

Chapter 5

Choice Experiments

Thomas P. Holmes, Wiktor L. Adamowicz and Fredrik Carlsson

Abstract There has been an explosion of interest during the past two decades in a class of nonmarket stated-preference valuation methods known as choice experiments. The overall objective of a choice experiment is to estimate economic values for characteristics (or attributes) of an environmental good that is the subject of policy analysis, where the environmental good or service comprises several characteristics. Including price as a characteristic permits a multidimensional, preference-based valuation surface to be estimated for use in benefit-cost analysis or any other application of nonmarket valuation. The chapter begins with an overview of the historical antecedents contributing to the development of contemporary choice experiments, and then each of the steps required for conducting a choice experiment are described. This is followed by detailed information covering essential topics such as choosing and implementing experimental designs, interpreting standard and more advanced random utility models, and estimating measures of willingness-to-pay. Issues in implementing and interpreting random utility models are illustrated using a choice experiment application to a contemporary environmental problem. Overall, this chapter provides readers with practical guidance on how to design and analyze a choice experiment that provides credible value estimates to support decision-making.

Keywords Alternatives • Attributes • Choice set • Discrete-choice analysis • Experimental design • Nonmarket valuation • Passive-use value • Policy analysis • Questionnaire • Random utility model • Stated preference • Survey • Trade-offs • Use value • Welfare measures • Willingness to pay

T.P. Holmes (✉)

U.S. Forest Service, Southern Research Station, Research Triangle Park, NC, USA
e-mail: tholmes@fs.fed.us

W.L. Adamowicz

University of Alberta, Edmonton, AB, Canada
e-mail: vic.adamowicz@ualberta.ca

F. Carlsson

University of Gothenburg, Gothenburg, Sweden
e-mail: fredrik.carlsson@economics.gu.se

Stated-preference methods of environmental valuation where market transaction data have limitations or do not exist in a form useful for measurement of economic values have been used by economists for decades. There has been an explosion of interest during the past two decades in a class of stated-preference methods that we refer to as choice experiments (CEs).¹ The overall objective of a CE is to estimate economic values for characteristics (or attributes) of an environmental good that is the subject of policy analysis, where the environmental good or service comprises several characteristics. Including price as a characteristic permits a multidimensional, preference-based valuation surface to be estimated for use in benefit-cost analysis or any other application of nonmarket valuation.

CEs have gained popularity because they offer several potential advantages relative to other valuation methods.

- CEs can provide values for changes in a single characteristic or values for changes in levels of characteristics or values for multiple changes in characteristics, resulting in a response surface of values rather than a single value.
- Because characteristics are experimentally manipulated and presented to respondents, they are typically exogenous, not collinear, and can reflect characteristic levels outside the range of the current market or environment. This is in contrast with revealed preference data that are often collinear, may be limited in variation, and may be endogenous in explaining choices.
- Just like contingent valuation, CEs can be used to assess preferences or trade-offs in behavioral settings (e.g., recreation site choice) that are relevant for measuring use values or in settings that are used to measure passive-use values like voting or referenda.
- The CE presentation format makes choices relatively easy for respondents, and attributes and their levels can be customized such that they are realistic for respondents (i.e., reflecting specific conditions they face). These types of choices are often similar to those consumers face in markets.
- There is experience in using CEs in several disciplines, including marketing, transportation, and health economics. Environmental economists are typically more interested in welfare measures than are practitioners in other fields.
- The use of experimental design theory increases the statistical efficiency of the parameters estimated so that smaller samples may be used, which reduces implementation costs.

¹The label “choice experiment” is a source of controversy. The previous edition of this book used the phrase “attribute-based methods” (which included ratings and rankings), while others have referred to this approach as “attribute-based stated choice methods,” “choice-based conjoint analysis,” and a host of other names. Carson and Louviere (2011) recommended the term “discrete choice experiment” to reflect the fact that these methods elicit a discrete response to an experimentally designed set of choice alternatives. Their definition includes what would normally be viewed as binary contingent valuation questions, as well as other variants of elicitation processes. This chapter focuses on what they refer to as a “multinomial choice sequence” (a series of multialternative experimentally designed choice questions).

Of course, these potential advantages come at a price, including the following challenges:

- As is the case in contingent valuation, CE responses are stated preferences, and concerns about strategic behavior or hypothetical bias arise.
- The cognitive difficulty faced by respondents in considering alternatives with multiple attributes in new choice situations may be high. Requiring respondents to assess complex trade-offs may result in behavioral responses such as the use of decision heuristics that are not well understood and might not reflect how they would make actual market choices.
- Experimental design theory is becoming more complex, and a sound grasp of the basic design principles is required to construct an experiment.
- The econometric models used to analyze CE data are becoming more complex and require advanced econometric and programming skills to operationalize.

However, while there are challenges, new software programs and tools have been developed that help in experimental design, data collection, and econometric analysis, and relatively straightforward procedures can be used to generate estimates that can be used in policy analysis.

In this chapter, the goal is to provide practical guidance on how to design and analyze a CE that provides credible value estimates to support decision-making. The chapter begins by describing the rich historical antecedents of modern applications of CEs (Sect. 5.1). This is followed by an overview of the basic steps required for conducting a choice experiment (Sect. 5.2). The next section expands on a set of selected topics in experimental design that are important to understand when developing a choice experiment (Sect. 5.3). Next, a description of the standard random utility model that provides the conceptual foundation for empirical analysis of CE data is presented (Sect. 5.4), along with an explanation of how to compute various measures of willingness to pay (Sect. 5.5). That is followed by descriptions of empirical models that relax standard assumptions, which are the subject of much current research (Sect. 5.6). To illustrate issues in implementation and interpretation of the standard and more advanced models, an application of a CE to an environmental problem is provided (Sect. 5.7). Concluding comments are presented (Sect. 5.8), followed by two examples illustrating recent applications of choice experiments (Appendices 1 and 2).

5.1 Interpretive History

The origins of CEs are found in various social science disciplines. Within economics, the conceptual foundation for CEs finds its source in the hedonic method that views the demand for goods as derived from the demand for attributes. This approach can be traced to Court (1939), who used hedonic regressions to study the demand for automobiles, and Griliches (1971), who used hedonic regressions in the construction of hedonic price indices. In the psychology literature, the comparative

judgment approach (Thurstone 1927) and the judgment and decision-making literature (Hammond 1955; Anderson 1970) also include discussions of how consumers evaluate items and use these evaluations in choosing between items. Lancaster's (1966) theory of consumer demand provided the basic conceptual structure that underlies economic applications of CEs.

At the same time that Lancaster was writing about consumer demand being driven by commodity attributes, a new measurement technique in mathematical psychology was articulated for decomposing overall judgments regarding a set of complex alternatives into the sum of weights on attributes of the alternatives (Luce and Tukey 1964). The method was rapidly embraced by marketing researchers who recognized the value of information about the relative importance of commodity attributes in the design of new products (Green and Rao 1971; Green and Wind 1975). This new marketing research method became generally known as "conjoint analysis."

Many commercial applications for conjoint analysis were soon found, particularly the prediction of market share for new products (Cattin and Wittink 1982). The typical procedure would ask respondents to rate the attractiveness of a set of products and then model the preferences of each respondent.² This approach emphasized the importance of capturing individual-level preference heterogeneity as a key element in predicting market share.

Two primary concerns arose regarding the typical conjoint procedure. First, it was not clear that the information contained in rating data was the same as the information contained in choice data, which mimicked market transactions. Second, implementation of choice simulators, based on ratings data, was cumbersome and often confusing to managers who used the predictions of market share models.

A simpler, more direct approach to predicting choices in the marketplace was provided by discrete choice theory, particularly as formulated for econometric analysis by McFadden (1974). The conceptual foundation for McFadden's analysis of economic choice lay in Thurstone's (1927) idea of random utility. By positing that individuals make choices that maximize their utility and that not all determinants of choice are available for analysis, choice theory was placed on a strong economic foundation that included a richness of behavior not found in standard Hicks-Samuelson theory. In addition, starting with Luce's (1959) choice axiom, linked to the random utility model by Marschak (1960), McFadden developed an econometric model that combined hedonic analysis of alternatives and random utility maximization.³ This model is known as the multinomial logit (conditional logit) model.

A further advance identified by McFadden and others is the linkage between random utility models and welfare economics. The utility function in random utility

²Rating scale approaches, or traditional conjoint analysis, are based on Torgerson's (1958) Law of Comparative Judgment. This approach presents individuals with profiles (alternatives) or bundles of attributes and asks them to provide a rating of each profile (e.g., 1 to 10, where 10 is very good, and 1 is very poor). The development of rating-based conjoint is discussed in Green and Srinivasan (1978) and Louviere (1988b).

³See also subsequent work by Manski (1977) and Yellot (1977).

models is actually a conditional indirect utility function (conditional on the choice of the alternative). Thus, including price as an attribute in the conditional indirect utility function allows one to assess economic welfare measures (Small and Rosen 1981).

The conceptual richness of random utility theory and the practical advantages of the multinomial logit model were embraced by marketing researchers who promoted the use of multinomial logit to analyze aggregate marketing data or choice data aggregated up to frequencies of choice (Louviere and Hensher 1983; Louviere and Woodworth 1983; Louviere 1988a). The random utility model also found wide application in transportation demand (Ben-Akiva and Lerman 1985; Louviere et al. 2000; Hensher et al. 2005). While initial work using the multinomial logit model was based on the analysis of aggregate or frequency of choice data, recent methodological developments have focused on understanding choices and sources of individual preference heterogeneity in random utility models, reminiscent of the focus on individual-level modeling used in conjoint analysis.

The first application of hedonic, stated-preference methods to environmental valuation was by Rae (1983), who used rankings (most preferred alternative, secondmost preferred, etc.) to value visibility impairments at U.S. national parks. Other environmental valuation studies using rankings include Smith and Desvousges (1986), who evaluated changes in water quality, and Lareau and Rae (1989), who estimated values for diesel odor reductions. Subsequent to these studies, ratings methods for environmental valuation, based on a Likert scale, grew in popularity (Gan and Luzar 1993; Mackenzie 1993; Roe et al. 1996). The popularity of ranking and rating methods for environmental valuation has diminished due to difficulties in linking such responses to responses consistent with economic theory (Louviere et al. 2010).

In the early 1990s, a number of applications of stated-preference experiments that used choices, rather than ratings or ranking, began to appear in the environmental economics literature. Among the first applications was Adamowicz et al. (1994), who demonstrated how revealed and stated-preference (choice experiment) data can be combined. Since then, the use of CEs in the literature has grown rapidly with applications to use values and passive-use or total values. At present, CEs are probably the most commonly used approach in the peer-reviewed literature within the class of discrete choice experiments. The literature on applications of CEs and on methodological issues surrounding CE implementation (particularly in the areas of experimental design and econometric analysis) has increased steadily, and “standard practice” changed dramatically over the last 30 years.

5.2 Steps in Conducting a Choice Experiment

Before deciding to conduct a choice experiment, it is essential to consider whether this method is the most appropriate or whether another technique, such as contingent valuation, would be better. The essence of this decision is whether it makes sense to frame a policy question in terms of the attributes and whether marginal

values of the attributes are required for policy analysis. If a policy question, for example, seeks to identify forest management options that will provide the greatest benefit to moose hunters, then consumer choices between alternative moose hunting sites with different levels of attributes (such as moose abundance, road quality, and travel distance) provide a reasonable framework for analysis (Boxall et al. 1996). In contrast, if the policy question focuses on the value that hunters place on a moose hunting experience given current conditions, then a contingent valuation study may be a better approach (Boyle et al. 1996).

The second issue to consider is the technical composition of alternatives and the perception of attribute bundles by consumers. In the moose hunting example (Boxall et al. 1996), moose abundance, road quality, and travel distance can reasonably be considered to be independent attributes. This may not be the case for a suite of ecological characteristics that are technically linked in production (Boyd and Krupnik 2009).

If it is decided that a CE is the best approach for conducting policy analysis, then implementation should follow the seven steps outlined in Table 5.1 (based on Adamowicz et al. 1998). Each step is briefly described following the table.

5.2.1 *Characterize the Decision Problem*

The initial step in developing a CE is to clearly identify the dimensions of the problem. This requires thinking about two key issues: (1) the geographic and temporal scope of potential changes in policy attributes, and (2) the types of values that are associated with those changes. The geographic scope of a CE would include consideration of whose values are to be included in the valuation or benefit-cost analysis. If the value of a change in an endangered species management program is being considered, for example, should the CE be applied to people living in the region, province/state, country, or internationally? It is essential to identify who will be impacted by changes in policy attributes as well as to articulate how they will be impacted. In addition, if the policy context is specific to a geographic site, the location of substitute sites will be important in the design, as demonstrated in a tropical rainforest preservation study reported by Rolfe et al. (2000).

Table 5.1 Steps in implementing a choice experiment

Step 1	Characterize the decision problem
Step 2	Identify and describe the attributes
Step 3	Develop an experimental design
Step 4	Develop the questionnaire
Step 5	Collect data
Step 6	Estimate model
Step 7	Interpret results for policy analysis or decision support

Temporal considerations will also be important. There may be a need to include an attribute for program duration or when the benefits will accrue to the public (e.g., Qin et al. 2011).

The second issue is the type of value arising from the policy under consideration. Is the choice to be examined one that reflects use value or behavior (such as recreation site choice or choices of market goods), or is the choice best represented as a public choice (referendum) on a set of attributes arising from a policy change? The latter may contain both use and passive-use values—or it may reflect total economic value.

5.2.2 Attribute Identification and Description

Once the decision problem is characterized, it is necessary to identify and describe the relevant attributes, including the levels to be used for each attribute. Holding structured conversations (focus groups) with resource managers, scientists, and people who typify the population that will be sampled will help identify the important attributes. At this stage, it is often challenging to decide how many attributes to include in the experiment as well as the particular levels that each attribute can take.

