A Pseudo-Sequential Choice Model for Valuing Multi-Attribute Environmental Policies or Programs in Contingent Valuation Applications

Dmitriy Volinskiy, John C. Bergstrom, Christopher M. Cornwell, and Thomas P. Holmes

The assumption of independence of irrelevant alternatives in a sequential contingent valuation format should be questioned. Statistically, most valuation studies treat nonindependence as a consequence of unobserved individual effects. Another approach is to consider an inferential process in which any particular choice is part of a general choosing strategy of a survey respondent. A stochastic model is suggested, consistent with the reflexivity, transitivity, and continuity axioms of utility analysis. An application of this theoretical model to the valuation of watershed ecosystem restoration demonstrates that an empirical model recognizing reflexivity and transitivity, and also allowing for continuity, shows the highest in-sample predictive ability.

Key Words: contingent valuation, sequential choices, modeling approaches, watershed ecosystem service valuation

Modern-day environmental policies or programs, such as watershed ecosystem restoration, are designed to improve multiple ecosystem services and consist of multiple components or parts. The valuation of such policies or programs should address the multi-dimensionality of the problem. A hybrid of the contingent valuation method (CVM) and attribute-based analysis (Holmes and Adamowicz 2003) is often used. Several related policy options are included in the survey, which are valued in a sequential manner. An example of such a sequence may be valuing a bare-bones policy first and subsequently augmenting it with more attributes or higher levels of the already included ones, building up to the most comprehensive package of management actions.

When multiple items are valued using the dichotomous choice format, a binary discrete-response data set with a sequence of observations per individual is generated. The sequential nature of the choice gives rise to concerns that the probability of observing the choice outcome for a particular policy option may depend on observable or unobservable components of other choice options in the survey (Holmes and Boyle 2005). The dependence due to anchoring and framing effects, both related to the monetary bid, has received some attention (Herriges and Shogren 1996, DeShazo 2002). Giraud, Loomis, and Johnson (1999) provide evidence of sequencing and instrument context effect in sequential valua-

Dmitriy Volinskiy is Postdoctoral Fellow in the Department of Rural Economy at the University of Alberta in Edmonton, Alberta, Canada; John C. Bergstrom is Russell Professor of Public Policy and Professor in the Department of Agricultural and Applied Economics at The University of Georgia in Athens, Georgia; Christopher M. Cornwell is Department Head and Professor in the Department of Economics at The University of Georgia in Athens, Georgia; and Thomas P. Holmes is Research Social Scientist at the Southern Research Station of the USDA Forest Service in Research Triangle Park, North Carolina. This research was supported by the USDA Cooperative State, Research, Education, and Extension Service (CSREES) under the W2133 Regional Research Project entitled, "Benefits and Costs of Natural Resources Policies Affecting Public and Private Land," and the USDA Forest Service Southern Research Station.

tion. Holmes and Boyle (2005) explain the dependence by generalizing the notion of anchoring to include anchoring to the context of a specific valuation question, i.e., information in other choice sets.

In this paper, we argue that, should the econometric investigator choose to adhere to canons of utility theory, choices in the sequential valuation format cannot be considered independent if they are conditioned on concurrent observables only. Further, a particular composition of the entire sequence entails a particular pattern of their dependence. The remainder of the paper is organized as follows. A general conceptual valuation model is developed in the next section. On the grounds of dynamic consistency we argue that, as long as the commodity information the respondent possesses remains unchanged, the exact, albeit unobservable, utility levels attainable at all restoration programs involved should be thought of as the same throughout the valuation process. This conjecture leads to the equivalence of the sequential and simultaneous elicitation formats and makes the model consistent with the utility reflexivity and transitivity axioms. We further build on this reasoning and posit that, for the utility continuity axiom to be maintained, the degree of dependence between the utility shocks in any pair of items should increase as the items get closer together attribute-wise. This constitutes the main research hypothesis of the study.

Following the conceptual model section, we introduce the specifics of survey data for the Little Tennessee River watershed empirical application. We provide several alternative stochastic specifications for our valuation model. Model parameters estimated by maximum likelihood are presented and discussed. We discuss empirical evidence in support of the continuity hypothesis. Willingness to pay (WTP) values for restoration program components based on median voter equilibrium are presented and compared to the results from an earlier CVM study with the same data. The paper concludes by discussing the ability of our model to produce economically and statistically valid welfare change estimates from data generated by the sequential, multiple valuation question per respondent format. Further enhancements are also discussed.

