A Split-Sample Test of Experimental Designs for Stated Choice Models of Environmental Valuation

Frank Lupi,[1,2] Michael D. Kaplowitz,[3] and John P. Hoehn[1]

lupi@msu.edu, kaplowit@msu.edu, & hoehn@msu.edu

[1]Department of Agricultural Economics
[2]Department of Fisheries and Wildlife
[3]Department of Resource Development
Michigan State University
East Lansing, MI 48824

May 14, 2003

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, July 27-30, 2003

Abstract

A stated choice model is used to estimate wetland mitigation preferences. In a split sample mail survey, a main effects design is compared to a randomized design. Although randomized designs estimate main effects less efficiently, several policy relevant interactions were found to be significant, suggesting some merits of randomized designs.

Introduction

Stated choice methods aimed at valuing the attributes of non-market goods, rather than the goods and services themselves, are increasingly popular among valuation researchers. Stated choice techniques (sometimes referred to as choice experiments or conjoint) can be thought of as an extension of the contingent valuation method for nonmarket goods. Both approaches ask respondents to state a preference over alternatives. Stated choice models are also widely used in marketing applications to estimate preferences over products with attributes that are not currently available and researchable through market data (Carson et al, 1994).

In stated choice studies, respondents are typically presented with descriptions of two or more recreation sites, environmental goods, or environmental programs and asked to indicate which of the alternatives they prefer. By varying the attributes of the alternatives, econometric methods can be used to estimate respondents´ preferences for the attributes. Typically, this is done by estimating a logit or probit model within the framework of a random utility model. This paper uses a study of wetland restoration to address the question of how to define attribute levels and combine the attributes that make up the alternatives that enter a stated choice study.

Experimental Designs

In implementing a stated choice study, one of the fundamental decisions a researcher faces is how to arrange the attributes of the goods to be presented to respondent? This is essentially a question of what experimental design to use. The experimental design is the set of attribute levels, how these are then bundled into alternative goods or programs, and how the alternatives are combined into the choice set presented to a respondent.   To fix concepts, refer to Table 1 which depicts a stylized binary choice question. The two alternatives, A and B, are made up of the attributes and are combinations of the available attribute levels. The columns represent the alternatives, and the rows represent attributes. For each attribute, there is a set of attribute levels that are pre-defined by the researchers. The matrix of the pairs of alternatives across all the respondents in referred to as the experimental design.

A typical stated choice study will select the attributes to be valued, and then select the levels of attributes that will be used in the study. While alternatives can then be formed by taking the full factorial combination of each of the levels, this often results in a very large number of alternatives. To reduce the number of alternatives, some researchers rely on fractional factorial designs such as a main-effects plan. An main-effects plan, wherein each of the attributes are combined into alternatives such that the attributes are orthogonal, permits the estimation of the independent effect of each attribute. While such designs are reasonably efficient, they do not permit the estimation of higher order effects such an interactions among variables (Lazari and Anderson, 1994). Estimation of interaction effects requires more complex designs that increase in size. Potential interaction effects might be very important for a policy analysis. In addition, if a researcher is interested in examining non-linear effects for a single variable (attribute) in the utility function, then one must include multiple levels for that variable. While feasible in a main-effects design plan, this too, increases the size of the resulting design plan.   The research reported here uses a split-sample survey to assess two alternative experimental designs: 1) a typical main-effects design, and 2) a fully randomized design that permits nonlinear marginal utility and interaction effects.

To generate the fixed design plan, we treat the attribute vectors for alternative A and for alternative B as separate design variables (this approach is discussed by numerous choice modelers including on page

133 of Louviere, Hensher and Swait 2000). In essence, this design approach draws a main-effects plan on the vector of attributes, $X = X^A | X^B$, defined by "stacking" the vectors for each alternative – see Figure 1. For the other sub-sample of individuals, the stated choice questions consisted of independent random draws from a joint uniform distribution over the integer values spanning the range of attribute levels used in the fixed design. Each draw resulted in a distinct set of attribute levels so that each of the choice scenarios in this sub-sample was unique.