Focus groups can be very useful in this case. Group members can be asked to describe what attributes they think of when considering the goods and services being affected by the policy. They can provide information on whether attributes and levels are credible, understandable, and clearly presented. Focus groups of policymakers and the public can be useful to identify whether the attributes being considered by policymakers coincide with those being evaluated by members of the public. However, focus groups will often provide long lists of attributes that could result in complex choice tasks. Because not much is known about how people respond to highly complex survey questions (Mazzotta and Opaluch 1995; Swait and Adamowicz 2001a, b), it is a good idea to keep the set of attributes and levels as simple as possible. Overall, focus groups are a very important and effective way to construct attributes, levels, and the appropriate framing of a choice task.

Describing attributes that represent passive-use values (such as the value of biodiversity conservation) can be particularly challenging. Boyd and Krupnick (2009) suggested that attributes should be thought of as endpoints that directly enter the utility functions or household production functions of consumers, or—if intermediate inputs are being considered—the pathway to the endpoint needs to be made clear. Thus, passive-use values associated with forest biodiversity, for example, can be described using indicators of species richness (Horne et al. 2005). However, because forest biodiversity can be influenced by forest management processes that are under the control of decision-makers, attributes could be described in terms of those processes so long as the linkages between processes and outcomes are made clear. Because individuals might be interested in the processes associated with the endpoint, it is important to clarify the things that people value,

what decision-makers can affect, and the description of the attributes during this stage of survey development. In addition to identifying utility endpoints, Schultz et al. (2012) recommended further standards for attributes in stated-preference studies that include measurability (endpoints are quantifiable), interpretability (endpoints can be understood by a nonscientist), and comprehensiveness (all relevant endpoints are described).

Once the attributes have been defined, attribute levels must be specified. In some cases this is simple, such as the presence or absence of some attribute. In other cases the assignment of levels is more difficult, such as determining the appropriate levels and ranges used to specify forest species richness (Horne et al. 2005). This issue is also faced when specifying price or cost levels. Because the price/cost attribute provides control over the key factor that determines welfare measures, it is important that this attribute can be estimated precisely in the econometric model and also be reasonable in the policy context. Much as in contingent valuation, we would like low-price/cost alternatives to be frequently purchased and high-price alternatives to be rarely purchased.⁴ Price levels should not be so high or low that they do not appear to be credible, but it may be informative for prices/costs to lie outside the range of existing market prices (such as travel costs) or be reasonable costs for the provision of public programs. Pilot studies play an important role in testing price or cost levels, as well as all other attributes and levels, to ensure that they have sufficient variation to identify the parameters and to ensure that welfare measures can be calculated.

These first two steps, which are critical to the successful implementation of CEs, are often not given the due consideration they require. Practitioners are encouraged to spend significant time and effort in scoping the problem, using focus groups and pretests, and making sure the choice context and scenario descriptions are carefully developed.

5.2.3 Develop an Experimental Design

Once attributes and levels have been determined, the researcher must determine the number of alternatives to present in each choice set (two, three, four, etc.), and the number of choice sets to present to the respondents (one, four, eight, 16, etc.). The number of alternatives could depend on the type of value being measured and/or on the context of the study. At a minimum, choice questions should contain a status quo alternative and an alternative indicating a change from the status quo. A status quo alternative is required in each choice set so that estimated utility functions represent changes from baseline conditions. Total value (or passive-use value) studies often employ only two alternatives because of the incentive compatibility of a two-alternative choice or referendum (Carson and Groves 2007). The number of

⁴A useful graphical tool for visualizing the role of price on choice in a multiattribute context is described by Sur et al. (2007).

alternatives in some studies depends on the number of alternatives that occur in the real world.

The number of choice questions to ask depends in part on the complexity of the task and is often a judgment the researcher must make based on focus groups, pilot tests, and expert judgment. In general, the number of choice sets included in the design depends on the number of degrees of freedom required to identify the model. The use of multiple choice sets can also have implications for incentive compatibility (Carson and Groves 2007, 2011).

Experimental design procedures are used to assign attribute levels to the alternatives that form the basis for choices and to construct the sets of choices that will be presented to respondents. Alternatives presented to the respondents must provide sufficient variation over the attribute levels to allow one to identify preference parameters associated with the attributes. In most cases, presenting all combinations of attributes and levels will be impossible. Thus, experimental design procedures are used to identify subsets of the possible combinations that best identify attribute preferences. Because of the importance of this topic to the success of any CE (Scarpa and Rose 2008), it is discussed in detail in Sect. 5.3.

5.2.4 Questionnaire Development

As with other stated-preference methods, CEs involve surveys, and various questionnaire formats can be used for collecting data (see Chap. 3), including:

- Mail-out, mail-back surveys.
- Telephone recruitment, mail-out, mail-back surveys.
- Telephone recruitment, mail-out, telephone surveys.
- Computer-assisted surveys at centralized facilities or in person.
- Intercept surveys that could be paper and pencil or computer-assisted.
- Internet-based surveys, including Internet panels.

The selection of the questionnaire format is usually based on pragmatic concerns, such as availability of a sample frame and budget limitations. In the case of CEs, Internet modes, particularly Internet panels, are becoming increasingly popular. Because CEs present respondents with complex sets of choice questions and randomization of the order of these questions is desirable, mail and telephone surveys can be more difficult to use relative to Internet or computer-based in-person surveys (e.g., using tablets to collect information from respondents). Also, in some cases information from early parts of a survey is used in the design of attributes and/or levels in the choice tasks, making computer-based Internet or in-person surveys more convenient. While Chap. 3 discusses survey mode comparisons and trade-offs in general, specific features of CEs, such as the incorporation of the experimental design into the survey, are facilitated by the use of Internet panels. Concerns about the social context induced by in-person surveys (social desirability bias) and the cost of in-person surveys result in less use of this mode.

Lindhjem and Navrud (2011) reviewed survey mode effects in stated-preference models and found relatively little difference between Internet modes and other modes. However, they raised concerns about the representativeness of Internet modes. Some Internet panels have very good properties in terms of representativeness because they are based on random samples of the population, while others are opt-in panels that raise questions about selection bias.

Various methods can be used to communicate information about the attributes of a valuation problem. In addition to verbal descriptions, maps, and photographs, other graphic displays should be considered. As in any survey-based research, pretesting of the questionnaire is absolutely necessary to assure that respondents clearly understand the information being communicated (see Chap. 3 for more detail on survey methods). Choice experiments are typically presented as matrices with alternatives as the columns and attributes as the rows, but there are various other forms of presentation that can be used. Often researchers will include graphics or other visual aids within the choice matrix to represent attributes and levels.

An issue that is critical in the design of stated-preference surveys, including CEs, is the inclusion of methods to address hypothetical bias (strategic behavior) and, in some cases, to assess scope effects. CEs are probably as prone to strategic behavior or hypothetical bias as are contingent valuation tasks. Carson and Groves (2007, 2011) outlined the theory associated with strategic behavior and emphasized the need to construct instruments that are “consequential.” They also described the ways that CEs can differ from other stated-preference methods in terms of strategic behavior. For example, strategic behavior can arise from the examination of choice sequences to look for the “best deal” (Holmes and Boyle 2005; Day et al. 2012), as well as from the design of the CE (Vossler et al. 2012).

Three major approaches for dealing with hypothetical bias in stated-preference surveys have been used. The first is to include “cheap talk scripts” (Cummings and Taylor 1999; List 2001) that describe to respondents that hypothetical values or bids are often higher than they would be when there are real consequences. This approach is no longer being recommended because it may influence values by suggesting that respondents’ values are often too high. Reminders of substitutes represent good practice, but statements that may influence values before the valuation question is asked are questionable (see Vossler, 2016, for details).

The second approach is to ask respondents how certain they are about their choice (Blumenschein et al. 2008). Uncertain preferences for a program (particularly regarding a nonstatus quo program in a passive-use context) could lead to status quo choices, and adjustments for uncertain preferences in CEs have been investigated (Ready et al. 2010). Finally, Vossler et al. (2012) outlined how choice experiments can generate strategic responses and showed that when respondents think that the program being presented is consequential (could actually be used in policy), responses are more likely to correspond to preferences elicited in an incentive-compatible fashion.

5.2.5 Data Collection

Data collection should be carried out using the best survey practices (e.g., Dillman 1978). Chapter 4 outlines a number of issues in data collection for contingent valuation studies that apply as well to the implementation of CEs. One unique feature arising in CEs is that multiple choice sets are presented to individuals with the intent that choice sets be considered independently and without comparing strategically across choice sets. This means that it is desirable to prevent respondents from reading ahead or going back and changing responses. It is also valuable to randomize the order of the presentation of the choice sets so that the first task, in a large enough sample, can be used to estimate values that are not affected by repeated choices. In a mail survey (paper and pencil), this is very difficult to accomplish because respondents can flip through the survey booklet. Computer-based surveys (Internet and in-person) can achieve this through the design of the survey implementation program. Computer-based methods also capture the amount of time spent on each question, which tells researchers if respondents are taking time to consider the choice set carefully.

5.2.6 Model Estimation

Once data have been collected, the next step is to estimate preference parameters using a random utility model. A growing number of econometric specifications have been used to analyze choice data. These models typically vary over how the error term is interpreted, particularly in the context of heterogeneity in preferences across respondents. Due to the complexity of these models and the variety of econometric specifications available, estimation is discussed in detail in Sects. 5.4 through 5.6.

5.2.7 Policy Analysis and Decision Support

Most CE applications are targeted to generating welfare measures (see Sect. 5.5), predictions of behavior, or both. These models are used to simulate outcomes that can be used in policy analysis or as components of decision support tools. CEs provide the opportunity to evaluate the welfare effects of multiple policy options involving combinations of attributes and levels. They also allow for calibration to actual policies or outcomes when these conditions become known. For example, choice experiments on park visitation have been calibrated using actual visitation information when measuring nonmarginal welfare impacts (Naidoo and Adamowicz 2005). As such, they can provide a richer set of policy information than most other valuation approaches.

Yellowstone National Park Case Study

PROBLEM

Debates over how national parks within the United States should be managed have been very divisive, with some groups arguing in favor of strict wilderness conservation while other groups have stressed the need for various in-park recreational activities. This debate has been especially intense in Yellowstone National Park, the oldest and arguably most well-known national park in the U.S. During the winter months, snowmobiles are the primary means of accessing the park. People who oppose the use of snowmobiles in the park claim that they are noisy and smelly, cause congestion and interfere with other recreational uses such as hiking and cross-country skiing, and threaten the park's wildlife. Proponents of snowmobile use in the park argue that the machines are safe, convenient, and often the only way to access the extraordinary winter landscape. Designing efficient and equitable winter use regulations for Yellowstone National Park has been a challenge for several decades.

APPROACH

A winter visitation survey was designed to obtain information relevant to a benefit-cost analysis of winter management alternatives being considered by the National Park Service. The sampling frame was based on users of the park during the winter of 2002-03. In addition to information on trip-related expenditures and winter recreation activities pursued, the survey implemented a choice experiment that was designed to solicit preferences for attributes of Yellowstone National Park under different winter management alternatives, including various snowmobile restrictions.

RESULTS

Study results indicated that some restrictions on snowmobile access in Yellowstone National Park could improve social welfare. Welfare losses to snowmobile riders could be offset by welfare gains to other park users although the net benefits depend on the number of riders and nonriders using the park as well as the specific regulations imposed. Further, heterogeneous preferences were found regarding the restriction for snowmobilers to be on a group tour—experienced snowmobilers did not like the restriction while novice snowmobilers did not mind group tours. The results of the survey have been used in several benefit-cost analyses to support rulemaking in Yellowstone National Park.

SOURCE

Mansfield, C., Phaneuf, D. J., Johnson, F. R., Yang, J.-C. & Beach, R. (2008). Preferences for public lands management under competing uses: The case of Yellowstone National Park. *Land Economics*, 84, 282-305.

5.3 Experimental Design

The basic problem addressed in the experimental design literature for CEs—given the selected attributes and their levels—is how to allocate attribute levels to alternatives and choice sets. Several approaches to experimental design for CEs have been proposed, and the best approach to use depends on which preference parameters need to be estimated and whether or not prior information on the parameters is available. The researcher also needs to think about the complexity of the design because the inclusion of many alternatives and choice sets can cause respondents to use decision-making shortcuts (heuristics) that might not reflect their true preferences.

An experimental design must contain sufficient independent variation among attribute levels within and across alternatives so that each preference parameter can be identified. For example, if the levels of an attribute are always identical across alternatives, it will not be possible to identify the effect of that attribute on responses. A good design is also statistically efficient, meaning it minimizes (maximizes) the standard errors (precision) of the preference parameter estimates.

Traditional experimental designs were constructed to support linear-in-parameters statistical models, and orthogonal designs were popularized because they eliminated correlation between attributes so that the independent influence of each variable on outcomes could be estimated. Researchers have come to realize that orthogonal designs might not, in most situations, provide optimal statistical efficiency when nonlinear-in-parameters models are used to analyze CE data. This section provides an overview of traditional orthogonal designs as well as statistically efficient designs that are being adopted by CE practitioners.

5.3.1 Orthogonal Full Factorial Designs

The most complete experimental design is a full factorial design, which combines every level of each attribute with every level of all other attributes (Hensher et al. 2005). The primary advantage of a full factorial design is that all main and interaction effects are statistically independent (orthogonal) and can be identified when estimating a model.⁵ The major drawback of this design is that a very large number of alternatives are generated as the numbers of attributes and levels are increased. For example, suppose that a recreation agency is evaluating plans for developing

⁵A main effect is the direct effect of an attribute on a response variable (choice), and it reflects the difference between the average response to each attribute level and the average response across all attributes (Louviere et al. 2000). An interaction effect occurs if the response to the level of one attribute is influenced by the level of another attribute. Interaction effects are represented by parameter estimates for the interaction (cross product) of two or more variables and can account for more complex behavioral responses to combinations of attribute levels.