Conceptual Model

Consider this admittedly contrived example. The investigator is interested in comparing the agent's preferences between three hypothetical states of the world yielding deterministic utility levels V_0 , V_1 , V_2 . One way to elicit the preferences is to let the agent pick the preferred state from all possible pairs (three in this case). An alternative is to ask the agent to rank the three states at once. From an economic theory perspective, choosing the latter simultaneous format over the former sequential one or vice versa is immaterial as long as preferences remain unchanged. But it is not so when it comes to empirical modeling. If choices are arranged in pairs, the following random utility model (RUM) arises:

$$
\begin{aligned}
&(1) \qquad U_{jt} = V_j + \varepsilon_{jt} \\
&\qquad\quad U_{kt} = V_k + \varepsilon_{kt},
\end{aligned}
$$

where (j, k) are (1,0), (2,0), (2,1) respectively for t = 1, 2, 3; V represents the deterministic components of the respective random utility levels, and ε represents utility shocks. There are six random quantities involved in the comparison and there are eight possible outcomes, of which two are intransitive. There are two distinct utility realizations for each of the three states, which means the choice generally is not reflexive. Meanwhile, there are only three random quantities in the simultaneous ranking format, and there are six guaranteed transitive and reflexive outcomes. As a result, estimates from the simultaneous model will differ systematically and to an unknown extent from those coming from the sequential model, even if the deterministic parts are identically specified.

Let us now switch to a choice experiment setup that has more practical relevance. Consider this common T-period sequential binary choice model. A utility maximizing agent i chooses between two states of the world at each period t; t = 1, ..., T, in a sequence. These states are a period/individual-specific "alternative" (e.g., a particular environmental policy) and the "status quo," e.g., a no-action baseline policy. To simplify notation, we will omit the individual subscript i; however, it is important to keep in mind that we discuss choosing behavior of the *same* individual. These are the utility levels involved in the paired comparisons at each period t:

(2) $$U_{jt} = V_{jt} + \varepsilon_{it}$$
$$U_{0t} = 0 + \varepsilon_{0t},$$

where $V_{jt} = V(x_{it})$ is the deterministic utility of the alternative with attributes x_{jt} ; the deterministic utility of the status quo is zero; and $(\varepsilon_{it}, \varepsilon_{0t})$ are the respective error terms. The above model implies the following marginal probabilities of choice outcomes:

(3) $$\Pr[U_{jt} > U_{0t}] = \Pr[V_{jt} > \varepsilon_{0t} - \varepsilon_{jt}]$$
$$= F_{\varepsilon_{0t}-\varepsilon_{jt}}(V_{jt}),$$

where $F_{\varepsilon_{0t}-\varepsilon_{jt}}$ is the distribution function of the difference of utility shocks at time t.

The standard practice is to use 2 x T independently and identically distributed (i.i.d.) errors (Hoehn 1991). As already mentioned, this leads to the nonequivalence of elicitation formats and potential problems with transitivity and reflexivity. In their review of statistical methods with CVM data, Hanemann and Kanninen (1999) consider a model where $\varepsilon_{0t} = \varepsilon_0$, $\forall t$. The authors do not put forward any justification for this restriction, but technically the restriction means that an unobserved utility level of the "status quo" state is the same no matter where in the survey this state is invoked, i.e., $U_{0t} = U_0$, $\forall t$. To be consistent, one should extend this to all states: if any pair of states (x_{js}, x_{jt}), $s \neq t$, are identical so that these describe the same alternative j that individual i faces at two different points in time, then $U_{js} = U_{jt}$, $\forall(s, t)$, $s \neq t$. It is apparent that, if all T alternative states are different, this specification restricts the number of latent random quantities to $T + 1$ and statistically forces the equivalence of the sequential and parallel choice representations from our earlier example. We will term this model the *pseudo-sequential choice*, to emphasize its atemporal nature.

The pseudo-sequential structure of a choice model makes the latter consistent with the transitivity and reflexivity axioms, so that the agent is assumed to be able to order her preferences among policy alternatives in a consistent manner, with no preference reversal allowed. Another refinement can make the model consistent with the continuity axiom as well. Loosely stated, the principle of continuity postulates that any two states that are infinitely close cannot be far apart in terms of their respective utility levels. Considering environmental policies as bundles of services to the consumer, continuity is critically important, for it allows for the possibility of substitution between policy components, which, in turn, permits comparing the relative importance of these components. To deal with continuity formally, let metric d be the Euclidean distance between two sets of attributes x_{js} and x_{kt}. Continuity will then require that, as $d(x_{js}, x_{kt})$ becomes infinitesimally small, the error terms should become perfectly statistically dependent, in order to maintain our earlier assumption of a single utility level per alternative, $U_{js} = U_{jt}$, $\forall(s, t)$, $s \neq t$. Using Pearson's product-moment coefficient (the standard correlation coefficient) $\rho(\varepsilon_{js}, \varepsilon_{kt})$ to measure the departure of two random variables from independence, as, $d(x_{js}, x_{kt}) \to 0$, $\rho(\varepsilon_{js}, \varepsilon_{kt}) \to 1$. This should also hold true when using any other measure of dependence among random variables. Notably, unlike transitivity, any sequential model with i.i.d. error terms will always violate continuity.

The (above) assumptions behind the pseudo-sequential choice model have an important consequence. As all utility realizations of an individual can potentially be interdependent no matter how they are arranged in choice sets, the property of independence of irrelevant alternatives can no longer be maintained.

