Journal of Environmental Economics and Management **44**, 391–425 (2002)
doi:10.1006/jeem.2001.1210

Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?[1]

Trudy Ann Cameron

Department of Economics, University of California at Los Angeles, Los Angeles, California 90095-1477

Gregory L. Poe

Department of Applied Economics and Management, Cornell University, Ithaca, New York 14853; and Jackson Environmental Institute, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Robert G. Ethier

ISO–New England, Holyoke, Massachusetts 01040

and

William D. Schulze

Department of Applied Economics and Management, Cornell University, Ithaca, New York 14853

Received April 15, 1999; revised June 30, 2000; published online December 6, 2001

We advocate a more formal structural approach for comparing WTP for non-market or pre-test-market goods conveyed by fundamentally different preference elicitation mechanisms. Seven independent samples of respondents were asked to value the identical good. Elicitation methods include one actual purchase and six widely used hypothetical choice formats. Using a common underlying indirect utility function (and stochastic structure) allows data for different elicitation methods to be used independently, compared pair-wise (as in much of the earlier literature) or pooled across all samples in one unified model with heteroscedasticity across elicitation methods. Our differences in estimated WTP for the individual models are typical of earlier findings. However, pooled-data models that allow for heteroscedasticity reveal that while there are substantial differences in the amount of noise in the different samples, a common underlying systematic component of the preference structure cannot be rejected for at least four (and possibly five) of these seven elicitation methods.

Key Words: stated/revealed preference; non-market value elicitation; referendum contingent valuation; open-ended; payment card; multiple-bounded; conjoint analysis; willingness to pay; scale factor; telephone/mail survey.

[1] This paper has benefited from the helpful comments of participants at the AERE session at the ASSA meetings, Chicago, January 1998, at a conference of NSF/EPA research grant recipients in April 1998, at a department seminar at the University of New Mexico, Albuquerque, May 1998, at UCLA's brown-bag econometrics seminar during the Spring of 1998, at the University of Sydney in September 1998, and at the W-133 Annual Meeting in Tucson, February 1999. Matt Neidell and Jordi Prat allowed us to sound out one troublesome empirical conundrum. This research was supported by Grant R824688 from the National Science Foundation/U.S. Environmental Protection Agency Partnership for Environmental Research program. Any errors remain our own.

0095-0696/02 $35.00

1. INTRODUCTION

Stated preference information is employed by researchers using conjoint analysis (in the marketing and transportation literatures) and contingent valuation methods (in the environmental and health literatures). A longstanding observation in both these contexts is that different methods of eliciting values often appear to provide systematically different value estimates for the pre-test-market or non-market good in question. Furthermore, these stated preference values are frequently at odds with the values implied by revealed preferences—the preferred source of value information where markets exist.

Each of a number of commonly used non-market value elicitation methods is intended to elicit the same underlying preferences.[2] It would be reassuring if different elicitation formats led to statistically indistinguishable estimates of the systematic parameters of the assumed preference function. The early empirical literature addressing this issue is somewhat fragmented by a reliance mostly on pair-wise, and often ad hoc, comparisons. It is also incomplete because of the tendency of such comparisons to focus on differences between the estimates of mean willingness to pay (WTP) from unrelated models, rather than the structure of underlying preferences.

In this paper, we argue for a more formal structural approach to comparing preferences across samples. We employ a unique survey, designed specifically to allow simultaneous comparison of choices elicited by seven alternative preference elicitation methods (one revealed-preference and six stated-preference) that have been employed elsewhere in the literature. The type of elicitation method is randomly assigned across respondents, which provides an opportunity to directly test the statistical equivalence of the implied preference functions for respondents in each group. Although a telephone survey medium was used for the revealed-preference elicitation and a mail survey medium for the other six, the instruments were otherwise almost identical except for the elicitation method, so we can be reasonably confident that any differences in values across methods do not stem from different descriptions of the good, from distinctly different time frames for the survey, or from different populations being sampled.[3]

In this paper, we first consider the various threads that have been evolving in the expanding literature that addresses apparent systematic differences in willingness to pay implied by different valuation methods. This review highlights the tendency toward pair-wise comparisons of methods, where each member of the pair is one of the seven methods we consider in the present paper. Many of the earlier pair-wise comparisons also employ separate ad hoc specifications for either utility functions or WTP functions and hence are not able to offer formal statistical tests of the comparability of the preferences elicited by each method. The review also draws attention to growing concerns about heteroscedasticity across methods (sometimes referred to as "different scale factors").

A brief description of our own survey instrument, with its seven different elicitation methods, is followed by a discussion of how one might formulate appropriate empirical models, for each type of data, that are based on the same

[2] This has been termed "procedural invariance" (Tversky *et al*. [48], Kahneman and Tversky [33]).

[3] One other minor feature shared by the telephone survey and one of the six mail surveys (but not the others) will be discussed later.

underlying preference structure. We employ a simple indirect utility-difference model and its corresponding WTP function, compatible in both their systematic and stochastic components. This formal preference modeling strategy is critical to our ability to use each of the seven samples either individually, in pairs, or all together in one fully pooled specification that allows a rigorous seven-way statistical test of the equivalence of the underlying systematic preference parameters across these different samples.

The empirical findings, described in Section 5, contain the main points of this paper. Our differences in the implied WTP for each sample, taken alone, are consistent with the sorts of results observed in the earlier literature. However, we show that when the different types of data are pooled, it is not possible, at least for our data, to reject the hypothesis of identical underlying systematic preference function for at least four, and possibly five, of these seven widely used preference elicitation methods. However, a few of caveats are in order. First, there are only limited options among tractable indirect utility functions that can be adapted to all seven types of data. Second, we employ a convenient but non-standard generalization of our model (to systematically varying preferences) that is necessary to permit us to identify key parameters in two of our data sets. Finally, despite the respectable size of the entire sample, some of our sub-sample sizes are smaller than would have been ideal. Because of these constraints, failure to reject the equality of preference parameters may not guarantee that preferences are indeed identical. However, our main contribution is conceptual. We seek to demonstrate that it is entirely possible to compare choice information from fundamentally different elicitation methods within a single unified framework.

Our results accommodate distinctly different amounts of noise (different scale factors or error variances) in the otherwise common preference models used for each type of data. The fact that there remains pronounced heteroscedasticity across elicitation methods is relevant to the choice among elicitation methods for future valuation surveys. As other recent studies have begun to reveal, some methods appear to be much more precise than others. The robustness of our results concerning error variances across methods is supported by the finding that similar elicitation methods display similar amounts of noise, while dissimilar methods display dissimilar noise.

The role of the error variance (or scale parameter) in choice models has been the subject of considerable attention in the recent literature. We propose and employ a novel (but we argue very useful) parameterization of the indirect utility function that precludes negative expected values for latent WTP but allows positive probability in the negative domain. This specification means that identical parameters in the systematic portion of the indirect utility function lead to identical expected values for latent WTP. Nevertheless, different scale parameters in the indirect utility function still mean different variances for the associated fitted latent WTP distributions. Furthermore, if the cumulative fitted probability in the negative domain of latent WTP is converted to a spike at zero (as a way of disallowing negative WTP in the population), then even the expected value of the distribution with this spike will depend upon the magnitude of the estimated dispersion.

To illustrate the effect of the utility function error dispersion parameter on likely observable demand behaviors, we consider the different proportions of our sample that would be predicted to be willing to pay a specified benchmark amount (larger than the fitted mean latent WTP) for the non-market good in question. This proportion differs across elicitation methods. Thus, to be able to predict demand behavior, knowledge of the dispersion associated with the elicitation method appears to be essential. What would be most valuable for predicting actual demand behavior from stated preference choice data would be some means of using common underlying systematic preference parameters, but mapping the dispersion parameter from the particular stated preference method into the likely corresponding dispersion parameter for a revealed preference choice context. This might allow prediction of the distribution of WTP for real market choices. Our data on real choices are unfortunately too limited to warrant such simulations here, but the potential for such exercises in future applications is clear.

Section 6 considers priorities for future research, and Section 7 concludes.

2. REVIEW OF PRIOR FINDINGS

2.1. Comparing Pairs of Elicitation Methods

Many previous studies have undertaken pair-wise comparisons between different elicitation methods. These studies have contributed to our conventional wisdom about the differences that can be expected among the values of non-market goods derived using these different methods.

Carson *et al.* [16] review 616 comparisons between contingent valuation and revealed preference valuations, derived from 83 different studies, and find that stated preference contingent valuation estimates average about 75–90% of corresponding revealed preference values.

With some exceptions (e.g., Kramer and Mercer [34]), most comparisons of dichotomous choice (DC) contingent valuation methods with open-ended (OE) methods suggest that the dichotomous choice methods produce estimates that tend to be larger—sometimes much larger. The DC/OE WTP ratio generally seems to range between 1.1 and 5, although there are a number of exceptions (see Schulze *et al.* [42] and Brown *et al.* [11] for recent reviews). As a caveat, Huang and Smith [31] review the differences between DC and OE elicitation using Monte Carlo methods and conclude that most of the difference seems to arise from specification errors common to the empirical models in the literature. Furthermore, Bohara *et al.* [8] find that OE values are more sensitive to information about overall program costs than are DC values.

Other studies have compared DC elicitation methods with payment card (PC) methods (e.g., Holmes and Kramer [28], Ready *et al.* [38], and Welsh and Poe [49]). The DC/PC willingness-to-pay ratio appears to range between 2.7 and 4.4 in these studies.

There are fewer instances of comparisons between multiple bounded (MB) elicitation methods and DC, PC, or OE methods. “Multiple bounded” is the term used here to describe an elicitation technique where the respondent is allowed to choose among categories that describe the extent to which he or she might be willing to pay a stated amount (e.g., Ready *et al.* [39]). Sometimes, each respondent is also asked to rate their WTP at not just one but each of several different “bid” values. This increases the available information about preferences. Welsh and Poe [49] have attempted some comparisons using this latter elicitation method, finding that DC-, OE-, and PC-elicited values fall within the range of values implied by alternative interpretations of the data generated by the MB approach.

Multiple-choice hypothetical methods, now commonly associated with conjoint analysis (CA) methods originating in the marketing and transportation literatures (and referred to occasionally as "choice analysis" methods), ask respondents to choose among a set of alternatives that differ along several dimensions. Johnson and Desvousges [32] and Stevens *et al*. [46] have explored these techniques in comparison with earlier CV formats and find that CA value estimates seem to exceed other types of CV estimates.

2.2. Comparing Pairs of Elicitation Methods; Explicitly Differing Variances

Boyle *et al*. [10] compare DC with OE valuation responses, using independent samples and corresponding probit and tobit specifications allowing for differing error variances in the two sub-models. A common central tendency cannot be statistically rejected for two of their three pairs of data sets, but the estimated standard deviations are significantly different for all three pairs of data. Both the means and the standard deviations from the referendum-style samples exceed those from comparable OE data sets. Lacking a revealed-preference benchmark value, these authors can only speculate that either OE questions underestimate values, and/or referendum-style questions overestimate them.

Halvorsen and Sælensminde [23] compare DC and OE responses collected from the same individuals and estimated simultaneously as a discrete/continuous bivariate probit-like specification with a variety of alternative assumptions. They find heteroscedasticity across elicitation modes (and within the DC mode), but also different expected WTP. But they do not employ split samples, so the endogeneity of the OE follow-up question may be contaminating the comparison of the two methods.

The present paper differs from Boyle *et al*. [10] in that we employ the underlying utility-difference function as the basis from which all of our sub-model specifications are derived, and we adopt the logistic error assumption most common in random utility models. Furthermore, we combine not just pairs of samples, but seven independent samples. In contrast to Halvorsen and Sælensminde [23], our analysis also benefits from a randomized split-sample design for different elicitation methods, so that joint endogeneity across methods is avoided.

2.3. Comparing Choice Data with Different "Scales"

In the conjoint analysis literature, several papers have recognized the possibility of different scales of the latent variables underlying choices in different samples. Horowitz [29, 30] finds that random-utility choice models are relatively robust to scale differences, but Ben-Akiva and Morikawa [6] find that recognition of differing scales can make a difference to estimation results. Ben-Akiva and Lerman [5] demonstrate the explicit incorporation of scale differences into random-utility choice models. Econometric tests of the equality of the scale parameters across samples were introduced into the marketing literature by Swait and Louviere [47] and have been adapted to the environmental valuation context by Adamowicz *et al*. [1–3] and Boxall *et al*. [9]. Additional marketing applications have included Gupta *et al.* [21], Bell *et al.* [4], and Bockenholt and Dillon [7]. The most comprehensive recent citations are Hensher *et al.* [26] and Louviere *et al.* [36].

The idea of different scales across real and hypothetical elicitation methods also appears in a comment by Haab *et al.* [22] concerning a paper by Cummings *et al.* [17], wherein the incentive compatibility of real and hypothetical responses is assessed. Haab and his collaborators re-estimate the Cummings *et al.* empirical model allowing for different error dispersions and find that implied values are not different. In their approach, they estimated and tested the models (with and without heteroscedasticity) using packaged maximum likelihood discrete choice models. This is accomplished by conducting a line-search across possible values of the unknown error-variance proportionality parameter.

