Testing for the Internal Consistency of Choice Experiments Using Explicit Rankings of Quality Attributes

Christopher Azevedo
Associate Professor of Economics, University of Central Missouri, Warrensburg, MO

Jay R. Corrigan
Associate Professors of Economics, Kenyon College, Gambier, OH

John Crooker
Associate Professor of Economics, University of Central Missouri, Warrensburg, MO

October 2008

ABSTRACT

Choice experiments (CEs) are an increasingly important tool in the environmental valuation literature because of their ability to deal with multifaceted environmental issues and their basis in random utility theory. One particularly useful aspect of the CE method is that it allows researchers to estimate marginal rates of substitution between an environmental resource's various attributes. These marginal rates of substitution provide an implicit ranking of the attributes, which can be compared with other ranking mechanisms. In this paper we describe a method for testing for the internal consistency of choice experiments by comparing the implicit attribute ranking generated by a CE with that generated by an explicit attribute-ranking exercise. The analysis uses data gathered through a unique survey in which respondents completed both a CE exercise and an attribute-ranking exercise indicating their preferences over pollution abatement and water quality improvement strategies for a freshwater lake in north-central Iowa. Comparisons are made on a sample-wide basis as well as an individual basis.

Key words: Environmental valuation, choice experiments, internal consistency.

Testing for the Internal Consistency of Choice Experiment Using Explicit Rankings of Quality Attributes

INTRODUCTION

Choice experiments (CEs) have become increasingly prominent in the environmental valuation literature because of their ability to deal with multifaceted environmental issues and their basis in random utility theory. One particularly useful aspect of the CE method is that it allows for the estimation of the marginal rates of substitution (MRS) between the attributes of the environmental resource. These marginal rates of substitution provide an implicit ranking of the attributes, which can be compared with other ranking mechanisms. In this paper we describe a method for testing for the internal consistency of choice experiments by comparing the implicit attribute ranking generated by a choice experiment with that generated by an explicit attribute-ranking exercise. The analysis uses data gathered through a unique survey in which respondents completed both a CE exercise as well as an attribute-ranking exercise. Comparisons are made on a sample-wide basis as well as an individual basis.

CONTEXT OF THE RESEARCH

The travel cost model has gained widespread acceptance in the environmental valuation literature despite the limitations discussed by Randall. The main strength of the travel cost model is that it is based upon revealed preference (RP) data, or data that is revealed by actual behavior. An important limitation of the model is that it cannot be used in cases where there is no historical data. Stated preference (SP) methods are capable of overcoming this limitation since they are based not on revealed consumer behavior but rather on hypothetical statements contingent upon a scenario presented to the respondent by the researcher.

SP methods have seen widespread usage in the form of contingent valuation studies, but initially were viewed skeptically by many researchers who were justifiably wary of methods that rely on data based on hypothetical statements rather than on actual behavior (e.g., Arrow et al., Diamond and Hausman, Hanemann, Portney). A number of possible biases have been identified with SP methods (e.g., Loomis et al., Kemp and Maxwell, Desvousges et al., Kahneman and Knetsch, and Diamond and Hausman). Though they have gained wide acceptance, in part due to the National Oceanographic and Atmospheric Administration's Blue Ribbon Panel's qualified endorsement of the approach, they are still viewed warily by some.

The general category of SP methods contains several approaches: contingent valuation, contingent behavior, and multi-attribute techniques. In the contingent valuation approach, the resource is described in terms of its attributes and respondents are asked about their willingness to pay a dollar amount, referred to as the bid, for that bundle of attributes.[1] The only attribute of the resource that varies across respondents is the bid. In the contingent behavior approach, respondents are asked about how their behavior might change in response to a change in an attribute of a resource. For example, they might be asked how many additional trips they would take to a particular lake if the water quality were improved to a particular level.