As far as the conventional utility theory goes, there does not seem to be a reason not to have the canons of utility theory enforced, bearing in mind also that, if the model is supposed to produce WTP estimates, those can only be considered valid for as long as the utility theory behind them can. But it cannot be presupposed that choosing agents comply with the investigator's theoretical assumptions. Relative to the point in hand, theories of rational dynamic choice generally uphold dynamic consistency. The agent should be dynamically consistent in her actions so that, if the agent's present "self" embarks on a course of action, all later "selves" should abide by that commitment (McClennen 1990).

Dynamic consistency has a timing invariance property: a sequential choice problem and a planned choice problem should be equivalent to the agent, given they are strategically equivalent. Behavioral studies do not seem to offer any definitive results with respect to timing invariance (Read et al. 2001). Read and Loewenstein (1995) consider an undesirable "diversification bias," referring to a demonstrated excess variety in items selected in the simultaneous design. In contrast, Read, Loewenstein, and Rabin (1999) argue that simultaneous choice enables agents to diversify their assets to reduce the overall risk, thus giving preference to the simultaneous choice. Information that becomes available in between choices may also be a factor to consider.

Participants of CVM experiments are likely to have no experience with programs or policies to be valued. Different ways of supplying commodity-related information or different amounts of information supplied have led to a significant variation in valuation results (Bergstrom et al. 1989). Arguably, the time dimension and sequencing of choice sets can only be omitted in situations where information about the programs is supplied to respondents strictly prior to elicitation, and no additional information is given in between elicitation questions.

Continuity in an empirical context has its own fair share of concerns. If a small change in the scope or attributes of a commodity led to considerable changes in utility, CVM responses would have tended to demonstrate hypersensitivity to such changes. However, more studies seem to be suffering from abnormally low sensitivity, which has led to a broad scholarly interest in the phenomena of embedding and scope insensitivity (Kahneman 1986, McFadden and Leonard 1995). There is also evidence of lexicographic preferences in environmental amenity valuation; see the review of studies in Spash (2000).

A pragmatic approach for the investigator, to address both theoretical requirements and empirical concerns, would be to fit several models (e.g., the standard sequential choice model and a pseudo-sequential choice model) and then go with the one that provides the best fit for the data. Nonetheless, the pseudo-sequential choice framework, as outlined above, has a testable implication: the investigator should be able to observe whether the estimated dependence between error terms increases as the choice options become more similar, and vice versa.

For the T-period binary choice setup with no repeated choices, our pseudo-sequential choice model for actions of individual i is given by:

$$\text{(4)} \qquad \begin{aligned} U_t &= V_t + \varepsilon_t \\ U_0 &= \varepsilon_0 \end{aligned}$$

where t now indexes T alternative choice options, $t = 1, 2, \ldots, T$, and the joint distribution of $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T)$ may have $\binom{T+1}{2}$ parameters of dependence for all possible pairs of shocks. If the measure of dependence is bounded, as is the case with the correlation, then normalization is required. One should select at least one pair, either actual or imaginary, for which no dependence is allowed. We suggest setting dependence to zero for all $(x_t, 0)$ pairs, since the baseline option is by default most different from the rest of the policies. This results in the availability of $\binom{T}{2}$ dependence parameters.

To estimate the model in equation (4) by maximum likelihood, one should be able to obtain probabilities of all 2^T choice outcomes. Getting the outcome probabilities for a two-period model $T = 2$ is fairly straightforward:

(5)

$$\Pr[U_1 < U_0, U_2 < U_0] = \Pr[U_0 = \max(U_0, U_1, U_2)]$$

$$\Pr[U_1 > U_0, U_2 < U_0] = \Pr[U_2 < U_0] - \Pr[U_0 = \max(U_0, U_1, U_2)]$$

$$\Pr[U_1 < U_0, U_2 > U_0] = \Pr[U_1 < U_0] \quad \Pr[U_0 = \max(U_0, U_1, U_2)]$$

$$\Pr[U_1 > U_0, U_2 > U_0] = 1 - \Pr[U_1 < U_0, U_2 < U_0]$$

$$- \Pr[U_1 > U_0, U_2 < U_0]$$

$$- \Pr[U_1 < U_0, U_2 > U_0].$$

For larger numbers of periods, outcome probability formulae become unwieldy, which implies having quite a complex likelihood function. While algebraic expressions for outcome probabilities grow prohibitively complex, the computation of those is easily automated, using the fact that the outcome probability for a subset of

$$\mathrm{Y} = \{ y_1 = 1[U_1 > U_0],\ y_2 = 1[U_1 > U_0],\ \ldots, y_T = 1\,[U_T > U_0] \}$$

can be expressed as a sum of probabilities of the mutually exclusive joint events that constitute it. All that is needed is a facility to calculate

$$\Pr[U_0 = \max(U_0, U_A)],$$

where U_A are the utility levels of options in the subset A of indices $\{1, 2, \ldots, T\}$. The objective now is to solve a linear system $A_p = b$ of 2^T equations. Let T_r be a set of $\binom{T}{r}$ unique ordered combinations of subscripts, $r \in \{1,2,\ldots, T\}$, and refer to the algorithm presented in the Appendix. In the next section, we provide specifications for both deterministic and stochastic parts of the model. We then fit several alternative specifications with actual survey data to assess the reasonability of our conjectures about an agent's rationality.