Our paper differs from papers in the conjoint tradition in that the appropriate log-likelihood function differs across our independent samples in many more ways than just the magnitude of the error dispersion parameter. In the conjoint tradition, the types of elicitation methods usually differ only in that some choices are observed and some are hypothetical, although completely analogous. We have one observed discrete choice, one analogous hypothetical discrete choice, but also five additional, very different, hypothetical choice formats. All of these data from different elicitation formats must be accommodated within one unified specification for rigorous statistical testing of the equivalence of the underlying preferences.

Our paper also differs from the Haab *et al.* [22] strategy in that we avoid the device of a grid search and instead specify and estimate an appropriate (although unfortunately much more complicated) log-likelihood function. With seven pooled data sources, we normalize one dispersion parameter to unity and estimate dispersion parameters for the other data types as multiples of the first. These multiplicative dispersion factors are estimated jointly with all of the other parameters in our pooled-data maximum likelihood model.

3. THE SURVEY

Our data consist of responses to one telephone survey and six mail surveys, all of which were conducted with the cooperation of the Niagara Mohawk Power Company (NMPC) in Erie County of New York State. The topic of these surveys was the NMPC Green Choice^{tm} program, wherein randomly selected households within NMPC's service territory were invited to consider either real or hypothetical additional charges on their utility bills for NMPC to plant trees and/or provide energy from renewable sources.[4] Across and within the telephone and mail modes, the survey instrument was essentially identical except for the manner in which consumer values for the proposed Green Choice^{tm} program were elicited.[5] Each of the seven surveys employed a different value elicitation method. Thus, except for possible telephone-versus-mail mode effects, there can be a presumption that the

[4] Poe *et al.* [37] and Ethier *et al.* [19] use a portion of the data from this same study.

[5] The only difference in the telephone survey was that its scenario mentioned that 12,000 subscribers were needed. The five main mail survey variants did not mention any specific number of subscribers, so an additional mail survey was used that employed the same constant bid ($6) as the revealed preference telephone survey and also mentioned the 12,000-subscriber requirement. The existence of this extra sample forms a link between the telephone survey and a subset of our mail survey sample that also received a $6 bid but did not hear about the subscriber requirement.

only explanation for systematic differences in the implied preferences is the difference in elicitation methods.

Overall, adjusted response rates to the different survey variants ranged from about 55% to 62%.[6] The survey instrument was titled "Clean Energy and You." The preamble was common to all seven versions of the survey. First, the Green Choice^{tm} program was introduced as a voluntary partnership between NMPC and its residential customers designed to reduce air pollution and to improve the environment in "our local communities." The program involved two parts: (i) using renewable energy and (ii) planting trees. First, the distinction between non-renewable and renewable energy sources was described. Potential renewable energy sources were identified as wind, solar power, and gas recovered from landfill sites. It was pointed out that these energy sources do not produce air or water pollution, they will conserve resources, but they also tend to cost more than other types of power.

If implemented, the second part of the Green Choice^{tm} program was to plant thousands of trees on public lands throughout upstate New York. These planting projects would be developed with American Forests, the nation's oldest citizen conservation group. Respondents were then informed about the role of trees as "natural air filters, absorbing carbon dioxide (a contributor to global warming) and releasing oxygen into the atmosphere. When planted near buildings, trees help conserve energy by providing shade in summer and windbreaks in winter." In addition, as described by Poe *et al*. [37] a provision point mechanism with a money-back guarantee and proportional rebate was used in both the hypothetical and actual payment surveys. Rondeau *et al*. [40], and Rose *et al*. [41] suggest that this mechanism, in a laboratory setting, can be approximately demand-revealing in the aggregate.

The survey variants for this study were designed specifically to allow comparisons across elicitation methods with minimum ambiguity. An appendix, available from the authors, provides details on the elicitation format for the portion of the questionnaire that differs across subsamples.

4. THEORETICAL AND ECONOMETRIC MODELING OF CHOICES

Whereas many early comparisons of WTP values across elicitation methods employed unrelated models for separate data sets, our objective in this paper is to create one unifying model that can subsume all the different types of choice data produced by our different survey variants. Only when a common preference structure and stochastic specification underlie all of the choice models is it possible to combine them in a single model with constrainable parameters. In what follows, we will describe the components of this unified model individually. Since we employed a split-sample design, we will then be able simply to sum the components to create a single log-likelihood function for the pooled data. This function can then be maximized with respect to a common set of utility parameters that show up (differently) in each component. In this way, utility parameters and error distribu-

[6] An appendix, available from the authors, provides a discussion of our non-response analyses and descriptive statistics on response proportions by survey variant.

tion parameters can be restricted or unrestricted across elicitation modes, and likelihood ratio test statistics can be used to conduct formal hypothesis tests regarding these utility and error distribution parameters.

4.1. The Common Preference Model

There are potentially five distinct indirect utility function parameters in the simplest model we use to compare the preferences implied by respondents' choices under our seven different elicitation methods. Each of these parameters could be assumed to be a simple constant, which would imply a common preference function for all respondents. If it is desired to allow preferences to vary systematically across respondents, each of these five constants can be generalized to a function of a (different) set of exogenous explanatory variables, such as sociodemographic characteristics.

The complete set of five possible environmental enhancement scenarios is considered only in Version 6 of the survey, the conjoint analysis (6-CA) variant:

Option *A*: pay nothing, get no environmental goods
Option *B*: plant 50,000 trees
Option *C*: plant 50,000 trees, provide renewable energy to 1200 homes
Option *D*: plant 100,000 trees, provide renewable energy to 1200 homes
Option *E*: plant 100,000 trees, provide renewable energy to 2400 homes

All of the other versions are concerned only with the choice between "doing nothing and paying nothing" (Option A), or choosing 50,000 trees and renewable energy for 1200 homes at a price (Option C).

The individual's level of indirect utility is presumed to depend upon the numbers of trees and houses affected, and the price of the proposed program. However, due to the minimal number of design points for numbers of trees and homes, the four possible "do something" program alternatives are captured here by four dummy variables—B_i, C_i, D_i, and E_i—rather than by continuous variables for the numbers of trees and homes involved. Let V_i^j be the level of indirect utility if one of these programs is chosen, and let V_i^A be utility with Option A (no program and no payment). We choose for our indirect utility function a form that is convenient for combining the choice data from our various elicitation methods. Indirect utility with and without program participation can be captured, in a simple linear model, by

$$V_i^j = \beta_0^* + \beta_1^* C_i + \beta_2^* B_i + \beta_3^* D_i + \beta_4^* E_i + \beta_5^* (Y_i - \text{price}_i) + u_i^j, \quad j = B, C, D, E \tag{1}$$

$$V_i^A = \beta_0^* + \beta_5^* (Y_i) + u_i^A.$$

Consistent with the usual practice for random utility choice models, we will assume that the error terms, u_i^j and u_i^A, have an extreme value distribution, which will lead to logit-type models in the empirical phase.

We will also assume that β_1^* through β_4^* are strictly positive, implying in the non-stochastic case that the commodities being valued are "goods," not "bads."[7] We will further assume that β_5^* is positive, which is required for rationality in that indirect utility should not decrease with income (or increase with price). This restriction is imposed by estimating β_5^* as $\exp(\beta_5)$, where β_5 is a (potentially systematically varying) parameter which can take on any value dictated by the data. We can similarly restrict the intercept-shifting parameters of the indirect utility function to be positive by estimating each β_j^* as $\exp(\beta_j)$, $j = 1, \ldots, 4$. Then the underlying parameters β_j can take on any value. This is particularly important if β_j is to be converted to a systematic varying parameter that depends on the observed levels of respondent attributes. Why are these parameter restrictions useful? In counterfactual simulations, we do not wish to find predicted values of β_j^* that are less than zero for outlying values of these attributes. While the indirect utility gleaned from the various combinations of the two environmental goods (trees and houses) may be very small (or even essentially zero), we choose to preclude the possibility that it might be negative, on average. It is reassuring that empirically, in the case with homogeneous preferences, these restrictions turn out to be non-binding.

An individual's choice regarding program participation is assumed to depend on whether the difference in indirect utility between participation and non-participation is positive. In our case, with systematically varying parameters with sign restrictions, the indirect utility-difference function for an individual's choice between Option C and the do-nothing alternative, Option A, will be

$$(V^C - V^A)_i = \exp(x'_{1i}\beta_1) - \exp(x'_{5i}\beta_5)\text{price}_i + e_i, \tag{2}$$

where β_1 and β_5 are now vectors of coefficients, and x_{1i} and x_{5i} are (possibly different) vectors of sociodemographic variables that are allowed to affect preferences systematically. The error term $e_i = u_i^C - u_i^A$ is distributed logistic(0, κ).

Vital to the pooled-data model in this paper is the correspondence between the indirect utility-difference function (which determines the discrete choices) and the continuous maximum WTP function (which forms the modeling basis for two of the seven types of data—the open-ended and payment card responses). If we set the indirect utility-difference in Eq. (2) equal to zero and solve for price_i, this dollar value will represent the predicted maximum WTP by any individual in the sample. The requirement that it be possible to solve the indirect utility function for a closed-form expression for WTP represents a non-trivial limitation on the types of indirect utility functions that can be entertained in this study.

The formula for fitted WTP for Option C is

$$\text{WTP}_i = \exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5) + e_i/\exp(x'_{5i}\beta_5). \tag{3}$$

The first term in this expression is constrained to be positive and will equal $E[\text{WTP}]$ since the transformed error term is just a scaled version of the underlying mean-zero logistic (0, κ) error. However, the support for this error distribution is

[7] In all but Survey Version 6, only Option C is being compared to Option A, so we are estimating only β_1^* and β_5^* (or, in more general specifications, systematically varying versions of these parameters). The dummy variables $B_i = D_i = E_i = 0$ for all observations for these versions.

the entire real line. The error term in Eq. (3) will of course be heteroscedastic if the marginal indirect utility of income, $\beta_5^* = \exp(x'_{5i}\beta_5)$, varies across individuals.

Our strategy for precluding negative fitted WTP estimates differs substantively from the approach taken elsewhere. The most common earlier strategy was to specify log(WTP) as a linear function of regressors in a logit or probit model. For referendum contingent valuation responses, when conventional maximum likelihood binary discrete choice estimation methods were being used, this strategy was convenient because the researcher merely needed to use the log of the bid value (rather than its level) as a right-hand-side variable. The exponentiated value of the fitted conditional log(WTP) provided an estimate of the fitted conditional median of WTP, but the distribution of WTP in this case is skewed. Calculating the fitted expected value of WTP involved scaling the median by $\exp(\sigma^2/2)$ in the probit variant (where σ^2 is the fitted error variance) and by $\Gamma(1-\kappa)\Gamma(1+\kappa)$ in the logistic case (where Γ is the mathematical gamma function, and κ is the estimated dispersion parameter). Furthermore, for some values of κ in the logit case, the mean is undefined (Cameron [12]).

Here, the variable we call WTP is conceived as the type of latent variable that underlies a conventional tobit model. When latent WTP is positive, it is assumed to reflect the individual's true WTP for the good. When latent WTP is negative, WTP is manifested as zero, since the option for the individual to receive compensation to accept the proposed change was not an explicit possibility. We preclude negative fitted latent WTP by forcing the systematic portion of the latent WTP function to be non-negative via exponentiation. Non-negative fitted latent WTP is guaranteed if we force the indirect utility function to be non-decreasing in the quantity of the environmental good and non-increasing in the price of that good.[8] The error distribution in Eq. (3) is symmetric around zero. None of the usual exponentiation is necessary, so we do not run afoul of fitted latent WTP in the logistic case depending upon $\Gamma(1-\kappa)\Gamma(1+\kappa)$ as it did with earlier strategies. As Eq. (3) demonstrates, fitted E[latent WTP_i] is simply $\exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5)$ and will be, even if there is heteroscedasticity, since the expected error, $E[e_i]$, is equal to zero. The value of E[latent WTP_i] is independent of the error variance.

