

Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions

Thomas P. Holmes and Kevin J. Boyle

ABSTRACT: *A hybrid stated-preference model is presented that combines the referendum contingent valuation response format with an experimentally designed set of attributes. A sequence of valuation questions is asked to a random sample in a mail-out mail-back format. Econometric analysis shows greater discrimination between alternatives in the final choice in the sequence, and the vector of preference parameters shifts. Lead and lag choice sets have a structural influence on current choices and unobserved factors induce positive correlation across the responses. These results indicate that people learn about their preferences for attribute-based environmental goods by comparing attribute levels across choice sets. (JEL Q23, Q26)*

I. INTRODUCTION

Contingent valuation (CV) has been the most commonly employed stated preference method used to estimate nonmarket values (Boyle 2003; Mitchell and Carson 1989). The most common response format, dichotomous choice, frames valuation questions as a policy alternative that has a specified cost (bid); respondents answer “yes,” they will pay the bid amount, or “no,” they will not pay that amount for the policy alternative. The policy alternative is typically fixed, but the bid amounts vary across individuals.

A new class of stated-preference methods has more recently been adopted for estimating nonmarket values, which are referred to as attribute-based methods, or ABMs (Holmes and Adamowicz 2003). Although the origins of ABMs are found in various social science disciplines such as psychology, transportation, and marketing research, the conceptual foundation for

this class of models within economics finds its source in Lancaster’s (1966) theory of demand for differentiated products. Rather than focusing attention on a total, or “holistic,” value as is done in standard CV, ABMs focus attention on a set of attributes, including the cost, that have management or policy relevance (Adamowicz, Louviere, and Swait 1998; Bennett and Blamey 2001).

As with contingent valuation, a number of response formats exist for use with stated preference ABMs (e.g., Holmes and Adamowicz 2003). The most common response format, the “choice experiment,” asks survey respondents to choose their most preferred alternative from a choice set containing two or more alternatives that differ in terms of the levels of one or more attributes (e.g., Louviere, Hensher, and Swait 2000). Using an attribute-based commodity description and a choice between one alternative and the status quo effectively merges contingent valuation and ABMs into a hybrid stated-preference question.

In order to improve the statistical efficiency of response data for a given number of questionnaires, it is typical for researchers using ABMs to include a sequence of preference questions, and to pool responses within and across individuals when conducting analysis. Concern with the issue of design efficiency has led researchers to include a sequence of 4, 8, or even 16 choice questions in a single survey instrument. By pooling

The authors are, respectively, research forest economist, Southern Research Station, USDA Forest Service, and Libra Professor of Environmental Economics, Department of Resource Economics and Policy, University of Maine. The authors would like to acknowledge the suggestions made by John Loomis and Patty Champ, as well as two anonymous referees, which have greatly improved the quality of this work. All errors remain the responsibility of the authors.

these data for analysis, three strong, but generally untested, hypotheses are maintained: (1) preferences are stable across the sequence of responses; (2) intra-individual responses are independent with respect to structural components of preferences; and (3) intra-individual responses are independent with respect to unobserved (random) components of preferences. If any of these maintained hypotheses are untrue, then resulting estimates of Hicksian surplus will be biased.

The first maintained hypothesis, that preferences are stable across a sequence of responses, has not received much attention in ABM experiments utilizing a discrete choice format. Stability of preferences has been evaluated, however, in ranking experiments (Ben-Akiva, Morikawa, and Shiroishi 1992), and in "data fusion," or combination of stated-preference and revealed-preference data (Swait and Louviere 1993; Adamowicz, Louviere, and Williams 1994; Louviere et al. 1999). The basic idea relies on the identification of a relative scale parameter (for one data set relative to another) that can be isolated from the vector of preference parameters. Then, by adjusting for differences in scale, hypothesis tests can be conducted to evaluate whether two (or more) parameter vectors are equivalent. Utilizing this procedure, Ben-Akiva, Morikawa, and Shiroishi (1992) found that scale-adjusted parameters were not stable across all ranks in a ranking experiment and recommended that data not be pooled unless preference parameters are equivalent up to a scaling constant. In a choice experiment, the relevant analogy is pooling the data from a sequence of choice questions.

The second maintained hypothesis is that preferences for alternatives in the current choice set do not depend on alternatives contained in prior or posterior choice sets. The assumption that choices between alternatives do not depend on the presence or absence of other alternatives is known as the independence of irrelevant alternatives (IIA) axiom. In an important paper, Simonson and Tversky (1992) argued that violations of IIA can be identified in two contexts. The first context is "local" and refers to alternatives within a choice set.

This is the context within which violations of IIA in the multinomial logit (MNL) model have typically been investigated. In addition, they draw attention to the "background context," wherein the choice between alternatives in the current choice set is conditioned by alternatives present in other choice sets. Although experiments identifying the influence of the background context on consumer choice have been reported in the consumer and marketing research literatures (Huber, Payne, and Puto 1982; Simonson and Tversky 1992; Tversky and Simonson 1993), this potentially important phenomenon has not received attention by economists interested in choice modeling. This oversight is addressed in this paper.

