A Pseudo-Sequential Choice Model for Valuing Multiple Environmental Policy or Program Components in Contingent Valuation Applications

Dmitriy Volinskiy[a]
John Bergstrom
and
Christopher Cornwell[b]

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24–27, 2005

[a]Dmitriy Volinskiy (dvolinskiy@agecon.uga.edu) is a graduate research assistant and John Bergstrom (jbergstrom@agecon.uga.edu) is Professor in the Department of Agricultural and Applied Economics, College of Agricultural and Environmental Sciences, University of Georgia.

[b]Christopher Cornwell (cornwl@terry.uga.edu) is Professor in Economics Department, Terry College of Business, University of Georgia.

A Pseudo-Sequential Choice Model for Valuing Multiple Environmental Policy or Program Components in Contingent Valuation Applications

Dmitriy Volinskiy
John Bergstrom
Christopher Cornwell

April 20, 2005

Abstract

The study proposes a discrete-choice model for environmental policy/program valuation, to be used in cases when several policies are valued sequentially. The stochastic specification of the model is consistent with the transitivity and continuity axioms of utility analysis. An empirical methodology for the model is suggested. An application of this model to WTP estimation for Little Tennessee River watershed ecosystem restoration is provided. Findings from the application agree with hypothesized agent's behavior.

1 Introduction and Objectives

Modern-day environmental policies or programs such as watershed ecosystem restoration are designed to improve multiple ecosystem services and naturally consists of multiple components or parts. Clearly, the valuation of such policies or programs should address the multi-dimensionality of the problem; that is, the relative importance of the program's components needs to be studied. If the contingent valuation method (CVM) is used, a commonly practiced approach is to include several policy options in the survey, which are to be 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[1] are valued using the dichotomous choice format — the respondents are asked "take it or leave it" questions for one item at a time — a binary discrete-response data set with a sequence of observations per individual is generated.

The focus in scholarly literature has been mainly on deterministic utility functions to be used with multiple items. Considerable effort has been made to suggest economically meaningful functional representations for deterministic utility. Modelling the stochastic part has received far less attention. In many applied studies, utility shocks are modelled as independently and identically distributed random variables. Some studies recognize the fact that observations on one and the same individual must be related to each other. Unobserved effects models are then used, that range in complexity from very simple random or fixed effects to sophisticated mixed logit or latent class models.

[1]Since our analysis essentially applies to valuing any composite non-market goods, the terms "commodity," "good," "policy," and the like are used interchangeably.

Whatever the level of sophistication, distributional assumptions with respect to utility shocks are typically made on the basis of intuition or statistical convenience. This makes the role of error terms rather ambiguous, from a utility analysis perspective.

Deterministic and stochastic terms determine unobserved utility, working as one whole. If an applied CVM study employs utility theory to estimate a welfare measure for the representative agent, then the estimate would only be valid provided the whole model is in agreement with the postulates of utility theory, for the latter need to be axiomatically assumed in order to arrive at any conventional welfare measure. All parts of the model should thus agree with utility theory. Conversely, if there is no holistic treatment of the model components, the validity of valuation results may be called into question on economic grounds and their usefulness for informing policy and management decisions hindered.

Concepts behind such a holistic model is the *raison d'être* of this paper. Our overall objective is to provide the rationale for, conceptually develop, and test empirically a stochastic model for valuing several multi-attribute environmental policies or programs, one at a time, that accounts for such principles of rational behavior as utility transitivity and continuity. Unlike most of the models suggested in CVM literature to date, our model is inherently invariant to changing from the sequential to simultaneous choice format. The model we consider can be used for valuation purposes as such, or it can serve as a default model, should one wish to test for instrument format effects.

A general conceptual valuation model is developed in Section 2. 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 transitivity axiom.

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.

In Section 3, we introduce the specifics of survey data for the Little Tennessee River 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 in Section 4 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.

2 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 and v_2. One way to elicit the preferences is to let the agent pick the preferred state from all possible pairs. An alternative is to ask the agent to rank the three states at once. From an economic theory perspective, choosing the simultaneous format over the sequential or vice versa is immaterial as long as preferences remain unchanged. But it is not so when it comes to empirical modelling.