For the expected actual (as opposed to latent) WTP to be $\exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5)$, positive error distribution density in the negative domain must be plausible. It is unknown whether any respondents in the current study would have expressed negative WTP, had that option been presented to them. If negative values for WTP are not plausible, E[latent WTP_i], if positive, will remain the median of the distribution of actual WTP, but E[actual WTP_i] will be larger and its value will depend upon the error variance. In this way, the error variance in the indirect utility function must be viewed as a consequential parameter of that function. Here, we can convert the portion of the fitted unbounded conditional density for latent WTP that lies in the negative domain to a point mass at zero before the implied E[actual WTP_i] is calculated. This operation requires calculation of the

[8] In specifications with more elaborate systematic varying parameters, specifications including, for example, both age_i and age_i^2 can wander into unacceptable territory for respondents who are very old. Exponentiation allows the effects of age/cohort to be fully reflected in preferences but precludes fitted negative demands.

conditional expectation of a logistic random variable truncated at zero. There is no simple analytical formula for this expectation, so we calculate it numerically.[9]

What are the derivatives of E[latent WTP_i] with respect to individual characteristics? If we ignore any positive error density in the negative domain, the derivatives of the expectation of the latent WTP variable are easy to calculate. If some particular individual attribute x_j is an element of both x_1 and x_5, with corresponding estimated coefficients β_{1j} and β_{5j}, the derivative of fitted latent WTP_i with respect to x_j is

$$\partial E[\text{latent WTP}_i]/\partial x_j = (\beta_{1j} - \beta_{5j})E[\text{latent WTP}_i]. \tag{4}$$

E[latent WTP_i] is non-negative, so the sign of ∂E[latent WTP_i $]/\partial x_j$ is determined by the sign of the parameter difference $(\beta_{1j} - \beta_{5j})$.[10] Elasticities of latent WTP with respect to x_j can also be calculated at the means of the data using E[latent WTP_i].

4.2. Separate Samples: Specific Forms

Survey Version 0-*ACT*: *Actual Dichotomous Choice*, \$6 *Bid* (*Sample* 0), *Telephone Survey*

Each individual in this sample is presented with the same \$6 bid value and is invited to actually sign up for the program described (Option C, the main option). Of course, no researcher (given a choice) would opt to rely exclusively on demand information generated in a context with no price variation to derive welfare estimates. Why is there only one price? NMPC is a regulated natural monopoly, so there was no latitude for charging different prices to different customers for the same option. Nevertheless, there remains valuable demand information in respondents' decisions with respect to this offer, and we exploit this information to the fullest extent possible. We can readily identify a corresponding WTP function if we are willing to make strong assumptions about functional form. Fortunately, we have the benefit of six other types of demand information to employ in judging the credibility of the demand function that can be extracted from the limited information in this sample.

Since Options B, D, and E are not being considered in this sample, the difference in indirect utilities, $V_i^C - V_i^A$, which drives this pair-wise choice is simply

$$(V^C - V^A)_i = \exp(x'_{1i}\beta_1) - \exp(x'_{5i}\beta_5)\text{price}_i + e_i = Z_{0i} + e_i, \tag{5}$$

[9] The density for a logistic(b,k) random variable truncated at zero is $\{\exp(-((x-b)/k))[1+\exp(-b/k)]\}/\{k[1+\exp(-((x-b)/k))]^2\}$. To verify the calculations we use for the conditional expectations for the truncated logistic, we take advantage of the fact that a logistic(b,k) distribution is well approximated through most of its range by $N(b, \pi k/3)$. For a standard normal random variable, X, $E[X|X>c] = \phi(c)/(1-\Phi(c))$, which is easily calculated with packaged integration functions.

[10] Point estimates of the individual parameters are produced in the estimation process. Estimates of their difference and the standard error of this difference must be constructed explicitly from the point estimates and the asymptotic variance–covariance matrix of these estimates. If an individual attribute x_j appears in only one "index," (either β_{1j} or β_{5j} but not both) as will be the case in our example, then the derivative of WTP with respect to this parameter will be either $\beta_{1j}\,\text{WTP}_i$ or $-\beta_{5j}\,\text{WTP}_i$, respectively.

where $e_i = (u_i^C - u_i^A)$, and the coefficients β_2 through β_4 do not appear. The variable price_i does not change across this sample (it is $6 for everyone). If x_1 and x_5, the vectors of respondent characteristics used to allow heterogeneous preferences, each consisted solely of an "intercept" term, all that could be estimated would be the sample average value of the index $Z_{0i} = \exp(\beta_1) - 6\exp(\beta_5) = b_0$. Instead of a WTP function, the best that could be achieved would be a function that describes the probability of being willing to pay the constant $6 bid.

The distinct scalar parameters $\exp(\beta_1)$ and $\exp(\beta_5)$ cannot be separately identified without additional information. As one possibility, this necessary extra information may take the form of sociodemographic variables that make $x_{1i}'\beta_1$ and $x_{5i}'\beta_5$ vary across respondents. When the parameter β_5 is generalized to $x_{5i}'\beta_5$, for example as $(\beta_{51} + \beta_{52}\,\text{inc}_i)$ as we will use later in the paper, it is possible to estimate β_1^* and β_5^* separately, even if β_1 remains a scalar. The utility-difference equation becomes $(V^C - V^A)_i = \exp(\beta_1) - 6\exp(\beta_{51} + \beta_{52}\,\text{inc}_i) + e_i$. The presence of the inc_i variable in the expression for β_5^* renders $\exp(\beta_1)$ distinguishable from $\exp(\beta_{51} + \beta_{52}\,\text{inc}_i)$. This means that we can recover separate estimates of β_1^* and β_5^* to use in calculating a fitted value for $E[\text{WTP}_i]$, even if price_i remains constant at $6. As a second possibility for identifying β_1^* and β_5^* separately, we can use additional information from other elicitation methods, pooling these 0-ACT data with the choices from other samples.

For this single sample, if a respondent indicates they are willing to pay the bid amount, then let $I_{0i} = 1$, otherwise $I_{0i} = 0$. The contribution of Survey Version 0 to the log-likelihood function is

$$\text{Log } L_0 = \sum_{i=1}^{n_0} I_{0i} \log\{\exp(Z_{0i}/\kappa_0)/[1 + \exp(Z_{0i}/\kappa_0)]\} + (1 - I_{0i}) \log\{1/[1 + \exp(Z_{0i}/\kappa_0)]\}. \tag{6}$$

Note that if $\kappa_0 = 1$, then $\text{Log } L_0$ is just the familiar binary logit discrete choice model, except for the nonlinearity-in-parameters of Z_{0i}.

Survey Version 1-*PDC*: *Hypothetical Dichotomous Choice*, $6 *Bid* (*Sample* 1), *Mail Survey*

The acronym PDC stands for "like the phone dichotomous choice" scenario. Each individual is again presented with a common $6 bid value, but in this case they are only asked if they would sign up for the program if it was available. Like the 0-ACT telephone sample, it was mentioned that 12,000 subscribers to the program would be needed. The algebra is identical, with an analogous systematic term Z_{1i}. However, we do allow a κ_1 dispersion parameter that is different from κ_0.

Survey Version 2-*MDC*: *Hypothetical Dichotomous Choice*, *Varying Bids* (*Samples* 2 *through* 8), *Mail Survey*

Individuals in this sample are presented with single (but different) randomly assigned bid values and are invited to indicate whether they would be willing to pay this amount for Option *C*. In this and all subsequent samples to be described, there was no mention of a 12,000-subscriber requirement. The indirect utility-difference formula is the same as before. The important modification is that price_i now varies across observations. The indirect utility-difference index Z_{0i} in Eq. (5) above is thus replaced by $Z_{2i} = \exp(x'_{1i}\,\beta_1) - \exp(x'_{5i}\,\beta_5)\text{price}_i$, yielding

$$(7) \qquad \left(V^C - V_i^A\right) = Z_{2i} + e_{2i},$$

where e_{2i} is distributed logistic(0, κ_2), with κ_2 not necessarily identical to κ_1. Since price_i does vary in this sample, $\exp(x'_{1i}\,\beta_1)$ and $\exp(x'_{5i}\,\beta_5)$ can be separately identified, even without relying on nonlinear systematic varying parameters or other types of choice information.

With $I_{2i} = 1$ defined analogously to I_{0i}, the contribution to the log-likelihood function for Survey Version 2 is

$$(8) \qquad \text{Log}\, L_2 = \sum_{i=1}^{n_2} I_{2i} \log\{\exp(Z_{2i}/\kappa_2)/[1 + \exp(Z_{2i}/\kappa_2)]\} + (1 - I_{2i}) \log\{1/[1 + \exp(Z_{2i}/\kappa_2)]\}.$$

The log-likelihood function Log L_2 is again the familiar binary logit discrete choice log-likelihood, except for its nonlinearity and the κ_2 parameter that allows the dispersion to differ from that for other survey versions.[11]

Survey Version 3-*OE*: *Open-Ended Willingness to Pay* (*Sample* 9), *Mail Survey*

Respondents are invited to state directly their highest WTP for the offered scenario. We need to assume that choice behavior is consistent with the simple economic theory of utility maximization, and that our model is appropriate. Then, this WTP should be the dollar price which would make the respondent just indifferent between paying the price and getting the offered scenario, or not paying and not getting the scenario.

The indirect utility difference will be rendered zero by the OE amount, WTP_i. Thus

$$(9) \qquad (V^C - V^A)_i = 0 = \exp(x'_{1i}\,\beta_1) - \exp(x'_{5i}\,\beta_5)\text{WTP}_i + e_{3i}$$

and

$$(10) \qquad \text{WTP}_i = \exp(x'_{1i}\,\beta_1)/\exp(x'_{5i}\,\beta_5) + e_{3i}/\exp(x'_{5i}\,\beta_5).$$

As before, when this sample is combined with other samples, we allow the analogous error term e_{3i} to be distributed logistic(0, κ_3), where $\kappa_3 \neq \kappa_j$, $j \neq 3$.

For the OE elicitation method, respondents are presumed to provide a point value for WTP_i. The conditional expected value of this distribution is $\exp(x'_{1i}\,\beta_1)/\exp(x'_{5i}\,\beta_5)$ and the dispersion parameter is $\kappa_3/\exp(x'_{5i}\,\beta_5)$. Define Z_{3i} fundamentally differently from Z_{1i} or Z_{2i} by using it now to denote the "standar-

[11] If only a single version of the data is being used for estimation, κ_j will always be normalized to unity.

dized" value of the respondent's stated WTP_i:

$$(11) \qquad Z_{3i} = \left[\mathrm{WTP}_i - \exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5)\right]/\left[\kappa_3/\exp(x'_{5i}\beta_5)\right].$$

To simplify the notation in what follows, define $s_3 = \kappa_3/\exp(x'_{5i}\beta_5)$.[12]

The assumption of logistically distributed regression errors can be adapted to a tobit-like regression-by-maximum-likelihood context. A tobit-like model is indicated because there is heaping of reported WTP_i at zero. Define $\mathrm{POS}_i = 1$ if a strictly positive value of WTP_i is reported for observation i. If a zero WTP_i is reported, then $\mathrm{POS}_i = 0$. The contribution of these responses to the log-likelihood function is

$$(12) \qquad \mathrm{Log}\, L_3 = \sum_{i=1}^{n_3} \mathrm{POS}_i\{Z_{3i} - \log(s_3) - 2^*\log[1+\exp(Z_{3i})]\} \\ + [1-\mathrm{POS}_i]\log\{\exp(Z_{3i})/(1+\exp(Z_{3i}))\}.$$

This component of the overall log-likelihood function can be viewed simply as an analog to the familiar tobit log-likelihood, but based on the logistic error distribution rather than the more-common normal distribution.

Survey Version 4-PC: Payment Card (Sample 10), Mail Survey

This version of the survey generates interval data for the true but unobserved WTP_i value. As in the OE case, the latent variable we must model is WTP_i. It can be defined as in Eq. (10), but we will now substitute e_{4i}, distributed logistic(0, κ_4), where $\kappa_4 \neq \kappa_j$, $j \neq 4$. Analogous to the normal-error PC model used by Cameron and Huppert [14], let t_{ui} be the upper bound of the payment card interval chosen by the respondent, and let t_{li} be the associated lower bound. Then define

$$(13) \qquad Z_{ui} = \left[t_{ui} - \exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5)\right]/\left[\kappa_4/\exp(x'_{5i}\beta_5)\right], \qquad \text{and}$$
$$Z_{li} = \left[t_{li} - \exp(x'_{1i}\beta_1)/\exp(x'_{5i}\beta_5)\right]/\left[\kappa_4/\exp(x'_{5i}\beta_5)\right].$$

The log-likelihood function involves the difference in the cumulative densities between the standardized upper bound (Z_{ui}) and the standardized lower bound (Z_{li}). For the assumed underlying logistic density function, these cumulative densities are $P_{ui} = 1/[1+\exp(-(Z_{ui}))]$ and $P_{li} = 1/[1+\exp(-(Z_{li}))]$. The contribution to the log-likelihood of the observations from Survey Version 4 is

$$(14) \quad \mathrm{Log}\, L_4 = \sum_{i=1}^{n_4} \log\left\{\left(1/\left[1+\exp(-(Z_{ui}))\right]\right) - \left(1/\left[1+\exp(-(Z_{li}))\right]\right)\right\} \\ = \sum_{i=1}^{n_4} \log\{P_{ui} - P_{li}\}.$$

If we are dealing with the lowest interval, we will use just $\log\{P_{ui}\}$, with the highest interval, $\log\{1-P_{li}\}$. This sub-model is just a variant of the common interval-data

[12] If x_{5i} includes variables in addition to a constant term, then s_3 would become s_{3i}.

model, also adapted to an underlying logistic error distribution instead of the usual normal distribution.