Empirical Application: Estimating WTP for Little Tennessee River Management Alternatives

The Little Tennessee River (LTR) watershed is located in Georgia, North Carolina, and Tennessee. The watershed encompasses 10,783 acres, including eighteen rivers and streams and twenty-six lakes. The LTR watershed is used by logging, agriculture, and mining industries; however, the aesthetically pleasing environment in the basin has brought about a tremendous increase in the population of people who visit or live within the watershed. In the last twenty years the population has doubled, leading to concerns about the future health of the watershed and the ecosystem services the watershed provides. The majority of land within the watershed is privately owned, and private land use decisions have a major impact on ecosystem functions and services. For example, agricultural activities such as watering cattle in streams, as well as housing and commercial developments along the streams and creeks, influence water quality, a key determinant of ecosystem health and services.

The objectives of a recent CVM study by Holmes et al. (2004) were to develop and test a general methodology for valuing ecosystem services and to identify and value particular ecosystem services present in the Little Tennessee River watershed. To place a value on ecosystem services, a computerized CVM survey instrument was designed and implemented. The present study uses the data set obtained through the above survey.

Four focus group sessions were conducted in the study area to facilitate the development of the computerized survey instrument. Concern about the ability of respondents to distinguish between different restoration programs led to the development of a matrix showing the level of ecosystem services associated with each program. The computerized survey format also permitted the extensive use of photographs and diagrams demonstrating restoration activities, as well as land use maps depicting land use change and the proximity of economic development to the LTR and its tributaries.

Survey panels were held in the study area using locally available computer labs, and each individual who participated in the final survey received an incentive payment. The citizen valuation panel was a non-probability sample made up of recruits from local civic organizations. Although a quota system was not used for recruiting, an attempt was made to recruit a diverse set of citizens to make up the panel. A comparison of socio-economic characteristics of the sample and the county showed that the income and education of the sample were higher than the values for the population (which is not uncommon for probability samples). The age and gender of the sample were quite similar to population values.

The sample included a larger proportion of people who owned property along the LTR than occurred in the general population.

The survey followed a close-ended, single-bounded format. Valuation questions were posed in the "take it or leave it" way: "If a local county sales tax were to reduce your annual household income by *$BID* each year for the next ten years to support Program *t*, would you vote in favor of it?" Fifty-eight respondents ($N = 58$) provided complete sequences of votes in the survey.

The survey included four different programs ($T = 4$). Program 1 offered an overall watershed protection plan, whereby buffer strips along all small streams and creeks running into the LTR would be created. Programs 2–4 included partial restoration of the stream bank along a 20-mile stretch of the LTR, in addition to the omnipresent watershed protection plan. The suggested scope of the restoration was 2 miles in Program 2, 4 miles in Program 3, and 6 miles for Program 4.

The computer-assisted bidding followed a simple adaptation structure. If the respondent had voted in favor of Program $\tau = 2, 3$, then the bid for Program $\tau + 1$ was increased, otherwise $\tau + 1$ was offered at the same bid amount as τ. The initial bids were randomly selected from the amounts \$1, \$5, \$10, \$50, or \$75. Bid amounts for Program 4 ranged from \$1 (resulting from a string of prior NO responses) to \$500 (resulting from a string of prior YES responses). The payment vehicle used represented an increase in the local sales tax (in the study area, local sales taxes must be approved by a public referendum and are a common means of financing local public goods and services).[1]

The conditional indirect utility function we use for this study is a linear combination of weighted policy attributes and the bid:

(6)
$$V_{it} = -\gamma_i\, BID_{it} + \beta_{wp} wp_t + \beta_{2m} 2m_t + \beta_{4m} 4m_t + \beta_{6m} 6m_t ,$$

where BID_{it} is an amount in \$100s, asked from respondent i for Program t, and $(wp_t,\ 2m_t,\ 4m_t,\ 6m_t)$ are indicators for attributes of the program. $wp_t = 1$ indicates the presence of the watershed protection plan, and $qm_t = 1$ indicates that the program provides for the restoration of a q-long stretch of the river, $q = 2$, 4, and 6 miles.

This specification admits an arbitrary dependence of utility on miles restored. To account for heterogeneity amongst respondents, we allow the coefficient t on bid, γ_i, to be varying across the panel. It is assumed to follow log-normal distribution with parameters μ and σ^2 to be estimated. Solving $V_t = 0$ for the bid value yields the compensating surplus welfare change measure for Program$_t$ as the ratio of the implicit price of its attributes to that of \$100 of extra income (Hanemann 1984):

(7)
$$WTP_t = \frac{\beta_{wp} wp_t + \beta_{2m} 2m_t + \beta_{4m} 4m_t + \beta_{6m} 6m_t}{\gamma_i} .$$

One option for estimating the four-period system is a multivariate normal distribution of utility shocks. It offers a general covariance structure and, accordingly, a full range of values for the dependence parameters, from independence to the perfect positive/negative correlation. Unfortunately, choice probabilities from a probit-type model are not closed-form expressions and must be simulated. Simulation is very computationally expensive and may result in a large variation of likelihood values when the sample size is small. Another candidate is a generalized extreme value (GEV) distribution. Despite being more restrictive in comparison with multivariate probit models, GEV models still allow sufficient flexibility. More importantly, GEV choice probabilities are directly computable, which substantially reduces the computational load and saves one from other problems related to simulation-assisted estimation.