Econometric literature offers a broad spectrum of panel data discrete-response model options to consider for both cases. If choices are arranged in pairs, the following random utility model (RUM) arises:

$$\begin{aligned} u_{jt} =& v_j + \varepsilon_{jt} \\ u_{kt} =& v_k + \varepsilon_{kt} \end{aligned} \tag{1}$$

where (j, k) are $(1, 0), (2, 0), (2, 1)$ respectively for $t = 1, 2, 3$; v are deterministic components of the respective random utility levels and ε are utility shocks.

Some stochastic specifications may assume that utility shocks are independently and identically distributed; an error components model would account for unobserved heterogeneity of individuals. Whatever model is used, however, it will operate implicitly assuming there are six random quantities involved and there are eight possible outcomes, of which two are intransitive. Evidently, reflexivity is also violated.

Meanwhile, there are only three random quantities in the simultaneous ranking format, and there are six transitive outcomes. Thus, results from a simultaneous model will differ systematically and to an unknown extent from those coming from a sequential model even if the deterministic parts are identically specified. To make things worse, if the study aims to address a possible instrument format effect, the investigator may erroneously conclude such an effect exists while in reality it does not.

Following this line of reasoning, the investigator will probably opt for the simultaneous format even though the necessity for the respondent to consider multiple options at once may bring about accuracy concerns. Questions will remain. Is it possible to attain an equivalent representation of both formats? Should intransitivity be excluded? And, more in general, is there a way to reconcile rationality axioms that the traditional utility theory imposes on the decision-maker's preferences within the stochastic setting of a stated-preference experiment?

Let us consider a T-period sequential binary choice model. A utility maximizer i chooses, at each period t, $t = 1 \ldots T$, in a sequence between two states of the world. These states are a period/individual-specific "alternative" (a particular environmental policy) and a no-action baseline policy, the "status quo", with the corresponding utility levels:

$$\begin{aligned} u_{it} &= v_{it} + \varepsilon_{it} \\ \tilde{u}_{it} &= \tilde{\varepsilon}_{it} \end{aligned} \tag{2}$$

where $v_{it} = v(\boldsymbol{x}_{it})$ is the deterministic utility of the alternative with attributes $\boldsymbol{x}_{it}$, the deterministic utility of the status quo is zero, and $(\varepsilon_{it}, \tilde{\varepsilon}_{it})$ are the respective error terms. The model implies the following marginal probabilities of choice outcomes:

$$\Pr(u_{it} > \tilde{u}_{it}) = \Pr(v_{it} > \tilde{\varepsilon}_{it} - \varepsilon_{it}) = F_{\tilde{\varepsilon}_{it} - \varepsilon_{it}}(v_{it}) \tag{3}$$

where $F_{\tilde{\varepsilon}_{it} - \varepsilon_{it}}$ is the distribution function of the difference of utility shocks at time t.

The standard practice is to use $2T$ independently and identically distributed (i. i. d.) errors (Hoehn 1991). As already mentioned, this leads to the non-equivalence of elicitation formats and potential problems with transitivity. In their comprehensive review of statistical methods with CVM data, Hanemann and Kanninen (1999) consider a model where $\tilde{\varepsilon}_{it} \equiv \varepsilon_{i0}$, $\forall t$. This condition means that an unobserved utility level of the "status quo" state, u_{i0}, is the same no matter where in the survey this state is invoked. To be consistent, one can extend this logic to all states. If $\boldsymbol{x}_{is}$ and $\boldsymbol{x}_{it}$ are identical, then:

$$\Pr(u_{it} > u_{i0} | u_{is} > u_{i0}) = 1, \;\; s \neq t \tag{4}$$

It is apparent that, if all 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 shall term this model the *pseudo-sequential choice* to emphasize its atemporal nature. The study (ibid.) does not put forward any justification for this restriction. Indeed, what are the reasons why one should consider imposing it?

We begin with presenting our understanding of the roles played by the deterministic and stochastic components of the structural RUM in Equation (2). Hanemann (1984a) provides the following definition for a generic RUM:

> A random utility model arises when one assumes that, although a consumer's utility function is deterministic for him, it contains some components which are unobservable to the econometric investigator and are treated by the investigator as random variables.