Survey Version 5-MB: Multiple Bounded (Samples 11 through 13), Mail Survey

Respondents were presented with 13 different bid values and were asked to indicate (in categories) how likely they would be to pay each of these amounts. The intuitive framework we employ for analyzing these responses is a multi-category generalization of the binary discrete choice model that applies for Survey Versions 0-ACT through 2-MDC—i.e., a panel of ordered logits instead of a single binary logit.[13]

The indirect utility-difference function is presumed to drive the ordered categorical response to each question on this survey. Again, let the relevant indirect utility difference be

$$(15) \qquad (V^C - V^A)_i = \exp(x'_{1i}\beta_1) - \exp(x'_{5i}\beta_5)\,\text{price}_i + e_{5i} = Z_{5i} + e_{5i}.$$

Respondents to this survey variant choose one of five ordered levels of certainty that they would be willing to pay for the program at each of the 13 different bid amounts. Let $Y_i = 1$ if the respondent chooses the "Definitely Yes" response, zero otherwise. Let $H_i = 1$ if the respondent chooses the "Probably Yes" response, zero otherwise. The indicators M_i, L_i, and N_i are defined similarly for the "Not Sure," "Probably No," and "Definitely No" responses.

For an ordered logit model with five response categories, there are four standardized estimated thresholds. Generally, researchers set one threshold to zero (usually the lowest one) since the location and scale of the underlying "propensity to be willing to pay" variable are unknown. But these data will sometimes be used in conjunction with other samples from this study (in particular, the binary logit data from Survey Versions 0-ACT through 2-MDC). In these pooled samples, we expect the "zero" level of the latent propensity variable in the ordered logit to lie somewhere in the middle interval (i.e., in the "Not Sure" category), so we do not normalize the lowest threshold to zero when using pooled data.

Let the thresholds between the five intervals be α_0, α_1, α_2, and α_3. If only this one sample is being used, then α_0 must be normalized to zero. Furthermore, these thresholds are known only up to a scale factor consisting of the dispersion parameter of the error term in this particular submodel, namely κ_5. We are able to estimate $\exp(x'_{1i}\beta_1)/\kappa_5$ and $\exp(x'_{5i}\beta_5)/\kappa_5$, as well as α_1/κ_5, α_2/κ_5, and α_3/κ_5. The dispersion parameter κ_5 can itself be estimated as a multiple of some κ_j normalized to unity for some other sample j if other samples are being used.[14]

We define five probabilities, one associated with each of the five categories within which each individual may have responded (for each of the 13 choice

[13] By treating the responses in each row of the multiple-bounded choice matrix as the basis for a separate ordered logit analysis, we depart in this paper from an alternative interpretation of multiple-bounded data employed by Welsh and Poe [49], for example. Their estimation strategy focuses on the columns of the matrix as the basis for separate PC-type analyses at different certainty-of-payment levels.

[14] In this paper, we constrain the vector of α parameters to be identical across all 13 questions (or portion thereof answered by any respondent). We also estimate just a single κ_5 parameter for all 13 questions.

opportunities afforded respondents to this survey variant):

$$
\begin{aligned}
(16)\qquad PY_i &= 1/[1 + \exp(\alpha_3/\kappa_5 - Z_{5i})],\\
PH_i &= \exp(\alpha_3/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_3/\kappa_5 - Z_{5i})]\\
&\quad - \exp(\alpha_2/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_2/\kappa_5 - Z_{5i})],\\
PM_i &= \exp(\alpha_2/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_2/\kappa_5 - Z_{5i})]\\
&\quad - \exp(\alpha_1/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_1/\kappa_5 - Z_{5i})],\\
PL_i &= \exp(\alpha_1/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_1/\kappa_5 - Z_{5i})]\\
&\quad - \exp(\alpha_0/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_0/\kappa_5 - Z_{5i})], \text{ and}\\
PN_i &= \exp(\alpha_0/\kappa_5 - Z_{5i})/[1 + \exp(\alpha_0/\kappa_5 - Z_{5i})].
\end{aligned}
$$

The contribution to the log-likelihood function of the set of m $(= 1, \ldots, 13)$ categorical responses for complete observations[15] from Survey Version 5 can then be expressed as

$$
\begin{aligned}
(17)\quad \operatorname{Log} L_5 = \sum_{i=1}^{n_5} (1/13) \sum_{m=1}^{13} \{&Y_i \log[PY_i] + H_i \log[PH_i] + M_i \log[PM_i]\\
&+ L_i \log[PL_i] + N_i \log[PN_i]\}.
\end{aligned}
$$

When this sample is used alone, the α parameters cannot be distinguished from the intercept term in the indirect utility-difference function, and without a separate estimate of the intercept, it is not possible to calculate a point estimate of fitted WTP. This is an artifact of the presence of the "not sure" category in our survey instrument. In multiple-bounded elicitation formats where there is no such category (e.g., "probably no" is directly adjacent to "probably yes"), it is possible to separately identify the intercept term and then to solve for a point estimate of WTP.[16] When this sample is used alone, we can say only that WTP lies between $[\exp(x'_{1i}\beta_1) - \alpha_2]/\exp(x'_{5i}\beta_5)$ and $[\exp(x'_{1i}\beta_1) - \alpha_1]/\exp(x'_{5i}\beta_5)$.

Survey Version 6-CA: Conjoint Analysis (Samples 14 and 15), Mail Survey

This version of the survey introduces three other possible scenarios in addition to the binary choice scenario that is offered to respondents in Versions 0 through 5 of the survey. We now have five different indirect utility levels: V^A, V^B, V^C, V^D,

[15] However many completed responses are provided to these 13 questions by respondents receiving Survey Version 5, we scale each Version 5 respondent's contribution to the log-likelihood so that his or her choices have equal weight. At the next level, when the choices from Version 5 are pooled with other samples, each sample is equally weighted, so that no elicitation format, as a consequence of its number of respondents, has a disproportionate influence upon the parameter estimates in the pooled model. We do not address the joint endogeneity of the 13 multiple-bounded responses in the present paper.

[16] This suggests that if a "not sure" category is desired for a multiple-bounded study, a portion of the sample should receive a variant of the instrument that does not offer the opportunity to express uncertainty. Error dispersions could be allowed to differ to capture some of the consequences of a precluded "not sure" option, but the pooled data would allow fitted WTP to be calculated as a point estimate, rather than just an interval.

and V^E. Indirect utility under each of these choices can then be expressed as in Eq. (1) and simplified using Z_{ji} to denote the systematic portion of the indirect utility:

$$V_i^j = Z_{ji} + u_i^j, \qquad j = A, B, C, D, E. \tag{18}$$

For each of the pair-wise choices of Options B through E against simply the do-nothing Option A, the choice probabilities are

$$P_{ji} = \exp(Z_{ji}/\kappa_{6P}) / \left[\exp(Z_{ji}/\kappa_{6P}) + \exp(Z_{Ai}/\kappa_{6P})\right], \qquad j = B, C, D, E. \tag{19}$$

We will again allow for a different (common) error dispersion parameter from those that apply to the other samples. We will refer to this common dispersion parameter as κ_{6P}, where the P denotes "pair-wise" choices.[17]

Let $I_{Bi} = 1$ if scenario B is chosen over A by respondent i; $I_{Bi} = 0$ otherwise, with similar definitions for I_{Ci}, I_{Di} and I_{Ei}. The contribution to the log-likelihood function of the set of four pair-wise choices is

$$\text{Log } L_{6P} = \sum_{i=1}^{n_{6P}} \sum_{j=B,C,D,E} \left\{ I_{ji} \log(P_{ji}) - (1 - I_{ji}) \log(1 - P_{ji}) \right\}. \tag{20}$$

These same respondents were also asked to choose their single most preferred option from the full set of options A through E. We allow for a separate dispersion parameter for this choice, denoted κ_{6J}, with the J subscript denoting "joint choice." Define $SUMP_i$ as $\exp(Z_{Ai}/\kappa_{6J}) + \exp(Z_{Bi}/\kappa_{6J}) + \exp(Z_{Ci}/\kappa_{6J}) + \exp(Z_{Di}/\kappa_{6J}) + \exp(Z_{Ei}/\kappa_{6J})$. Under the logistic model, the probability of choosing each specific alternative from this set of five is then given by

$$P'_{ji} = \exp(Z_{ji}/\kappa_{6J}) / SUMP_i, \qquad j = A, B, C, D, E. \tag{21}$$

Define $I'_{Ai} = 1$ if scenario A is most-preferred; $I'_{Ai} = 0$ otherwise. Define I'_{Bi}, I'_{Ci}, I'_{Di}, and I'_{Ei} similarly. The contribution to the log-likelihood of this first-choice program from among the five alternatives is then

$$\text{Log } L_{6J} = \sum_{i=1}^{n_{6J}} \sum_{j=A}^{E} I'_{ji} \log(P'_{ji}). \tag{22}$$

A single respondent to Survey Version 6 should have, in total, only unit weight in determining the maximized value of the log-likelihood function using pooled data. If all questions are answered, each respondent would provide four binary choices (between each of Options B, C, D, E, and Option A), and one multiple choice (among Options A through E). Some respondents did not provide answers to all five choice questions. Thus, the number of pieces of choice information extracted from each respondent is used to divide the total contribution of each respondent to the overall log-likelihood function.[18] For complete responses, we use $\text{Log } L_6 = (\text{Log } L_{6P} + \text{Log } L_{6J})/5$.

[17] It would be possible to specify different dispersion parameters for each of the four pair-wise choices made by respondents between each of Options B through E and Option A. Due to a relatively small number of observations, however, we elect to impose the restrictions $\kappa_{BA} = \kappa_{CA} = \kappa_{DA} = \kappa_{EA}$.

[18] We do not address the joint endogeneity of the five choices from the 6-CA sample in the present paper.

4.3. Pooled Specifications

The independence of the seven samples (one actual and six hypothetical choices) means that the log-likelihood function for the pooled data is just the sum of the seven component likelihood terms. The parameters are found by maximizing

$$\text{Log } L = \sum_{k=0}^{6} \text{Log } L_k. \tag{23}$$

The same utility parameters appear in each of the seven components, and these can be restricted or left unrestricted across the seven additive log-likelihood terms as desired.

The log-likelihood functions for the individual samples, L_0 through L_6, are in some cases similar to maximum likelihood models already programmed in commercially available econometric software packages (for example, the binary logit model for the 2-MDC sample). However, our models feature exponentiation of their systematically varying indirect utility-difference parameters in order to (a) enforce the sign restrictions implied by consumer theory and (b) aid in the identification of separate slope and intercept parameters even with the non-varying bids of the 0-ACT and 1-PDC samples. This modeling strategy renders the usual "index" for the choice model nonlinear-in-parameters, thereby precluding the use of packaged software for optimizing the individual sample likelihoods.

A further complication is that testing our key hypotheses about preference commonality across elicitation methods using the pooled log-likelihood requires that we be able to restrict the corresponding parameters to be identical across the individual sample submodels. Thus all seven sub-models must be incorporated into one very complex log-likelihood function.[19]

A final consideration with respect to pooled models is weighting. We do not wish to allow the unequal sizes of the samples for each different elicitation method to exert undue influence on the parameter estimates for the pooled model. Therefore, we weight observations so that each sample, rather than each respondent, has equal representation. In this application, the decision to weight each elicitation method equally, rather than to weight each respondent equally, has no apparent qualitative effect on our findings. Since the model variant with elicitation methods weighted equally seems the most defensible choice, we report these results.[20]

5. EMPIRICAL RESULTS

We allow for preferences that are heterogeneous across sociodemographic groups by allowing β_5 to differ across income brackets. The income information in this data set is unfortunately limited to income brackets. As in virtually all survey data, these brackets are wide relative to the central tendency of WTP for the good in question. They seem rather too wide to be used to allow the indirect utility

[19] We use the GQOPT package of FORTRAN subroutines for generic function optimization (Goldfeld and Quandt [20]).

[20] A referee has suggested the possibility of using equal weights on the revealed and stated preference samples. For a study with more revealed preference data to exploit, this would be a very sensible strategy to explore.

function have its marginal indirect utility of income, β_5 in Eq. (1), be a systematic varying parameter that depends on the same continuous income variable, Y_i, that currently appears in the indirect utility function in Eq. (1).

Thus, we let (unmeasured exact continuous) income Y_i drop out of the linear-in-income indirect utility-difference function, but we use a variable called inc_i, the midpoint of the individual's income bracket, merely as a discrete-valued indicator of sociodemographic status that captures heterogeneity in the preference function. To a first approximation, this allows for the marginal utility of net income to differ with a gross measure of income level. This modeling strategy allows diminishing marginal utility of income to be captured, albeit crudely, by our model.[21] It also permits a glimpse at the likely construct validity of the estimated model, in that the marginal utility of income ought to be, if anything, decreasing in the level of this sociodemographic status variable.