Consider a GEV distribution that underlies the paired combinatorial logit (PCL) (Chu 1989):

(8)
$$F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J) = \exp\left[-G\left(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \ldots, e^{-\varepsilon_J}\right)\right]$$
$$= \exp\left[-G\left(a_1, a_2, \ldots, a_J\right)\right] ,$$
$$G = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left(a_k^{1/\lambda_{kl}} + a_l^{1/\lambda_{kl}}\right)^{\lambda_{kl}},$$

[1] For an additional description of the survey, see Holmes et al. 2004.

where J is the total number of options. Each $(k, l: k \neq l)$ pair of error terms in this distribution forms a nest, with the total number of nests equal to $\binom{J}{2}$ and λ_{kl} being a measure of independence for the members to the respective nest. When $\lambda_{kl} = 1$, members of the nest do not exhibit any dependence; when $\lambda_{kl} \to 0$, the dependence becomes perfect. The distribution thus provides the dependence parameters that meet our needs. Besides, if one set $\lambda_{kl} \equiv 1, \forall k, l,$ this GEV model reduces to a multinomial logit (MNL).

In our case $J = T+1$. Since the status quo option is assumed to be different from the others to the utmost extent, we restrict $\lambda_{0t} = 1$, $t = 1, \ldots, T$; that is, we will not allow any covariation between the error term of the status quo and those of the alternative options. This restriction conforms with a PCL identification requirement to have at least one λ set to unity. The option of "no action" appears to be the most different from any available course of action (i.e., restoration) because of its very nature. Human decision-making is often done in stages. A decision whether or not to act, to intervene, normally precedes any choice of the way to act. A brief analysis of the Little Tennessee River data shows that more than 50 percent of subjects either supported all programs or rejected them all. It would then be reasonable to speculate that a significant percentage of the sample was eager to contribute to the restoration—they may have made a choice to intervene—while a significant percentage of our respondents simply might have been not interested in restoring the watercourse at all. Hence the "no action" option being different from all, as one may speculate on similar courses of action. The restriction also has a useful consequence: the model becomes the standard binary logit for any cross-section.

Using the PCL choice probability formula,

$$\Pr[U_{it} = \max(\mathbf{U}_i)] = \frac{\sum_{j \neq t} e^{V_{it}/\lambda_{tj}} \left(e^{V_{it}/\lambda_{tj}} + e^{V_{ij}/\lambda_{tj}}\right)^{\lambda_{tj}-1}}{\sum_{k=0}^{T-1} \sum_{l=k+1}^{T} \left(e^{V_{ik}/\lambda_{tj}} + e^{V_{il}/\lambda_{kl}}\right)^{\lambda_{kl}}}, \tag{9}$$

where $\mathbf{U}_i$ is a collection of the individual i utility realizations, and the algorithm in the Appendix, one can apply the regular maximum likelihood to estimate parameters in V_{it} and all λ.

The adaptive nature of the bid generation leads to the endogeneity of *BID* for Programs 2 and 3. It is important to emphasize, however, that since outcome probabilities are obtained in the simultaneous choice framework, it is equivalent to conditioning the probabilities on all values of *BID* for a given individual, which makes the endogeneity of *BID* immaterial.

Table 1 summarizes the three versions of the model that we estimated with the specification of V_{it} given by equation (6). The PCL specification applies no restrictions to the model in equations (4) and (9); that is, the pseudo-sequential choice framework is used to ensure transitivity, and six dependence parameters are estimated to see whether they are related to changes in the mileage of riverbank restoration in the manner hypothesized in the previous section. The MNL specification is also built on the pseudo-sequential choice framework but it excludes any dependence amongst utility shocks, so that choice probabilities are obtained from MNL. This model roughly corresponds to the one suggested by Hanemann and Kanninen (1999). Finally, the logit

Table 1. Three Model Specifications

Mnemonic	Description
PCL	PCL choice probabilities; λ_{st} unrestricted
MNL	PCL/MNL choice probabilities; $\lambda_{st} = 1, \forall (s,t)$
Logit	All errors are *i.i.d.* standard Gumbel

specification is simply a panel logit model with a random price coefficient, which addresses neither utility theory axiom. There are 8 *i.i.d.* standard Gumbel errors in this specification, 2 for each of 4 pairs of choices. Logit was chosen as a mainstream discrete-choice model. All three models reduce to binary logit for any cross-section.

Table 2 summarizes model parameter estimates for all specifications. Comparing the estimates, one can notice that respondents did not quite distinguish between Programs 1–3.

Estimated coefficients $\hat{\beta}_{2m}$ and $\hat{\beta}_{4m}$ are not significantly different from zero in all specifications. The restoration of 6 miles of the river produces a spectacular effect. A possible cause of such a dramatic increase may be the "bet big, win big" maxim. In each management program, the survey identified category values for a set of ecosystem services, such as habitat for fish, wildlife, water purity, etc. Levels of those services were defined as "low," "moderate," or "high." While other programs featured differing service levels, Program 4 has all levels at "high." It seems to be likely that the maximum improvement was the threshold to trigger both attention and considerable spending.