Table 5.2 Orthogonal codes illustrating the properties of full and fractional factorial designs

Attribute combination	Main effects			2-way interactions			3-way interactions
	Picnic shelters (A1)	Boat ramps (A2)	Camping fee (A3)	A1 × A2	A1 × A3	A2 × A3	A1 × A2 × A3
<i>First fraction</i>							
1	-1	-1	+1	+1	-1	-1	+1
2	-1	+1	-1	-1	+1	-1	+1
3	+1	-1	-1	-1	-1	+1	+1
4	+1	+1	+1	+1	+1	+1	+1
<i>Second fraction</i>							
5	-1	-1	-1	+1	+1	+1	-1
6	-1	+1	+1	-1	-1	+1	-1
7	+1	-1	+1	-1	+1	-1	-1
8	+1	+1	-1	+1	-1	-1	-1

lakeside campgrounds and that agency managers are considering whether or not to install picnic shelters and boat ramps at each location as well as how much to charge for an overnight camping fee. If each of these three attributes takes two levels (install or not for facilities, \$10 or \$20 for camping fees), there are $2^3 = 8$ possible combinations of attribute levels in the full factorial design (the eight alternatives shown in Table 5.2).⁶ If the number of levels associated with each attribute increases from two to three, the full factorial design increases to $3^3 = 27$ possible combinations of attribute levels.

The properties of an experimental design can be understood using a helpful coding scheme known as “orthogonal coding.” Under this scheme, attribute levels are assigned values so that the sum of values in each column (representing main or interaction effects) equals zero (Hensher et al. 2005). For an attribute with two levels, for example, this is accomplished by assigning a value of 1 for the first level of the attribute and a value of -1 for the second level. As is illustrated in Table 5.2, the sum of values for each main (or interaction) effect is zero because each level appears equally often. A design with this property is said to be balanced, and it ensures that preference parameters are well-estimated across the range of each attribute.⁷ Further, Table 5.2 shows that the inner product of any two column vectors equals zero. This is because each pair of levels appears equally often (in

⁶More generally, the number of possible combinations of attribute levels in a full factorial design is $\pi_{k=1}^K L_k$, where L_k is the number of attribute levels associated with attribute k .

⁷Attribute level balance leads to larger experimental designs when the number of attribute levels differs across attributes.

one-quarter of the attribute combinations) across all columns and indicates that all of the main effects are orthogonal.⁸

Orthogonal codes are also useful for identifying the design columns associated with interaction effects. In Table 5.2, let Attribute 1 (A1) represent picnic shelters, Attribute 2 (A2) represent boat ramps, and Attribute 3 (A3) represent camping fees. Then, the design of all interaction effects is found by multiplying together the appropriate orthogonal codes associated with each attribute. In the first alternative, for example, the two-way interaction between picnic shelters and boat ramps ($A1 \times A2$) is computed as $(-1 \times -1 = 1)$. By computing the inner products of the columns of orthogonal codes, one can see that all interaction effects are statistically independent of each other and also independent of the main effects.

5.3.2 Orthogonal Fractional Factorial Designs

The number of attribute combinations needed to represent a full factorial design increases rapidly as the number of attributes and levels increases, and fractional factorial designs can be used to reduce the design size. The simplest method of generating a fractional factorial design is to select subsets of attribute combinations from the full factorial design using higher order interaction terms (Hensher et al. 2005). For example, in Table 5.2, two fractions of the full factorial are shown based on the three-way interaction ($A \times B \times C$) which takes the value +1 for the first half-fraction and -1 for the second half-fraction. Note that for each half-fraction (four alternatives), the design remains balanced and orthogonal for the main effects.

However, in reducing the design size, fractional factorial designs omit some or all information on interaction effects. If the omitted interactions are important in explaining responses, the preference parameter estimates may be biased due to confounding an omitted variable with a main effect. To see this (Table 5.2), note that the vector of two-way interactions $A1 \times A2$ [+ 1, -1, -1, +1] is identical to the vector of main effects for A3. Thus, $A1 \times A2$ is perfectly collinear with A3. If econometric analysis of data collected using this fractional factorial design showed that the parameter estimate on A3 was significantly different than zero, we could not be sure whether camping fees (A3) were significant, the interaction of picnic shelters and boat ramps was significant ($A1 \times A2$), or both. Thus the interpretation of the parameter estimates (mean and standard error) on the key economic variable (camping fee) would be ambiguous because a potentially important interaction between facility attributes was not identified.

⁸In general, orthogonality occurs when the joint occurrence of any two attribute levels, for different attributes, appear in attribute combinations with a frequency equal to the product of their individual frequencies. In Table 5.2 for example, each attribute level (-1 or 1) for each attribute appears in one-half of the attribute combinations. Therefore, the joint combination of any two attribute levels (say, -1 and -1) must occur in $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ of the attribute combinations for the design to be orthogonal.

Due to the possibility that a fractional factorial design can induce biased parameter estimates on the main effects and can fail to identify meaningful interactions, it is essential to identify which attribute interactions might be important during the design stage of survey development. For example, a CE that is investigating alternative transportation routes might anticipate that some combination of attributes, such as low travel cost and low road congestion, are strongly preferred. These potentially important interactions could be evaluated by asking focus group participants if some combinations of attribute levels are particularly desirable or undesirable. If so, a main effects plus selected interactions plan should be used. In general, this is accomplished using orthogonal codes to examine the correlations between the main and interaction effects and assigning attributes to design columns that are orthogonal to the specified interaction effects (Hensher et al. 2005).

5.3.3 Generating Choice Sets for Orthogonal Designs

The key issues to consider in creating an experimental design for a CE are how to place alternatives into choice sets and how many choice sets are needed. Several choice sets are typically included in a CE, and the number of choice sets depends on the number of degrees of freedom (the number of parameters plus one) required to identify the parameters of the specified model. In turn, the number of degrees of freedom depends on whether the alternatives are described using a label to differentiate the alternatives (such as transportation modes or recreational locations) or whether they are unlabeled (generic). Labeled alternatives are used when the researcher wants to estimate a utility function for each alternative. The ability to identify the independent effect of each attribute in each alternative requires that attributes are orthogonal within and between alternatives. Unlabeled alternatives are used when the researcher is only interested in estimating a single utility function. Because labeled designs require more parameters to be estimated, more degrees of freedom are required, resulting in a larger design.

Note that labeled designs permit the analyst to estimate a constant term specific to each alternative, known as an alternative specific constant. If respondents derive utility from unobserved attributes associated with the labels of alternatives, the alternative specific constants provide a means for measuring that component of utility that is independent of the experimentally designed attributes. It is also common to test for status quo bias in unlabeled designs by including an alternative specific constant for the status quo alternative. If the alternative specific constant is statistically significant, it suggests that respondents have a preference for (or against) the status quo option independent of the designed attributes.

The number of degrees of freedom also depends on whether parameters are estimated for the levels of an attribute (referred to as nonlinear effects) or whether a

single parameter is estimated for an attribute (referred to as a linear effect).⁹ For categorical variables, it is necessary to estimate nonlinear effects, and the number of nonlinear effects that can be estimated equals the number of levels (such as low, medium, or high) minus one (hence two nonlinear effects). For continuous variables such as price, a single parameter is usually estimated.

The number of attribute combinations required to estimate main effects must equal or exceed the number of degrees of freedom, which can be simply computed (Hensher et al. 2005). For unlabeled alternatives in which the analyst wants to estimate nonlinear effects, the minimum degrees of freedom required is $(L - 1) \times A + 1$, where L is the number of attribute levels and A is the number of attributes. If only one parameter is estimated for each attribute, the number of degrees of freedom is reduced to $A + 1$. For labeled alternatives, the comparable formulas are $(L - 1) \times NA + 1$ and $NA + 1$, where N is the number of alternatives.¹⁰

Experimental designs for unlabeled alternatives can be created starting with an orthogonal plan of attribute combinations. Each attribute combination (such as combinations 1 through 4 in Table 5.2) provides a design for the first alternative in one of four choice sets. Then a second alternative could be created by randomly pairing nonidentical attribute combinations (Table 5.3). This design is suitable for estimating the three desired parameters (one parameter for each attribute) because there are four degrees of freedom (the number of parameters plus one) and four sets of attribute combinations. However, note that the design codes for specific attributes are identical in some attribute combinations (such as picnic shelters in the first combination). Because choice models are based on attribute differences, the lack of a contrast for attribute levels within choice sets reduces the statistical efficiency of the design. Also note that this design would be unsuitable for estimating the parameters of a labeled experiment because there is an insufficient number of attribute combinations relative to the degrees of freedom ($2 \times 3 + 1 = 7$) required to identify each parameter. Further, as can be seen by computing the inner product of the camping fee attribute levels in Alternatives *A* and *B* (Table 5.3), these columns are not orthogonal. In fact, they are perfectly (negatively) correlated. In general, randomizing attribute combinations will be inadequate for estimating independent utility functions for labeled alternatives because the attributes will not be orthogonal across alternatives, and main effects will be correlated (Hensher et al. 2005; Street et al. 2005).

⁹Nonlinear effects in this context should not be confused with functional forms of the variables, such as quadratic or logarithmic transformations. If the researcher is interested in whether continuous variables (such as price) are better described by nonlinear functional forms, nonlinear effects could be estimated and used to evaluate the functional form.

¹⁰If the number of attribute levels differs across attributes, then the formulas for computing the number of degrees of freedom required to estimate nonlinear effects must be adjusted. In particular, the value of $(L - 1) \times A$ must be computed for each set of attributes with a unique number of levels. Then these values must be summed before adding 1.

Table 5.3 Orthogonal codes illustrating properties of choice sets for random attribute combinations

Attribute combination	Alternative A			Random combination	Alternative B		
	Picnic shelters	Boat ramps	Camping fee		Picnic shelters	Boat ramps	Camping fee
1	-1	-1	+1	2	-1	+1	-1
2	-1	+1	-1	4	+1	+1	+1
3	+1	-1	-1	1	-1	-1	+1
4	+1	+1	+1	3	+1	-1	-1

An alternative approach is to use a design based on a collective full factorial in which the design of choice sets occurs simultaneously with the design of alternatives, and the attribute levels are orthogonal within and across alternatives. A collective full factorial design is referred to as an L^{NA} design (Louviere et al. 2000). Using the campground example with $A = 3$ and $L = 2$, if one decides to include two alternatives plus the status quo alternative in each choice set, then the collective full factorial design includes $2^{(2 \times 3)} = 64$ rows, and each row contains a design for both of the new campground alternatives.¹¹

Given a collective full factorial, a main effects fractional factorial design can be selected as the smallest orthogonal plan, and it depends on the number of degrees of freedom required to estimate all of the main effects. The smallest orthogonal design available is usually larger than the number of degrees of freedom required to estimate the parameters, and finding the smallest orthogonal design for differing levels of attributes and levels is mathematically challenging (Louviere et al. 2000). Fortunately, orthogonal main effects plans are available in published literature, on the Internet, and by using software programs.

In the campground example, the number of degrees of freedom required to estimate a main effects only collective fractional factorial design is $(2 - 1) \times (2 - 3) + 1 = 7$. An example of a labeled campground design is shown in Table 5.4, and as can be seen, the smallest collective fractional factorial design includes eight choice sets. Each attribute (design column) is orthogonal to all other attributes, both within and across alternatives. Note that the design codes for specific attributes (such as picnic shelters) are identical in some sets (such as in Choice sets 1, 2, 5, and 6). Also note that some attribute combinations (Combinations 1 and 6) have identical orthogonal codes for each labeled alternative. In general, this design would not provide any information for those attribute combinations if the attribute levels associated with each code were identical. However, it is common practice when using labeled designs to assign different values for the attribute levels associated with each alternative. For example, the camping fees associated with Label A might be \$10 (coded -1) versus \$20 (coded +1), and the camping fees associated

¹¹As the attribute levels of the status quo alternative are held constant across choice sets, the status quo alternative is not included in N (the number of alternatives).

Table 5.4 Orthogonal codes illustrating properties of choice sets for a labeled, collective fractional factorial design

Attribute combination	Label A			Label B		
	Picnic shelters	Boat ramps	Camping fee	Picnic shelters	Boat ramps	Camping fees
1	+1	+1	+1	+1	+1	+1
2	+1	+1	-1	+1	-1	-1
3	+1	-1	+1	-1	-1	-1
4	+1	-1	-1	-1	+1	+1
5	-1	+1	+1	-1	-1	+1
6	-1	+1	-1	-1	+1	-1
7	-1	-1	+1	+1	+1	-1
8	-1	-1	-1	+1	-1	+1

Table 5.5 A labeled main effects campground choice set taken from a $2^{(2 \times 4)}$ collective fractional factorial

	Campground A	Campground B	Status quo alternative
Picnic shelters	Yes	Yes	Stay at home: I would not choose either campground <input type="checkbox"/>
Boat ramps	Yes	Yes	
Camping fee	\$20	\$25	
I would choose: (please check one box)	<input type="checkbox"/>	<input type="checkbox"/>	

with Label B might be \$15 (coded -1) versus \$25 (coded +1). Thus, these alternatives would be contrasted on the camping fee attribute.

Once the design of the alternatives to be included in choice sets has been accomplished, the analyst needs to translate the orthogonal (or other design) codes into a format that makes sense to respondents and to add the status quo alternative. An example of a labeled choice set (based on the first attribute combination in Table 5.4) as it might appear in a questionnaire is shown in Table 5.5.

If the alternatives shown in Table 5.5 were unlabeled (generic), the first alternative would dominate the second alternative because both alternatives offer identical facilities and the first alternative is less expensive. Dominated alternatives are undesirable because they do not provide any useful information and therefore reduce the statistical efficiency of the design. However, for labeled alternatives, if unobserved attributes associated with the second alternative (Campground B) contributed substantially to respondent utility, that alternative might be chosen. Therefore, dominated alternatives are more difficult to detect for labeled alternatives.