In Section 5.1, we describe the empirical results for each of our seven samples estimated independently and consider their distinct implications for WTP at the means of the overall data. Section 5.2 briefly reviews results for all possible pair-wise comparisons of our seven methods. What happens when we pool all seven types of data and statistically test the equality of the corresponding parameters across all data types? Section 5.3 describes the search for a subset of elicitation methods for which common preferences cannot be rejected. Section 5.4 examines the main results for this subset of compatible elicitation methods.

5.1. Independent Estimates Based on Different Methods with Separate Samples

First, we use each of our individual samples in an entirely separate empirical model. The "0-ACT" through "6-CA" columns of Table I provide results for these separate models. Estimates of the indirect utility-difference parameters for the specification shown in Eq. (2) are provided in the first horizontal section of this table. Here, $\exp(x'_{1i}\beta_1) = \exp(\beta_1)$ and $\exp(x'_{5i}\beta_5) = \exp(\beta_{51} + \beta_{52}\text{inc}_i)$. The absolute levels of these utility parameters are not expected to be comparable because of the arbitrary normalization to unity of the dispersion parameters for each of the separate samples.

The nonlinearity-in-parameters of the indirect utility-difference function means that almost all individual elicitation methods can produce a point estimate of expected WTP (at the overall mean of the income data). The only exception is the multiple-bounded method. The 5-MB sample can only provide bounds for an interval within which its "predicted WTP at the data means" must lie, since we elect not to impose the assumption that the middle pair of thresholds is symmetric around zero.

A referee has pointed out that our strategy for identifying the parameters in the 0-ACT and 1-PDC samples, used individually, relies upon conditioning the scale parameter estimate upon a variable, income measured in categories, "which is known for being difficult to rely upon." We agree wholeheartedly that the WTP estimates produced by these two samples are tenuous because they are identified only as a consequence of our choice of functional form. For completeness,

[21] We have also entertained models that allow β_1 to vary with age. These are detailed in a separate appendix, available from the authors. Unfortunately, they cannot be estimated for the 0-ACT sample, so we stick with the simpler model.

TABLE I

Systematically Varying Utility Parameters[a]

	0-ACT[b]	1-PDC[b]	2-MDC	3-OE	4-PC	5-MB[d]	6-CA	Pooled-0123456	Pooled-01256	Pooled-0126
Utility par. (βs):										
$\exp(\beta_{11}) * \text{Opt } C_i +$	−0.969	−1.598	−0.533	−1.767	−0.346	0.146	−0.423	−0.839	−0.403	−0.575
	(−0.55)	(−1.14)	(−1.76)*	(−3.11)**	(−2.48)**	(1.09)	(−1.01)	(−5.17)**	(−1.81)*	(−2.13)**
$\exp(\beta_{21}) * \text{Opt } B_i +$	—	—	—	—	—	—	0.177	0.170	0.245	0.223
							(0.827)	(0.61)	(0.84)	(0.76)
$\exp(\beta_{31}) * \text{Opt } D_i +$	—	—	—	—	—	—	−1.717	−1.803	−1.665	−1.941
							(−0.985)	(−1.01)	(−1.01)	(−0.89)
$\exp(\beta_{41}) * \text{Opt } E_i -$	—	—	—	—	—	—	−9.371	−19.36	−15.28	−18.74
							(−0.122)	(−0.07)	(−0.07)	(−0.08)
$\exp(\beta_{51} +$	−0.618	−0.887	−1.174	−1.871	−1.079	−2.488	−1.206	−1.415	−1.133	−1.055
	(−3.13)**	(−4.00)**	(−5.57)**	(−20.1)**	(−18.54)**	(−10.4)**	(−4.32)**	(−11.04)**	(−6.76)**	(−6.25)**
$\beta_{52} * \text{inc}_i) * \text{bid}_i$	−19.50	−41.19	−9.965	−4.215	0.125	2.201	−10.66	−5.228	−10.78	−14.49
	(−1.95)*	(−2.52)**	(−2.13)**	(−3.09)**	(0.87)	(0.58)	(−1.63)	(−5.86)**	(−4.67)**	(−4.60)**
Ord. logit thresh:										
α_0	—	—	—	—	—	0.0	—	−2.029	−1.987	—
								(−4.79)**	(−4.05)**	
α_1	—	—	—	—	—	0.514	—	−0.9222	−0.7675	—
						(6.00)**		(−3.14)**	(−2.34)**	
α_2	—	—	—	—	—	0.963	—	0.0431	0.2976	—
						(8.54)**		(0.15)	(0.93)	
α_3	—	—	—	—	—	1.816	—	1.882	2.335	—
						(12.18)**		(3.66)**	(3.67)**	
κ multiples:[c]										
κ_{ACT}	1.0	—	—	—	—	—	—	0.573	0.496	0.550
								(−3.29)**	(−3.12)**	(−2.87)**
κ_{PDC}	—	1.0	—	—	—	—	—	1.269	1.030	1.118
								(1.03)	(0.11)	(0.46)
κ_{MDC}	—	—	1.0	—	—	—	—	1.0	1.0	1.0
κ_{OE}	—	—	—	1.0	—	—	—	1.487	—	—
								(2.83)**		

κ_{PC}	—	—	—	—	1.0	—	—	0.552 (−4.45)**	—	—
κ_{MB}	—	—	—	—	—	1.0	—	2.164 (3.80)**	2.414 (3.90)**	—
κ_{CA}	—	—	—	—	—	—	1.0	1.006 (0.03)	1.072 (0.27)	1.058 (0.22)
$\kappa_7(\mathrm{CA}_{\mathrm{joint}})$	—	—	—	—	—	—	1.887 (2.205)**	1.928 (2.21)**	2.031 (2.28)**	2.009 (2.24)**
Fitted WTP at means for Option C	\$1.64[b]	\$2.92[b]	\$2.92	\$1.33	\$2.07	\$2.12–\$7.02	\$3.47	\$2.23	\$3.31	\$3.03
∂(latent WTP)/∂ inc (inc in \$mill)	\$0.84*	\$1.78**	\$0.43**	\$0.18**	\$ − 0.005	\$ − 0.10	\$0.46	\$0.23**	\$0.47**	\$0.63**
Prop. WTP \$6										
ACT	0.27	—	—	—	—	—	—	0.22	0.25	0.27
PDC	—	0.45	—	—	—	—	—	0.36	0.37	0.38
MDC	—	—	0.35	—	—	—	—	0.33	0.37	0.37
OE	—	—	—	0.35	—	—	—	0.38	—	—
PC	—	—	—	—	0.21	—	—	0.21	—	—
MB	—	—	—	—	—	0.23–0.41	—	0.42	0.44	—
CApair	—	—	—	—	—	—	0.38	0.33	0.38	0.37
CAjoint	—	—	—	—	—	—	0.44	0.41	0.43	0.43
# choices	117	262	735	241	271	4828	1268	7722	7210	2382
# people	117	262	735	241	271	437	303	2366	1737	1417
Max. Log L	−165.119	−210.122	−205.397	−987.398	−852.166	−429.010	−279.827	−3153.928	−1298.534	−864.683

[a] Exogenously weighted so that each data type has an equal effective number of observations; total = 2366.

[b] Caution must be exercised with respect to the fitted E[latent WTP] values for the 0-ACT and 1-PDC samples. Recall that these samples exhibit no variation in prices, and the β parameters in this case are identified only because of differences in inc_i across respondents, in combination with the exponentiation of the systematic varying parameters in order to impose the regularity conditions that indirect utility be increasing with income and with the levels of the "goods" in question.

[c] Non-unitary κ multiples are estimated as powers of e to ensure that they remain positive. For ease of interpretation, the point estimates are exponentiated. The t-test statistics, however, remain tests of the hypotheses that the estimated exponents are zero. If the exponent is zero, the κ multiple is unity (exp(0)) and the dispersion for the sample in question is not different from that of the numeraire sample, which is 2-MDC.

[d] Due to the presence of a "not sure" category, if no other data are used to identify the location of the ordered logit thresholds, the location of the indirect utility-difference function in the MB case is determined by the normalization of α_0 to zero. Without additional information to establish location, the 5-MB sample, if used alone, only allows us to locate E[latent WTP] within an interval. Likewise, we can only establish an interval for the proportion of the MB sample would have been willing to pay \$6: 0.23–0.41. In contrast, WTP derivatives can be calculated because these quantities depend only upon the slopes.

however, the independent model results for these two samples are shown alongside those for the other, better-identified models.

The disparities among the fitted WTP estimates in Table I should be familiar to researchers who have compared alternative elicitation methods: The point estimates of WTP differ rather markedly across our different samples. Furthermore, these differences seem to be attributable almost exclusively to the different elicitation methods, since all other features of the sampling frame and survey instrument were controlled as rigorously as possible. (Recall that 0-ACT and 1-PDC surveys mentioned that 12,000 customers were needed, because only the $6 bid was used in these two cases. The surveys were otherwise identical except that 0-ACT was conducted by telephone and 1-PDC by mail. A test of the equivalence of the 1-PDC and 2-MDC (using only the $6 bid subsample) is possible. The likelihood ratio test for equivalence of the three utility parameters with the dispersions unrestricted is only 0.718. The analogous test when the dispersions are also constrained to be identical is only 1.46. Thus we see no evidence that the incidental mention of the 12,000-subscriber requirement in the 0-ACT and 1-PDC versions made any measurable difference to the utility parameters implied by the data.)

The relative magnitudes of the different WTP point estimates are roughly consistent with the orderings that have been observed elsewhere in the literature. A referee has pointed out that WTP from the 0-ACT sample should converge from below to the truthful preference revelation of WTP, with convergence depending upon population size and the total cost of the project. Some free-riding is to be expected. The 3-OE and 4-PC methods of elicitation have been argued not to be incentive compatible and therefore should not be expected to yield the same implications about preferences as the discrete choice methods, namely the 2-MDC and its generalization, the 6-CA elicitation method. (For some theoretical results on this point, see Hoehn and Randall [27] and Carson *et al.* [15].)

Regarding the more-recent 5-MB method, there is less evidence. To the extent that this method can be viewed as a generalization of the 2-MDC method to include more answer categories than just YES/NO and to involve repetitions for each individual at different parametric costs, one would expect this method to retain some of the desirable properties of the 2-MDC method. On the other hand, it is possible that the 5-MB method may be viewed as a generalization of the 4-PC method, with five possible answer categories, rather than just YES/NO at each listed bid level. To the extent that this is a relatively more open-ended elicitation method, incentive incompatibility may creep in.

Confidence intervals for these distinct WTP estimates are somewhat difficult to calculate because of the nonlinearities in their formulae. Thus, we use a nonparametric bootstrap procedure to generate a sampling distribution for these point estimates across 1000 draws, with replacement, from the true data. Figure 1 begins by displaying box-whisker plots summarizing these distributions for the 0-ACT through 6-CA samples, including the lower and upper bounds on WTP from the 5-MB sample.

The 0-ACT and 1-PDC results are the least robust, undoubtedly due to their constant bids and identification via functional form. While some readers might be disappointed by the large bootstrap confidence intervals for the 0-ACT and 1-PDC results, others may consider the results from these independent models to be remarkably disciplined, in view of the fact that the demand information has been

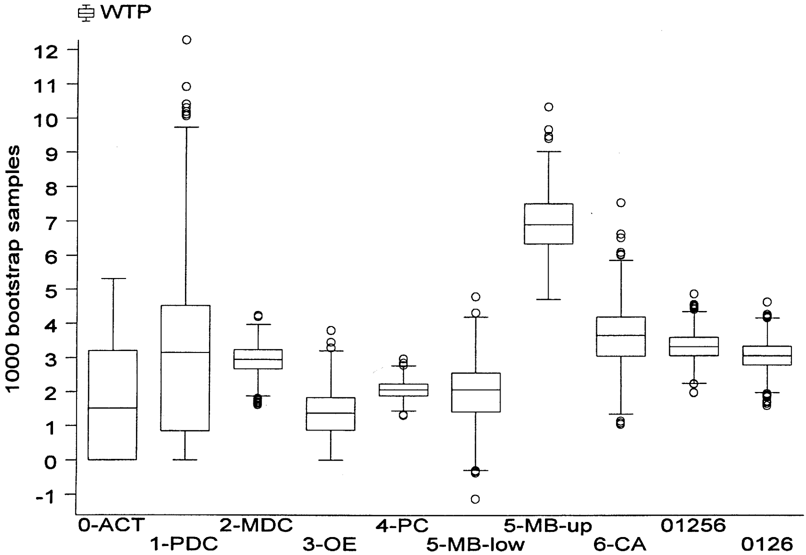

FIG. 1. Nonparametric bootstrap replications: fitted **median** WTP (= E[**latent WTP**]) at means of data, individual samples and pooled samples where equivalent utility parameters are not rejected. The "box" in each plot captures the median and interquartile range and the ticks on the whiskers show the upper and lower "adjacent values." See Stata [45].

obtained exclusively as an artifact of convenient functional from assumptions that impose the usual regularity conditions on demand in a simple fashion. Far from being random numbers, the WTP estimates from these two samples, used separately, are remarkably concordant with results from the other samples.