Binary Choice Models: Logit or Probit

Stated choice models and the theory underlying them is well developed in the literature. We briefly review it here. In a typical specification of a random utility model, utility is a function of the attribute values that make up an alternative, and the utility function is assumed to have random errors. When the underlying errors have an extreme value distribution, the probability of a respondent choosing alternative A among the alternatives A and B in choice scenario j is given by:

$$P_j = \frac{\exp(D_j)}{1 + \exp(D_j)} \tag{1}$$

The vector of attributes, q, enter the utility function linearly and thus:

$$D_j = \beta_1 (x_{1j}^A - x_{1j}^B) + \beta_2 (x_{2j}^A - x_{2j}^B) + \cdots + \beta_m (x_{mj}^A - x_{mj}^B) \tag{2}$$

where $(x_{mj}^{A} \& x_{mj}^{B})$ represents the difference in the level of site attribute m between the two alternatives *A* and *B* in choice scenario j. Similarly, had the underlying errors been normally distributed we would define the binary choice model using a probit model as a function of *D*.

That the above parameter vector is estimated on the difference of the attribute vectors for the two alternatives is one of the key features distinguishing designs for binary choice models from designs for the typical linear models with continuous dependent variables. The distinction is worth noting because most of the available design plans are actually derived based on the assumption that good designs for standard linear models are also good designs for binary choice models (Kuhfeld). A main-effects design for the logit model, i.e., a design that sought to minimize the number of distinct scenario combinations for identifying the main effects in (1), would need to be based on these attribute differences (Kaninnen). Alternatively, the commonly used main-effect plans we examine for the "stacked" design based on $X=X^A|X^B$ are constructed for identifying the main-effects of X in a simple linear model, but in the context of a logit defined on the difference X^A-X^B, such designs are actually somewhat redundant in terms of minimally identifying the main effects of (1).

Data

We estimated the effects of using alternative design plans within a study of preferences over wetland mitigation projects. In our mail survey, we elicited a choice between an impaired and a restored wetland that were described by attributes which include wetland acreage and wetland habitat quality for various species of flora and fauna. The choice question asked people if a restrored wetland offset the loss of the impaired wetland, and the choice questions and context are the same as presented in Lupi et al. (2002).

In the survey, each respondent was given five wetland choice questions. Our stated choice survey was implemented via the mail and received a 46% response rate (despite being conducted during the anthrax attack of late 2001). The survey was sent to a statified random sample of Michigan residents drawn from the state drivers license data base. The stratification ensured that the geographic distribution across counties for the sample matched that of the state population. Our focus here is on the analysis of the split sample test of the two experimental designs. One subset of individuals was sent a survey booklet with the stated choice questions based on a main-effects design plan for X, while the other subset was sent an individually produced survey booklet with distinct alternatives constructed of random attribute levels.

The variables used in the wetland mitigation choices are presented in Table 2. The variables took between three and five levels each. As a result, our fixed design involves selecting a main-effects orthogonal plan from the full factorial for the "stacked" X with 4^2 by 3^12 possibilities. The full factorial design for this set of attributes and levels contains over eight million alternatives. Using Addelman's classic set of design plans, the resulting main-effects plan contains 64 distinct choice scenarios. The resulting data from the fixed-design sub-sample contains 1,463 binary choices over alternative wetland mitigation projects.

The other sub-sample of individuals was sent a survey booklet in which the stated choice questions consisted of independent random draws from a joint uniform distribution over the integer values spanning the range of attribute levels used in the fixed design. Each draw resulted in a distinct set of attribute levels so that each of the choice scenarios in this sub-sample was unique. The surveys for this group were custom-produced on a color laser printer by creating a spreadsheet with thousands of draws from the uniform distribution over the joint parameter values, and using the spreadsheet in a merge file to individually create each distinct randomized booklet. The data resulting from this fully randomized design contains 1,146 binary choices over alternative wetland mitigation projects.