To assess in-sample predictive ability of the three models, we used the percentage of correctly predicted sequences on 1,000 draws with replacement from the sample. The results were compared to a benchmark success rate of 17 percent attainable by indiscriminately guessing the outcome on each trial, solely based on the proportions of outcomes in the sample. Notably, three outcomes out of the possible sixteen make up 68 percent of the sample. These are: "yes" to all programs (29 percent); "no" to all Programs (22 percent); and "yes" only to Program 4 (17 percent). The PCL specification performed the best, marginally improving on the MNL results. At the same time, the logit specification proved to be falling short of the results of simple guessing.

The estimation of a PCL model is an intricate process because of the possibility of corner solutions: $\lambda_{kl} \to 0$ or $\lambda_{kl} \to 1$. The logistic transform was used with all lambdas; i.e., the optimization was performed on

$$\lambda^*_{kl}, \lambda_{kl} = \lambda^*_{kl}(1+\lambda^*_{kl})^{-1}.$$

Table 2. Estimated Coefficients

	Estimate (estimated standard deviation)		
	PCL	**MNL**	**Logit**
Watershed protection, β_{wp}	**0.8973 (0.4368)	**1.0687 (0.4206)	***1.6371 (0.1537)
2 miles of restoration, β_{2m}	-0.1078 (0.4680)	-0.1390 (0.54005)	-0.2602 (0.4743)
4 miles of restoration, β_{4m}	0.0934 (0.4876)	0.0959 (0.4434)	0.1196 (0.3755)
6 miles of restoration, β_{6m}	**1.8613 (0.7908)	***1.9798 (0.5567)	***2.9450 (0.2538)
Distribution of $\ln(\gamma_i)$			
μ	**1.1751 (0.5462)	***1.3838 (0.4101)	***1.3134 (0.3488)
σ^2	***4.4105 (0.5652)	***4.0044 (0.3877)	***0.7121 (0.1963)
Predictive ability, %	20	18	16

Significance level: *** 99%, **95%, *90%.

Three corner solutions were encountered, λ_{12}, λ_{14}, and λ_{34}. To avoid numerical problems, the optimization program (written in SAS IML) was set to fix any lambda at 0.01 or 0.99 if the optimization algorithm attempted to go beyond the boundaries of the [0.01, 0.99] interval after 50 iterations.

Estimated λ are presented in Table 3. These are, at a glance, consistent with our hypothesis that the dependence between unobserved utility levels decreases as the items grow farther apart with respect to attributes.

Table 3. Estimated PCL Dependence Parameters

Program	2	3	4
1	<0.01	0.78	>0.99
2		0.05	0.98
3			<0.01

Estimates of λ for neighboring programs are very close to zero; in other words, the respective utility shocks are highly dependent. The degree of dependence plummets to almost nothing for non-adjacent options. The result is peculiar yet not necessarily puzzling. Consider the following parable. A group of people was asked to pick two cities that are closer to each other geographically, with the options being Washington, DC; New York; and Boston. We can reasonably expect people to pick Washington, DC–New York and New York–Boston pairs. Hardly anybody will choose Washington, DC and Boston. So, by a preponderance of evidence, the investigator will probably conclude that while Washington, DC and New York, and New York and Boston are "similar," Washington, DC and Boston are not. The peculiar aspect is that it seems respondents considered adjacent programs to be a sort of "Red Bus/Blue Bus" tradeoff. Note from Table 3 that all lambdas but one (Programs 1 and 3) are either corner solutions or very close to being such. This, per se, is still consistent with the hypothesis of the paper, but this makes the proportionality work in a highly nonlinear fashion. The dependence between the difference in miles restored and estimated λ was tested with Kendall's τ nonparametric test. The value of the statistic was 0.77, which has a p-value less than 0.01. This provides a statistical confirmation to the observed pattern.

Since the coefficient on bid is assumed to be following a log-normal distribution, WTP calculated according to equation (7) is distributed as a weighted reciprocal of this log-normal variate. Table 4 presents selected WTP quantiles for all models. We report the single WTP value for Program 1, assuming the insignificant estimates of β_{2m} and β_{4m} to be zero and, thus, a difference in WTP for these Programs to be undetectable. The WTP distributions in Table 4 were obtained through Monte-Carlo simulation, using the sampling distribution of $(\hat{\beta}_{6m}, \hat{\mu}, \hat{\sigma}^2)$. The distribution is (asymptotically) normal with the mean $(\hat{\beta}_{6m}, \hat{\mu}, \hat{\sigma}^2)$ and the variance Σ. Draws from the distribution of $(\hat{\beta}_{6m}, \hat{\mu}, \hat{\sigma}^2)$ were made first; then, draws from $N(\hat{\mu}, \hat{\sigma}^2)$ were made to obtain ln(γ). Lastly, WTP was calculated as

$$WTP = \hat{\gamma}^{-1}\hat{\beta}_{6m}.$$

The process was repeated many times to obtain a sufficiently large number of draws from the WTP distributions. WTP values therefore come from the distribution of a nonlinear, stochastic function of $(\hat{\beta}_{6m}, \hat{\mu}, \hat{\sigma}^2)$; the variability in the estimated WTP is driven by all parameters in the distribution of $(\hat{\beta}_{6m}, \hat{\mu}, \hat{\sigma}^2)$.