When these models are estimated separately for each sample, the dispersion parameter in each case must be normalized to unity. Recall that the error term is unbounded, so there will always be some positive probability associated with negative values of the latent indirect utility-difference variable. Furthermore, when the indirect utility-difference function is solved for the latent WTP variable, the dispersion of the latent WTP variable depends upon the systematically varying marginal indirect utility of income parameter. Since the formula for this parameter differs across samples, each sample used separately will have a different variance for its fitted latent WTP variable, as well as a different expectation.

If we preclude negative WTP and heap the probability in the negative domain at zero in a tobit-like manner, the fitted E[WTP] can be calculated numerically. For each model estimated separately, Fig. 2 displays the fitted E[WTP] for each of our 1000 nonparametric bootstrap samples (as opposed to the fitted E[latent WTP] of Fig. 1). Table II collects descriptive statistics for both the median WTP (the mean of fitted E[latent WTP]) and mean WTP when negative WTP is disallowed. The vagaries of identification via functional form for the 0-ACT and 1-PDC samples are more apparent in E[WTP] than in E[latent WTP].[22] Our tobit interpretation is

[22] To avoid distorting the vertical scale, the graph does not show 6 replications for 0-ACT (between \$34 and \$52 and one at \$220) and 60 replications for 1-PDC (between \$25 and \$102).

(6 observations of 0_ACT >25 and 60 observations on 1-PDC >25 are not displayed.)

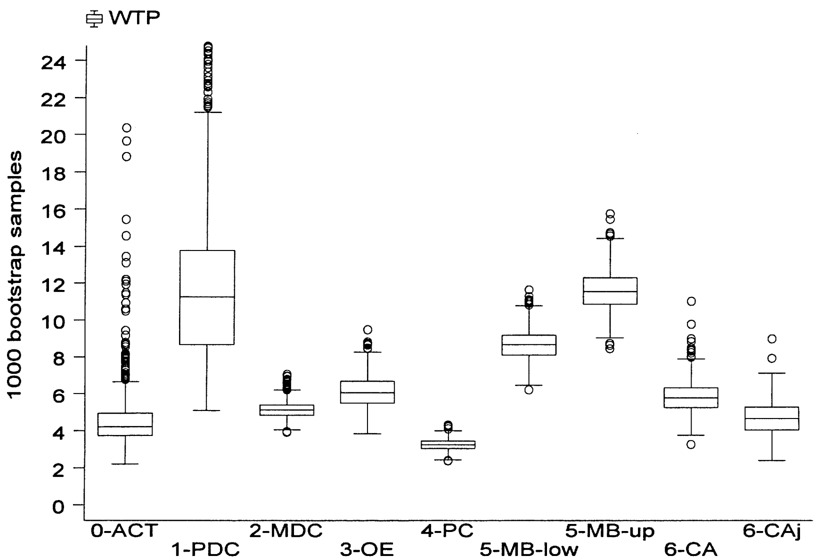

FIG. 2. Nonparametric bootstrap replications: fitted E[WTP] for individual samples at means of data. The "box" in each plot captures the median and interquartile range and the ticks on the whiskers show the upper and lower "adjacent values." See Stata [45]. (Six observations of 0_ACT > 25 and 60 observations on 1-PDC > 25 are not displayed.)

an alternative to a "spike" model (Kristrom [35]). Our data are not suitable for a Kristrom-type "spike" model because our survey does not separately establish whether each individual has a positive or zero WTP. In our model, zero WTP is implied by a latent WTP that is non-positive.

Goodness-of-fit must be measured differently for different types of models. With preferences varying systematically only with income bracket, the degree of fit cannot be expected to be very high. Nevertheless, we achieve a 76.9% prediction success rate for the 0-ACT sample, a 66.0% prediction success rate for the 1-PDC sample, and a 67.5% prediction success rate for the 2-MDC sample. These success rates are all in terms of the frequency with which "yes" and "no" responses were correctly predicted as having the highest probability for each individual. A pseudo R-squared value for the 3-OE sample was only 0.1277. For the implicit payment card intervals of the 3-PC sample, we sum across respondents the fitted probability of selecting the interval actually chosen. Expressed as a fraction of the sample size, this measure is only 0.04 (but as in the case of any prediction success measure, these types of statistics do not recognize "near misses"). The overall prediction success rate for the five categories in the 5-MB sample is 58.0% and for the five alternatives in the 6-CA sample, it is 48.42%. An available appendix contains a much more detailed breakdown of prediction success, as well as summed probability measures of goodness-of-fit.

TABLE II

Distribution of WTP Estimates (at Mean Income) across 1000 Nonparametric Bootstrap Replications, Median and Mean with Density in the Negative Domain Set to Zero

Model	Bootstrap Median WTP Avg.	Bootstrap Median WTP 95% CI	Bootstrap Median WTP 90% CI	Bootstrap Mean WTP Avg.	Bootstrap Mean WTP 95% CI	Bootstrap Mean WTP 90% CI
Separate Samples						
0-ACT	$1.73	$0.00–$5.06	$0.00–$4.73	$5.00	$2.88–$8.78	$3.10–$7.46
1-PDC	2.97	0.00–7.88	0.00–6.63	13.51	5.96–37.79	6.35–25.97
2-MDC	2.92	2.03–3.72	2.21–3.56	5.15	4.33–6.18	4.45–5.97
3-OE	1.37	0.02–2.76	0.16–2.56	6.11	4.59–7.88	4.83–7.64
4-PC	2.07	1.60–2.60	1.68–2.49	3.25	2.67–3.88	2.74–3.77
5-MBl	2.00	0.24–3.69	0.47–3.51	8.65	7.13–10.33	7.31–10.04
5-MBu	6.93	5.45–8.52	5.67–8.28	11.61	9.51–13.76	9.91–13.47
6-CApair[a]	3.65	1.93–5.52	2.30–5.10	5.85	4.42–7.83	4.58–7.50
6-CAjoint				4.67		
Pooled-01256						
0-ACT[b]	$3.33	$2.49–$4.15	$2.63–$4.01	$3.96	$3.37–$4.75	$3.44–$4.58
1-PDC				5.63	4.63–7.88	4.71–7.08
2-MDC				5.42	4.57–6.44	4.72–6.17
5-MB				10.14	8.06–12.71	8.37–12.26
6-CApair				5.62	4.54–6.93	4.66–6.68
6-CAjoint				9.10		
Pooled-0126						
0-ACT[b]	$3.06	$2.19–$3.93	$2.34–$3.76	$4.08	$3.33–$5.27	$3.41–$4.98
1-PDC				6.13	4.80–9.13	5.00–8.08
2-MDC				5.59	4.57–7.00	4.76–6.67
6-CApair				5.75	4.53–7.42	4.64–7.09
6-CAjoint				9.72		

[a] The joint 6-CA portion of the conjoint choice data is allowed to have a different dispersion than the pairwise 6-CA choices. While the median WTP will be the same for both types of choices, the means (expected values) will be different.

[b] For the pooled data model, since the indirect utility-difference parameters are identical across these samples, the median WTP is identical across these samples. However, since the dispersions differ, the means are different when probability in the negative domain is disallowed and moved to zero.

5.2. Pair-wise Comparisons across Individual Value-Elicitation Methods

Most previous studies drawing comparisons between alternative elicitation methods have considered different methods two at a time. Table III reports selected estimation results for the 21 possible pairings of samples. Identical indirect utility-difference function parameters can be rejected at the 5% level for the 3-OE/0-ACT and 3-OE/1-PDC pairs, for the 4-PC/0-ACT and 4-PC/1-PDC pairs, and for the 5-MB/1-PDC pair. There is also some evidence (at the 10% level) that the underlying utility parameters differ across the 4-PC/2-MDC pair and the 5-MB/0-ACT pair. The 0-ACT and 1-PDC samples have in common their constant $6 bid values and mention of the 12,000-subscriber requirement. The 3-OE and 4-PC samples have in common their estimation directly in terms of WTP, rather than in terms of the indirect utility-difference function.

TABLE III

Pair-wise Comparisons of WTP Estimates by Different Elicitation Methods

Alone[a]	With 1-PDC	With 2-MDC	With 3-OE	With 4-PC	With 5-MB	With 6-CA
0-ACT	−377.367[b]	−370.963	−1159.303	−1028.592	−597.325	−445.397
−165.119	−375.241	−370.516	−1152.516	−1017.285	−594.129	−444.946
	$p = 0.235$	$p = 0.827$	$\mathbf{p = 0.004}$**	$\mathbf{p = 0.000}$**	$\mathbf{p = 0.094}$*	$p = 0.825$
1-PDC		−418.610	−1207.126	−1074.286	−645.198	−491.997
−210.122		−415.519	−1197.520	−1062.288	−639.132	−489.949
		$p = 0.103$	$\mathbf{p = 0.000}$**	$\mathbf{p = 0.000}$**	$\mathbf{p = 0.007}$**	$p = 0.251$
2-MDC			−1195.008	−1061.134	−636.541	−485.302
−205.397			−1192.795	−1057.563	−634.407	−485.224
			$p = 0.219$	$\mathbf{p = 0.068}$*	$p = 0.234$	$p = 0.984$
3-OE				−1842.028	−1417.490	−1268.800
−987.398				−1839.564	−1416.408	−1267.225
				$p = 0.177$	$p = 0.539$	$p = 0.369$
4-PC					−1281.298	−1134.270
−852.166					−1281.177	−1131.993
					$p = 0.970$	$p = 0.208$
5-MB						−710.508
−429.010						−708.837
						$p = 0.342$
6-SP						
−279.827						

[a] Maximized value of the log-likelihood for this sample when used alone.

[b] First number in each set is the maximized value of the log-likelihood when this pair of samples is pooled and the utility parameters are constrained to be identical; second number is the sum of the maximized values of the log-likelihood for each sample in the pair used separately (with utility parameters unconstrained); third number is the p-value for the χ^2 test of the three parameter restrictions in the constrained model.

5.3. Pooled Models (Multiple Samples)

We now compare the indirect utility-difference function implied by each sample (used individually) with the indirect utility-difference function estimated when preferences are constrained to be identical across multiple samples. The sum of the separate log-likelihood functions for the seven independent samples, reported in the 0-ACT through 6-CA columns of Table I, is -3129.03. For the fully encompassing joint model (Pooled-0123456 in Table I), where the three corresponding utility parameters, β, are restricted to be the same, the maximized log-likelihood is only -3153.93, which rejects these 18 restrictions overwhelmingly.

We can look back to the pair-wise comparisons for clues as to the identity of samples for which preferences may be inconsistent. Leaving out the worst-offending 4-PC sample and testing utility-parameter equivalence across the remaining samples rejects preference parameter equivalence at the 0.00003 level, likewise leaving out just the 3-OE sample rejects at the 0.004 level, and similarly leaving out just the 5-MB sample rejects parameter equivalence at the 0.0004 level. The remaining samples appear to be pair-wise consistent among themselves, but not with the 4-PC, 3-OE, or 5-MB samples, so there is no point in exploring other one-by-one omissions of other individual samples.

We can now consider subsets of five data types. For pooled data excluding both the 3-OE and 4-PC samples (Pooled-01256 in Table I), we can reject the hypothesis of common indirect utility parameters only at the 0.11 level. If we also exclude the 5-MB sample (Pooled-0126 in Table I), common β parameters can be rejected only at the 0.49 level of significance. Thus, there is very little evidence that distinct preferences are being elicited by the 0-ACT revealed preference method and three of the stated preference methods: 1-PDC (identical to the 0-ACT instrument, but delivered by mail), the 2-MDC (mail dichotomous choice survey), and the 6-CA (mail conjoint analysis survey). There is only weak evidence suggesting that the 5-MB preferences may be different than these others, given that we fail to reject equivalence at the 10% level of significance.[23]

5.4. Specific Parameter Estimates for Admissible Pooled Model

Some of the specific results in Table I deserve comment. The point estimates of the intercept parameters for Options B, D, and E might be worth scrutiny, but these are based solely upon the choices by the 303 members of the 6-CA sub-sample, both in the 6-CA individual-sample results and in the pooled-sample results. The absence of statistical significance on all but one of the six estimates unfortunately does not make it appropriate to speculate further.

The estimated ordered-logit threshold values are relevant whenever we use the multiple-bounded value information from the 5-MB sample. For the 5-MB sample alone, α_0 is normalized to zero, although it can be estimated if the 5-MB sample is combined with other samples. Recall that an ordinary binary logit model is simply a special case of an ordered logit with only two categories of response. In the ordinary logit case, the threshold between the two categories (willing to pay, not willing to pay) is set arbitrarily to zero. The corresponding zero-level for the five-category multiple-bounded data should fall somewhere in the "middle" interval if data for this sample are being combined with data from the 2-MDC sample. This middle interval is bounded by α_1 and α_2. We would therefore expect to see α_0 and α_1 negative, and α_2 and α_3 positive. This expectation is borne out.