After controlling for the potential effects of alternatives presented in prior and posterior choice sets, it is possible that intra-individual responses remain correlated through variables that are not observed by the researcher. If the sampled population has heterogeneous preferences, then the response behavior of any particular individual may differ systematically from the average behavior of the sample. If respondent heterogeneity is due to unobserved factors, it is possible to decompose equation errors using an error-components model in which the error term is comprised of a permanent component that captures idiosyncratic behavior of the individual, and a transitory random shock. If the permanent component is treated as an individual-specific intercept that shifts the indirect utility function, and if it is assumed that the intercept variables are randomly distributed over the population, then a random-effects probit model can be estimated (Greene 1997). Failure to include a specification for the permanent component in the non-linear probit model can lead to omitted-variable bias (Hsiao 1986).

In this study, we utilize a hybrid stated-preference question that uses the ABM framework to present the valuation commodity and then frames the response format as a dichotomous referendum between one alternative and the status quo, which we refer to as an "attribute-based referenda" (ABR) question. In our experiment, a se-

quence of four policy packages is presented to each respondent, and they are asked to vote on each policy package. The experimental design allows us to test whether preferences are stable across a sequence of policy packages, whether prior and posterior choice sets in the valuation sequence influence the preference expressed in any specific choice, and whether unobserved, random components of individual responses are independent across the sequence of choices. These issues have not been sufficiently investigated in nonmarket valuation applications of ABMs.

II. THE ATTRIBUTE-BASED REFERENDA MODEL

In the ABR format, attributes for a public good are clearly defined and then their levels are randomly varied, along with bid price, across a sequence of valuation alternatives; one for each of the four valuation questions in the current application. Attributes are represented by a fixed number of attribute levels, and a bundle of attributes is referred to as an "alternative." This framing of the stated-preference questions embeds the bid amount as one of the attributes for each policy package.

In this study, we used a "completely randomized" experimental design, wherein policy alternatives were designed using randomly sampled attribute levels for each person in the sample. Random selection of attribute levels reduces potential multicollinearity among the explanatory variables of the model, thereby improving the efficiency of the preference parameter estimates. The design included the constraint that alternatives in each of the four valuation questions presented to individuals must differ in the level of at least one attribute. All four packages were described on facing pages of the survey and respondents were then asked to vote sequentially "yes" or "no" for each package.

Econometric Model

We assume that respondents' indirect utility can be expressed as a function of

the vector of policy attributes (\mathbf{Z}_j) and the bid amount (bid_j):

$$U_j = v(\mathbf{Z}_j, bid_j; \boldsymbol{\beta}, \lambda) + e_j \quad [1]$$

where U_j is indirect utility for policy package j ($j = 1, 2, 3, 4$), v is the nonstochastic part of utility, $\boldsymbol{\beta}$ is a vector of preference parameters, $-\lambda$ is the marginal utility of money, and e_j is a stochastic error term. If there are no sources of response bias, the probability that an individual would vote "yes" to a policy proposal with attributes \mathbf{Z}_j and cost bid_j is described by the difference in utility between the proposed policy and the status quo:

$$\begin{aligned} \Pr[YES_j] &= \Pr[v(\mathbf{Z}_j, bid_j; \boldsymbol{\beta}, \lambda) - v(\mathbf{Z}_{status\ quo}; \boldsymbol{\beta}) \\ &= \Delta v_j \geq \varepsilon_j = e_{status\ quo} - e_j], \end{aligned} \quad [2]$$

where, in a probit model, it is assumed that the e_j , and therefore the ε_j , are normally distributed. There is no cost associated with maintaining the status quo and $bid_{status\ quo}$ is zero.

If utility is a linear function of the policy attributes, the probability of a YES vote on a referendum with policy attributes \mathbf{Z}_j and cost bid_j is

$$\Pr[YES_j] = \Phi\left(\frac{\sum_k b_k z_{kj} - \sum_k b_k z_{k, status\ quo} + \lambda bid_j}{\sigma}\right), \quad [3]$$

where $\Phi(\cdot)$ denotes the standardized cumulative normal distribution, b_k is the preference parameter associated with the attribute z_k , σ is the standard deviation of ε_j , and $\mu = 1/\sigma$ is the scale of the utility function. In cases where only a single data set is available, scale is set equal to an arbitrary positive value, typically $\mu = 1$. Additionally, note that the utility of the status quo, $\sum_k b_k z_{k, status\ quo}$, can be set to zero with no loss of generality.

Testing Preference Parameter Stability

The scale parameter in equation [3] has an important interpretation in choice models, which is revealed in the limiting cases

of the scale parameter. As $\mu \rightarrow \infty$ (or, $\sigma \rightarrow 0$), the choice model becomes deterministic (Ben-Akiva and Lerman 1985, 72). That is, people perfectly discriminate between alternative j and the status quo as the standard deviation of ϵ_j vanishes. At the other limit, as $\mu \rightarrow 0$, people do not discriminate at all between alternative j and the status quo, and the model predicts equal probability of choice between the alternatives (probability = 1/2 for YES or NO). An increase in the scale parameter over a sequence of choices, then, indicates that people are learning about the task and how to formulate responses, resulting in better discrimination between alternatives. Conversely, a decrease in the scale parameter μ over a sequence of choices indicates a loss of discrimination, which may be attributed to factors such as mental fatigue or confusion. The realization that scale is intimately related to the preference parameters, and that estimates of scale can provide insight into the effects of learning, fatigue and task complexity on choice behavior, is an emerging topic in stated preference research (Johnson and Desvousges 1997; Swait and Adamowicz 2001; DeShazo and Fermo 2002).

As implied in equation [3], the scale parameter and the preference parameters are always represented in multiplicative form, so it is not possible to identify scale in any single data set. However, it is possible to recover an estimate of relative scale where more than one data set (or sequence of preference questions) is available. The ability to estimate a relative scale parameter for each step in a sequence of preference questions provides information on potential changes in respondent discrimination among alternatives as they respond to a series of questions.