Multi-attribute techniques encompass a range of methods, some of which may or may not be consisted with utility theory. These techniques differ from contingent valuation and contingent behavior techniques in that rather than varying a single attribute, numerous attributes are simultaneously varied. This variation of multiple attributes allows for the estimation of the marginal rates of substitution between the attributes. If one of those attributes is a dollar bid, the wil-

[1] Willingness to pay information can be gathered either through the use of an open-ended question such as, "How much would you be willing to pay for this quality improvement?" or through the use of a referendum-style question such as, "Would you be willing to pay $X for this quality improvement?" Open-ended questions can lead to strategic answers on the part of the respondent. As a result, referendum-style questions are the generally-accepted approach.

lingness to pay for the particular attributes can be estimated. Often, multi-attribute techniques are mistakenly characterized as conjoint techniques, though this characterization is inaccurate. As discussed by Louvierer, conjoint analysis is actually a specialized form of multi-attribute technique.

Multi-attribute techniques include contingent rating, contingent ranking, and choice experiments. With the contingent rating method, the respondents are presented with a number of bundles, each varying in terms of the resource attributes, and are asked to give either a numerical or categorical rating. With the contingent ranking method, respondent are also presented with a number of bundles, but are only asked to rank them in order of preference. Finally, with the choice experiment method, respondents are presented with a number of bundles and asked to choose their most preferred bundle. The contingent rating method is not consistent with utility theory since the respondent is not directly comparing the alternatives. The contingent ranking and choice experiment methods are consistent with utility theory as long as the choice set includes the status quo option. For a complete discussion of the choice experiment method, see Adamowicz et al.

The increasing popularity of SP approaches has naturally led to an investigation of the consistency (or lack thereof) between RP and SP methods. Many researchers have found evidence in favor of consistency between the two methods, while others have found only limited evidence in favor of consistency. If the two methods were found to be consistent, this would indicate that they are separate, legitimate methods of gathering information about preferences. On the other hand, if the two methods are found to be inconsistent, the researcher is left to speculate about the root of the inconsistency. Possible explanations for a rejection of consistency might be that SP and RP approaches are measuring inherently different processes, the particular functional form used in the estimations, or limitations in the data gathering process. For a detailed treatment of RP-SP consistency, see Azevedo et al., Cameron, Adamowicz et al., Cummings et al., or Carson et al.

Aside from the RP and SP consistency debate, researchers have also focused on how the various SP approaches compared to each other (i.e., are the various SP approaches consistent with each other). The results of these studies have been mixed. Numerous researchers have compared the contingent valuation method with the choice experiment method (e.g., Christie and Azevedo, Boxall et al., Hanley et al., Mogas et al., and Colombo et al.), while others have compared the choice experiment with the contingent ranking method (Foster and Mourato).

The approach we take in this paper is to reduce the complexity of the comparison by examining the attribute rankings derived from a choice experiment with those from a simple attribute-ranking exercise. The choice experiment method has come to be regarded as a reliable, utility-consistent value estimation technique. One output of the estimation of the choice experiment method is an implied ranking of the attributes of an environmental resource. At the most basic level, this ranking should be consistent with an exercise that asks respondents to explicitly rank attributes. It would certainly be problematic if respondents explicitly state that they or prefer attribute A over attribute B, yet the choice experiment exercise indicates that they value B over A.

Our methodology adds to the literature in that it allows for a clean, intuitive comparison between ranking mechanisms—one implicit and the other explicit. As noted above, comparisons will be made on a sample-wide basis (i.e., do the two mechanisms result in rankings that are consistent on average?), as well as on an individual basis (i.e., do the two mechanisms result in rankings that are consistent for each individual respondent?).

In order to make these comparisons, we use data from two separate survey questions: (1) a choice experiment which can be used to estimate marginal rates of substitution for the attributes, and (2) an importance-point question that asks respondents to allocated points to various attributes. We should note that though importance-point questions are often used in other fields to imply a cardinal ranking of the attributes, we only rely on the ordinal ranking given by the allocation of points.