Estimation Results

For the basic main-effects specification of preferences, Table 3 presents the estimated wetland mitigation preference parameters from a random effects probit on the pooled dataset. The negative parameter on the constant indicates that all else equal, people tend not to find that the restored wetland offsets the loss. However, the positive parameter on the acres variable indicates that if the restored wetland was larger people were more likely to choose the restored wetland. The wetland "type" variables did not have a significant effect on mitigation choices. Gaining public access and trails at the restored wetland made people more likely to favor it, although the effect for trails is not significant at the 5% level. The "Low_1" to "Low_4" variables represent the low habitat quality for the habitat attributes listed in Table 2 while the "Hi" variables represent the excellent habitat quality for these same attributes. If a restored wetland lowered habitat quality people were less likely to find it acceptable. Conversely, if a restored wetland increased habitat quality, people were more likely to find it acceptable. Put differently, people require an additional acreage premium to compensate them for any declines in habitat quality. The above results are consistent with previous findings from a pilot survey (Lupi et al, 2002). The rho parameter indicates a significant correlation across choices made by an individual.

Interestingly, the habitat variables seem to exhibit a form of non-linearity in their effects. That is, for all of the habitat attributes, the effect of moving from the low habitat quality to the medium quality is about double the benefit of moving from the medium level to the high level. These effects are graphed by the bold lines in Figure 2. However, we must note that the habitat quality variables are not based on an underlying cardinal scale. That is, there is no reason to assume that respondents perceive the distance between the low and medium levels to be the same as the distance between the medium and high levels. For example, in the horizontal axis of Figure 2, when considering the scale from"poor" to "excellent," individuals may simply place "good" closer to "excellent" than to "poor." This effect is plotted in Figure 2 and illustrates the difficulty of discerning non-linear effects when dealing with qualitative discrete variables.

Design Performance for Estimating Main Effects

The effectiveness of the two designs for estimating the main-effects was compared. To assess design efficiency, one usually takes some function of estimated covariance matrix to measure the performance of a design. For example, D-optimal designs minimize the determinant of the Fisher information matrix (i.e., the inverse of the covariance matrix evaluated at the estimated parameters). Here, we follow this approach and compare the designs using the ratio of the determinant of the information matrix, I, evaluated using $\hat{\beta}$, for each of the two design matrices. The ratio of the measures is,

$$|I(X^{M};\ \hat{\beta})| \,/\, |I(X^{R};\ \hat{\beta})| = 0.79$$

where X represents the design matrix for the design being evaluated (the main-effects plan, M, or the randomized plan, R). Both matrices must be evaluated at a parameter vector so the estimated parameter vector was selected as this is our best estimate of the true parameters. The resulting ratio equals 0.79 suggesting that the efficiency loss of using the randomized design plan to estimate the main effects is not pronounced.

Interaction Effects

The randomized design provides an opportunity to investigate the importance of interaction effects. For our application, there are 62 possible interactions in this model. When all of these interactions were run at once, none of the interactions were significant by themselves. A variety of models were then run to see if subsets of the variables mattered. Across all model runs, we observed that the main effect parameters are robust to interactions, both in term of sign, magnitude, and significance levels. Moreover, only four of the interaction effects were found to be significant at $p<0.1$ across a variety of models. Of these, two involve an interaction between "type" and "habitat." Since wetland "type" was not a significant variable by itself, these could be policy relevant interactions. Of course, without the more general designs, one could not test for such interactions.

Discussion

How practical is it to implement a fully randomized design? The answer depends on the mode for collecting the survey data. For internet surveys, which are increasingly popular, the added complexity and cost of the fully randomized design is trivial. The same is true for face-to-face interviews conducted using computer assisted interviewing software. Randomized designs are even feasible with simple paper instruments (An et al., 2002). Although we´ve shown it is feasible in the context of a mail survey, we judge the difficulty of using the randomized design to be the highest when a survey is implemented via the mail. The reason for this is that the randomized design requires individualized survey instruments for each person in the survey. We accomplished this task by using the merge feature of our wordprocessor, and then printing batches of the surveys on a high speed laser-printer. None the less, this approach is labor and time intensive, and can cost more than standard photocopying. We experienced paper and cartridge costs in the range of 10 cents a page. Our survey included several pictures of wetlands, and image quality from the laser printer was superior to standard photocopies. Factoring in the labor, the resulting production costs for the randomized design are somewhat higher than the costs of a high quality print-job for the 64 versions of the main-effects design.