Although it is not generally known how many choice sets respondents can evaluate, it is common to present four to six choice sets per respondent; sometimes 16 or more sets are included. If it is thought that the number of designed choice sets will place too great a cognitive burden on respondents, the burden can be reduced

by assigning respondents to blocks, or subsets, of the fractional factorial design. One method used for blocking is to list the choice sets in random order and then subdivide the list to obtain blocks of reasonable size. A second method of blocking is to consider blocks as a separate attribute in the experimental design, letting the number of levels represent the desired number of blocks. This second method is preferred because including blocks as attributes in an orthogonal design assures that every level of all attributes will be present in every block (Adamowicz et al. 1998).

5.3.4 Statistical Efficiency for CEs

Traditional orthogonal designs were developed for linear-in-parameters statistical models and meet two criteria for good designs: (1) they remove multicollinearity among attributes so that the independent influence of each attribute can be estimated, and (2) they minimize the variance of the parameter estimates so that t -ratios (based on the square roots of the variances) are maximized. These design criteria are met when the elements of the variance-covariance matrix for the linear model are minimized (Rose and Bliemer 2009).¹²

For linear-in-parameters statistical models, the information used to estimate the variance-covariance matrix depends on the design matrix (data on the explanatory variables of the model) and a constant scaling factor. In contrast, the variance-covariance matrix for the nonlinear models used to analyze CE data contains information on the design matrix and the preference parameters (McFadden 1974).¹³ A statistically efficient CE design has the best variance-covariance matrix. Various definitions of “best” have been proposed, and they depend on the assumptions made about the preference parameters as well as the method chosen to summarize information in the variance-covariance matrix.

A commonly used summary statistic for the information contained in the variance-covariance matrix is the determinant as it uses information on the main diagonal (variances) and the off-diagonals (covariances). The determinant of a variance-covariance matrix, scaled by the number of parameters to be estimated in the model, is known as the D -error. Designs that minimize the D -error are considered to be D -efficient.¹⁴

¹²The variance-covariance matrix is the inverse of the Fisher information matrix and is based on the second derivative of the log-likelihood function.

¹³In particular, McFadden (1974) showed that $VC = \left[\sum_{n=1}^N \sum_{j=1}^{J_n} x'_{jn} P_{jn}(Z, \beta) x_{jn} \right]^{-1}$, where P_{jn} is the probability that an individual will choose Alternative j in Choice set n , which is a function of the attribute design matrix (Z) and a vector of preference parameters (β). Also, $x_{jn} = z_{jn} - \sum_{i=1}^{J_n} z_{in} P_{in}$, where z_{jn} is a row vector describing the attributes of Alternative j in Choice set n .

¹⁴Other criteria for design efficiency have been proposed in the literature. For example, the A -error minimizes the trace of the variance-covariance matrix, which is computed as the sum of the elements on the main diagonal.

5.3.4.1 Optimal Orthogonal Designs

One approach for finding D -efficient designs for CEs is to assume that all alternatives contained in choice sets are equally attractive or, equivalently, that all preference parameters equal zero.¹⁵ These designs are referred to as optimal orthogonal designs.

An optimal orthogonal CE design is initialized with an orthogonal design for the first alternative in a choice set; it then makes systematic attribute level changes in the design to generate other alternatives (Street et al. 2005; Street and Burgess 2007).¹⁶ Optimality for these designs is defined by two criteria: (1) the attributes within alternatives are orthogonal, and (2) the number of times an attribute takes the same level across alternatives in a choice set is minimized (known as the minimal overlap property). Under the second criterion, survey respondents must make trade-offs on all attributes in a choice set that, presumably, provides more information about preferences and avoids dominated alternatives, which can arise in traditional orthogonal designs.¹⁷

An optimal orthogonal design for our campground example is illustrated in Table 5.6. Beginning with an orthogonal design for Alternative A (as in Table 5.3), Alternative B was created by multiplying the levels in Alternative A by -1 .¹⁸ This fold-over procedure maintains the orthogonality of the design, which is also balanced, while providing a contrast for each level of each attribute. This procedure obviously would not work for labeled designs because the attributes in each alternative are perfectly (negatively) correlated.

In general, a D -efficient optimal orthogonal design is constructed by minimizing the following expression:

$$D_0\text{-error} = \det(\text{VC}(Z, 0))^{1/k}, \quad (5.1)$$

where Z represents the attributes in the experimental design, 0 indicates that $\beta = 0$ for all model parameters, and k is the number of parameters used in the scaling factor. The efficiency of the fold-over design (Table 5.6) relative to the design obtained using random attribute combinations (Table 5.3) can be demonstrated by computing Eq. (5.1) for each design. In particular, the authors find that the D_0 -error

¹⁵Huber and Zwerina (1996) showed that, under the assumption that $\beta = 0$, the variance-covariance matrix simplifies to $\left[\sum_{n=1}^N \frac{1}{J_n} \sum_{j=1}^{J_n} x'_{jn} x_{jn} \right]^{-1}$, where $x_{jn} = z_{jn} - \frac{1}{J_n} \sum_{i=1}^{J_n} z_{jn}$.

¹⁶This procedure, referred to as a “shifted design,” was initially proposed by Bunch et al. (1996). In general, these designs use modulo arithmetic to shift the original design columns so they take on different levels from the initial orthogonal design.

¹⁷This approach implicitly assumes that the cognitive burden imposed by making difficult trade-offs does not influence the error variance and, therefore, does not bias parameter estimates.

¹⁸To use modulo arithmetic in constructing Table 5.6, begin by recoding each of the -1 values as 0 . Then, add 1 to each value in Alternative A except for attributes at the highest level (1), which are assigned the lowest value (0).

Table 5.6 Orthogonal codes illustrating optimal orthogonal pairs using a fold-over design

Attribute combination	Alternative A			Alternative B		
	Picnic shelters	Boat ramps	Camping fee	Picnic shelters	Boat ramps	Camping fee
1	-1	-1	+1	+1	+1	-1
2	-1	+1	-1	+1	-1	+1
3	+1	-1	-1	-1	+1	+1
4	+1	+1	+1	-1	-1	-1

equals 0.79 for the random attribute combinations, and it equals 0.5 for the fold-over design, indicating the superiority of the latter design.¹⁹

5.3.4.2 Nonzero Priors Designs

A second approach to the efficient design of CEs using the variance-covariance matrix is based on the idea that information about the vector of preference parameters might be available from pretests or pilot studies and that this information should be incorporated in the design (Huber and Zwerina 1996; Kanninen 2002; Carlsson and Martinsson 2003; Hensher et al. 2005; Scarpa and Rose 2008; Rose and Bliemer 2009).²⁰ This approach, which we call a nonzero priors design, seeks to minimize the following expression:

$$D_p\text{-error} = \det(\text{VC}(Z, \beta))^{1/k}, \quad (5.2)$$

where p stands for the point estimates of the (nonzero) β 's. The constraints imposed on the optimal orthogonal model (orthogonality, attribute level balance, and minimal overlap) are relaxed in minimizing the D_p -error. However, if reasonable nonzero priors are available, relaxing these constraints can result in efficient designs that greatly reduce the number of respondents needed to achieve a given level of significance for the parameter estimates (Huber and Zwerina 1996). Note that designs that minimize the D_p -error do not generally minimize the D_0 -error and vice versa.

If the nonzero priors used in Eq. (5.2) are incorrect, however, the selected design will not be the most efficient. One method for evaluating this potential shortcoming is to test the sensitivity of a D -efficient design to alternative parameter values, which can provide the analyst some degree of confidence about the robustness of a design (Rose and Bliemer 2009). Another approach that can incorporate the analyst's uncertainty about parameter values is to specify a distribution of plausible values

¹⁹Although the covariances equal zero in both designs, the efficiency of the fold-over design is gained by the minimal overlap property.

²⁰One approach to developing nonzero priors is to use an orthogonal design in a pilot study to estimate the β vector, which is then used to minimize the D_p -error (Bliemer and Rose 2011).

that reflects subjective beliefs about the probabilities that specific parameter values occur (Sándor and Wedel 2001; Kessels et al. 2008). This Bayesian approach to experimental design proceeds by evaluating the efficiency of a design over many draws from the prior parameter distributions $f(\tilde{\beta})$. The design that minimizes the expected value of the determinant shown in Eq. (5.3) is a D -efficient Bayesian design:

$$D_b\text{-error} = \int_{\tilde{\beta}} \det(\text{VC}(Z, \tilde{\beta})) f(\tilde{\beta}) d\beta \quad (5.3)$$

The distribution of $f(\tilde{\beta})$ is typically specified as normal or uniform.

Note that a nonzero priors design that is efficient for estimating one model (such as a multinomial logit model) is not necessarily efficient for estimating other models (such as random parameter logit or latent class models), and efforts are being made to identify designs that are robust to alternative model types (Ferrini and Scarpa 2007; Rose and Bliemer 2009). Also of interest is the construction of efficient experimental designs for the estimation of willingness to pay (WTP) measures, which are computed as the ratio of two parameters (Scarpa and Rose 2008). Because efficient designs can increase the cognitive burden faced by respondents by requesting them to make difficult choices, understanding the trade-offs between statistical efficiency and response efficiency is an emerging area of concern (Louviere et al. 2008; Johnson et al. 2013).

5.3.5 *Selecting a Design*

Given a suite of alternative design options, which design should a researcher choose? Although this will depend on considerations specific to each study, the authors recommend the following general guidelines. First, use a design that is statistically efficient in the context of the nonlinear-in-parameters models used to analyze random utility models. If reasonable information is available on preference parameters from sources such as pretests or pilot studies, the authors recommend using a nonzero priors design. In general, these designs reduce the number of respondents needed to achieve a specific precision (standard error) for the parameters specified in the utility function(s) and can therefore help reduce the cost of survey implementation. In cases where no prior information is available or where parameter estimates from other CE studies do not provide a good match, an optimal orthogonal design should be considered. This recommendation is based on evidence that optimal orthogonal designs can produce good results where prior information on parameter values is of poor quality or when the model specification chosen by the analyst is inconsistent with the underlying data generating process (Ferrini and Scarpa 2007). The construction of statistically efficient designs is greatly facilitated by the availability of software programs (such as SAS and Ngene).

5.3.6 Attribute Coding Schemes

While orthogonal codes are used to study the properties of an experimental design, other codes are used at the data analysis stage. Recall that attributes can be coded to estimate either linear or nonlinear effects (Sect. 5.3.3). For continuous variables, such as the cost of an alternative, it is common to estimate a linear effect by using the level of the quantitative variable as the code. However, it is also possible to estimate the nonlinear effects of a continuous variable using a nonlinear effects coding method that is also used for coding qualitative attributes. Two coding methods are available for estimating nonlinear effects. First, dummy variables, using 0 – 1 codes, can be defined for $L - 1$ levels of an attribute. However, no information is recovered about preferences for the omitted level because all of the omitted levels are confounded. This limitation can be overcome using a second method known as “effects codes,” which creates a unique base level for each attribute (Louviere et al. 2000).

Begin by creating an effects-coded variable EC_1 , for the first attribute using the following steps:

- If the profile contains the first level of the attribute, set $EC_1 = 1$.
- If the profile contains the L th level of the attribute, set $EC_1 = -1$.
- If neither Step 1 nor Step 2 apply, set $EC_1 = 0$.

If an attribute has two levels, one only needs to create one effects-coded variable using the preceding three steps for that attribute. However, if an attribute has three levels, one continues the coding process by creating a second effects-coded variable, EC_2 , for that attribute using three additional steps:

- If the profile contains the second level of the attribute, set $EC_2 = 1$.
- If the profile contains the L th level of the attribute, set $EC_2 = -1$.
- If neither Step 4 nor Step 5 apply, set $EC_2 = 0$.

If an attribute has more than three levels, one continues creating effects codes in this manner until $(L - 1)$ effects codes are created for each L -level attribute. Using this coding scheme, the parameter value for the omitted attribute level can be simply computed. For example, the value of the parameter for the L th level of an attribute is the sum $b_1(-1) + b_2(-1) + \dots + b_{L-1}(-1)$, where b_n is the parameter estimate on the n th level ($n \neq L$) of an effects-coded variable.

For labeled experiments, as well as for the status quo alternative in a generic experiment, it is important to include a code for the alternative specific constant. Alternative specific constants are coded using dummy variables and, if there are N alternatives in the choice set, then $(N - 1)$ alternative specific constants can be included in the econometric specification. Because the status quo alternative will typically set the attributes at their current level (unless the status quo is an alternative such as “stay at home”), the status quo levels are coded with the same codes used for the other alternatives.

5.4 The Random Utility Model

The analysis of responses to a choice experiment is based on an extension of the random utility maximization (RUM) model that underlies discrete choice contingent valuation responses (Chap. 4) and recreation site choices between competing alternatives (Chap. 6). The CE format focuses the respondent's attention on the trade-offs between attributes that are implicit in making a choice. As shown below, model estimates are based on utility differences across the alternatives contained in choice sets.

The RUM model is based on the assumption that individuals know their utility with certainty, but analysts are unable to perfectly observe respondent utility so the unobservable elements are part of the random error. This assumption is formalized in a model where utility is the sum of systematic (v) and random (ε) components for individual k :

$$V_{ik} = v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik}, \quad (5.4)$$

where V_{ik} is the true but unobservable indirect utility associate with Alternative i , Z_i is a vector of attributes associated with Alternative i , p_i is the cost of Alternative i , y_k is income, and ε_{ik} is a random error term with zero mean.

For simplicity, let's consider an individual who is faced with a choice between mutually exclusive alternatives, where each alternative is described with a vector of attributes, Z_i . We assume that this individual maximizes their utility when making a choice. Therefore the individual will choose Alternative i if and only if

$$v_{ik}(Z_i, y_k - p_i) > v_{jk}(Z_j, y_k - p_j); \quad \forall j \in C, \quad (5.5)$$

where C contains all of the alternatives in the choice set. However, from an analyst's perspective, unobserved factors that influence choice enter the error term and, thus, individual k will choose Alternative i if and only if

$$v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik} > v_{jk}(Z_j, y_k - p_j) + \varepsilon_{jk}; \quad \forall j \in C. \quad (5.6)$$

The stochastic term in the random utility function allows probabilistic statements to be made about choice behavior. The probability that a consumer will choose Alternative i from a choice set containing competing alternatives can be expressed as

$$P_{ik} = P[v_{ik}(Z_i, y_k - p_i) + \varepsilon_{ik} > v_{jk}(Z_j, y_k - p_j) + \varepsilon_{jk}; \quad \forall j \in C]. \quad (5.7)$$

Equation (5.8) is very general, and assumptions need to be made about the specification of the utility function and the probability distribution of the error terms in order to estimate a model.