Given an estimate of relative scale, scale-adjusted preference parameters for each step in the sequence of valuation questions can be isolated. The hypothesis that scale-adjusted parameters are stable across a sequence of valuation questions can be tested using a likelihood ratio test (Swait and Louviere 1993). If the null hypothesis that preference parameters are identical for each stage in a sequence is rejected, then

data should not be pooled for cases where the null is rejected. Further, if the parameter stability hypothesis is not rejected, it is possible to test the null hypothesis that the scale parameter for each step in the sequence of valuation questions is identical to the scale parameter for the first question, again using a likelihood ratio test.

Identifying Context-Dependence

Context independence has been viewed as a fundamental requirement for internal consistency of choice as used in the development of rational choice theory. A fundamental condition for internal consistency of choice is known as Property α (basic contraction consistency), or the IIA axiom (Sen 1993). Property α states that an alternative (x) chosen from set T , and which belongs to a subset S of T , must also be chosen from S :

$$[x \in C(T) \ \& \ x \in S \subseteq T] \Rightarrow x \in C(S) \quad [4]$$

where $C(T)$ is a choice function that specifies the choice from set T . Violation of Property α would occur if the following statements were true: (1) $\{x\} = C(\{x, y\})$; and (2) $\{y\} = C(\{x, y, z\})$. Property α implies that the probability of choosing an alternative from a choice set cannot increase if a new alternative is added to the choice set (the regularity property).

Although Property α can be imposed on choices between alternatives within a choice set (leading to the typical tests for violation of IIA as conducted by economists), social psychologists have drawn attention to the effect of decision context, or the global set of alternatives under consideration, on choices. In particular, Tversky and Simonson (1993) distinguish between local context, or the set of alternatives within a choice set, and the background context, containing alternatives that have been previously considered, and present a theoretical model that accommodates both components. The local context is not investigated in the analysis reported here, as all four valuation questions are posed as binary choices. However, the Tversky-Simonson

(henceforth, TS) model is investigated and extended to include the potential impact of lead and lag choice sets on preferences for alternatives contained in the current choice set.

The TS model is based on the assumption that context-dependence can be represented by terms that are linearly additive in the utility function. Integrating the TS model of background context dependence with the linear indirect utility model shown above, an empirical model can be specified:

$$U_j = \sum_k b_k z_{kj} + \lambda bid_j + \sum_k \sum_m \delta_{kj+m} z_{kj+m} + \sum_m \lambda_{j+m} bid_{j+m} + e_j \quad [5]$$

where δ_{kj+m} are the preference parameters induced by attribute levels of alternatives contained in prior and posterior choice sets, λ_{j+m} represents the parameters induced by costs in prior and posterior choice sets, and m represents the number of lead or lag choice sets relative to the current choice set j . For example, in the experiment reported here, m represents the number of background lags ($m = -1, -2, -3$) and leads ($m = 1, 2, 3$). If the parameters δ_{kj+m} and λ_{j+m} are not statistically different than zero, then equation [5] reduces to the standard economic model where the utility associated with an alternative is a function only of the attributes of that alternative, and is not related to the attributes of alternatives contained in other choice sets.

Within the framework of a sequence of attribute-based referenda questions, background context-dependence occurs when a person's vote on a policy alternative is conditioned by attribute levels contained in alternatives other than the current valuation question. In this paper, the following background effects are evaluated

• *Lag dependence*—A one-question lag occurs where information from the immediately preceding alternative affects the evaluation of the current alternative. This could occur where information in the first, second, and third alternatives affect responses to the second, third, and fourth

valuation questions, respectively. A two-question lag could occur when the votes on the third and fourth referendum questions are affected by the respective attributes of the first and second alternatives. Finally, a three-referendum lag could occur if information from the attribute levels in the first alternative affect responses to the fourth valuation question.

• *Lead dependence*—The notion here is the same as for lag dependence but forward-looking information is used. One, two, and three-question leads would arise if the information in the second, third, and fourth alternatives affected responses to the first valuation question.

When considering lag and lead affects for the bid variable (bid_{j+m}), relative changes in bid amount were modeled as a gain or loss for each respondent. This method implements the prospect theory approach proposed by Kahneman and Tversky (1979) to explain loss aversion in choices involving risk, and later modified to include loss aversion in riskless choice (Tversky and Kahneman 1991). Prospect theory is based on the idea that, when faced with a choice that would either increase or decrease personal holdings of property or money, people form reference points from which they consider gains and losses, and potential losses are weighted more heavily than potential gains. Allowing for lag and lead gains and losses provides the opportunity to investigate whether context-dependence induces asymmetric affects around the current alternative reference point. DeShazo (2002) used this idea to examine responses to double-bounded WTP questions, and concluded that people form reference points when they respond "yes" to the first WTP question in the double-bounded format.

