpret regional changes as larger than park-specific changes (e.g., they may assume that policies affecting air pollution that impacts a specific park would also affect nearby parks because of the physical processes determining visibility). Without specific information on how people conceptualize ''more'' visibility, we cannot resolve this issue.[14] However, it is clear that the positive and significant relationship between WTP and the proportionate change in visibility is not affected by the treatment of this variable.

The Balson *et al*. [3] study was the only one to use a direct-question format. The coefficient estimate for ''direct question'' indicates MWTP responses were significantly lower than other studies, but this effect cannot be distinguished from the lower number of days specified in this study as well. The Rowe *et al*. [24] study generally elicited responses that are significantly higher than the other studies. We cannot say this result is due to their use of ''iterative bidding,'' because they were the only study to use this format. The MWTP values from respondents interviewed at a park are significantly lower than responses elicited elsewhere. Of course, as in the other attributes, this effect may be confounded by the issues specific to the individual studies, and it may not be a reflection of the effects of on-site interviews.

[13] The qualitative variable indicating whether the visibility change was for the whole region was also entered as an interaction term with the proportionate change in visibility in each model. This effect was never statistically significant at any conventional level of significance.

[14] To interpret this shift variable (''region = 1''), it is important to recognize how the meta-analysis is using the CV results. When studies considered regional changes in visibility, the variable identifying these studies implies a lower intercept, all else equal, for the log (WTP) function. We could easily interpret this variable as reflecting a substitution effect between parks and not a scope effect of the commodity offered. This is why resolution of this issue cannot be accomplished without further information on people's understanding of what was asked.

Only two studies conducted interviews at a park site, MacFarland *et al*. [22] and Rowe *et al*. [24], which contributed 3 observations in this category.

Finally, the estimated parameter for a variable identifying respondents as residents of the state where the park (being studied) was located has a positive but not statistically significant effect on models (1a) through (1c). However, in the log-linear model with no intercept (see Table III) this coefficient estimate is statistically significant, suggesting use-related effect.

In terms of our overall objectives, the consistent link between the size and sign of the relationship between the visibility change and the WTP is not affected by changes in the specified determinants of WTP, the functional form, the sample composition, or the treatment of the intercept. These findings imply the CV estimates for visibility change are related to the changes in the visible range described with the photos used in the original CV studies. Because some of the CV analyses (especially Chestnut and Rowe [5] and Balson *et al*. [3]) were undertaken with respondents who could be expected to have use and passive-use (or nonuse) values, these estimates are responsive to the NOAA Panel's initial focus on measuring passive-use values and are not the result of a large sample accentuating statistical significance (Arrow *et al*. [1]).

The statistical association between WTP and changes in visible range may not, by itself, suggest that the WTP estimates are ''plausibly'' responsive in the manner Arrow *et al*. [1] propose for scope tests. A second approximate gauge of the plausibility of the WTP estimates is to consider how important the various aspects of the economic commodity (or object of choice) are to the WTP models' ability to detect a relationship simply between the proportionate change in visible range and WTP. Focusing solely on aspects of the economic commodity, Table IV compares estimates of the scope effect based on a common model that includes only factors that describe the economic commodity, but vary according to which subsample is

TABLE IV

Sample Composition and WTP—Visible Range Responses

Sample	Visibility coefficient[a]		Initial conditions[b]	Change[b]
	Proportionate change	Eastern park proportionate change		
Full sample	0.685	−0.353	76.1	43.8
(n = 115)	(2.34)	(−2.08)	[56–96]	[25–50]
Omitting Balson *et al*.	0.660	−0.498	87.3	48.5
(n = 106)	(34.97)	(−16.72)	[70–96]	[25–59]
Omitting eastern sites	0.661	—	75.4	41.6
(n = 97)	(2.02)		[56–96]	[25–50]
Omitting eastern site	0.636	—	87.7	46.3
and Balson *et al*.	(54.01)		[70–96]	[25–59]
(n = 88)				

[a] The numbers in parentheses are ratios of the estimated coefficient to the Huber standard errors for the semi-log models relating WTP to visibility conditions, region, and change to be relevant to a future visit with an intercept.

[b] The first statistic corresponds to the mean in miles, and the values in the brackets are the inter-quartile range for the relevant sample.

included to estimate the parameters.[15] While this was implicitly done with the models in Table II, we also changed the independent variables included in those models, thus confounding the effects of sample changes with covariate changes. Table IV also includes measures of the base visibility condition for the relevant sample, the average change in visibility for the relevant sample, and the interquartile ranges for these statistics.