5.4.1 Specification of the Utility Function

A common assumption is that utility is a linear function of the attributes included in the experimental design so that the utility of choosing Alternative i is

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}, \quad (5.8)$$

where β is the vector of preference parameters for nonmonetary attributes and λ is the marginal utility of money. When choosing a specification, there is a trade-off between the benefits of assuming a less restrictive formulation (e.g., including interaction terms) and the complications that arise from doing so. Furthermore, the specifications that can actually be identified depend on the experimental design (see Sect. 5.3).

Consider an experiment with three attributes, including a monetary attribute. A utility function that is a linear function of the attributes would then be written as

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.9)$$

However, if the experiment allows for an interaction term between the two nonmonetary attributes, the utility function could be specified as

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 z_{i1} z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.10)$$

Note that this function remains linear in parameters, but it is not a linear function of the attributes.

One important property of discrete choice models is that only the differences in utility between alternatives affect the choice probabilities—not the absolute levels of utility. This can be shown by rearranging the terms in Eq. (5.7):

$$P_{ik} = P[\varepsilon_{ik} - \varepsilon_{jk} > v_{jk}(Z_j, y_k - p_j) - v_{ik}(Z_i, y_k - p_i); \quad \forall j \in C]. \quad (5.11)$$

Here one sees that choices are made based on utility differences across alternatives. Thus, any variable that remains the same across alternatives, such as respondent-specific characteristics like income, drops out of the model. Although Eq. (5.11) indicates that there must be a difference between attribute levels for competing alternatives in order to estimate the preference parameters for the attributes, the levels of some attributes could be equal in one or several of the choice sets.²¹

The property that there must be a difference between alternatives also has implications for the possibility of including alternative specific constants. Because alternative specific constants capture the average effect on utility of factors that are

²¹As discussed in Sect. 5.3, when attribute levels are the same across alternatives within a choice set, they do not elicit respondent trade-offs and therefore are uninformative.

not explicitly included as attributes, only differences in alternative specific constants matter. As was described in Sect. 5.3.6, a standard way of accounting for this is to normalize one of the constants to zero so that the other constants are interpreted as relative to the normalized constant.

The fact that only utility differences matter also has implications for how socioeconomic characteristics can enter a RUM model. Socioeconomic characteristics are used to capture observed taste variation, and one way of including them in the model is as multiplicative interactions with the alternative specific constants. Otherwise, these characteristics could be interacted with the attributes of the alternatives.

5.4.2 The Multinomial Logit Model

The next step is to make an assumption regarding the distribution of the error term. Alternative probabilistic choice models can be derived depending on the specific assumptions that are made about the distribution of the random error term in Eq. (5.11). The standard assumption in using a RUM model has been that errors are independently and identically distributed following a Type 1 extreme value (Gumbel) distribution. The difference between two Gumbel distributions results in a logistic distribution, yielding a conditional or multinomial logit model (McFadden 1974). This model relies on restrictive assumptions, and its popularity rests to a large extent on its simplicity of estimation. The multinomial logit model is introduced first and its limitations are discussed before introducing less restrictive models.

For simplicity, suppose that the choice experiment to be analyzed consists of one choice set containing N alternatives ($i = 1, \dots, N$). If errors are distributed as Type 1 extreme value, the multinomial logit model applies, and the probability of respondent k choosing Alternative i is

$$P_{ik} = \frac{\exp(\mu v_{ik})}{\sum_{j=1}^N \exp(\mu v_{jk})}, \quad (5.12)$$

where μ is the scale parameter that reflects the variance of the unobserved part of utility (Ben-Akiva and Lerman 1985). In basic models, the scale parameter is typically set equal to one, although other formulations will be discussed below.

There are two important properties of the multinomial logit model: (1) the alternatives are treated as independent, and (2) the modeling of taste variation among respondents is limited. The first problem arises because of the independently and identically distributed assumption about the error terms and results in the famous independence of irrelevant alternatives property. This property states that the ratio of choice probabilities between two alternatives in a choice set is unaffected by other alternatives in the choice set. This can be seen in the expression for the ratio of choice probabilities for the multinomial logit model:

$$\frac{P_{ik}}{P_{nk}} = \frac{\exp(\mu v_{ik}) / \sum_{j=1}^N \exp(\mu v_{jk})}{\exp(\mu v_{nk}) / \sum_{j=1}^N \exp(\mu v_{jk})} = \frac{\exp(\mu v_{ik})}{\exp(\mu v_{nk})}. \quad (5.13)$$

This expression only depends on the attributes and the levels of the attributes for the two alternatives and is assumed to be independent of other alternatives in the choice set. This is a strong assumption that might not always be satisfied.

Fortunately, the assumption about independence of irrelevant alternatives can easily be tested. If independence of irrelevant alternatives is satisfied, then the ratio of choice probabilities should not be affected by whether another alternative is in the choice set or not. One way of testing independence of irrelevant alternatives is to remove one alternative and re-estimate the model and compare the choice probabilities. This type of test was developed by Hausmann and McFadden (1984) and is relatively simple to conduct (see Greene, 2002, for details). If the test indicates that the assumption of independence of irrelevant alternatives is violated, an alternative model should be considered. One type of model that relaxes the homoscedasticity assumption of the multinomial logit model is the nested multinomial logit model (Greene 2002). In this model, the alternatives are placed in subgroups, and the error variance is allowed to differ between the subgroups but is assumed to be the same within each group. Another alternative specification is to assume that error terms are independently, but nonidentically, distributed Type I extreme value, with a scale parameter (Bhat 1995). This would allow for different cross elasticities among all pairs of alternatives. Furthermore, one could also model heterogeneity in the covariance among nested alternatives (Bhat 1997).

The second limiting property of the multinomial logit model is how the model handles unobserved heterogeneity. As we will see, observed heterogeneity can be incorporated into the systematic part of the model by allowing for interaction between socio-economic characteristics and attributes of the alternatives or constant terms. However, the assumption about independently and identically distributed error terms is severely limiting with respect to unobserved heterogeneity.

5.5 Welfare Measures

The goal of most CEs is to estimate economic welfare for use in policy analysis. Because CEs provide quantitative measures of tradeoffs between attributes (including price), they can be used to estimate how much money respondents would be willing to pay for a change in attribute levels while remaining as well off after the change as they were before the change, which provides estimates of compensating variation (Chap. 2). The fact that CEs provide estimates of the indirect utility function allows one to calculate willingness to pay for gains or losses for any combination of change in attribute levels.

Section 5.3.3 described methods for generating choice sets for “state-of-the-world” (generic, unlabeled) experiments and alternative specific (labeled) designs, and it is

important to understand that estimates of WTP are computed differently for these two designs. This is because in a state-of-the-world experiment, only one alternative can be realized at the end of the day. In an alternative specific experiment, several of the alternatives may exist at the same time. This means, in turn, that an additional problem arises because when one wants to make welfare evaluations, an assumption needs to be made about which alternative a respondent would choose. For example, changes in a policy attribute (such as adding a boat ramp at Campground A) might cause some respondents to choose that alternative instead of a different alternative, while some others already chose that alternative before the change and, finally, others will still not choose that alternative.

Another concept that is somewhat related to the difference between a state-of-the-world experiment and an alternative specific experiment is the difference between generic labels (such as Alternative A, Alternative B, and so forth) and explicit labels (such as “car,” “bus”). The generic labels do not convey any information about the alternatives, so WTP is simply a function of preferences over the levels of the policy-related attributes. In contrast, explicit labels convey information about fixed attributes that are not associated with the experimentally designed attributes, and WTP is therefore a function of preferences regarding both the policy-related attributes and the fixed attributes associated with the label.

5.5.1 Willingness to Pay: State-of-the-World Experiments

In the case of a state-of-the-world CE, one can think that there is only a single alternative to consider that can exhibit various attribute conditions or “states of the world.”

Assume a simple linear utility function for Alternative i , where the alternative simply represents a certain state of the world, and respondent k :

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.14)$$

Suppose one wishes to estimate the respondent’s WTP for a change in the attribute vector from initial conditions (Z_0) to altered conditions (Z_1). To estimate the compensating variation for a new state of the world versus a base case, one does not have to consider the probability of choosing different alternatives. Therefore, the compensating variation (CV) associated with this change is

$$CV = \frac{1}{\lambda} \{V^1 - V^0\}, \quad (5.15)$$

where V_1 and V_0 are expressions of utility for the new and base case states of the world. For example, suppose one conducts a choice experiment with three attributes, including the cost attribute, and the following utility function is estimated:

$$v_{ik} = \beta_1 z_{i1} + \beta_2 z_{i2} + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.16)$$

Based on the estimated model, one wishes to calculate the WTP for changes in the two nonmonetary attributes relative to the base: Δz_{i1} and Δz_{i2} . WTP would then be

$$\text{WTP} = -\frac{\beta_1 \Delta z_{i1} + \beta_2 \Delta z_{i2}}{\lambda}. \quad (5.17)$$

This expression shows the maximum amount of money an individual is willing to pay in order to obtain the improvement in the two attributes.

So far we have discussed WTP for a discrete change in multiple attributes. However, what is often reported from generic choice experiments is the marginal WTP. Using a simple linear utility function (Eq. 5.14), the marginal rate of substitution between any of the attributes and money is simply the ratio of the coefficient of the attribute and the marginal utility of money:

$$\text{MRS} = -\frac{\partial v_{ik} / \partial z_i}{\partial v_{ik} / \partial y_k} = -\frac{\beta_i}{\lambda} = \text{MWTP}. \quad (5.18)$$

Marginal WTP (also known as the implicit price) shows how much money an individual is willing to sacrifice for a marginal change in the attribute. Note that because the expression is a ratio of the coefficients, the scale parameter cancels from the expression.

However, in many instances the attributes are not continuous. For example, the attribute could be a dummy variable indicating if the attribute is present or not. In that case, the ratio of the attribute coefficient and the marginal utility of money is strictly not a marginal WTP because one cannot talk about a marginal change of the discrete attribute. The interpretation of this WTP measure is instead the amount of money a respondent is willing to pay for a change in the attribute from, say, not available to available.

Two extensions of the estimation of marginal WTP can now be considered. The first concerns nonlinear utility functions for which marginal WTP would have to be evaluated at a certain level of the attribute. Suppose one includes an interaction term in an experiment with two attributes so that the utility function is

$$v_{ik} = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2 + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.19)$$

The marginal WTP for attribute z_1 is then

$$\text{MWTP} = -\frac{\partial v_{ik} / \partial z_1}{\partial v_{ik} / \partial y_k} = -\frac{\beta_1 + \beta_3 z_2}{\lambda}. \quad (5.20)$$

Therefore, the marginal WTP for Attribute 1 depends on the level of the other attribute, and one would have to decide at what values to calculate the WTP.

The second extension is to allow for observed heterogeneity in WTP. This can be done by interacting attributes of the choice experiment with a set of socioeconomic characteristics (see Sect. 5.7.2). This way one would obtain the marginal WTP for different groups of people with a certain set of socioeconomic characteristics. Note that in many choice experiments, the socio-economic characteristics are interacting with the alternative specific constants. In that case they will not affect the marginal willingness to pay.

5.5.2 Sources of Variation in Marginal WTP and Comparison of Models

The expressions derived for marginal willingness to pay are point estimates. It is important to report the uncertainty of the estimates as well. Moreover, in many cases one would like to make tests between different models or data sets with respect to differences in marginal WTP. However, the preference parameters used to compute marginal WTP result from maximum likelihood estimation, and standard errors are associated with each point estimate. The problem is that one wants to find the standard deviation of an expression that is a nonlinear function of a number of parameters. There are several methods for doing this (Kling 1991; Cooper 1994). One common approach is the Delta method which involves a first-order Taylor series expansion of the WTP expression. For example, if $MWTP = -\frac{\beta_i}{\lambda}$, then the variance is approximately

$$\text{var}(MWTP) = \frac{1}{\lambda^2} \text{var}(\beta_i) + \frac{\beta_i^2}{\lambda^4} \text{var}(\lambda) - 2 \frac{\beta_i}{\lambda^3} \text{cov}(\lambda, \beta_i). \quad (5.21)$$

Most econometric programs include routines for estimating the variance of nonlinear functions of estimated parameters using the Delta method.

Another method is the so-called Krinsky-Robb method (Krinsky and Robb 1986). This method is based on a number of random draws from the asymptotic normal distribution of the parameter estimates, and the welfare measure is then calculated for each of these draws. The standard deviation or the confidence interval is then constructed based on these draws.

A third method for computing the variance of marginal WTP is “bootstrapping,” where a number of new data sets are generated by resampling, with replacement, from the estimated residuals. For each of these new data sets the model is re-estimated and welfare measures are calculated. The confidence intervals or standard deviation is then constructed based on this set of welfare measures. The Krinsky-Robb method is less computationally burdensome than bootstrapping, but its success critically depends on how closely the distribution of errors and asymptotically normal distribution coincide. Kling (1991), Cooper (1994), and Chen and Cosslett (1998) find that bootstrapping and the Krinsky-Robb method give quite similar standard deviations.