To evaluate whether gains and losses had an asymmetric impact on choices, the following variables were defined:

$$lag_gain_{j+m} = \begin{cases} -1*(bid_j - bid_{j+m}) \\ 0 \end{cases}$$

if $bid_{j+m} > bid_j, m = (-1, -2, -3)$ otherwise

[6a]

$$\begin{aligned} \text{lead_gain}_{j+m} &= \begin{cases} -1*(bid_j - bid_{j+m}) \\ 0 \end{cases} \\ &\text{if } bid_{j+m} > bid_j, m = (1,2,3) \\ &\text{otherwise} \end{cases} \quad [6b] \end{aligned}$$

$$\begin{aligned} \text{lag_loss}_{j+m} &= \begin{cases} bid_j - bid_{j+m} \\ 0 \end{cases} \\ &\text{if } bid_j > bid_{j+m}, m = (-1, -2, -3) \\ &\text{otherwise} \end{cases} \quad [6c] \end{aligned}$$

$$\begin{aligned} \text{lead_loss}_{j+m} &= \begin{cases} bid_j - bid_{j+m} \\ 0 \end{cases} \\ &\text{if } bid_j > bid_{j+m}, m = (1,2,3). \\ &\text{otherwise} \end{cases} \quad [6d] \end{aligned}$$

That is, relative gains and losses are computed as the relative change in bid amounts from reference bids. Although a full specification would include up to three lags and leads, the empirical specification reported here only included two lags and leads because additional lags and leads were not statistically significant.

Lag and lead variables for the non-price attributes enter the model specification directly, not as differenced variables. This specification is used because the levels of the non-price attributes are each modeled as binary (0/1); it is not clear, a priori, which levels of the non-price attributes would contribute positive or negative marginal utility.

Given these considerations, indirect utility can now be fully specified:

$$\begin{aligned} U_j &= \sum_k b_k z_{kj} + \lambda bid_j + \sum_k \sum_m \delta_{k,j+m} z_{k,j+m} \\ &+ \sum_m \lambda_{\text{lag_gain},j+m} \text{lag_gain}_{j+m} \\ &+ \sum_m \lambda_{\text{lead_gain},j+m} \text{lead_gain}_{j+m} \\ &+ \sum_m \lambda_{\text{lag_loss},j+m} \text{lag_loss}_{j+m} \\ &+ \sum_m \lambda_{\text{lead_loss},j+m} \text{lead_loss}_{j+m} + e_j. \end{aligned} \quad [7]$$

Note that, in equation [7], the change in utility when the current bid amount changes is not simply the parameter estimate on current bid, λ . If the lag and lead, gain and loss parameters are statistically significant, then

the first derivative of U_j with respect to bid_j is the sum ($\lambda + \sum \lambda_{\text{lag_gain},j+m} + \sum \lambda_{\text{lead_gain},j+m} + \sum \lambda_{\text{lag_loss},j+m} + \sum \lambda_{\text{lead_loss},j+m}$), and the marginal utility of money is a function of the cost of the current alternative and the relative cost of alternatives contained in other choice sets.

Estimates of compensating variation for a specific policy package relative to an alternative package, usually the status quo, is typically calculated as the difference between the inner product of the implicit prices (b_k/λ) and attribute levels for the two policy packages:

$$CV_1 = \left[\frac{\sum_k b_k z_k^1 - \sum_k b_k z_k^0}{-\lambda} \right], \quad [8]$$

where z_k^0 and z_k^1 are the attribute levels for the base and altered-policy scenarios. This specification assumes that lag and lead, gain and loss effects are not present. If bid prices contained in lag and lead choice sets have a statistically significant effect on current choices, then compensating variation is computed by including the relative components of the marginal utility of money:

$$CV_2 = \left[\frac{\sum_k b_k z_k^1 - \sum_k b_k z_k^0}{(-\lambda - \sum \lambda_{\text{lag_gain},j+m} - \sum \lambda_{\text{lead_gain},j+m} - \sum \lambda_{\text{lag_loss},j+m} - \sum \lambda_{\text{lead_loss},j+m})} \right] \quad [9]$$

Of course, it is possible that only selected lag and lead, gain and loss effects are statistically significant.

Identifying Idiosyncratic Effects with a Panel Model

Even after controlling for dependence in the structural components of responses to a sequence of questions, responses may be linked due to correlation (ρ) induced by the permanent, unobserved component of individual preferences. Alberini, Kanninen, and Carson (1997) showed how a random-effects model could be used to identify correlation in the stochastic component

TABLE 1
FOREST MANAGEMENT ATTRIBUTES AND LEVELS

Attribute	Variable Name	Levels
Road density	ROADS	<i>One road every mile</i>
	n.a.	One road every ½ mile
Post-harvest dead and dying trees	n.a.	Remove all
	DEAD5	<i>Leave 5 per acre</i>
	DEAD10	Leave 10 per acre
Post-harvest live trees	n.a.	Remove all > 6" diameter
	LIVE153	153 per acre
	LIVE459	<i>459 per acre</i>
Maximum size harvest	n.a.	< 5 acres
	HOPEN35	35 acres
	HOPEN125	125 acres
Percentage of land available	PERH80	80% harvest, 20% natural
	PERH50	<i>50% harvest, 50% natural</i>
	n.a.	20% harvest, 80% natural
Width of riparian protection	H2OZONE500	At least 500 feet
	n.a.	At least 250 feet
Slash disposal	n.a.	Leave it where it falls
	DSTSLASH	<i>Distribute along skid trails</i>
	REMSLASH	Remove all

Note: Bold letters indicate the status quo level (most common practice) and italics indicate a more environmentally benign practice, for use in estimating Hicksian surplus.

of double-bounded CV questions. Unlike the double-bounded CV method, where the commodity description is held constant in both the first- and second-stage questions, the commodity description changes for each step in the sequence of ABR questions. Because each commodity description faced by respondents is independent of all other commodity descriptions they face (due to the completely randomized experimental design), the identification of a positive, statistically significant ρ parameter would indicate a correlated sequence of YES-YES or NO-NO responses that is independent of the commodity attributes.