An immediate implication of this definition is that there is conceptually only one source of uncertainty in the model and it is the investigator's uncertainty. The respondent consistently applies a deterministic yet unknown decision rule throughout the elicitation process. Under the general guidance of economic theory the investigator subjectively formulates a specification of the respondent's deterministic utility. There is no theory to substantiate any parametric assumption with respect to errors in the model, and both deterministic and stochastic parts need to be functionally specified before the actual survey data are incorporated into the model[2]. If so, then a distributional assumption with respect to utility shocks, whatever it turns out to be, determines the subjective rule the investigator will apply to specify the likelihood of a choice outcome for any parameter values in both parts of the model. That is, conditional on v_{it}, the investigator presents her subjective odds that $u_{it} > \tilde{u}_{it}$ by making a parametric assumption about the distributions of ε_{it} and $\tilde{\varepsilon}_{it}$.

Because the investigator has made subjective judgements with respect to both parts of the RUM, she has set the modelling rules which must be followed. This leads to determinism in her understanding of the respondent's behavior. Thus, we predicate non-volatility of both agent and investigator in their decision making. This determines dynamic consistency of the parties. 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[3].

Participants of CVM experiments are likely to have no experience with programs or policies to be valued. A number of empirical studies found that different ways of supplying commodity-related information or different amounts of information supplied led to in a significant variation in valuation results (Bergstrom et al. 1989). The time dimension and sequencing of choice sets can only be reasonably omitted in situations where information about the programs is supplied to

---

[2]In a number of studies, researchers attempted to obviate the issue of supplying parametric specifications for either or both systematic and random terms by using semiparametric (Klein and Spady 1993) or nonparametric approaches (Matzkin 1992; Matzkin 1993). The resultant models, however, either replace the problem of parametric specification with that of selecting a kernel and its parameters or are unsuitable to obtain a welfare change estimate.

[3]Experimental revealed-preference studies conducted by behavioral scientists do not appear to have come up with definitive results with respect to timing invariance (Read et al. 2001). Read and Loewenstein (1995) introduced the term "diversification bias," referring to a demonstrated excess variety in items selected in the simultaneous design. 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. In these and other studies, reviewed by Read et al. (2001) agents showed some psychological phenomena which admit various interpretations. The issue of an empirical validity of timing invariance largely remains a moot point.

respondents strictly prior to elicitation, and no additional information is given in between elicitation questions.

The pseudo-sequential structure of a choice model makes the latter consistent with the transitivity axiom, so that the consumer is supposed to be able to order her preferences amongst policy alternatives in a consistent manner. The essence of this model is that, if $\boldsymbol{x}_s = \boldsymbol{x}_t$[4] for Program_s and Program_t, then $\varepsilon_s = \varepsilon_t$; that is, the respective utility shocks are perfectly dependent. But suppose there is an infinitesimally small difference between $\boldsymbol{x}_s$ and $\boldsymbol{x}_t$. Now one deals with two random quantities, ε_s and ε_t. But the utility continuity axiom, however, asserts that in this case the departure from perfect dependence should be small as well. Loosely stated, the principle of continuity postulates that any two states which 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 allows for the possibility of substitution between policy components, which, in turn, permits comparing the relative importance of these components.

In real-world situations, agents themselves select relevant attribute sets for items they compare. The investigator may correctly guess a large or small subset of these sets. The bare fact that two options have the same observable attribute levels does not give the investigator much information on whether or not the bundles lie close together in one's consumption space.

Policy alternatives in CVM studies are different. These are made distinct in a number of key attributes which are communicated to respondents. Accordingly, respondents determine their preferences on the basis of what they have been told and/or shown about the choice options. If the attributes of two options are the same, there can be nothing else to distinguish them one from another. It also follows that, provided the deterministic utility is not badly misspecified, the errors are likely to reflect some unknown effect of option attributes that could not be modelled within the deterministic utility specification. An example of such an effect is the agent's overall subjective perception of a policy alternative.

Assume the investigator has a measure of dissimilarity between two states, $\lambda_{st} = \lambda(\boldsymbol{x}_s, \boldsymbol{x}_t)$, increasing as dissimilarity grows. If pairs of choices can be compared on the basis of that measure, continuity would imply that unobserved utility terms for a pair of adjacent options are, on average, closer to each other compared to either out of the pair and a non-neighboring third. For example, if, for a chosen λ, $\lambda(\boldsymbol{x}_r, \boldsymbol{x}_s) \ll \lambda(\boldsymbol{x}_r, \boldsymbol{x}_t)$ and $\lambda(\boldsymbol{x}_r, \boldsymbol{x}_s) \ll \lambda(\boldsymbol{x}_s, \boldsymbol{x}_t)$, then one should expect that $|\varepsilon_r - \varepsilon_s| < |\varepsilon_r - \varepsilon_t|$ and $|\varepsilon_s - \varepsilon_t| < |\varepsilon_r - \varepsilon_t|$ to be (subjectively) more probable events than their respective complements.