By estimating marginal WTP and the variance with any of the approaches above, one can make direct tests of difference between, say, two different groups of respondents or between different experiments. However, if one would like to pool the data from two different experiments and estimate a joint model, he or she would need to be cautious. For example, one could have conducted the same type of CE in two different countries. Now he or she wants to test whether the preferences in the two countries are the same. There are, however, two factors that could differ between the two countries: the utility parameters and the scale parameter. Remember that one cannot simply compare the estimated coefficients from two sets of data because the coefficient estimates are confounded with the scale parameter. Nevertheless, it is actually possible to construct tests of both these differences using an approach proposed by Swait and Louviere (1993). They showed that it is possible to estimate the ratio of scale parameters for two different data sets. This procedure can then be used to compare different models or to pool data from different sources (Adamowicz et al. 1994; Ben-Akiva and Morikawa 1990).

5.5.3 Willingness to Pay: Alternative Specific Experiments

When evaluating WTP in an alternative specific experiment, it is important to understand that the researcher is not certain which alternative an individual would chose. Consequently, WTP in alternative specific experiments is based on the probability of choosing the various labeled alternatives, and this is accomplished using the so-called log-sum formula (Hanemann 1999; Morey 1999), which is an expected utility version of the welfare measures used in contingent valuation. Again assume a simple linear utility function for Alternative i and respondent k :

$$v_{ik} = \beta Z_i + \lambda(y_k - p_i) + \varepsilon_{ik}. \quad (5.22)$$

Suppose a researcher wishes to estimate the respondent's WTP for a change in the attribute vector from initial conditions (Z_0) to altered conditions (Z_1). The compensating variation (CV) associated with this change is obtained by solving the equality

$$V_k(Z^0, p^0, y_k) = V_k(Z^1, p^1, y_k - CV), \quad (5.23)$$

where V is the unconditional utility function. Recall that in the RUM model, respondents are assumed to choose the alternative that maximizes their utility: $\max(V_k) = \max(v_{ik} + \varepsilon_{ik}) \forall i$. This expression can be written in alternative form as

$$V_k[Z, p, y_k] = \lambda y_k + \max[\beta Z_1 - \lambda p_1 + \varepsilon_1, \dots, \beta Z_N - \lambda p_N + \varepsilon_N], \quad (5.24)$$

where N is the number of alternatives. Inserting this expression into Eq. (5.23) for the compensating variation, we obtain the following equality:

$$\begin{aligned} \lambda y_k + \max[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \dots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0] \\ = \lambda(y_k - \text{CV}) + \max[\beta Z_1^1 - \lambda p_1^1 + \varepsilon_1^1, \dots, \beta Z_N^1 - \lambda p_N^1 + \varepsilon_N^1]. \end{aligned} \quad (5.25)$$

Now, solve Eq. (5.25) for the compensating variation:

$$\begin{aligned} \text{CV} = \frac{1}{\lambda} \left\{ \max[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \dots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0] \right. \\ \left. - \max[\beta Z_1^1 - \lambda p_1^1 + \varepsilon_1^1, \dots, \beta Z_N^1 - \lambda p_N^1 + \varepsilon_N^1] \right\}. \end{aligned} \quad (5.26)$$

The final step requires expressions for the expected value of the maximum indirect utility of the alternatives. In order to do this, an assumption about the error terms in Eq. (5.22) is needed. It turns out that if the errors have an extreme value Type 1 distribution (as generally assumed for RUM models), then the expected value of the maximum utility is the so-called log-sum (or inclusive) value. For example, the log-sum value for Alternative i under initial attribute conditions can be written as

$$\max[\beta Z_1^0 - \lambda p_1^0 + \varepsilon_1^0, \dots, \beta Z_N^0 - \lambda p_N^0 + \varepsilon_N^0] = \ln \sum_{i=1}^N e^{v_i^0}. \quad (5.27)$$

This result leads to a very convenient expression for computing the compensating variation:

$$\text{CV} = \frac{1}{\lambda} \left\{ \ln \sum_{i=1}^N e^{v_i^0} - \ln \sum_{i=1}^N e^{v_i^1} \right\}, \quad (5.28)$$

which is simply the difference between the expected values of maximum utility for the initial and altered attribute levels for Alternative i , divided by the marginal utility of money.

5.6 Relaxing the Assumptions of the Conditional Logit Model

Up to this point, two assumptions have been made to simplify the econometric analysis of the conditional logit model. First, it was assumed that everyone in the population has the same preference structure. This assumption restricts the β 's to be the same for all members of the population. Second, it was assumed that the ratio of choice probabilities between any two alternatives was unaffected by other alternatives in the choice set. This property (independence of irrelevant alternatives) results in limited substitution possibilities.

This section looks at a few models that relax these assumptions. In particular, it will focus on models that relax the assumption of identical preference parameters for all respondents, and it will look at three modifications: (1) including interaction effects, (2) estimating a latent class/finite mixture model, and (3) using a random parameter/mixed logit approach. Regarding the independence of irrelevant alternatives property, the main approach to address this issue has been the nested logit model (Ben-Akiva and Lerman 1985; Louviere et al. 2000).

5.6.1 Interaction Effects

Individual- (respondent-) specific variables (age, wealth, etc.) cannot be examined directly in a conditional logit model because these variables do not vary across alternatives. Thus, individual-specific variables drop out of the utility difference. However, individual-specific variables can interact with alternative specific attributes to provide some identification of attribute parameter differences in response to changes in individual characteristics. For example, interacting age with the price attribute would generate information on the marginal utility of money (price) as a function of age. This is a simple approach that provides insight into heterogeneity of consumers, but it assumes we already know the elements that lead to heterogeneity (those items included as interaction effects) and results in many parameters and potential collinearity problems.

5.6.2 Latent Class/Finite Mixture Model

A more advanced approach is to use a latent class/finite mixture model in which it is assumed that respondents belong to different preference classes that are defined by a small number of segments. Suppose S segments exist in the population, each with different preference structures and that individual k belongs to segment s ($s = 1, \dots, S$). The conditional indirect utility function can now be expressed as $V_{ik|s} = v_{ik|s} + \varepsilon_{ik|s}$. For simplicity, one can write the deterministic part of utility as $v_{ik} = \beta Z_i$, where again Z_i is a vector of attributes that now includes the monetary attribute. The preference parameters (β) vary by segment, so that one can write the indirect utility function as $V_{ik|s} = \beta_s Z_i + \varepsilon_{ik|s}$. The probability of choosing Alternative i depends on the segment one belongs to and can be expressed as

$$P_{ik|s} = \frac{\exp(\beta_s Z_i)}{\sum_{j=1}^N \exp(\beta_s Z_k)}, \quad (5.29)$$

where the β 's are segment-specific utility parameters (and scale is fixed at 1).

Now let there be a process describing the probability of being included in a particular segment as a function of demographic (and other) information. Following Boxall and Adamowicz (2002), Swait (1994), and Gupta and Chintagunta (1994), that process can be specified as a separate logit model to identify segment membership as

$$P_{ks} = \frac{\exp(\delta_s X_i)}{\sum_{j=1}^N \exp(\beta_s Z_k)}, \quad (5.30)$$

where X is a set of individual characteristics and δ is a vector of parameters.

Let P_{iks} be the joint probability that individual k belongs to segment s and chooses Alternative i . This is also the product of the probabilities defined in Eqs. (5.29) and (5.30): $P_{iks} = P_{ik|s} \times P_{ks}$. The probability that individual k chooses i becomes the key component in the finite mixture or latent class approach:

$$P_{ik} = \sum_{s=1}^S P_{ik|s} P_{ks} = \sum_{s=1}^S \frac{\exp(\beta_s Z_i)}{\sum_{j=1}^N \exp(\beta_s Z_k)} \frac{\exp(\delta_s X_i)}{\sum_{j=1}^N \exp(\beta_s Z_k)}. \quad (5.31)$$

The joint distribution of choice probability and segment membership probability is specified and estimated in this model. Note that this approach provides information on factors that affect or result in preference differences. That is, the parameters in the segment membership function indicate how the probability of being in a specific segment is affected by age, wealth, or other elements included in the segment membership function. Further details on this approach to heterogeneity can be found in Swait (1994), Boxall and Adamowicz (2002), or Shonkwiler and Shaw (1997).

Note that the ratio of probabilities of selecting any two alternatives would contain arguments that include the systematic utilities of other alternatives in the choice set. This is the result of the probabilistic nature of membership in the elements of S . The implication of this result is that independence of irrelevant alternatives need not be assumed (Shonkwiler and Shaw 1997).

One issue with latent class models is the choice of number of classes, S . The determination of the number of classes is not part of the maximization problem, and it is not possible to use conventional specification tests such as a likelihood ratio tests. Some sort of information criteria are sometimes used (Scarpa and Thiene 2005), as well as stability of the parameters in the segments as tools to assess the best number of classes to represent the data.

5.6.3 Random Parameter/Mixed Logit Model

Another advanced approach to identifying preference heterogeneity is based on the assumption that parameters are randomly distributed in the population. Then, the

heterogeneity in the sample can be captured by estimating the mean and variance of the random parameter distributions. This approach is referred to as random parameter logit or mixed logit modeling (Train 1998). In order to illustrate the random parameter logit model one can write the utility function of Alternative i for individual k as

$$v_{ik} = \beta Z_i + \varepsilon_{ik} = \bar{\beta} Z_i + \tilde{\beta}_k Z_i + \varepsilon_{ik}, \quad (5.32)$$

where, again, Z_i is a vector of attributes, including the monetary attribute. With this specification, the parameters are not fixed coefficients, but rather they are random. Each individual's coefficient vector, β , is the sum of the population mean, $\bar{\beta}$, and an individual deviation, $\tilde{\beta}_k$. The stochastic part of utility, $\tilde{\beta}_k Z_i + \varepsilon_{ik}$, is correlated among alternatives, which means that the model does not exhibit the independence of . It is assumed that the error terms are independently and identically distributed Type I extreme value.

Assume that the coefficients β vary in the population with a density distribution $f(\beta|\theta)$, where θ is a vector of the underlying parameters of the taste distribution. The probability of choosing Alternative i depends on the preferences (coefficients). The conditional probability of choosing Alternative i is

$$P_{ik|\beta} = \frac{\exp(\beta Z_i)}{\sum_{j=1}^N \exp(\beta Z_k)}. \quad (5.33)$$

Following Train (1998), the unconditional probability of choosing Alternative i for individual k can then be expressed as the integral of the conditional probability in (5.33) over all values of β :

$$P_{ik|\theta} = \int P_{ik|\beta} f(\beta|\theta) d\beta = \int \frac{\exp(\beta Z_i)}{\sum_{j=1}^N \exp(\beta Z_k)} f(\beta|\theta) d\beta. \quad (5.34)$$

In general, the integrals in Eq. (5.34) cannot be evaluated analytically, so one has to rely on simulation methods (Train 2003).

It is important to point out the similarities between the latent class model and the random parameter logit model. The probability expression (Eqs. 5.31 and 5.34) are both essentially weighted conditional logit models. Equation (5.31) reflects a finite weighting or mixture, whereas Eq. (5.34) is a continuous mixture.

The random parameter logit model requires an assumption to be made regarding the distribution of the coefficients. Note that it is not necessary for all parameters to follow the same distribution, and not all parameters need to be randomly distributed. The choice of distribution is not a straightforward task. In principle, any distribution could be used, but in previous applications the most common ones have been the normal and the log-normal distribution. Other distributions that have been applied are the uniform, triangular, and Raleigh distributions.

There are several aspects that one could consider when determining the distribution of the random parameters. First, one might want to impose certain restrictions. The most natural one might be that all respondents should have the same sign for the coefficients. Of the previously discussed distributions, only the log-normal distribution has this property. For example, if one assumes that the cost coefficient is log-normally distributed, it ensures that all individuals have a nonpositive price coefficient. In this case, the log-normal coefficients have the following form:

$$\beta_k = \pm \exp(b_k + \eta_k), \quad (5.35)$$

where the sign of coefficient β_k is determined by the researcher according to expectations, b_k is constant and the same for all individuals, and η_k is normally distributed across individuals with mean and variance equal to 0 and σ_k^2 , respectively. This causes the coefficient to have the following properties: (1) median = $\exp(b_k)$; (2) mean = $\exp(b_k + \sigma_k^2/2)$; and (3) standard dev = $\exp(b_k + \sigma_k^2/2) (\exp(\sigma_k^2) - 1)^{0.5}$. While the log-normal distribution seems like a reasonable assumption, there may be some practical problems in its use. First, experience has shown that this distribution often causes difficulty with convergence in model estimation, likely because of the restriction it places that all respondents have the same sign on the associated coefficient. Another problem with the log-normal distribution is that the estimated welfare measures could be extremely high because values of the cost attribute close to zero are possible.

5.7 Application to Swedish Wetlands

To illustrate how the econometric models that have been described can be interpreted and inform policy decisions, an empirical example based on data collected in a mail survey regarding Swedish citizens' valuation of wetland attributes is presented. For purposes of this chapter, the following example is modified from data descriptions and analyses presented elsewhere (Carlsson et al. 2003). This example is used because it illustrates various econometric specifications in a concise way. However, it is acknowledged that the attribute specifications do not reflect current best practices regarding utility endpoints, and if the experiment had been conducted today, the specifications would have been improved (especially regarding the design of the biodiversity attribute). Given this caveat, the example is first used to illustrate the basic multinomial logit model, and then it is extended to the more advanced econometric models.

In Sweden and elsewhere, there is an increasing interest in the restoration and construction of wetlands. The purpose of the choice experiment was to identify public preferences for the characteristics of a wetland area located in southern Sweden. The attributes and their levels are presented in Table 5.7.