Table 4. Estimated WTP

Specification	WTP Quantile, $		
	25%	**50%**	**75%**
PCL:			
Programs 1–3	7	31	128
Program 4	20	86	361
MNL:			
Programs 1–3	9	25	74
Program 4	30	80	229
Logit:			
Programs 1–3	10	41	147
Program 4	34	121	452

There are no large differences between the WTP values resulting from the three model specifications. Yet the differences of 20 percent to 50 percent of the values' magnitudes are in no case trifling, either. Parenthetically, the obtained WTP estimates are several times higher than estimates arising from the random-effects probit model by Holmes et al. (2004), while the conclusion with respect to the WTP overall super-additivity remains. The results do not quite satisfy the scope test (Arrow et al. 1993), since WTP values do start increasing until the program size reaches 6 miles of river restoration. However, the triggering effect of the maximal action package in Program 4 makes this result fairly logical. The individual demand therefore appears to be more of a step-function rather than a conventional downward-sloping schedule.

Discussion and Conclusions

Listed below are three net results from the empirical part of the study:

(a) The PCL specification that enforces transitivity and allows for continuity provides a moderately better fit, *ceteris paribus*, than others that exclude either or both continuity and transitivity.

(b) Whatever specification is used, Program 4, being the most extensive management package, has a super-additive effect on WTP.

(c) Estimated dependence parameters from the PCL specification appear to exhibit the pattern that the continuity hypothesis suggests: when the difference between values of an attribute increases for two policy options, the dependence between the respective utility levels diminishes.

The immediate implication of the results (a) through (c) for WTP estimates is that a model that adheres to the principles of utility analysis is capable of providing more reliable WTP estimates both economically and statistically. Even though no dramatic differences between estimates from different models have been found, these differences are still non-negligible and may be important for policy decisions.

Why do PCL and MNL specifications do a better job of predicting choice outcomes for the Little Tennessee River data set as compared to logit? As mentioned earlier, more than half of all observations in the data set are invariable sequences of "yes" and "no" votes given for all the alternatives. Roughly 50 percent of respondents had not changed their mind with respect to whether or not they would want any restoration of the Little Tennessee River watershed. The utility from the status quo level of the river's protection had a great deal of influence on people's choices. Knowing the respondent's vote for any given program, one could flip a coin to predict the voting outcome for another program without any valuation model. The logit specification totally disregards this fact. It allows the utility of the baseline state to change so that, after conditioning on the person-specific marginal utility of income, any previous or subsequent choices bear no additional information. Meanwhile, PCL and MNL specifications anchor the utility from the alternatives to the unvarying individual point of reference and thus make use of this information. These specifications provide a better control for individual heterogeneity rather than imputing the series persistence to the "warm glow" or protest voting phenomena. The PCL specification goes further and reaps a reward. Based on utility continuity, it allows the utility levels from similar states to be also similar. This lets the model extract even more information from the unobserved utility components, while doing so in a manner consist with utility theory.

We do not intend to promote the use of paired combinatorial logit or any other particular distribution of utility shocks. For example, a general mixed logit model may be a more general way to handle the problem, since the PCL specification can be easily mimicked using mixed logit. One (seeming) advantage of the mixed logit model is that the dependence is modeled explicitly as correlation. Yet, the fact that a correlation matrix must be non-negative definite is a restriction that may turn the above advantage into a disadvantage. The matrix of lambdas (see Table 3) is in no way non-negative definite. If a mixed logit model were used to obtain a correlation matrix, it would probably have produced a matrix with correlations decreasing in magnitude as we go away from the main diagonal. On one hand, this would be even more consistent with the hypothesis of the paper. On the other hand, this might actually force the compliance. Thus, a mixed logit model is unlikely to have caught the aforementioned "Red Bus/Blue Bus" peculiarity for adjacent pairs.

The message of our entire exercise is simple. We would want to stress the importance of specifying a stochastic CVM model in such a way that the investigator can attach theoretically based meaning to all parameters in the empirical model. Modern statistical software allows fitting a variety of flexible probabilistic choice models. But if a particular chosen model accounts for unobserved phenomena only mechanistically, then the researcher is left with the need for ex post facto interpretation of estimates. This limits the extent of quality control, since one would never know whether the observed pattern in estimates is what one should reasonably expect or is a mere sporadic occasion. The mechanism of a dose-response statistical model ultimately reflects on the welfare estimates. If utility shocks are allowed to follow whatever process, then welfare change estimates have whatever meaning. This is definitely not what a CVM investigator would intend to produce.