In our results, as expected, the randomized design was strictly less efficient than the main effects plan for estimating a model with only main effects. However, our results also found a few significant interaction effects among the habitat quality variables. These interaction effects cannot be identified in the main-effects plan. Thus, while less efficient, the randomized designs might be preferred due to their to ability to detect such interaction effects – effects which would be difficult to guess at a priori. Moreover, the fully randomized design approach is particularly tractable and extremely simple to implement for internet and computer-based collection of stated choice data.

Acknowledgments

The research reported here was supported by an EPA STAR Grant (R827922), by the Michigan Department of Environmental Quality, and the Michigan Sea Grant College Program, project number R/EP9 from the Office of Sea Grant, NOAA, U.S. Department of Commerce and funds from the state of Michigan. The U.S. Government is authorized to produce and redistribute reprints for government purposes notwithstanding any copyright notation appearing herein. The views expressed here are those of the authors who are solely responsible for any errors.

References

An, Li, Frank Lupi, Jiangou Liu, M. Linderman, and J. Huang, "Modeling the Switch from Fuelwood to Electricity: Implications for the Conservation of Giant Panda Habitat," *Ecological Economics*: 42:445-457, 2002.

Carson, R., J. Louviere, D. Anderson, P. Arabie, D. Bunch, D. Hensher, R. Johnson, W. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, J. Wiley, 1994, "Experimental Analysis of Choice," *Marketing Letters*, 5, 351-67.

Kanninen, B. J. "Optimal Design of Choice Experiments for Nonmarket Valuation." Paper presented at EPA's Environmental Policy and Economics Workshop, *Stated Preference: What Do We Know? Where Do We Go?*, Washington, DC, October, 2000.

Kuhfeld, W.F., R.D. Tobias, and M. Garratt. "Efficient Experimental Design with Marketing Applications". *Journal of Marketing Research*. 31 (1994): 545-557.

Lazari, A.G. and D.A. Anderson, 1994, "Designs of Discrete Choice Experiments for Estimating Both Attribute and Availability Cross Effects," *Journal of Marketing Research*, 31, 375-83.

Louviere, J., D. Hensher and J. Swait 2000. *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.

Lupi, F., M. D. Kaplowitz, and J. P. Hoehn. 2002. The Economic Equivalency of Drained and Restored Wetlands in Michigan. *American Journal of Agricultural Economics* 84 (5):1355-1361.

Table 1: Stylized Binary Stated Choice Question

	Alternative A	Alternative B
Attribute X_1	X_1^A	X_1^B
Attribute X_2	X_2^A	X_2^B
	!	!
Attribute X_K	X_K^A	X_K^B
Which is Best?	T	

Table 2: Wetland Attributes and Levels

Variable	Levels	#
Baseline acres	5, 7, 9, 12, 15	5
Restored acres (multiple of baseline)	0.8, 1, 1¼, 1½, 2	5
Public assess/trails	No, yes, yes with trails	3
Habitat: Frogs/turtles	poor, good, excellent	3
Habitat: Song birds	poor, good, excellent	3
Habitat: Wading birds	poor, good, excellent	3
Habitat: Wild flowers	poor, good, excellent	3

Table 3: Estimated Wetland Mitigation Preference Parameters from Random Effects Probit

Variable	**Parameter**	***p-value***
Constant	-0.199	0.0009
Acres	0.056	0.0000
Marsh	-0.059	0.3240
Wooded	-0.058	0.3277
Public	0.262	0.0000
Trails	0.096	0.0882
Low_1	-0.349	0.0000
Low_2	-0.327	0.0000
Low_3	-0.365	0.0000
Low_4	-0.190	0.0008
Hi_1	0.151	0.0112
Hi_2	-0.177	0.0025
Hi_3	0.180	0.0031
Hi_4	0.088	0.1129
Rho	0.475	0.0000

N = 532; Choices = 2,609;

Overall, 61% predicted correctly; 59% "yes" predicted correctly; 65% "no" predicted correctly.