Of particular interest are the estimated factors of proportionality in the error dispersions for the indirect utility-difference function across each data type. These results are pertinent to the emerging debate about the effects of survey instrument complexity on the cognitive difficulty experienced by respondents (DeShazo and Fermo [18]). Because this is the largest sample, and because of the popularity of this method, sample 2-MDC (the mail dichotomous choice sample) is defined as the numeraire sample, with dispersion parameter κ_2 normalized here to unity. In order to ensure positive dispersions, we estimate the logs of the multiples of this dispersion factor for each sample. For ease of interpretation, Table I reports the corresponding levels of these estimated dispersion factors. (However, the associated t-test statistics still refer to the logged parameters.) The asymptotic t-ratios

[23] We know from a more elaborate utility specification, with age as a shifter on the intercept term of the indirect utility-difference function, that the 5-MB preference parameters become consistent with the Pooled-126 parameters. Unfortunately, our small sample of 117 0-ACT observations does not allow this richer utility specification to be estimated for that sample individually. It is unfortunate that the actual NMPC Green Choice[tm] program terminated prematurely and therefore restricted the size of this sample.

can be used to test the hypothesis that the logarithm of the relevant parameter is zero (or, equivalently, that the estimated factor of proportionality for the dispersion is one, so that the dispersion is the same as for the numeraire sample).[24]

Based on the Pooled-0126 results, the constant-bid actual choice 0-ACT sample appears to have a indirect utility-difference error dispersion that is about half the size of the analogous 2-MDC error dispersion. The hypothetical dichotomous choice variant (1-PDC) has an error dispersion that is just a little larger than the 2-MDC sample. Contingent DC preference information again appears to be noisier than data on actual choices, a result consistent with experimental economics research. Harrison [25] and Smith and Walker [43, 44] have suggested, based on their laboratory results, that valuations will be more variable the lower the opportunity cost to respondents of deviating from the rational decision. Smith and Walker [44] show that bid function slopes are insensitive to these opportunity costs, but response variability can be decreased markedly by imposing larger opportunity costs.

The error dispersions for the 3-OE and 4-PC samples, and possibly the 5-MB sample in the first two pooled models, are not reliable because we reject the equality of their preference parameters that is a necessary maintained hypothesis for identifying this difference in error dispersions. In the Pooled-01256 model, the error dispersion in the multiple-bounded (5-MB) data is more than twice as large as that for the 2-MDC sample. This may belie greater cognitive challenges associated with this elicitation format or could be an artifact of the pooled ordered-logit approach used here, as opposed to the separate PC interpretation used by others (e.g., Welsh and Poe [49]). Further research on appropriate modeling strategies for multiple-bounded data, and their consequences, is clearly warranted.

The dispersion measure for the pair-wise conjoint analysis (6-CA) sample is virtually identical to the 2-MDC case, as one would expect. For the joint stated preference sample, however, the error dispersion is about twice as large. It appears that when an alternative is embedded in a wider range of choices, additional noise is created. Certainly, the cognitive challenge for respondents has increased.

The common indirect utility parameters from the Pooled-0126 model produce a point estimate of latent WTP equal to \$3.03. Referring to Fig. 1, the bootstrap mean of \$3.01 (which is similar to the bootstrap median for this model) lies well within the lower and upper adjacent values for the sampling distributions for WTP for each of the constituent data sets used alone. However, \$3.01 lies outside the

[24] We do not report the results of hypothesis tests concerning the equality across elicitation methods of all parameters, including the dispersion parameters. These hypotheses are overwhelmingly rejected. The only reason the hypothesis of equal variance is an interesting one is because of the history of comparisons of logit parameter point estimates across samples where researchers have failed to allow for different dispersion in the different samples and have therefore rejected “identical preferences.” The recent conjoint analysis literature has been very vocal on the necessity of allowing for different variances before testing the equality of the underlying utility parameters. Early contingent valuation researchers were not in the habit of doing this, because it was easy to overlook the fact that ordinary logit model parameters are normalized on the implicit dispersion of the underlying latent variable’s distribution. Different variances could thus lend the appearance of different utility parameters when these parameters were in fact identical. Noting that failure to allow for different variances in our specifications does indeed make it appear that vastly different preferences underlie choices is a replication of these early results.

adjacent-value interval for the 3-OE and 4-PC samples, which tend to produce lower values for fitted latent WTP. We do not know the exact location of the 5-MB value, only bounds on the interval where this value lies. The interval formed by the medians of the distributions of the upper and lower bounds lies strictly above \$3.01. However, \$3.01 is well within the range of bootstrapped values for the lower bound, so it is possible that the WTP underlying the 5-MB data is consistent with the Pooled-0126 model.

However, we must again consider the consequences of precluding negative actual WTP. Across all sub-samples of the Pooled-01256 or Pooled-0126 models (including both the pair-wise and joint 6-CA data), there is a common E[latent WTP] equal to the median of the distribution of actual WTP values. However, if we wish to compute the E[actual WTP], the differing fitted dispersions for each sub-sample will come into play. Figure 3 shows the distribution of these distinct expected values across our 1000 bootstrap replications for the Pooled-01256 model. The means of these distributions (as distinct from the medians shown in the diagram) are between \$5.54 and \$5.76 for the pairwise choices, \$4.00 for the 0-ACT data, \$10.50 for the 5-MB data, and about \$9.40 for the joint 6-CA data. As expected, the sub-samples with larger fitted dispersions show larger expected values. Figure 4 gives the analogous results for the Pooled-0126 model. The means for these distributions are between \$5.74 and \$6.30 for the pairwise choices, \$4.14 for the 0-ACT sample, and \$10.04 for the joint 6-CA data. Again, these differences stem, statistically, from the varying amounts of noise in the estimated indirect-utility difference function and the fact that this carries over into the WTP function (and will affect E[WTP] if negative WTP is disallowed).

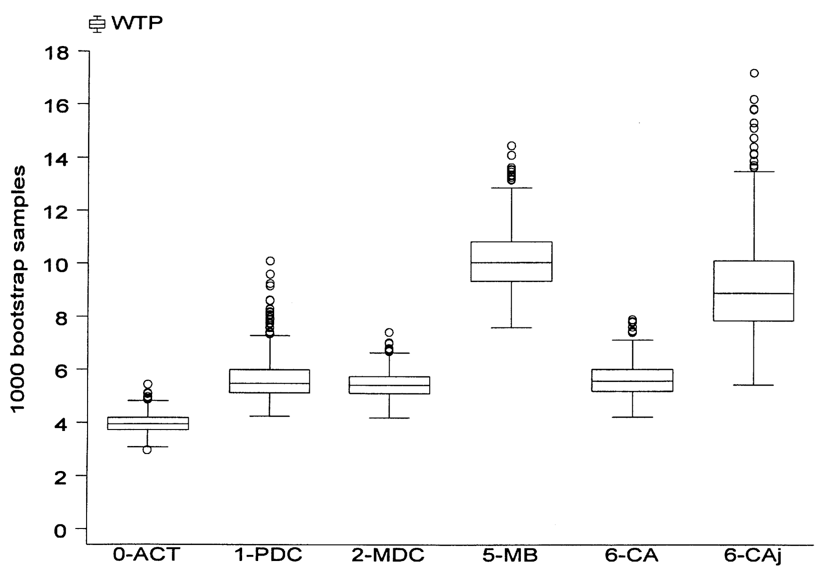

FIG. 3. Nonparametric bootstrap replications: fitted E[WTP] for subsamples of Pooled-01256 model at means of data. The "box" in each plot captures the median and interquartile range and the ticks on the whiskers show the upper and lower "adjacent values." See Stata [45].

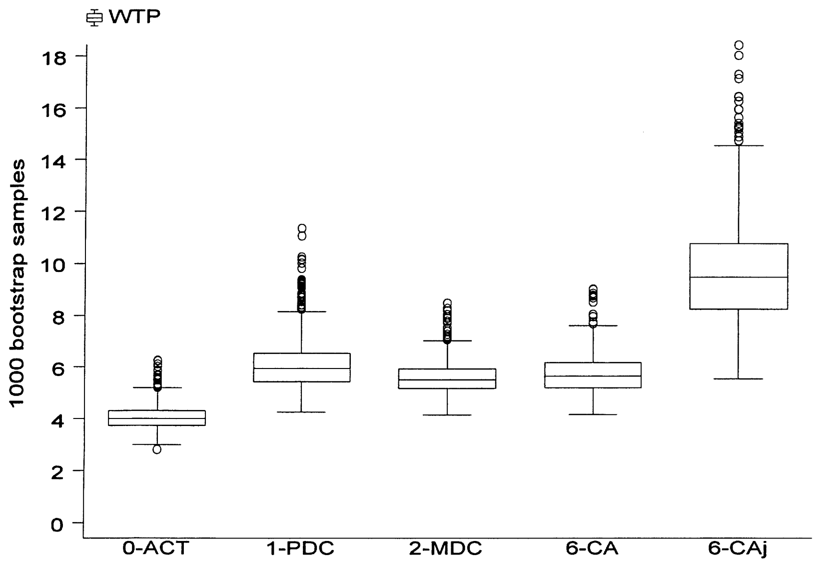

FIG. 4. Nonparametric bootstrap replications: fitted E[WTP] for subsamples of Pooled-0126 model at means of data. The "box" in each plot captures the median and interquartile range and the ticks on the whiskers show the upper and lower "adjacent values." See Stata [45].

The fifth horizontal section of Table I shows the estimated derivative of latent WTP at the data means (i.e., income = \$0.04329 million = \$43,290). Asterisks emphasize where the relevant parameter estimates make this derivative statistically significantly different from zero. These effects, when significant, are all positive, suggesting normality, although they differ by an order of magnitude across individual samples. But the absolute effects are small. For the Pooled-01256 sample, a \$10,000 higher income corresponds only to a slightly more than half-cent greater willingness to pay for Option C, implying elasticities of WTP with respect to income on the order of only about 0.006.

For the fitted distributions of WTP for each sub-sample in the pooled cases, the median WTP values at the means of the data are the same, but the dispersions are different. Different dispersions mean that the proportion of the population that would be predicted to vote in favor of the policy (at any cost other than this common median WTP) will differ across elicitation methods. The last main horizontal section of Table I shows the consequences of differing fitted dispersions for the predicted portion of the represented population that would be willing to pay at least \$6 for Option C. For Pooled-0126, despite identical mean WTP, these proportions are 27% for the actual 0-ACT data, 38% for the 1-PDC data, 37% for the 2-MDC data, 37% for the pair-wise 6-CA, and 43% for the multiple-choice "joint" 6-CA data. Like the differences in E[actual WTP] considered above, these differences are another reflection of the different fitted dispersions for each type of data. Despite the fact that the indirect utility-difference parameters from all four types of data in the Pooled-0126 model are statistically indistinguishable, the different dispersions will affect predicted market behaviors.

6. DIRECTIONS FOR FUTURE RESEARCH

We have kept our example of a formal structural specification as clean as possible by adopting a very simple form for the indirect utility difference. More exotic functional forms could certainly be explored, subject to the requirement that the corresponding formula for maximum willingness to pay have an empirically tractable functional form. However, the more complicated the indirect utility difference, the more unwieldy the associated WTP formula, especially since the stochastic structure must be carried along.

We also lean heavily on the tradition of using logit-type models to analyze discrete-choice contingent valuation data. However, the assumption of logistic errors derives only from the usual assumptions of random utility models and represents a maintained hypothesis. In this application, propagating the logistic errors through to the tobit-type regression model used for the open-ended data maintains the utility-theoretic functional form and the stochastic structure throughout the model but may not provide the best possible fit to the open-ended data. This sub-sample is the only one that directly gives us an array of point values for WTP, although these values are heavily "clumped." Still, it may be fruitful to begin with the open-ended data and to choose a functional form for the distribution of WTP values that is more approximately consistent with the observed marginal (or conditional) distribution of WTP values for this sample.

Heteroscedasticity could also be explored further. The present analysis shows that the implied error dispersions vary systematically with elicitation modes, and since modes were varied randomly across respondents, we do not expect the different modes to be picking up systematic differences in respondent characteristics across samples. However, even within a mode, there may be systematic heteroscedasticity in WTP across individuals according to their characteristics. Our models could be generalized as in Cameron and Englin [33]. It will be important to determine whether there are important interactions between respondent characteristics and elicitation methods that affect the magnitude of the heteroscedasticity across methods. Evidence will continue to accumulate concerning systematic differences in scale factors across elicitation methods. We will begin to understand more clearly what will be the expected difference in scale factors between particular stated choice models (involving choice scenarios with particular properties) and actual market choices. This expected difference may well differ systematically with respondent characteristics, so greater understanding of the determinants of heteroscedasticity both within and across elicitation methods will be important. Then the predicted choices from conjoint choice models can be better calibrated to predict actual market behavior.