III. DATA

The data for our analysis were taken from a forest management study in the state of Maine. The valuation scenarios were structured around the (hypothetical) proposed sale of a tract of forestland (23,000 acres) from a large forest-land management company to the State. Seven forest management attributes were used to differentiate the policy scenarios (Table 1), and attri-

butes were coded using dummy variables. The payment vehicle used was a one-time increase in state taxes, and bid amounts ranged from \$1 to \$1,600. Compensating variation was estimated for purchase of the land, given the status quo level of attributes, and conversion to a more environmentally benign forest management plan.

Information about the attributes was presented to respondents in a booklet that contained line drawings for two levels of each attribute, provided descriptions of positive and negative impacts for each attribute level, and described the current forest management conditions in the study area. In the questionnaire, respondents were quizzed on their understanding of the information contained in the booklet, and were asked to rate each attribute level on a Likert scale. These steps were undertaken to familiarize respondents with the attributes under consideration and force them to think carefully about the attribute levels. They were then asked to answer the preference questions.

The initial sample was composed of 2,500 individuals who were randomly sampled from records of Maine drivers' licenses and

TABLE 2
ATTRIBUTE-BASED REFERENDA PARAMETER ESTIMATES, BY ORDER IN SEQUENCE

Variable	1st Question	2nd Question	3rd Question	4th Question
CONSTANT	-0.079 (0.186)	-0.431** (0.189)	-0.182 (0.184)	-0.589*** (0.190)
ROADS	0.266*** (0.101)	0.074 (0.104)	0.040 (0.100)	0.234** (0.106)
DEAD5	0.008 (0.127)	0.277** (0.128)	0.226* (0.121)	0.3412*** (0.130)
DEAD10	0.084 (0.126)	0.232* (0.126)	0.137 (0.125)	0.230* (0.131)
LIVE153	0.435*** (0.127)	0.486*** (0.127)	0.288** (0.124)	0.435*** (0.128)
LIVE459	0.379*** (0.120)	0.341*** (0.123)	0.292** (0.132)	0.296** (0.130)
HOPEN35	0.059 (0.123)	0.109 (0.125)	-0.102 (0.122)	0.248* (0.130)
HOPEN125	-0.250** (0.123)	0.213* (0.128)	-0.013 (0.125)	0.258** (0.128)
H2OZONE	-0.148 (0.100)	0.125 (0.103)	0.078 (0.100)	(0.118) (0.106)
PERH80	-0.262** (0.123)	-0.425*** (0.127)	-0.573*** (0.122)	-0.376*** (0.127)
PERH50	-0.072 (0.123)	-0.124 (0.122)	-0.148 (0.125)	0.076 (0.126)
REMSLASH	-0.305** (0.123)	-0.203 (0.125)	-0.041 (0.126)	-0.331*** (0.127)
DSTSLASH	0.084 (0.121)	-0.047 (0.134)	0.080 (0.122)	-0.236* (0.127)
λ	-0.00056*** (0.00012)	-0.00077*** (0.00014)	-0.00051*** (0.00012)	-0.00087*** (0.00012)
<i>N</i>	680	668	678	663
Pseudo- <i>R</i> ²	0.070	0.087	0.060	0.110
CV1	\$1,164	\$1,142	\$1,758	\$1,286

Note: standard errors in parentheses.

*** Significant at $p < 0.01$; ** significant at $p < 0.05$; * significant at $p < 0.10$.

state identification cards. The survey was administered as a mail-out, mail-back survey in early 1997. Of the initial sample, 451 had addresses that were undeliverable (18%). A total of 926 surveys were completed and returned for a usable response rate of 45%.

IV. RESULTS

Seven of the eight attributes included in our experiment were statistically different than zero at the 10% significance level or better in at least one of the estimated models for each of the four preference questions, and several attribute levels were significant at the 5% level or better in every equation (Table 2). In general, we found

that respondents were willing to pay a substantial amount to reduce the environmental and aesthetic impacts of timber harvesting practices relative to the status quo. Respondents favored selective harvest systems over the practice of clear-cutting and preferred moderate-intensity (LIVE153) over low-intensity (LIVE459) selective harvest systems. Respondents also preferred leaving some standing dead trees after harvest, a practice that mimics old-growth forest structure. A lower road density was also preferred, which decreases recreational access to the forest, but also reduces forest fragmentation. Finally, respondents preferred leaving the slash resulting from harvest operations in the forest, which benefits soil pro-

TABLE 3
TEST STATISTICS FOR HYPOTHESES REGARDING SCALE PARAMETERS AND
PREFERENCE PARAMETER STABILITY^a

Test	Q1 vs. Q2	Q1 vs. Q3	Q1 vs. Q4
Relative scale $\mu_j = 2,3,4/\mu_j = 1$	1.05	0.825	1.23
Log-L, Q1 data	-436.282	-436.282	-436.2818
Log-L, Q2, Q3, or Q4 data	-416.037	-432.592	-397.2972
Log-L, optimally scaled, pooled data	-862.906	-879.178	-849.017
LR test statistic (χ^2)	21.174	20.609	30.876
Reject H_0 ?: $\beta_j = 1 = \beta_j = 2,3,4$	NO ^b	NO ^b	YES ^b
Log-L, unadjusted pooled data	-862.923	-879.429	N/A ^c
LR test statistic (χ^2)	0.035	0.50	N/A ^c
Reject H_0 ?: $\mu_j = 1 = \mu_j = 2,3,4$	NO ^d	NO ^d	N/A ^c

^a Tests based on Swait and Louviere (1993).