MODEL SPECIFICATION

The basis of our comparison is the choice experiment model, so it will be useful to briefly review the mechanics of that approach. The choice experiment approach can essentially be thought of as a more complex version of the contingent valuation approach. With the contingent valuation approach, the respondent is asked to choose between the status quo and an alternative. In the choice experiment approach, the respondent is asked to choose between the status quo and multiple alternatives that differ in terms of the attributes that make up the bundle. The same information gathered in the choice experiment approach could also be gathered by asking multiple contingent valuation question, but an intractably large number of contingent valuation questions would be needed.

As with the contingent valuation model, random utility theory (RUT) forms the behavioral foundation of the choice experiment model. RUT postulates that the respondent n's utility function for alternative i takes the form

$$U_{ni} = V_{ni} + \varepsilon_{ni}.$$

Though the respondent has full knowledge of her utility function, U_{ni}, utility cannot be directly observed by the researcher. While the researcher can view the deterministic component of the utility function, V_{ni}, a portion, ε_{ni}, is unobservable.

The choice of alternative i represents a discrete choice from a set of alternatives. The probability of respondent n choosing alternative i from choice set C_n can be expressed as

$$\begin{aligned}\Pr\left(i \mid C_n\right) &= \Pr\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}\right) \qquad \forall j \in C_n \\ &= \Pr\left(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni}\right) \\ &= \Pr\left(V_{ni} - V_{nj} > \xi_n\right)\end{aligned}$$

where $\xi_n = \varepsilon_{nj} - \varepsilon_{ni}$. By assuming that the error terms ξ_n are independently, identically distributed Gumbel random variables, we can express the probability of choosing alternative i from choice set C_n as

$$\Pr\left(i \mid C_n\right) = \frac{\exp\left(V_{ni}\right)}{\sum_{j \in C_n} \exp\left(V_{nj}\right)},$$

which is the familiar multinomial logit model (MNL).

The researcher must specify the functional form of V_{ni}. A common parameterization is

$$V_{ni} = \alpha_i + \beta_T\left(Y_n - Bid\right) + \beta_Z \bar{Z},$$

where α_i is an alternative-specific constant that captures the effect of systematic but unobservable factors of the respondent's choice, Y_n is respondent n's income, the β s represent parameters to be estimated, and $\bar{Z}$ is composed of variables measuring attributes of the choice site.

With this model, the β s can be used to estimate the MRS between attributes. For example, if $\bar{Z}$ contained the attributes "water clarity" and "number of algae blooms per year," the rate at which respondents are willing to trade off clarity and algae blooms takes the form

$$MRS = -\frac{\beta_{Clarity}}{\beta_{Algae}}.$$

Assuming one of the attributes is measured in terms of dollars, the MRS between any attribute and dollars, or the willingness to pay for that attribute, can be calculated as

$$MRS = -\frac{\beta_{Attribute}}{\beta_{Dollars}}.$$

These marginal rates of substation are commonly referred to as part-worths in the choice experiment literature.

APPLICATION

The data we use to make ranking comparison comes from a survey of residents living in the vicinity of Clear Lake, Iowa. Clear Lake, located in north central Iowa, is the state's third largest lake.[2] The lake is used intensively for recreation due to its large size, its proximity to several communities, and the lack of substitute sites in the immediate area.

Water quality at Clear Lake has deteriorated significantly over the past several decades. This has been the result of phosphorus loading and the fact that the lake is rather shallow, having a maximum depth of 19 feet and an average depth of 9.5 feet. Because the lake is relatively shallow, wind and recreation activity (boat traffic) are continually disturbing the sediments located on the lake bottom. Contaminants located in the sediments are continuously resuspended into the water column, which has led to a drastic reduction in water clarity, increased numbers of algae blooms, deterioration of water color and odor, and a reduction in the quality and diversity of fish populations present in the lake.

[2] For a complete discussion of the Clear Lake survey see Azevedo, Herriges, and Kling 2001.