Recognizing that each model in Table IV has omitted variables, we focus on whether or not factors important to the commodity itself are omitted and how these omitted factors affect the estimates of WTP. Perhaps the most important omitted factors involve descriptions of the amount of time the changes in visible range were to occur and the vistas that are affected by these changes. All studies conveyed different information about the timing of changes in visible range. The Balson *et al*. study is the starkest example because it focused on a very small number of days visibility changes were to occur (as little as 10 days in the summer months). The second factor, the vistas affected by visibility changes, are captured at a very simple level with the interaction term between visible range and eastern sites. The photographs and associated vistas used to describe visibility conditions to respondents were different for all studies, even those focusing partially on the same parks. One way to attempt to appraise the effects of both these factors is by changing the sample composition used to estimate WTP. The results in Table IV suggest that including the Balson *et al*. study's estimates or WTP responses for visibility changes at eastern sites, without accounting for their influence, reduces the ability of the proportionate change in visibility alone to explain the results.[16] The estimated standard error (statistical significance) declines dramatically for the visibility measure when we attempt to include Balson *et al*. in the model. Further, omitting both eastern sites and Balson *et al*. results in the most significant association between WTP and proportionate changes in visible range. This is what we would expect because there are more aspects of the economic commodity that are different and unaccounted for in the models for WTP. Examination of the model's sensitivity to reasonable variations in the sample composition, hypothesized to reflect variations in the object of choice, provides our second gauge of the plausibility of our estimated responsiveness of WTP to visibility changes. Overall, the composite results of our meta-analysis for the case of visibility changes do seem to be responsive to the NOAA panel's call for both statistical significance and economic plausibility in the scope test.

[15] The independent variables used in all four models are the variables indicating ''the proportionate change in visibility,'' ''region,'' ''future use,'' and the interaction term between ''eastern sites'' and ''the proportionate change in visibility.''

[16] Levy *et al*. [21] suggested concerns about CV's ability to measure visibility benefits in general terms. Their evaluation was based on a simple comparison of extrapolations from the Chestnut and Rowe [5] study that were used in an EPA policy evaluation of the benefits of regulations on the emissions of the Navajo Generating Station versus what would be implied by the results of Balson *et al*.'s [3] pilot study. While they were careful to identify the multiple considerations in interpreting each study, the authors did not consider pooling the estimates. Our meta-analysis confirms that the large differences in the benefit estimates seem to arise from an economic characteristic of the situations described by Balson *et al*., distinguishing the amount of the economic commodity being offered.

IV. IMPLICATIONS

Recent tests of economic models—whether they relate to choices over lotteries (for laboratory experiments, see Harrison [17]) or they consider monetary measures for WTP based on stated preferences (i.e., contingent valuation)—have highlighted the difficulty in developing unambiguous tests for consistent (or rational) economic behavior. In each case, recognition of the set of maintained assumptions involved should condition the interpretations that can be drawn from seemingly direct tests. We have proposed an alternative criterion that acknowledges the potential to use past empirical literature to quantitatively evaluate the evidence for or against a hypothesis. This practice is common in medical research, where a composite judgment must be developed from a sequence of trials, each testing for a treatment effect under different conditions. Moreover, such composite studies often serve some regulatory action.

Our analysis suggests that it may be possible to use a meta-summary of CV studies to meet a comparable objective in nonmarket valuation. Of course, the crucial issue in applying the method is finding a common metric to measure the commodity. Research on the economic value of changes in the visible range meets this need and is one of the simplest cases one could consider. In the case of visibility changes, reasonably close links have been developed between the technical measurements of changes in the emissions of air pollutants related to visibility, the changes in visible range, and the photographs displaying the visibility conditions. As the complexity of changes in environmental resources increases, however, it is not clear whether the same strategy can be used. Nevertheless, for this case our results indicate clear responsiveness of CV based estimates for WTP to scope. Because these findings indicate that summary models also fit the set of data reasonably well, they are suggestive of a consistent economic relationship between WTP and the proportionate change in visible range.

REFERENCES

1. K. Arrow, E. Learner, H. Schuman, and R. Solow, ''Comments on NOAA Proposed Rule on Natural Resource Damage Assessment 69'' (1994).
2. K. Arrow, R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman, ''Report of NOAA Panel on Contingent Valuationo,'' 58 Fed Reg 46 (1993).
3. W. E. Balson, R. T. Carson, M. B. Conaway, B. Fischhoff, W. M. Hanemann, A. Hulse, R. J. Kopp, K. M. Martin, R. C. Mitchell, J. Molenar, S. Presser, and P. A. Ruud, ''Development and Design of a Contingent Valuation Survey for Measuring the Public's Value for Visibility Improvements at the Grand Canyon National Park,'' draft report, Decision Focus, Inc., Los Altos, CA (1990).
4. G. E. P. Box and D. R. Cox, An analysis of transformations, *J. Roy. Statist. Soc*., Series B, **26**, 211–243 (1964).
5. L. G. Chestnut and R. D. Rowe, ''Preservation Value for Visibility Protection at the National Parks,'' draft final report to U.S. Environmental Protection Agency, RCG/Hagler Bailley (1990).
6. L. G. Chestnut and R. D. Rowe, Economic valuation of changes in visibility: A state of the science assessment for NAPAP, *in* ''Acidic Deposition: State of Science and Technology, Report No. 27, Methods for Valuing Acidic Deposition and Air Pollution Effects,'' Washington, D.C., NAPAP (1990).
7. R. G. Cummings, D. S. Brookshire, and W. D. Schulze, ''Valuing Environmental Goods: An Assessment of the Contingent Valuation Method,'' Rowman and Allanheld, Totowa, NJ (1986).