Table 5.7 Attributes and attribute levels for a wetland choice experiment

Attribute	Description	Levels
Cost	The lump-sum cost for the individual if the alternative is chosen	SEK ^a 200, 400, 700, 850
Landscape vegetation	The land cover type surrounding the wetland	Forest, meadow
Biodiversity	Alternative levels of plant, animal, and insect species	Low, medium, high
Sport fish	Improved habitat for sport fish such as bass and pike	No, yes
Fence	The wetland is enclosed with a 1-m fence in order to prevent drowning accidents	No, yes
Crayfish	Swedish crayfish are established and harvesting is allowed	No, yes
Walking trails	Walking trails are constructed with signs describing the plant and animal life	No, yes

^aOne Swedish krona (SEK) = approximately \$0.15

Wetland attributes	Alternative 1	Alternative 2	Alternative 3
Landscape vegetation	Forest	Forest	Meadow
Biodiversity	Low	Low	High
Sport fish	No	Yes	No
Fence	No	No	Yes
Crayfish	No	Yes	No
Walking trails	No	No	Yes
Cost	SEK 0	SEK 850	SEK 400
I would choose: (please check one box)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 5.1 Example of a choice set for a wetland choice experiment

In the survey, respondents made selections from four choices sets, and in each set they had to choose between three alternatives, one of which was a base alternative (Alternative 1, simple ponds) with no improvements and low biodiversity. An example choice set is presented in Fig. 5.1.

The results are presented in sequence, beginning with the multinomial logit model and then moving on to the extensions.

5.7.1 Results from the Multinomial Logit Model

The results from the maximum likelihood estimation of a multinomial logit model are shown in Table 5.8; note that not all of the attribute coefficients are statistically significant.²² For the attribute “landscape vegetation,” the dummy variable for meadowland is negative but statistically insignificant, which indicates that there is no significant difference in preferences between meadowland and forest vegetation. For the biodiversity attribute, the dummy variables for medium and high biodiversity are both positive and statistically significant. This indicates that respondents prefer both medium and high levels of biodiversity relative to the baseline of low biodiversity. The dummy variables for improved sport fish habitat and walking trails are both positive and statistically significant, which indicates that respondents prefer having these attributes relative to a baseline of no sport fish habitat improvement and no walking trails, respectively. The dummy variables for crayfish and fence construction are both negative and statistically significant, which indicates that respondents dislike establishment of crayfish and construction of a fence around the area. The coefficient of the total cost attribute is negative and significant, which, as expected, indicates that respondents have a positive marginal utility of money. Finally, the coefficient of the alternative specific constant is negative and insignificant. Recall that, as defined here, the alternative specific constant represents the utility of choosing the status quo alternative (Alternative 1). This result suggests the absence of status quo bias (choosing the status quo regardless of attribute levels) and that respondents made choices strictly based on the level of the attributes.

Marginal WTP values are estimated by dividing the attribute coefficient by the marginal utility of money.²³ The marginal WTP values indicate the strength of respondent preferences for the attributes expressed in Swedish kronor. For example, on average, respondents are willing to pay almost 680 kronor (roughly \$102) in a lump sum to obtain a high biodiversity wetland compared with a low-biodiversity wetland. The marginal WTP to obtain a medium biodiversity wetland is 512 kronor. A simple *t*-test (using standard errors calculated using the Delta method) reveals that there is no statistically significant difference between respondents willingness to pay for high relative to medium biodiversity. Marginal WTP for improved sport fish habitat is 351 kronor relative to the base level of no action. The marginal WTP for the construction of a fence is -169 kronor. This indicates the marginal WTP is 169 kronor less than the marginal WTP for the base level, which was no fenced area.

²²In all of the tables, *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.10 level. Also, standard errors (s.e.) of the coefficients are shown in parentheses below the coefficients.

²³The coefficients shown in Model 1 (Table 5.8) have been rounded to three decimal places. However, the marginal WTP values shown in Table 5.8 were computed before rounding. Computation of marginal WTP values based on the coefficients shown in Model 1 will therefore result in slightly different values than reported in Table 5.8.

Table 5.8 Multinomial logit parameter estimates for the wetland choice experiment

	Model 1 coefficient (s.e.)	Model 1 marginal WTP (s.e.)
ASC (Alternative specific constant for status quo alternative)	-0.119 (0.125)	–
Landscape vegetation is meadowland	-0.052 (0.061)	-45 (53)
High biodiversity	0.784*** (0.086)	679*** (106)
Medium biodiversity	0.591*** (0.085)	512*** (94)
Sport fish, improved habitat	0.405*** (0.062)	351*** (65)
Fence constructed	-0.194*** (0.062)	-169*** (58)
Crayfish established	-0.130** (0.061)	-113** (55)
Walking trails constructed	0.753*** (0.063)	653*** (92)
Cost	-0.001*** (0.0001)	–

5.7.2 Results from Model Extensions

This section examines the results from models that relax the standard multinomial logit assumption, starting with a model using interaction terms (Table 5.9). This approach will help develop intuition regarding how key variables in the latent class and random parameter logit models might be identified. The results of two models are reported—one with interaction between one socio-economic characteristic—male (a dummy variable equal to one if the respondent is a male)—and the alternative specific constant (Model 2), and another model that interacts this characteristic with the alternative specific constants and all of the other attributes except cost (Model 3).

This section will not comment on all of the results in detail because the interpretation of the coefficient estimates and WTP values have been discussed already. The one aspect that is important to point out is the result in Model 2, where the alternative specific constant is interacted with the male dummy variable. In Model 1, without interaction terms, the alternative specific constant is negative but statistically insignificant. In Model 2, with an interaction term, there are significant differences between males and females. Females have a negative alternative specific constants (-0.275), which is statistically significant. The negative sign indicates that choosing the status quo decreases indirect utility for this group of respondents (choosing alternatives to the status quo increases indirect utility). The interaction

Table 5.9 Multinomial logit parameter estimates for the wetland example with interaction terms

	Model 2 coefficient (s.e.)	Model 3 coefficient (s.e.)
ASC (for status quo alternative)	-0.275** (0.138)	-0.001 (0.161)
ASC × male	0.317*** (0.114)	0.238 (0.224)
Landscape vegetation is meadowland	-0.051 (0.061)	0.219*** (0.085)
Landscape vegetation is meadowland × male	-	-0.580*** (0.125)
High biodiversity	0.784*** (0.086)	1.001*** (0.122)
High biodiversity × male	-	-0.418** (0.174)
Medium biodiversity	0.591*** (0.085)	0.737*** (0.118)
Medium biodiversity × male	-	-0.271 (0.172)
Sport fish, improved habitat	0.406*** (0.062)	0.340*** (0.084)
Sport fish, improved habitat × male	-	0.150 (0.125)
Fence constructed	-0.195*** (0.062)	-0.095 (0.084)
Fence constructed × male	-	0.234* (0.125)
Crayfish established	-0.131** (0.061)	-0.125 (0.084)
Crayfish established × male	-	-0.003 (0.125)
Walking trails constructed	0.753*** (0.063)	0.777*** (0.084)
Walking trails constructed × male	-	-0.005 (0.128)
Cost	-0.001*** (0.0001)	0.001*** (0.0001)

term between the alternative specific constant and the male dummy variable is positive. The alternative specific constant for males is the sum of the two terms (i.e., $-0.275 + 0.317 = 0.042$), which is positive.

When estimating marginal WTP values for the attributes in Model 3, note that it is possible to estimate multiple values for each attribute. First, one can estimate one value for women, which will be the ratio of the attribute coefficient to the marginal utility of money. One can estimate a second value for men, which will be the ratio of the sum of the attribute coefficient plus the interaction term to the marginal utility

of money. And third, one can estimate the sample mean WTP values as well. This is calculated as a weighted sum of female and male WTP values, where the weights are the percentage of females and males in the sample (Table 5.10).

Again, this section will not discuss all the results in detail. The first observation that can be made is that the average WTP is almost the same in the two models when both males and females are included, as would be expected. However, the model with interaction terms (Model 3) reveals that, in some instances, there are large differences between men and women. For example, females have a positive WTP for meadowland, while males have a negative WTP. Furthermore, females have a higher WTP for high and medium levels of biodiversity (compared with low levels) relative to males.

The results obtained using a random parameter logit model and a latent class model are shown in Table 5.11. In the random parameter logit model, all the nonmonetary attributes were specified as being normally distributed. In the latent class model, two classes were specified and the gender of the respondent was used to explain class membership.

For the random parameter logit model, there are two coefficients estimated for each of the random parameters, where the first is an estimate of the mean preference and the second is an estimate of the standard deviation of preferences across the sample. Three of the standard deviation parameters are statistically significant, which indicates that the model is capturing unobserved heterogeneity. Furthermore, the standard deviation estimates are large relative to the mean values and indicate that respondents had reverse preferences (opposite signs) for some attributes. For example, the coefficient for the fence attribute is -0.494 , and the standard deviation is 1.605 . This indicates that, on average, respondents disliked fencing in the

Table 5.10 Marginal WTP values (and standard errors) comparing the basic multinomial logit model results and a multinomial logit model with interaction terms

	Model 1	Model 3		Sample
	Sample	Female	Male	
Landscape vegetation is meadowland	-45 (53)	183** (72)	-301*** (83)	-47 (52)
High biodiversity	679*** (106)	836*** (134)	487*** (117)	671*** (102)
Medium biodiversity	512*** (94)	616*** (120)	389*** (114)	507*** (92)
Sport fish, improved habitat	351*** (65)	284*** (76)	410*** (88)	344*** (63)
Fence constructed	-169*** (58)	-79 (71)	-275*** (85)	-172*** (56)
Crayfish established	-113** (55)	-105 (72)	-102 (78)	-103* (54)
Walking trails constructed	653*** (92)	649*** (102)	645*** (105)	647*** (89)

Table 5.11 Parameter estimates of the random parameter and latent class logit models for the wetland choice experiment example

	Random parameter logit model		Latent class logit model	
	Coefficient (s.e.)	Coefficient (s.e.)	Class 1 (s.e.)	Class 2 (s.e.)
ASC (for status quo alternative)	0.048 (0.211)	–	–0.328 (0.324)	0.039 (0.381)
Landscape vegetation is meadowland	–0.220 (0.158)	1.489** (0.695)	–1.105*** (0.413)	0.782*** (0.288)
High biodiversity	1.402*** (0.309)	1.019 (0.944)	0.313 (0.276)	1.145*** (0.286)
Medium biodiversity	1.031*** (0.242)	0.859 (1.025)	0.136 (0.265)	1.127*** (0.269)
Sport fish, improved habitat	0.788*** (0.202)	0.689 (1.008)	0.913*** (0.221)	0.131 (0.188)
Fence constructed	–0.494*** (0.186)	1.605** (0.663)	–0.586*** (0.216)	0.021 (0.166)
Crayfish established	–0.477*** (0.218)	1.982*** (0.682)	0.297 (0.196)	–0.095 (0.159)
Walking trails constructed	1.472*** (0.362)	0.383 (0.662)	1.087*** (0.262)	0.839*** (0.177)
Cost	–0.002*** (0.0006)		–0.002*** (0.0004)	–0.001*** (0.0004)
<i>Latent class probability model</i>				
Constant			–0.595 (0.527)	–
Male			1.359*** (0.290)	–
Average class probabilities			0.510	0.490

wetland, but that a fraction of respondents had a positive preference for constructing a fence.

The latent class model was estimated with two classes (male and female), and one can see some distinct differences between the two classes. For example, in the first class, there are no statistically significant preferences for improvement in biodiversity, while in the second class, biodiversity is the most important attribute. In the class inclusion equation, the gender of the respondent is statistically significant, which means that it is more likely that a man belongs to Class 1 than to Class 2. Not surprisingly, these results are consistent with results reported in Model 3 (Table 5.9).

Next the marginal WTP values for the random parameter logit and latent class logit models are estimated (Table 5.12).

For the random parameter logit model, the point estimates of marginal WTP are found to be mostly similar to the values obtained from the standard logit model

Table 5.12 Marginal WTP values from random parameter logit and latent class logit models

	Random parameter logit (s.e.)	Latent class logit	
		Class 1 (s.e.)	Class 2 (s.e.)
Landscape vegetation is meadowland	-99 (65)	-724** (293)	565** (230)
High biodiversity	634*** (101)	205 (200)	1045*** (310)
Medium biodiversity	466*** (92)	89 (177)	813*** (268)
Sport fish, improved habitat	356*** (68)	598*** (194)	95 (139)
Fence constructed	-223*** (75)	-383** (165)	15 (120)
Crayfish established	-215*** (79)	-194 (144)	-69 (118)
Walking trails constructed	665*** (91)	712*** (238)	606*** (193)

(Model 1). However, it should be noted that the estimated unobserved heterogeneity has not been taken into account (i.e., there is an estimated standard deviation of WTP values as well). For the latent class logit model, the marginal WTP values are estimated for the two classes, and as can be seen, there are considerable differences between the two classes. Again, in the second class there is a considerable WTP for biodiversity, while in the first class priority is put on fishing and walking trails.

Note two important considerations. First, for the latent class logit model, mean WTP for the sample can be estimated by taking the weighted average of the two classes, where the weights are the average class probabilities in Table 5.11. By doing that, the estimated WTP values will be similar to the WTP values of the conditional logit model and the random parameter logit model. Thus, neither the random parameter logit nor latent class logit model results in very different overall WTP values, but they do provide much richer information about the distribution of preferences over the sample.

Second, the results of the latent class logit model are consistent with what was found in the conditional logit model where the attributes were interacted with the male dummy variable (Model 3). In that model, for example, it was found that men care less about biodiversity and more about fish, while women care more about biodiversity.

5.8 Conclusions

Choice experiments have emerged as the preferred method (relative to rankings and ratings) for conducting stated-preference studies when a good or service is best characterized by a suite of attributes. This result is primarily explained by the fact

that CEs mimic actual market behavior in a policy context where various dimensions of the policy attributes are under consideration. The use of CEs provides an opportunity to evaluate the welfare effects of multiple policy options and to calibrate ex ante value estimates to ex post policies. The realistic nature of well-designed CEs is complemented by the power of random utility maximization models to describe decision-making involving trade-offs among attributes and by recent advances that integrate RUM models with experimental design. This evolution of thinking and practice provides powerful tools for collecting and analyzing data that can assist the estimation of environmental (and other) values in a benefit-cost context and can assist benefit transfers when funds for original studies are scarce.