As part of a 2001 Iowa Department of Natural Resources Diagnostic and Feasibility study of Clear Lake, a random sample of 600 residents in the vicinity of Clear Lake were sent a mail survey designed to elicit information concerning local residents' willingness to pay for improving the water quality at Clear Lake. Survey recipients were told that they would be paid $5 upon the return of a completed survey. Of the 600 surveys sent, 134 were undeliverable. Among the deliverable surveys, 249 were returned, resulting in a 53% response rate. The CE data was incomplete in 19 cases. So only data from the 230 respondents who provided completed surveys is included in our analysis.

The attributes varied in the CE question were water color and clarity, number of algae blooms, lake odor, fish population, and the dollar bid. Table 1 summarizes the attributes and their levels of variations. The current water quality conditions are shown in italics. Qualitative attributes were accounted for using an effects code matrix, while quantitative attributes were accounted for by using the midpoint of the attribute levels.

Each respondent was asked to complete eight choice tasks. Figure 1 shows a typical choice task. The respondent was asked to select between Option A, Option B, or the status quo (i.e., no change in the quality of the lake and no additional tax bill). Options A and B differ from each other in terms of the levels of the various attributes.

The data gathered through the CE questions and analyzed with the MNL model can be used to estimate the parameters of the indirect utility. As described above, these parameter estimates can be used to further estimate the willingness to pay for changes in the attributes of the lake.

In addition to the CE questions, respondents were also asked to explicitly rank lake attributes in terms of their importance. Each respondent was told to, "Assume you have a total of

100 importance points to assign to the lake characteristics below. Please indicate the importance of each item by allocating your 100 points among the items on this list. To indicate one item is more important to you than another, you should allocate more points to it. You do not need to give points to all of the items, but remember that the total needs to equal 100." The list of attributes included water clarity, hard clean sandy lake bottom in swimming areas, lack of water odor, diversity of wildlife seen at Clear Lake, diversity of fish species/habitat, quantity of fish caught, and safety from bacterial contamination.

In the next section we summarize and compare the rankings generated by analysis of the data generated by these two questions.

RESULTS

Table 2 shows the parameter estimates generated from the maximum likelihood estimation of the choice experiment model. Almost all parameter estimates are statistically significant, and the signs of the coefficients confirm our *a priori* expectations in that attribute levels associated with deterioration in the lake condition from the current condition have negative signs, while those associated with improvements have positive signs. As expected, the sign of the tax coefficient is negative indicating that respondents were less likely to choose an option associated with higher tax payments. The overall fit of the model is good by conventional standards with a Rho^2 value of 0.134.

Prior to discussing the marginal rates of substitution, or part-worths, implied by the parameter estimates in Table 2, it is first necessary to explicitly describe the comparisons that we wish to make. Obviously, there is not complete symmetry between the attributes described in the CE question and the importance point (IP) question. The IP question asks the respondent to consider a more detailed list of attributes than the CE question. Despite the differences, the two questions do imply an ordinal ranking of two attributes that are included in both, namely water clarity and algae blooms.

Sample-wide Rankings

Using the CE parameter estimates, the part-worth for water clarity is $25.89 while the part-worth for algae blooms is -$105.20, which implies that, at the margin, a reduction in algae blooms is preferred to improved water clarity.[3]

The data gathered through the IP question tells a similar story. By averaging the number of importance points allocated by respondents to each attribute, we can calculate the average ranking of attributes. The average number of IP points allocated to water clarity is 21.29 while the average number of IP points allocated to algae blooms is 33.37. This too implies that, at the margin, a reduction in algae blooms is preferred to improved water clarity.

The results of this exercise indicate that on a sample-wide basis the ranking implied by the CE model is consistent with the IP question which asks respondents to explicitly rank the attributes in order of preference.

Individual Rankings

Since each respondent is asked eight CE questions, it is possible to estimate a separate CE model for each respondent (as opposed to using all data from all respondents to estimate a single CE model). In this manner we generate a CE ranking for each respondent, which can be compared to the IP ranking.

---

[3] The negative willingness to pay indicates that respondents would be willing to accept $105 for an additional algae bloom. In other words, they would be willing to pay $105 for one less algae bloom.