The α_0, α_1, α_2, and α_3 thresholds in the multiple-bounded data are here constrained to be identical for all 13 questions posed to each respondent, but they could be allowed to differ question by question. Or we might allow the 13 MB questions to exhibit a distinct error standard deviation. Specifications such as these would let the researcher check for a tendency for respondent fatigue (if precision declines systematically with consecutive questions). Alternately, if precision increases with questions, perhaps learning (or crystallization of values) is occurring and respondents are increasingly able to identify their latent values as they progress through this list of choice questions. (Concern regarding the consequences of respondent fatigue across repeated questioning is also voiced by Johnson and

Desvousges [32].) Multiple-bounded elicitation methods are still new and the information that they provide is not yet entirely well understood.

7. CONCLUSION

For almost two decades, researchers in several disciplines have been puzzled by discrepancies among the empirically estimated values of pre-test-market or non-market goods across different elicitation methods used to assess these values. Many past comparisons have been hampered by the need to use samples collected at substantially different points in time, or from different populations, or via survey instruments that differ in other ways besides just the elicitation method employed. The earliest comparisons involved simply an inspection of the different point estimates of value produced by different methods, without conformable models or any opportunity for rigorous assessment of whether the values are statistically significantly different. We use more than 7000 choices to fit an elaborate specification that pools the data for seven types of elicitation methods.

Despite the individual statistical significance of many of the utility function parameters estimated from separate samples, we are unable to reject the hypothesis of identical indirect utility-difference functions across four of these elicitation methods.[25] The two methods which appear to be the least consistent with the others are the only two methods that attempt to elicit WTP directly, rather than inferring WTP from choices. It may be the case that these methods—the open-ended and payment card methods—might lead respondents to think about the problem in rather different ways, as Hanemann [24] has suggested may be the case when values appear to differ. The jury may still be out on whether the multiple-bounded format yields a different preference structure.

Our approach differs from earlier ones in three important ways. First, we have the widest representation of different elicitation methods. Our seven different elicitation methods are part of one careful experimental design across split samples. With the exceptions that the survey instrument was delivered by telephone for the sample that was asked to make a real choice at a constant bid of $6, and that this survey and its linking constant-bid mail analog mentioned a 12,000-subscriber requirement, all other aspects of the survey are carefully controlled so that any differences in the valuation results across samples can be attributed, with the least possible ambiguity, to the different elicitation formats being used.

Second, we do not simply compare the marginal mean WTP values (for convenient ad hoc specifications) implied by the different samples with their varying "treatments." Instead, we focus our efforts on the estimation of the underlying preference function parameters. This structural emphasis allows us to specify functional forms and stochastic structures that are entirely compatible across all of the different types of consumer choice scenarios that correspond to each elicitation method, subject to the constraint that there be a relatively simple analytical formula for the corresponding WTP function. The key insight is that if two or more elicitation methods imply the identical preference function for individuals in each sample, then these individuals will also have identical expected latent WTP (since

[25] The usual important caveat, however, is that failure to reject a null hypothesis may also be an artifact of insufficient or ill-conditioned data, rather than an indication that the hypothesis is true. Further work in this vein is clearly required.

the E[latent WTP] is not a function of the dispersion of the WTP distribution). This is because the formula for expected WTP is derived from the indirect utility-difference function that is the common foundation for all of our models. However, if we shift to zero any density associated with the latent WTP variable that corresponds to negative values, E[WTP] again depends upon the dispersion associated with each type of data. Finding a clever way around this problem would certainly be useful.

Third, we explicitly allow for heteroscedasticity across methods. Our results confirm a growing consensus. It is known to be essential to accommodate heteroscedasticity across pre-test-market or non-market valuation samples that differ in fundamental ways but are examined using the same elicitation method (Swait and Louviere [47], Adamowicz *et al*. [2], Haab *et al*. [22], Hensher *et al*. [26]). It is not surprising, therefore, that the same should be true across very similar samples that are confronted by very different elicitation methods.

While it is interesting to confirm in such a general setting that the indirect utility-difference error terms appear to be heteroscedastic across elicitation techniques, this cannot be the end of the story. As researchers, we are now compelled to explain why error variances differ across methods. Competing (or perhaps complementary) theories are beginning to emerge to explain why we might expect to see characteristic differences in error variances across elicitation methods. DeShazo and Fermo [18] offer the interaction of choice set complexity with respondent cognitive capacity as an explanation for different variances. Carson *et al*. [15] suggest that some of these differences may be an artifact of optimal strategic behavior on the part of survey respondents. Development of utility-theoretic empirical models that can capture the sources of systematic variation in error variances across value-elicitation methods is clearly high on the research agenda.

REFERENCES

1. W. Adamowicz *et al*., Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation, *Amer. J. Agric. Econom.* **80**, 64–75 (1998).
2. W. Adamowicz *et al*., Combining revealed and stated preference methods for valuing environmental amenities, *J. Environ. Econom. Management* **26**, 271–292 (1994), doi:10.1006/jeem.1994.1017.
3. W. Adamowicz *et al*., Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation, *J. Environ. Econom. Management* **32**, 65–84 (1997), doi:10.1006/jeem.1996.0957.
4. D. R. Bell *et al*., Determining where to shop: fixed and variable costs of shopping, *J. Marketing Res.* **35**, 352–369 (1998).
5. M. Ben-Akiva and S. R. Lerman, "Discrete Choice Analysis: Theory and Application to Travel Demand," MIT Press, Cambridge, MA (1985).
6. M. Ben-Akiva and T. Morikawa, Estimation of travel demand models from multiple data sources, *in* "Transportation and Traffic Theory" (M. Koshi, Ed.), Elsevier, New York (1990).
7. U. Bockenholt and W. R. Dillon, Some new methods for an old problem: Modeling preference changes and competitive market structures in pretest market data, *J. Marketing Res.* **34**, 130–142 (1997).
8. A. K. Bohara *et al*., Effects of total cost and group-size information on willingness to pay responses: Open ended vs. dichotomous choice, *J. Environ. Econom. Management* **35**, 142–163 (1998), doi:10.1006/jeem.1998.1022.
9. P. C. Boxall *et al*., "Valuing Undiscovered Attributes: A Combined Revealed-Stated Preference Analysis of North American Aboriginal Artifacts," Department of Applied Economics and Statistics, University of Nevada, Reno (1998).

10. K. J. Boyle *et al*., Valuing public goods: Discrete versus continuous contingent-valuation responses, *Land Econom*. **72**, 381–396 (1996).
11. T. C. Brown *et al*., Which response format reveals the truth about donations to a public good?, *Land Econom*. **72**, 152–166 (1996).
12. T. A. Cameron, A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression, *J. Environ. Econom. Management* **15**, 355–379 (1988).
13. T. A. Cameron and J. E. Englin, Respondent experience and contingent valuation of environmental goods, *J. Environ. Econom. Management* **33**, 296–313 (1997), doi:10.1006/jeem.1997.0995.
14. T. A. Cameron and D. D. Huppert, OLS versus ML estimation of non-market resource values with payment card interval data, *J. Environ. Econom. Management* **17**, 230–246 (1989).
15. R. Carson *et al*., "Incentive and Informational Properties of Preference Questions," Department of Economics, UCSD, La Jolla, CA (2000).
16. R. T. Carson *et al*., Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods, *Land Econom*. **72**, 80–99 (1996).
17. R. G. Cummings *et al*., Are hypothetical referenda incentive compatible?, *J. Polit. Econ*. **105**, 609–621 (1997).
18. J. R. DeShazo and G. D. Fermo, Understanding the effects of choice set complexity on consumer's choice behavior: An application to recreational demand, *in* "W-133 Annual Meeting" (W. D. Shaw, Ed.), Tucson, AZ (1999).
19. R. Ethier *et al*., A comparison of hypothetical phone and mail contingent valuation responses for green electricity programs, *Land Econom*. **76**, 54–67 (2000).
20. S. M. Goldfeld and R. E. Quandt, GQOPT: General purpose numerical optimization package (1992).
21. S. Gupta *et al*., Do household scanner data provide representative inferences from brand choices: A comparison with store data, *J. Marketing Res*. **33**, 383–398 (1996).
22. T. C. Haab *et al*., Are hypothetical referenda incentive compatible? A comment, *J. Polit. Econ*. **107**, 186–196 (1999).
23. B. Halvorsen and K. Sælensminde, Differences between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods: The effects of heteroscedasticity, *Land Econom*. **74**, 262–282 (1998).
24. W. M. Hanemann, Theory versus data in the contingent valuation debate, *in* "The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs" (D. J. Bjornstad and J. R. Kahn, Eds.), Elgar, Brookfield, MA (1996).
25. G. W. Harrison, Theory and misbehavior of first-price auctions, *Amer. Econom. Rev*. **79**, 749–762 (1989).
26. D. Hensher *et al*., Combining sources of preference data, *J. Econometrics* **89**, 197–221 (1999).
27. J. Hoehn and A. Randall, A satisfactory benefit cost indicator from contingent valuation, *J. Environ. Econom. Management* **14**, 226–247 (1987).
28. T. P. Holmes and R. A. Kramer, An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation, *J. Environ. Econom. Management* **29**, 121–132 (1995), doi:10.1006/jeem.1995.1035.
29. J. Horowitz, Sampling, specification and data errors in probabilistic discrete-choice models, *in* "Applied Discrete-Choice Modeling" (D. A. Hensher and L. W. Johnson, Eds.), Wiley, New York (1981).
30. J. Horowitz, Sources of error and uncertainty in behavioral travel demand models, *in* "New Horizons in Travel-Behavior Research" (P. Stopher, A. Meyburg, and W. Brog, Eds.), Lexington, Lexington, MA (1980).
31. J. C. Huang and V. K. Smith, Monte Carlo benchmarks for discrete response valuation methods, *Land Econom*. **74**, 186–202 (1998).
32. F. R. Johnson and W. H. Desvousges, Estimating stated preferences with rated-pair data: Environmental, health, and employment effects of energy programs, *J. Environ. Econom. Management* **34**, 79–99 (1997), doi:10.1006/jeem.1997.1002.
33. D. Kahneman and A. Tversky, Choices, values, and frames, *Amer. Psychologist* **39**, 341–350 (1984).
34. R. A. Kramer and D. E. Mercer, Valuing a global environmental good: US residents' willingness to pay to protect tropical rain forests, *Land Econom*. **73**, 196–210 (1997).
35. B. Kristrom, Spike models in contingent valuation, *Amer. J. Agric. Econom*. **79**, 1013–1023 (1997).

36. J. J. Louviere *et al*., Combining sources of preference data for modeling complex decision processes, *Marketing Lett*. **10**, 205–217 (1999).
37. G. Poe *et al*., "Can Hypothetical Questions Predict Actual Participation in Public Programs? A Field Validity Test Using a Provision Point Mechanism," Cornell University (1997).
38. R. C. Ready *et al*., Differences between continuous and discrete contingent value estimates, *Land Econom*. **72**, 397–411 (1996).
39. R. C. Ready *et al*., Contingent valuation when respondents are ambivalent, *J. Environ. Econom. Management* **29**, 181–196 (1995), doi:10.1006/jeem.1995.1040.
40. D. Rondeau *et al*., Voluntary revelation of the demand for public goods using a provision point mechanism, *J. Public Econom*. **72**, 455–470 (1999).
41. S. K. Rose *et al*., "The Private Provision of Public Goods: Tests of a Provision Point Mechanism for Funding Green Power Programs," Cornell University (1997).
42. W. Schulze *et al*., Sources of bias in contingent valuation, *in* "The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs" (D. J. Bjornstad and J. R. Kahn, Eds.), Elgar, Cheltenham, UK (1996).
43. V. L. Smith and J. M. Walker, Monetary rewards and decision cost in experimental economics, *Econom. Inquiry* **31**, 245–261 (1993).
44. V. L. Smith and J. M. Walker, Rewards, experience and decision costs in first price auctions, *Econom. Inquiry* **31**, 237–245 (1993).
45. Stata, "Stata Graphics Manual," Stata Press, College Station, TX (1999).
46. T. Stevens *et al*., Conjoint analysis of groundwater protection programs, *Agric. Resour. Econom. Rev*. **26**, 229–236 (1997).
47. J. Swait and J. Louviere, The role of the scale parameter in the estimation and comparison of multinomial logit models, *J. Marketing Res*. **30**, 305–314 (1993).
48. A. Tversky *et al*., Contingent weighting in judgment and choice, *Psych. Rev*. **95**, 371–384 (1988).
49. M. P. Welsh and G. L. Poe, Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach, *J. Environ. Econom. Management* **36**, 170–185 (1998), doi:10.1006/jeem.1998.1043.