Much further research needs to be done in the valuation of multiple environmental policies. A rigorous testing of the timing invariance property in CVM applications is particularly desirable. An in-depth inquiry on specifications for the stochastic interaction of utility would be instrumental for the practitioner's needs. Considering similarities between the utility space and a geographical one, a direction for research can be spatial statistical models (Besag 1975). Those models adopt a conditional probability approach, the spatial Markovity, in formulating entire spatial systems and provide holistic schemes where deterministic and stochastic components are inherently interrelated.

References

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(10, January 15): 4601–4614.

Bergstrom, J., J. Stoll, and A. Randall. 1989. "Information Effects in Contingent Markets." *American Journal of Agricultural Economics* 71(3): 685-691.

Besag, J. 1975. "Statistical Analysis of Non-Lattice Data." *The Statistician* 24(3): 179-195.

Chu, C. 1989. "A Paired Combinatorial Logit Model for Travel Demand Analysis." In *Proceedings of Fifth World Conference on Transportation Research,* Vol. 4. Ventura, CA, pp. 295-309.

DeShazo, J. 2002. "Designing Transactions Without Framing Effects in Iterative Question Formats." *Journal of Environmental Economics and Management* 43(3): 360-385.

Giraud, K., J. Loomis, and R. Johnson. 1999. "Two Valuation Questions in One Survey: Is It a Recipe for Sequencing and Instrument Context Effects." *Applied Economics* 31(8): 957-964.

Hanemann, W. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(3): 332-341.

Hanemann, W., and B. Kanninen. 1999. "The Statistical Analysis of Discrete-Response CV Data." In I. Bateman and K.G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries.* New York, NY: Oxford University Press.

Herriges, J., and J. Shogren. 1996. "Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning." *Journal of Environmental Economics and Management* 30(1): 112-131.

Hoehn, J. 1991. "Valuing the Multidimensional Impacts of Environmental Policy: Theory and Methods." *American Journal of Agricultural Economics* 73(2): 289-299.

Holmes, T., and W. Adamowicz. 2003. "Attributed-Based Methods." In P. Champ, K. Boyle, and T. Brown, eds., *A Primer on Nonmarket Valuation.* Amsterdam, the Netherlands: Kluwer Academic Publishers.

Holmes, T., J. Bergstrom, E. Huszar, S. Kask, and F. Orr III. 2004. "Contingent Valuation, Net Marginal Benefits, and the Scale of Riparian Ecosystem Restoration." *Ecological Economics* 49(1): 19–30.

Holmes, T., and K. Boyle. 2005. "Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions." *Land Economics* 81(1): 114-126.

Kahneman, D. 1986. "Comments on the Contingent Valuation Method." In R.G. Cummings, D. S. Brookshire and W.D. Schulze, eds., *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method.* Totowa, NJ: Roweman and Allanheld.

McClennen, E. 1990. *Rationality and Dynamic Choice: Foundational Explorations.* Cambridge; New York, NY: Cambridge University Press.

McFadden, D., and G. Leonard. 1995. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." In J.A. Hausman, ed., *Contingent Valuation: A Critical Assessment*. Amsterdam, the Netherlands: North Holland.

Read, D., G. Antonides, L. van den Ouden, and H. Trienekens. 2001. "What is Better: Simultaneous or Sequential Choice?" *Organizational Behavior and Human Decision Processes* 84(1): 54-70.

Read, D., and G. Loewenstein. 1995. "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices." *Journal of Experimental Psychology: Applied* 1(1): 34-49.

Read, D., G. Loewenstein, and M. Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty* 19(1-3): 171-197.

Spash, C. 2000. "Ecosystems, Contingent Valuation and Ethics: The Case of Wetland Re-creation." *Ecological Economics* 34(2): 195–215.

Appendix: Obtaining Outcome Probabilities

```
Data: T, Y_i, T_r
Result: p_j
begin
    list all possible 2^T outcomes for Y_i ;
    /* probabilities of outcomes are the unknowns in p */
    Arrange all t_rs from all T_r , r = 1, 2, . . ., T in any array of sets A;
    /* A will then have Σ_{r=1}^{T} (T choose r) = 2^T − 1 elements of A_j */
    /* when T = 3, A = { {1}, {2}, {3}, {1,2}, {1,3}, {2,3}, {1,2,3} } */
    for each A_j do
        Calculate b_j = Pr (Y_{A_j} = 0);
    end
    /* For the T = 3 example */
    /* b_5 = Pr(Y_{A_5} = 0) = Pr(y_{i1} = 0, y_{i3} = 0) */
    for j = 1 to 2^T − 1 do
        for k = 1 to 2^T − 1 do
                if (Y_{A_j} = 0) event contains k-th outcome then
                    a_jk = 1;
                else
                    a_jk = 0;
                end
                /* calculate all a_jk elements of A, except for the last row */
        end
        end
        A_{2^T} = 0;
        b_{2^T} = 0;
        /* put 1 in all cells of the last row of A */
        /* and last cell of b - - - the sum of outcome probabilities must be one */
        Solve A_p= b for relevant outcome probability p_j with Cramer's rule;
        /* the determinant of A will be either 1 or -1, */
        /* which further simplifies calculations */
end
```