Of the 230 respondents who provided completed surveys, 62 allocated the same number of IP points to the water clarity and algae bloom attributes, indicating indifference. Of the remaining 168 respondents, the CE estimation routine would not converge in 23 cases, most likely due to the inefficient use of data when estimating a model for a single recipient. Of the remaining 145 respondents, the CE and IP rankings were consistent only 40% of the time.

CONCLUSION

Choice experiments are widely used in many different applications because of their ability to evaluate many different attributes of environmental amenities and their sound theoretical basis. In particular, CEs allow the researcher to estimate marginal rates of substitution between the attributes of the environmental resource. These MRSs provide an implicit ranking of the attributes, which we compare with other ranking mechanisms. In this paper we introduce a method for testing for the internal consistency of CEs by comparing the attribute ranking implicitly generated by a CE with that generated explicitly by an attribute-ranking exercise. Using data gathered through a unique survey in which respondents completed both a CE exercise and an attribute-ranking exercise, we find that sample-wide rankings are consistent across the two methods, but individual comparisons are only consistent across the two methods 40% of the time.

Table 1: Choice experiment attributes and levels

Attribute	Level
Water color and clarity	Brown color. Clarity = 1” to 5” *Brown color. Clarity = 6” to 1 foot* Blue color. Clarity = 2 feet to 4 feet Blue color. Clarity = 5 feet to 8 feet
Algae blooms	Almost constant algae blooms. Frequent swim advisories and beach closings *10 – 12 algae blooms / yr. Occasional short-term swim advisories and possible beach closings* 6 – 8 algae blooms / yr. Occasional swim advisories 3 – 4 algae blooms / yr. Very infrequent swim advisories
Lake odor	Always a strong ‘fishy’ odor *Mild odor, occasionally strong* Occasional mild odor No odor
Fish	Mostly bullhead (rough fish) *Mostly bullhead, with good walleye* Mostly desirable fish, but few walleye Mostly desirable fish with many walleye
Tax	$0 $225 (payable in five $55 instalments over a five year period) $700 (payable in five $140 instalments over a five year period) $1040 (payable in five $208 instalments over a five year period)

Attribute levels shown in italics relate to the current levels.

Table 2: Results of the choice experiment MNL model (t-statistics are shown in parenthesis)

Parameter	**Choice Experiment Model**
Intercept (β)	-0.323 (-2.87)**
Clarity (β_{CL})	0.021 (13.80)**
Algae (β_{AL})	-0.084 (-9.78)**
Odor A (β_{OD}^{1})	-0.196 (-2.93)**
Odor C (β_{OD}^{3})	0.138 (2.16)*
Odor D (β_{OD}^{4})	0.045 (0.67)
Fish A (β_{FI}^{1})	-0.587 (-7.85)**
Fish C (β_{FI}^{3})	-0.171 (-2.59)**
Fish D (β_{FI}^{4})	0.726 (9.96)**
Tax (β_{TAX})	$-0.080E^{-02}$ (-8.33)**
Log likelihood	-1757.185
Log likelihood (constants)	-2028.469
Chi-squared	542.569
Rho-squared	0.134

Figure 1: Typical choice scenario

Choice Scenario 1

Please indicate which of the two management options you prefer, if either.

Condition Indicator	Option A	Option B	Retain current water quality
Water color and clarity	Brown water. Visibility = 6" to 1 foot	Brown water. Visibility = 1 " to 5 "	I prefer to retain Clear Lake's current water quality See enclosed reference sheet for details
Algae blooms	6 – 8 algae blooms / yr. Occasional swim advisories	10 – 12 algae blooms / yr. Occasional short-term swim advisories and possible beach closings	
Lake Odor	Always a strong 'fishy' odor	Always a strong 'fishy' odor.	
Fish	Mostly desirable fish with many walleye.	Mostly desirable fish, but few walleye.	
Increase in your tax bill ($)	$700 (payable in five $140 installments over a five year period)	$225 (payable in five $55 installments over a five year period)	

Check only one box

❑ I prefer Option A	❑ I prefer Option B	❑ I prefer the current condition

REFERENCES

Adamowicz, Wiktor, Peter Boxall, Michael Williams, Jordan Louviere, “Stated Preference Approaches for Measuring Passive use Values: Choice Experiments and Contingent Valuation,” *American Journal of Agricultural Economics,* 80(1998): 64-75.