We have assumed the existence of a dissimilarity measure and its relationship to the hypothesized distribution of error terms. An empirical study would certainly require a specification of this measure. Utility theory abstractly defines continuity by asserting that, for any bundles $\boldsymbol{x}$ and $\boldsymbol{y}$ in the consumption space $\{\boldsymbol{x}|\boldsymbol{x} \succeq \boldsymbol{y}\}$ and $\{\boldsymbol{x}|\boldsymbol{x} \preceq \boldsymbol{y}\}$ are closed sets. Except for the values of attributes and the utility levels, there is nothing to use to assess the dissimilarity between a pair of bundles.

It is contended that, in general, the choice of a dissimilarity function can only be implemented on subjective grounds. A usual distance function or the gravity formula can be suitable candidates. We shall refrain from suggesting any particular specification. Instead, we shall first focus on the effects of increasing the level of one attribute while keeping the others constant. If our conjecture is true, then statistical dependence between errors should decrease as the goods are made more different in at least one dimension.

[4] The observation subscript i will be dropped to simplify notation.

This can be put in the form of a formal hypothesis. Let $\lambda_{st} \in [0,1]$ be a measure of statistical independence[5] between ε_s and ε_t, as above, and let $\triangle^k$ be the only different attribute k in $\boldsymbol{x}_s$ and $\boldsymbol{x}_t$, $\triangle^k = ||\boldsymbol{x}_s - \boldsymbol{x}_t||$, $\boldsymbol{x}_s^{-k} = \boldsymbol{x}_t^{-k}$. Then

Hypothesis. $\lambda_{st} \propto \triangle^k \,\big|\Im_s = \Im_t$, *where* $\Im_s, \Im_t$ *are the information the respondent has about the commodities at time* s *and* t, *respectively.*

When $\triangle^k = 0$, the agent will assign the same utility level to all occurrences of the same hypothetical state — this is transitivity. The monetary bid is not an attribute of a bundle of commodities, so it does not influence whether commodities are closer together or farther apart. Therefore we shall not be viewing the bid as part of $\boldsymbol{x}_t$. Finally, the role of constant information is to disallow the agent to reconsider expected experiences from the hypothetical commodities in light of any new information to come about in between choices.

Combining the major elements of our reasoning, we arrive at the following pseudo-sequential choice model:

$$\begin{aligned} u_{it} &= v_{it} + \varepsilon_{it} \\ u_{i0} &= \varepsilon_{i0} \end{aligned} \tag{5}$$

where the distribution of $(\varepsilon_{i0}, \varepsilon_{i1}, \ldots, \varepsilon_{iT})$ in general may have $\binom{T+1}{2}$ parameters of dependence for all possible pairs of shocks. If the measure of dependence is bounded, for example, between zero and one, 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 $(\boldsymbol{x}_t, \boldsymbol{0})$ 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.

In order to estimate the model in Equation (5) 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 is fairly straightforward:

$$\begin{aligned} \Pr(u_{i1} < u_{i0}, u_{i2} < u_{i0}) =& \Pr(u_{i0} = \max(u_{i0}, u_{i1}, u_{i2})) \\ \Pr(u_{i1} > u_{i0}, u_{i2} < u_{i0}) =& \Pr(u_{i2} < u_{i0}) - \Pr(u_{i0} = \max(u_{i0}, u_{i1}, u_{i2})) \\ \Pr(u_{i1} < u_{i0}, u_{i2} > u_{i0}) =& \Pr(u_{i1} < u_{i0}) - \Pr(u_{i0} = \max(u_{i0}, u_{i1}, u_{i2})) \\ \Pr(u_{i1} > u_{i0}, u_{i2} > u_{i0}) =& 1 - \Pr(u_{i1} < u_{i0}, u_{i2} < u_{i0}) \\ & - \Pr(u_{i1} > u_{i0}, u_{i2} < u_{i0}) \\ & - \Pr(u_{i1} < u_{i0}, u_{i2} > u_{i0}) \end{aligned} \tag{6}$$