Figure 1: Illustration of What is Meant by "Stacking" the X's and Then Drawing a Main Effects Plan for $X = X^A * X^B$

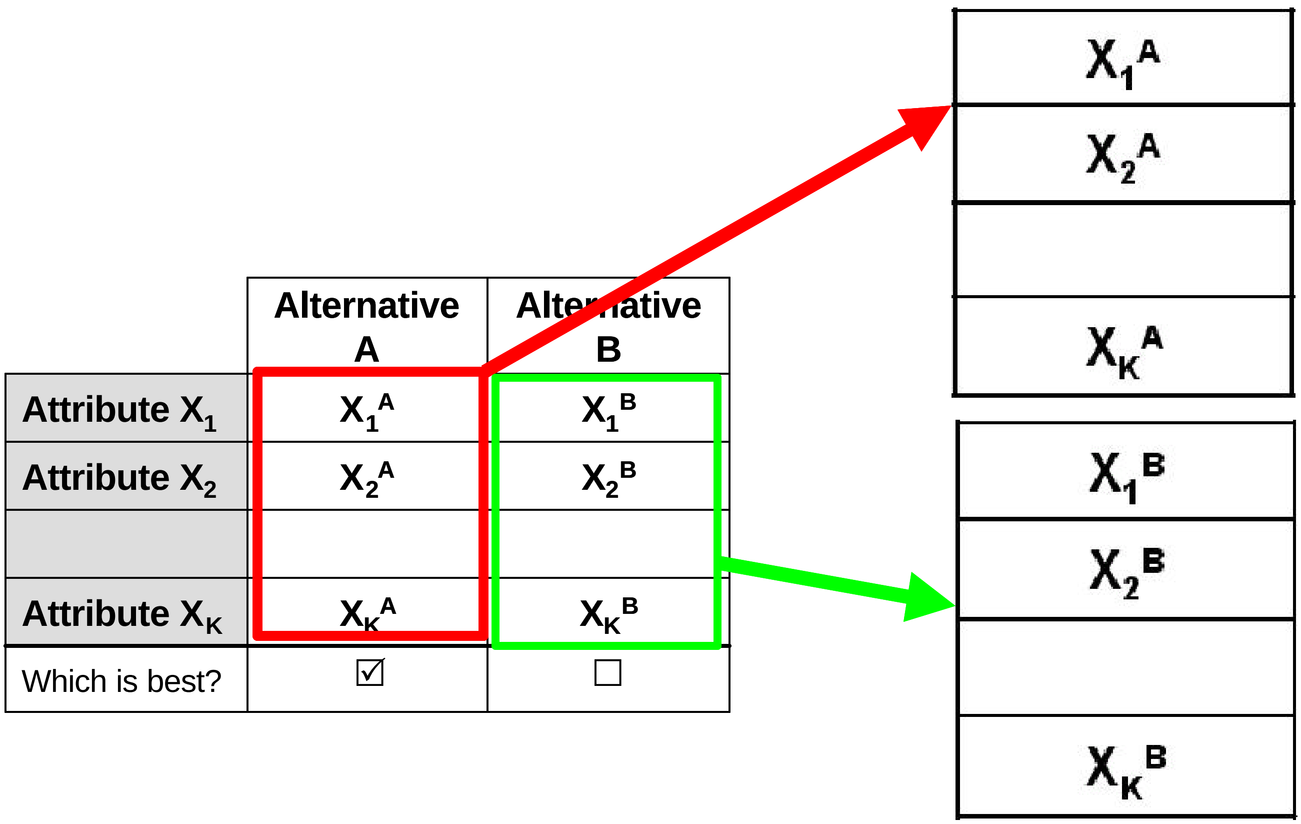

Figure 2: Non-Linear Habitat Variables