Adamowicz, Wiktor, Jordan Louviere, and Michael Williams. "Comparing Revealed and Stated Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 26(1994): 271-292.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Shuman, “Report of NOAA Panel on Contingent Valuation,” *Federal Register,* 58(1993): 4016-4614.

Azevedo, Christopher D., Joseph A. Herriges, and Catherine L. Kling, “Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations,” *American Journal of Agricultural Economics*, 85(2003): 525-537.

Azevedo, Christopher D., Joseph A. Herriges, and Catherine L. Kling, “Valuing Preservation and Improvements of Water Quality in Clear Lake,” Center for Agricultural and Rural Development, Staff Report 01-SR 94, March 2001.

Boxall, Peter, Wiktor Adamowicz, Michael Williams, Joffre Swait, and Jordan Louviere, “A Comparison of Stated Preference Approaches to the Measurement of Environmental Values,” *Ecological Economics,* 18(1996): 243-253.

Cameron, Trudy, "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics,* 68(1992a): 302-317.

Christie, Mike and Christopher D. Azevedo, "Testing the Consistency Between Standard Contingent Valuation, Repeated Contingent Valuation and Choice Experiments," Forthcoming in the *Journal of Agricultural Economics*.

Cummings, Ronald, Steven Elliott, Glenn Harrison, and James Murphy, "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, 105(1997): 609-21.

Carson, Richard, Nicholas Flores, Kerry Martin, and Jennifer Wright, "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics*, 72(1996): 80-99.

Colombo Sergio, Javier Calatrava-Requeno, and Nick Hanley, "Analyzing the Social Benefits of Soil Conservation Measures Using Stated Preference Methods," *Ecological Economics*, 58(2006): 850-861.

Desvousges, W., F. Johnson, R. Dunford, S. Hudson, K. Wilson, and K. Boyle. "Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy," *Contingent Valuation: A Critical Assessment*. J. Hausman, ed., pp. 91-164, New York: North Holland, 1993.

Diamond, Peter, and Jerry Hausman, "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives*, 8(1994): 45-64.

Foster, Vivien, and Susana Mourato, "Testing for Consistency in Contingent Ranking Experiments," *Journal of Environmental Economics and Management*, 44(2002): 309-28.

Hanemann, Michael, "Valuing the Environment through Contingent Valuation," *Journal of Economic Perspectives*, 8(1994): 19-43.

Hanley, N., D. Macmillan, R. Wright, C. Bullock, I. Simpson, D. Parrsisson, and R. Crabtree, "Contingent Valuation Versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland," *Journal of Agricultural Economics,* 49(1998): 1-15.

Kahneman, D., and J. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management*, 22(1992): 57-70.

Kemp, M., and C. Maxwell, "Exploring a Budget Context for Contingent Valuation." in *Contingent Valuation: A Critical Assessment*, J. Hausman, ed., pp. 217-70, New York: North Holland, 1993.

Loomis, J., A. Gonzalez-Caban, and R. Gregory, "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics*, 70(1994): 499-506.

Louviere, Jordan, "Why Stated Preference Discrete Choice Modeling is NOT Conjoint Analysis (and what SPDCM is)," Memetrics White Paper, #1, 2000.

Mogas, J, Riera P and Bennett J., "Accounting for Afforestation Externalities: A Comparison of Contingent Valuation and Choice Modelling," *European Environment* 15(2005): *44-58.*

Portney, P., "The Contingent Valuation Debate: Why Economists Should Care," *Journal of Economic Perspectives*, 8(1994): 3-17.

Randall, A., "A Difficulty with the Travel Cost Method," *Land Economics*, 70(1994): 88-96.