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 $\mathbf{Y}_i = \{y_{i1} = 1[u_{i1} > u_{i0}], y_{i2} = 1[u_{i2} > u_{i0}], \ldots, y_{i2} = 1[u_{iT} > u_{i0}]\}$ 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 $\Pr(u_{i0} = \max(u_{i0}, \boldsymbol{u}_{i,\mathcal{A}}))$, where $\boldsymbol{u}_{i,\mathcal{A}}$ are the utility levels of options in the subset $\mathcal{A}$ of indices $\{1, 2, \ldots, T\}$.

The objective is to solve a linear system $\mathbf{Ap} = \mathbf{b}$ of 2^T equations. Let $\mathbf{T}_r$ be a set of $\binom{T}{r}$ unique ordered combinations of subscripts, $\mathbf{t}_{rs}$, in $\{1, 2, \ldots, T\}$. The algorithm presented in Algorithm 1.

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.

[5] It can be, but is not limited to unity minus squared correlation.

```
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 an array of sets A;
    /*A will then have sum_{r=1}^{T} (T choose r) = 2^T - 1 elements A_j */
    /*when T = 3, A = {{1},{2},{3},{1,2},{1,3},{2,3},{1,2,3}} */
    foreach 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 Ap = 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
```

Algorithm 1: Obtaining Outcome Probabilities

3 Empirical Illustration: Estimating WTP for Little Tennessee River Management Alternatives

The Little Tennessee River watershed is located in Georgia, North Carolina, and Tennessee. The watershed encompasses 10,783 acres, including 18 rivers and streams and 26 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 structure and function. 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 parameter of ecosystem health.

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 CVM survey instrument was designed and implemented. The present study uses the data set obtained through the above survey.

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 10 years to support Program t, would you vote in favor of it?" 58 respondents ($N = 58$) provided complete sequences of votes in the survey.

The survey included 4 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 $\tau + 1$ would have increased, otherwise $\tau + 1$ would have been offered at the same bid amount as τ.

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

$$v_{it} = -\gamma_i BID_{it} + \beta_{wp} wp_t + \beta_{2m} 2m_t + \beta_{4m} 4m_t + \beta_{6m} 6m_t \tag{7}$$

where BID_{it} is an amount in \$100, 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 on bid, γ_i, to be varying across the panel; it is assumed to follow a 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$_t$ (Hanemann 1984b) as the ratio of the implicit price of its attributes to that of \$100 of extra income:

$$WTP_t = \frac{\beta_{wp} wp_t + \beta_{2m} 2m_t + \beta_{4m} 4m_t + \beta_{6m} 6m_t}{\gamma_i} \tag{8}$$

One option for estimating the stochastic version of Equation (8) 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 for estimating Equation (8) 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.

Table 1: 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. Extreme Value Type I

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

$$\begin{aligned} F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J) = \exp[-G(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \ldots, e^{-\varepsilon_J})] = \\ \exp[-G(a_1, a_2, \ldots, a_J)] \\ G = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left(a_k^{1/\lambda_{kl}} + a_l^{1/\lambda_{kl}}\right)^{\lambda_{kl}} \end{aligned} \tag{9}$$

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 significant covariation; when $\lambda_{kl} \to 0$, the dependence becomes perfect. The distribution provides the dependence parameters that meet our needs. Besides, if one sets $\lambda_{kl} \equiv 1, \forall k, l$, this GEV model reduces to 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 $\lambda_{0t} \equiv 1, \ t = 1 \ldots T$; that is, we shall 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. It also has a useful consequence: the model becomes the standard binary logit for any cross-section.

Using the PCL choice probability formula,

$$\Pr(u_{it} = \max(\mathbf{u}_i)) = \frac{\sum_{j \neq t} e^{v_{it}/\lambda_{tj}} \left(e^{v_{it}/\lambda_{tj}} + e^{v_{ij}/\lambda_{tj}}\right)^{\lambda_{tj}-1}}{\sum_{k=0}^{T-1} \sum_{l=k+1}^{T} \left(e^{v_{ik}/\lambda_{kl}} + e^{v_{il}/\lambda_{kl}}\right)^{\lambda_{kl}}} \tag{10}$$

and Algorithm 1, 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 given by Equation (7).