^b χ^2 statistic for 15 d.f. and 95% confidence = 25.00.

^c The Swait and Louviere (1993) method cannot test for identical scale parameters if the null hypothesis that preference parameters are identical has been rejected.

^d χ^2 statistic for 1 d.f. and 95% confidence = 3.84.

ductivity and provides habitat for small animals and insects.

Learning and Preference Parameter Stability

An examination of the parameter estimates in Table 2 shows that the model estimated on responses to the fourth referendum question was more informative than the models estimated on the other responses. Note that the number of preference parameters that were statistically significant at the 10% level or higher was 7 in the equation estimated from responses to the first question, 7 in the second equation, 5 in the third equation, and 11 in the final equation. Also note that the pseudo- R^2 increased from 0.07 in the first equation to 0.11 in the fourth equation. In concert with these findings, Johnson and Desvousges (1997), using a sequence of 26 rated pair questions, concluded that later responses in a sequence may provide better indicators of preference than earlier responses. In contrast, Swait and Adamowicz (2001), using a sequence of multiple-choice questions, found that most attributes were salient up to about the eighth choice, but that in subsequent questions, respondents tended to ignore attributes and relied more heavily on the brand name in order to make choices. Taken together, these results support the proposition that

learning occurs up to a point, but that there may be a threshold where fatigue sets in.

Results from the Swait and Louviere (1993) procedure for testing hypotheses regarding the equality of preference parameters showed that a structural change in preferences occurred during the valuation sequence. The hypothesis that preferences were identical in responses to the first and fourth valuation question was rejected at the 95% confidence level (Table 3). It is interesting to note that preference parameters changed for the fourth question, which is precisely the question for which responses were most informative. These results suggest that responses for question 4 should not be pooled with responses to the prior three questions, because a change occurred in the preference structure. The results in Table 3 also show that the relative scale parameter for question 4 (relative to the scale in question 1) was greater than the relative scale parameter for preceding questions. Although it is not possible to use the Swait and Louviere (1993) procedure to test whether this change in scale is statistically significant, given that the likelihood ratio test indicated that preference parameters were not identical, this result is consistent with the finding that responses to the fourth valuation question were the most informative, and that respondents demonstrated greater discrimi-

TABLE 4
PARAMETER ESTIMATES FOR MODELS WITH AND WITHOUT CONTEXT-DEPENDENT PARAMETERS

	Rescaled Q1-Q3	Rescaled Q1-Q3 with Context-Dependence	Q4 with Context-Dependence
CONSTANT	-0.257*** (0.103)	-0.166 (0.124)	-0.826*** (0.202)
ROADS	0.155*** (0.062)	0.144** (0.064)	0.231** (0.107)
DEAD5	0.187** (0.082)	0.183** (0.084)	0.396*** (0.131)
DEAD10	0.197** (0.079)	0.191** (0.081)	0.249* (0.132)
LIVE153	0.433*** (0.081)	0.449*** (0.083)	0.464*** (0.130)
LIVE459	0.367*** (0.077)	0.367*** (0.079)	0.326*** (0.132)
HOPEN35	0.008 (0.078)	0.007 (0.081)	0.255** (0.131)
HOPEN125	-0.040 (0.078)	-0.033 (0.082)	0.254** (0.129)
H2OZONE	0.028 (0.063)	0.032 (0.065)	0.121 (0.107)
PERH80	-0.443*** (0.073)	-0.473*** (0.076)	-0.410*** (0.129)
PERH50	-0.104 (0.076)	-0.091 (0.077)	0.070 (0.127)
REMSLASH	-0.237*** (0.078)	-0.245*** (0.081)	-0.342*** (0.129)
DSTSLASH	0.062 (0.076)	0.061 (0.079)	-0.229* (0.129)
λ	-0.00064*** (0.00007)	-0.00050* (0.00029)	-0.00074* (0.00045)
ROADS _{j+1}	-	-0.122** (0.058)	-
LIVE153 _{j+1}	-	-0.181*** (0.064)	-
DEAD10 _{j-1}	-	-0.170** (0.078)	-
H2OZONE _{j-1}	-	-0.149** (0.071)	-
HOPEN125 _{j-1}	-	-	0.254** (0.113)
$\lambda_{lead_loss,j+1}$	-	0.000009 (0.00019)	-
$\lambda_{lag_loss,j-1}$	-	-0.00034* 0.00018	-0.000014 (0.00040)
$\lambda_{lead_gain,j+1}$	-	0.00027*** (0.00008)	-
$\lambda_{lag_gain,j-1}$	-	0.00015* (0.000089)	0.00032** (0.00013)
$\lambda_{lead_loss,j+2}$	-	0.00006 (0.00027)	-
$\lambda_{lag_loss,j-2}$	-	0.00036 (0.00025)	-0.0.000029 (0.00035)
$\lambda_{lead_gain,j+2}$	-	0.00016 (0.00010)	-
$\lambda_{lag_gain,j-2}$	-	0.00034*** (0.00013)	0.00035** (0.00015)
ρ	0.104*** (0.035)	0.108*** (0.036)	-
N	2,026	2,026	663
Pseudo-R ²	0.061	0.080	0.128
CV1	\$1,477	\$1,985	\$1,629
CV2	-	\$754	\$829

Note: standard errors in parentheses.