8. R. Cornes, Dyke maintenance and other stories: Some neglected types of Public goods, *Quart. J. Econ*. **108**, 259–272 (1993).
9. W. H. Desvousges, F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson, ''Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy,'' RTI Monograph 92-1, Research Triangle Institute, Research Triangle Park, NJ, (1992).
10. P. A. Diamond, Testing the internal consistency of contingent valuation surveys. *J. Environ. Econom. Management*, in press (1966).
11. P. A. Diamond, and J. A. Hauseman, On contingent valuation measurement of nonuse values, ''Contingent Valuation: A Critical Assessment'' (J. A. Hausman, Ed.), North-Holland, Amsterdam (1993).
12. P. A. Diamond and J. A. Hausman, Contingent valuation: Is some number better than no number?, *J. Econom. Perspectives* **8**, 45–64 (1994).
13. P. A. Diamond, J. A. Hausman, G. K. Leonard, and M. A. Denning, Does contingent valuation measure preferences? Experimental Evidence, in ''Contingent Valuation: A Critical Assessment, (J. A. Hausman, Ed.) North-Holland, Amsterdam (1993).
14. W. M. Hanemann, Willingness to pay versus willingness to accept: How much can they differ? *Amer. Econom. Rev*. **81**, 635–647 (1991).
15. W. M. Hanemann, Preface: Notes on the history of environmental valuation in the USA, *in* ''Pricing the Environment: The European Experience'' (S. Navrud, Ed.), Oxford U. Press, Oxford (1992).
16. W. M. Hanemann, Valuing the environment through contingent valuation, *J. Econom. Perspectives* **8**, 19–44 (1994).
17. G. Harrison, ''Expected Utility Theory and the Experimentalists,'' working paper 90-04, University of South Carolina, September, 1990.
18. J. Hirshleifer, From weakest link to best shot: The voluntary provision of public goods,'' *Public Choice* **41**(3), 371–86 (1983); and correction in Public Choice (1985), Vol. 46, 221–223.
19. P. J. Huber, The behavior of maximum likelihood estimates under non-standard conditions, *in* ''Proceedings of the Fifth Symposium on Mathematical Statistics and Probability,'' Univ. California Press, Berkeley, CA (1967).
20. P. O. Johansson, ''The Economic Theory and Measurement of Environmental Benefits, Cambridge Univ. Press, Cambridge (1987).
21. D. S. Levy, J. K. Hammett, N. Duan, T. Downes-Le Guin, and D. Friedman, ''Conceptual and Statistical Issues in Contingent Valuation: Estimating the Value of Altered Visibility in the Grand Canyon,'' Rand, Santa Monica, CA (1995).
22. K. MacFarland, W. Malm, and J. Molenar, An examination of methodologies for assessing the value of visibility, *in* ''Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas'' (R. Rowe and L. Chestnut, Eds.), Westview Press, Boulder, CO (1983).
23. P. R. Portney, The contingent valuation debate: Why economists should care, *J. Econom. Perspectives* **8**, 3–18 (1994).
24. R. D. Rowe, R. d'Arge, and D. Brookshire, An experiment on the economic value of visibility, *J. Environ. Econom. Management* **7**, 1–19 (1980).
25. R. D. Rowe and L. G. Chestnut, ''The Value of Visibility: Economic Theory and Applications for Air Pollution Control,'' Abt Books, Cambridge, MA (1982).
26. W. D. Schulze, D. Brookshire, E. Walther, and K. Kelly, ''The Benefits of Preserving Visibility in the National Parklands of the Southwest, VIII,'' draft report prepared for the U.S. Environmental Protection Agency, Washington, D.C. (1983).
27. T. K. Stewart, P. Middleton, and D. Ely, Judgments of visual air quality: Reliability and validity, *J. Environ. Management* **3**, 129–145 (1983).
28. T. K. Stewart, P. Middleton, M. Downton, and D. Ely, Judgments of photographs versus field observations in studies of perception and judgment of the visual environment, *J. Environ. Psych*. **4**, 283–302 (1984).
29. H. White, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* **48**, 817–830 (1980).