The PCL specification applies no restrictions to the model in Equations (5) and (10), that is, the pseudo-sequential choice framework is used to ensure transitivity and 6 dependence parameters are estimated to see whether they are related to changes in the mileage of riverbank restoration in the

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 (0.4650)	−0.1390 (0.5005)	−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 $\ln(\gamma)$:			
μ	1.1751 (0.5462)**	1.3838 (0.4101)***	1.3134 (0.3488)***
σ^2	4.4105 (0.5682)***	4.0044 (0.3877)***	0.7121 (0.1963)***
Predictive ability, %	20	18	16

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

Table 3: Estimated Dependence Parameters from PCL Specification

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

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

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. Finally, the Logit specification is simply a mixed logit, which addresses neither transitivity nor continuity. There are 8 i. i. d. Extreme Value Type I 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 on β_{2m} and β_{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 an 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.

In order to assess the goodness-of-fit (in-sample predictive ability) of the three models, we used the percentage of correctly predicted sequences on a 1000 draws with replacement from the sample. The results were compared to a benchmark success rate of 17%[6] attainable by indiscriminately guessing the outcome on each trial, solely based on the proportions of outcomes in the sample.

The PCL specification performed the best, marginally improving on the MNL results. At the same time, the Logit specification proved to be useless in decision-making, even falling short of the simple guessing.

Estimated λ in Table 3 are, at a glance, consistent with our hypothesis that the dependence between unobserved utility levels decreases as the items grow farther apart attribute-wise.

$\hat{\lambda}$ for neighboring Programs are very close to zero, that is, the respective utility shocks are highly dependent. The degree of dependence plummets to almost nothing for non-adjacent options. 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 the 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-Nornal distribution, WTP calculated according to Equation (8) 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 Programs 1–3, 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.

There are no large differences between the WTP values resulting from the three model specifications. Yet the differences of 20–50% of the value's magnitude 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.

4 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 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.

[6] Such high a figure arises because 3 outcomes out of the possible 16 make up 68% of the sample. These are: "yes" to all programs (29%), "no" to all Programs (22%), and "yes" only to Program 4 (17%).

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 predicting choice outcomes for the Little Tennessee River data set than mixed logit? As earlier mentioned, 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% of respondents had not changed their mind with respect to whether or not they would want any restoration of 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 mixed logit model 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 then impute 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.

The authors do not intent to promote the use of paired combinatorial logit or any other particular distribution of utility shocks. Choosing such a distribution the investigator chooses her risk management technique for one thing, and a computational device, for the other. The message of our entire exercise is more general. It serves to stress the importance of specifying a stochastic CVM model in such a way that the investigator can attach theoretically-found 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 it 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 search 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. Portney, E. Leamer, R. Radner, and H. Schuman (1993, January). Report of the NOAA panel on contingent valuation. *Federal Register 58*(10), 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*, Volume 4, pp. 295–309.

Hanemann, W. M. (1984a). Discrete/continuous models of consumer demand. *Econometrica 52*(3), 541–562.

Hanemann, W. M. (1984b). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics 66*(3), 332–341.

Hanemann, W. M. and B. Kanninen (1999). The statistical analysis of discrete-response CV data. In I. J. Bateman and W. K. G. (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.

Hoehn, J. P. (1991). Valuing the multidimensional impacts of environmental policy: Theory and methods. *American Journal of Agricultural Economics 73*(2), 289–299.

Holmes, T. P., J. C. Bergstrom, E. Huszar, S. B. Kask, and F. Orr (2004). Contingent valuation, net marginal benefits, and the scale of riparian ecosystem restoration. *Ecological Economics 49*, 19–30.

Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for discrete choice models. *Econometrica 61*(2), 387–422.

Matzkin, R. (1992). Nonparametric and distribution-free estimation of the binary choice and the threshold-crossing models. *Econometrica 60*(2), 239–270.

Matzkin, R. (1993). Fully nonparametric estimation of polychotomous choice models. *Journal of Econometrics 58*, 137–168.

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

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*, 34–49.

Read, D., G. Loewenstein, and M. Rabin (1999). Choice bracketing. *Journal of Risk and Uncertainty 19*, 171–197.