*** Significant at < 0.01; ** significant at < 0.05; * significant at < 0.10.

nation between alternatives in the final question of the sequence.

Context-Dependence

Rescaled data were pooled for the cases where the hypothesis of parameter equality could not be rejected (questions 1, 2, and 3) and panel models were estimated to test for context-dependence. A context-dependent model specification was also estimated for question 4 alone.

The results of the tests for context-dependence provide abundant evidence that attribute levels of alternatives offered in lag and lead choice sets influenced current choices

(Table 4). This result holds for both price and non-price attributes. Examining the parameter estimates for gains and losses relative to lag and lead bid amounts (as defined in equations [6a] through [6d]), it is seen that bid prices had both direct (own bid) and indirect (lead and lag bids) effects on responses. As anticipated, results show that the bid price for the current policy scenario always had a negative impact on likelihood of purchase (the direct effect). In cases where the current bid amount was greater (less) than lagged or lead bid amounts, the likelihood of purchasing the current policy scenario decreased (increased). Significant parameter estimates on gains (i.e., current

bid price being less than lead and lag bid amounts) were generally more prevalent than significant parameter estimates on losses (i.e., current bid price being greater than lead and lag bid amounts.) In the one case where a significant parameter estimate was identified for losses ($\lambda_{lag_loss,j-1}$), the magnitude of the parameter was not much different than the statistically significant gain parameters ($\lambda_{lag_gain,j-1}$, $\lambda_{lag_gain,j-2}$, $\lambda_{lead_gain,j+1}$). These results do not provide much support for a prospect theory interpretation, which would hypothesize that losses are weighted more heavily than gains. However, the prevalence of significant parameter estimates on the gain variables ($\lambda_{lag_gain,j-1}$, $\lambda_{lag_gain,j-2}$, $\lambda_{lead_gain,j+1}$) suggests that respondents preferred policy alternatives with lower tax prices. This type of “comparison shopping” may mimic market behavior, where people seek bargains. In a policy context, respondents may be seeking cost minimizing alternatives—a process that may be used to justify their choices to themselves and others. The results (in terms of the number of significant parameter estimates) also indicate that respondents were more likely to be backward-looking than forward-looking when considering information from alternatives other than the current question they were answering. This result is consistent with the Tversky and Simonson (1993) model that emphasizes the conditioning effect of prior alternatives on current choice.

The indirect effects of lag and lead, price and non-price attributes on current choices made by survey respondents affected estimates of compensating variation in two ways (Table 4). First, including lag and lead, price and non-price attributes induced changes in the parameter estimates on the vector of policy attributes. Of particular importance for welfare estimation, the parameter estimates for the marginal utility of money were found to decrease, resulting in a modest increase in the estimates of compensating variation (CV1) relative to the under-specified models (Table 2). Second, adjusting the marginal utility of money to incorporate both the direct and indirect effects, as shown in equation [9], caused an overall increase in the mar-

ginal utility of money, resulting in large reductions in the estimates of compensating variation (CV2) relative to both CV1 and the under-specified models.

Evidence of context-dependence was also identified for five of the non-price attributes. Although all lead and lag attribute levels were tested for significance in current choices, the models shown in Table 4 only report significant parameter estimates (to simplify the presentation). Respondents were somewhat more likely to be backward-looking than forward-looking when considering information about non-price attributes from alternatives contained in other choice sets.

Preference Heterogeneity

As shown in Table 4, the panel models identified a positive, statistically significant ρ parameter (at the 0.01 level). Although the correlation across intra-individual responses was weak, these results demonstrate that sequences of intra-individual responses were not independent even though sequences of commodity descriptions were randomly designed. Unobserved respondent heterogeneity apparently contributed a degree of persistence to preference question responses by shifting individual indirect utility above or below mean values, irrespective of policy attribute levels. One explanation for this result is that a stochastic component of individual utility is associated with the underlying policy proposal, that is, with the proposal to transfer forest land from private ownership to the public.

V. CONCLUSIONS

The attribute-based referenda format is a promising approach for the valuation of environmental goods. This method combines an incentive compatible response format with a set of attributes that are under the control of managers or policymakers. This approach increases the informational efficiency of stated-preference surveys in that values can be computed for a variety of policy scenarios. For natural resource managers faced with the problem of evalu-

ating the costs and benefits of a suite of multi-attribute resources and policy packages, the attribute-based approach offers clear advantages.

However, this research shows that, in stated-preference experiments using a sequence of preference questions, dynamic effects must be considered. Two factors are highlighted in the analysis reported here. First, an increase in the scale parameter across the WTP sequence, indicating an increase in attribute discrimination, and an increase in the number of salient attributes, suggests that respondents learn about the task and how to formulate responses as they proceed through a series of valuation questions. The fact that the scale-adjusted preference parameters for the final WTP question were statistically different from the preference parameters for the first WTP question indicates that these preference data should not be pooled and that later responses may be more informative than earlier responses. Second, the results reported here suggest that survey respondents learn, to some degree, about their preferences for policy attributes by comparing price and non-price attributes across choice sets. This result is a violation of Property α and clearly demonstrates context-dependence.

The implication that consumers make inferences about the value of a good by juxtaposing it with other, similar goods bears some resemblance to the concept of anchoring as used in psychology and economics. An anchor point introduces non-essential information into a decision problem that causes a respondent to adjust their prior value toward the anchor (e.g., Herriges and Shogren 1996). In the cases of context-dependence and anchoring, information presented in the context of a specific valuation question (in the one instance—lag/lead alternatives, in the other—bid amounts) induces a rational inference about the value of the good. In concert with this view, McFadden (1999) refers to context-dependence and anchoring (as well as framing, prominence, and saliency) as occurring within a class of cognitive anomalies he labels “context effects.”

Researchers designing experiments with attribute-based commodity descriptions that

utilize a sequence of WTP questions need to be aware of the potential for dynamic response effects. One obvious solution to this problem is to ask a single preference question, although this approach would reduce the statistical efficiency sought by using a sequence of valuation questions. Moreover, if learning does occur, then response to a single question may not provide the best preference information. This leaves researchers facing the age-old issues of statistical efficiency and bias. A second approach, consistent with the models presented here, is to identify new variables and include them as additional independent variables in the specification of econometric models (Machina 1999; De Shazo 2002). Perhaps the most ambitious approach would be to develop a theory of choice that includes context effects as a component of rational behavior (Sen 1993, 1997), and then develop survey designs and empirical models consistent with such a theory (McFadden 1999). Given the popularity of attribute-based stated preference models, the prevalence of including sequential choices in the experimental design, and the increasing importance of using public preference information in the design of environmental programs, the development of such methods is warranted so that valuation researchers may detect underlying preferences despite the vagaries of human decision processes.

References

- Adamowicz, W., J. Louviere, and M. Williams. 1994. “Combining Stated and Revealed Preference Methods for Valuating Environmental Amenities.” *Journal of Environmental Economics and Management* 26 (3): 271–92.
- Adamowicz, W., J. Louviere, and J. Swait. 1998. *Introduction to Attribute-Based Stated Choice Methods*. Final Report to National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce. Washington, D.C: National Oceanic and Atmospheric Administration.
- Alberini, A., B. Kanninen, and R. T. Carson. 1997. “Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data.” *Land Economics* 73 (Aug.): 309–24.

- Ben-Akiva, M., and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Mass.: MIT Press.
- Ben-Akiva, M., T. Morikawa, and F. Shiroishi. 1992. "Analysis of the Reliability of Preference Ranking Data." *Journal of Business Research* 24 (2): 149–64.
- Bennett, J., and R. Blamey. 2001. *The Choice Modeling Approach to Environmental Valuation*. Northampton, Mass.: Edward Elgar.
- Boyle, K. J. 2003. "Contingent Valuation in Practice." In *A Primer on Non-Market Valuation*, ed. P. A. Champ, K. J. Boyle, and T. C. Brown. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- DeShazo, J. R. 2002. "Designing Transactions without Framing Effects in Iterative Question Formats." *Journal of Environmental Economics and Management* 43 (3): 360–85.
- DeShazo, J. R., and G. Fermo. 2002. "Designing Choice Sets for Stated-preference Methods: The Effects of Complexity on Choice Consistency." *Journal of Environmental Economics and Management* 44 (1): 123–43.
- Greene, W. H. 1997. *Econometric Analysis*. 3rd. ed. Upper Saddle River, N.J.: Prentice Hall.
- Herriges, J. A., and J. F. Shogren. 1996. "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning." *Journal of Environmental Economics and Management* 30 (1): 112–31.
- Holmes, T. P., and W. L. Adamowicz. 2003. "Attribute-Based Methods." In *A Primer on Non-Market Valuation*, ed. P. A. Champ, K. J. Boyle and T. C. Brown. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hsiao, C. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.
- Huber, J., J. W. Payne, and C. Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9 (June): 90–98.
- Johnson, F. R. and W. H. Desvousges. 1997. "Estimating Stated-preferences with Rated-Pair Data: Environment, Health, and Employment Effects of Energy Programs." *Journal of Environmental Economics and Management* 34 (1): 79–99.
- Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica* 47 (2): 263–91.
- Lancaster, K. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2): 132–57.
- Louviere, J. J., R. J. Meyer, D. S. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher, and J. Irwin. 1999. "Combining Sources of Preference Data for Modeling Complex Decision Processes." *Marketing Letters* 10 (3): 205–17.
- Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. *Stated Choice Methods: Analysis and Application*. New York: Cambridge University Press.
- Machina, M. J. 1999. "A Challenge to the 'Econoclasts': A Commentary on 'Rationality for Economists?'" *Journal of Risk and Uncertainty* 19 (1–3): 107–8.
- McFadden, D. 1999. "Rationality for Economists?" *Journal of Risk and Uncertainty* 19 (1–3): 73–105.
- Mitchell, R. C., and R. C. Carson. 1989. *Using Surveys to Value Public Goods—The Contingent Valuation Method*. Washington, D.C.: Resources for the Future.
- Sen, A. 1993. "Internal Consistency of Choice." *Econometrica* 61 (3): 495–521.
- . 1997. "Maximization and the Act of Choice." *Econometrica* 65 (4): 745–79.
- Simonson, I., and A. Tversky. 1992. "Choice in Context: Tradeoff Contrast and Extremeness Aversion." *Journal of Marketing Research* 29 (August): 281–95.
- Swait, J., and J. Louviere. 1993. "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research* 30 (Aug.): 305–14.
- Swait, J., and W. Adamowicz. 2001. "The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching." *Journal of Consumer Research* 28 (June): 135–48.
- Tversky, A., and D. Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics* 106 (4): 1039–61.
- Tversky, A., and I. Simonson. 1993. "Context-dependent Preferences." *Management Science* 39 (10): 1179–89.