Despite the mathematical and statistical nature of many of the topics presented in this chapter, the authors want to highlight the importance of the preliminary steps in designing a CE, which rely more on developing communication skills among economists, other scientists, policymakers, and members of the public. A CE should only be used for policy analysis when it is clearly the best method available for data collection and analysis. The ability to identify which attributes of a policy issue people really care about and how, where, and when environmental changes may impact use and passive-use values are topics that require careful deliberation and meaningful conversations in focus group settings. Also, as emphasized in the description of experimental design procedures, conducting meaningful survey pretests is essential for constructing an efficient experimental design. Without due care and attention to these steps, CE applications will not provide the desired information. However, by following the guidance provided in this chapter, researchers will be able to design and analyze a CE that provides credible value estimates that can provide meaningful support for decision-making.

Appendix 1: Choice Experiments and Behavior

Many choice experiments that appear in the literature examine passive-use values or “total economic values” in that they ask respondents to choose between options that may affect outcomes associated with passive-use values (e.g., endangered species) or use values (recreation enjoyment, etc.). These are often framed as referenda or social choices. The case study examined in this chapter is an example of this type of choice experiment. However, choice experiments can also be based on behavioral choices alone.

In other literature, such as transportation and marketing, choice experiments are typically used to assess how behavior such as transport mode choice or choice of a product will vary with different attributes. Indeed, the earliest applications of choice experiments in economics included cases of recreation site choice or property choices. There are a variety of reasons to use choice experiments in the analysis of behavior, even if revealed preference data on such choices are available. Choice experiments can present attributes that are outside the range of the existing set of attributes (e.g., higher levels of fishing catch rates or higher levels of congestion on

hiking trails) and reflect new attributes (e.g., environmental labels on consumer products), and experimental design can help in identifying parameters on attributes that are typically correlated in the real world (e.g., water quality and fish catch). Choice experiment data can also be combined with revealed preference data to help calibrate stated-preference responses or to compare stated and revealed preference information. Efforts in “data fusion” or the combination of stated and revealed preference information have included analyses of recreation, property choices, as well as the integration of perceived and objective measures of attributes. A key aspect of the use of choice experiments in behavioral contexts is that the collection of stated choice data support a model based on economic theory such as the travel cost model of recreation choice behavior (Bockstael and McConnell 2007) or the random utility approach to property choice and valuation (Phaneuf et al. 2013). To help illustrate these approaches we provide two examples of choice experiments that involve linkages to behavioral choices—recreation site choice and property choices.

In the recreation case, the respondent is asked to choose between two moose hunting sites in which the attributes include distance (the travel cost) and other characteristics of the hunting site (Fig. 5.2). The attributes are described in a way that hunters typically view them, and in a way that they can be translated to “objective” measures of wildlife populations and site descriptions. This case study, drawn from Adamowicz et al. (1997), included information on actual site choices and elicited information from hunters on their perceptions as well as actual measures of attributes. This data set has been used in other applications including the assessment of unobserved heterogeneity in stated and revealed preference data (von Haefen and Phaneuf 2008). Note that in this case, the hunting sites are described as “generic” sites (Site A and Site B). In some contexts these can be described as actual sites with “labels” such as the name of the area or the administrative label

Assuming that the following areas were the ONLY areas available, which one would you choose on your next hunting trip, if either?

Features of hunting area	Site A	Site B	
Distance from home to hunting area	50 kilometers	150 kilometers	
Quality of road from home to hunting area	Mostly gravel or dirt, some paved	Mostly paved, some gravel or dirt	
Access within hunting area	No trails, cutlines, or seismic lines	Newer trails, cutlines or seismic liens passable with a four wheel drive vehicle	Neither Site A nor Site B
Encounters with other hunters	No hunters, other than those in my hunting party, are encountered	Other hunters, on all terrain vehicles, are encountered	I will NOT go moose hunting
Forestry activity	Some evidence of recent logging found in the area	No evidence of logging	
Moose population	Evidence of less than 1 moose per day	Evidence of 3-4 moose per day	
Check ONE and only one box	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 5.2 Example of a recreational hunting site choice

(see Boxall and Adamowicz, 2002, for an application to canoeing sites). Similar types of choice experiments have been used to elicit tradeoffs in the context of subsistence resource use (trading off distance with traditional use hunting or caloric expenditures for fuelwood collection (see Adamowicz et al., 2004).

The second example (Fig. 5.3) is similar to the property choice cases used in Phaneuf et al. (2013) and is based on Kim (2014). In this case, information about a respondent’s current house is elicited. This information is used as the “base” case for the choice experiment, and attributes are presented that describe changes to the house (larger area, different water quality in the adjacent lake, etc.). This choice experiment uses an experimental design referred to as a “pivot design” in that it pivots the choices around the currently held option or behavior (Hess and Rose 2009).

Appendix 2: Choice Experiments and the Value of Health Risk Reduction

A nonmarket value that is very important in policy analysis is the value of mortality risk reduction, often referred to as the value of statistical life (see Cameron, 2010, for a review of the issues and a thorough critique of the term “value of statistical life”). The value of mortality risk reductions often comprises over 80% of the monetary value of air pollution reduction policies such as the assessment of the U.S. Clean Air Act Amendments. Mortality risk values have typically been measured using hedonic wage models (see Chap. 7) wherein the impact of changing job risk characteristics are reflected in higher wages, all else held constant. Over the past few decades however stated-preference methods have been increasingly used to elicit the value of risk reductions. In a typical setting a respondent is informed about baseline risk levels and then presented with a treatment that offers a reduction in health risks, but at a cost. The tradeoff between cost and risk change provides a measure of the monetary value of risk reduction.

Characteristics	Your current house	House A	House B
House size		30% smaller	30% larger
House age		Newer	20 years older
Distance from the lake		More than 500 meters farther	250 meters farther
Water quality		10% worse	20% better
House price		30% more	30% less
Which house would you choose?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 5.3 Example of a property choice

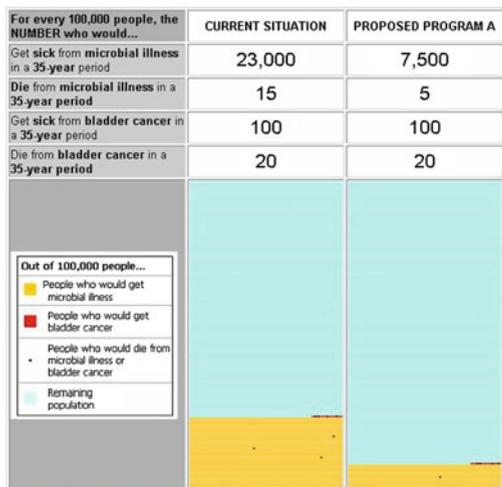
Here's the first program we want you to vote on.

THE BENEFITS OF MUNICIPAL WATER TREATMENT PROGRAM A

Based on current water drinking patterns in your community this program would have the following benefits to every 100,000 people:

- 15,500 fewer people will develop microbial illness over a 35-year period. Another way to say this is that the average person in a community of 100,000 people will see their risk of getting microbial illness from drinking the water fall from 23,000 in 100,000 to 7,500 in 100,000
- With fewer people developing microbial illness, 10 fewer people will die from getting the disease. Another way to say this is that the average person in this community will see their risk of dying from microbial illness reduced from 15 in 100,000 to 5 in 100,000
- Bladder cancer illness and deaths will not be affected by the program

Here is a table showing these benefits:



THE COST OF THE MUNICIPAL WATER TREATMENT PROGRAM A

If the majority of voters support this program your household will share in the cost starting January 2005 by paying an additional amount on your household water bill.

PLEASE VOTE NOW:

CVM21 If the estimated addition to your household's water bill was **\$25 per year (\$2.08 per month)** starting in January 2005, and a vote were held today, would you vote **FOR** or **AGAINST** the proposal?

- FOR
- AGAINST

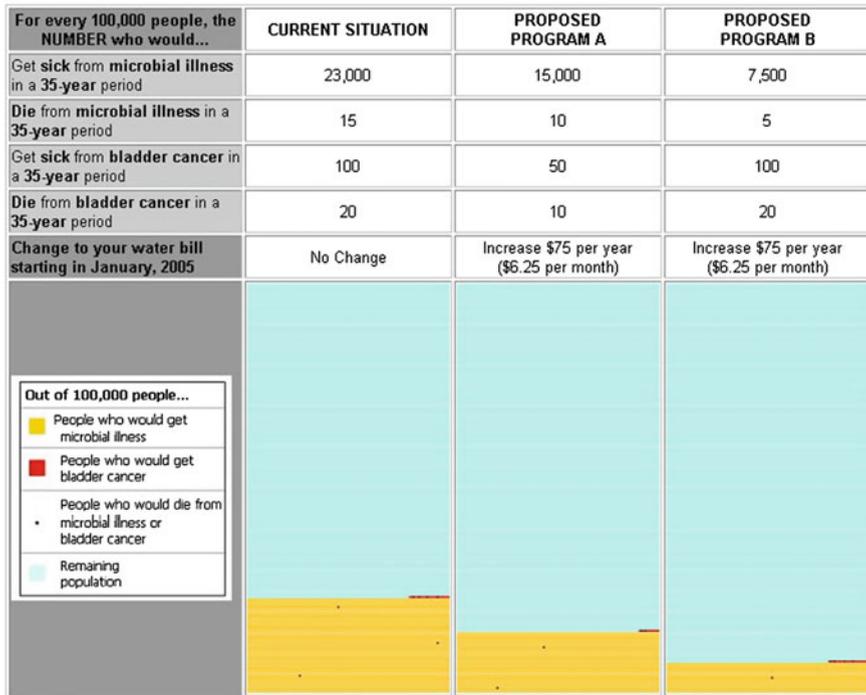


Fig. 5.4 Example of a two-alternative choice task eliciting values of health risk reductions

While contingent valuation has typically been used to measure mortality risks (e.g., Krupnick et al. 2002), choice experiments are increasingly being used to separate mortality from morbidity risks (Adamowicz et al. 2011) or to include risk context elements within the valuation tasks, such as latency (a delay in the timing of the benefits from the risk reduction), type of risk, or other elements (Alberini and Ščasný 2011).

The value of mortality risk reduction is expressed as the willingness to pay for a small reduction in the probability of mortality. For example, if individuals are willing to pay \$10,000 for a 1% reduction in their risk of dying in a year, this would translate into a \$1,000,000 value of statistical life (100 people valuing a 1% risk reduction would equal one “statistical life” and 100 times \$10,000 is \$1,000,000). A choice

This is the first scenario we want you to vote on.



DC1 If there were a referendum, I would vote for...

CHECK ONE ONLY

- Current Situation
- Proposed Program A
- Proposed Program B

Fig. 5.5 Example of a three-alternative choice experiment eliciting values of health risk reduction

experiment, therefore, can be designed to elicit trade-offs between a current situation (with the mortality risk presented) and an “improved” situation with the risks reduced.

The figures that follow illustrate these choices based on the case of water risks in Adamowicz et al. (2011). The respondent faces a base or status quo set of risks (mortality and morbidity from cancer and microbial impacts) and trades them off against a set of new programs. A two-alternative case (Fig. 5.4) and a three-alternative case (Fig. 5.5) are presented to illustrate that in this context, the choice experiment can be presented like a contingent valuation referendum task as well as in the context of a multiple alternative choice experiment (see, however, Zhang and Adamowicz, 2011). Note also that the risks are presented numerically (number of illnesses and deaths) as well as graphically, using grids of points to represent the risks.

Risk communication is a particularly important aspect of the assessment of health risk reductions. The random utility model that arises from these choices provides the

marginal utility of risk and the marginal utility of money, and thus the value of a change in risk can be derived. Adamowicz et al. (2011) examined risks in water, while other researchers have examined climate change risks (Ščasný and Alberini 2012), risks from nuclear versus fossil fuel based energy (Itaoka et al. 2006), and mortality risks from different risk contexts, including transport, respiratory illness, and cancer (Alberini and Ščasný 2011). One of the most sophisticated choice experiments examining the value of health risk reductions, from Cameron and DeShazo (2013), examined latency, timing of illness, type of illness, and other factors that affect health.

References

- Adamowicz, W., Boxall, P., Haener, M., Zhang, Y., Dosman, D. & Marois, J. (2004). An assessment of the impacts of forest management on aboriginal hunters: Evidence from stated and revealed preference data. *Forest Science*, 50, 139-152.
- Adamowicz, W., Dupont, D., Krupnik, A. & Zhang, J. (2011). Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. *Journal of Environmental Economics and Management*, 61, 213-226.
- Adamowicz, W., Louviere, J. & Swait, J. (1998). Introduction to attribute-based stated choice methods. Final report to the Resource Valuation Branch, Damage Assessment Center, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Edmonton, AB, Canada: Advantis.
- Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining stated and revealed preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26, 271-292.
- Adamowicz, W., Swait, J., Boxall, P., Louviere, J. & Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of Environmental Economics and Management*, 32, 65-84.
- Alberini, A. & Ščasný, M. (2011). Context and the VSL: Evidence from a stated preference study in Italy and the Czech Republic. *Environmental and Resources Economics*, 49, 511-538.
- Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, 77, 153-170.
- Ben-Akiva, M. & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.
- Ben-Akiva, M. & Morikawa, T. (1990). Estimation of travel demand models from multiple data sources. In M. Koshi (Ed.), *Transportation and traffic theory* (pp. 461-476). New York: Elsevier.
- Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29, 471-483.
- Bhat, C. R. (1997). Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel. *Transportation Research Part B: Methodological*, 31, 11-21.
- Bliemer, M. C. J. & Rose, J. M. (2011). Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transportation Research Part A: Policy and Practice*, 45, 63-79.
- Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N. & Freeman, P. (2008). Eliciting willingness to pay without bias: Evidence from a field experiment. *Economic Journal*, 118, 114-137.
- Bockstael, N. E. & McConnell, K. E. (2007). *Environmental and resource valuation with revealed preferences: A Theoretical Guide to Empirical Models*. Dordrecht, The Netherlands: Springer.
- Boxall, P. C. & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach. *Environmental and Resource Economics*, 23, 421-446.

- Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M. & Louviere, J. (1996). A comparison of stated preference methods for environmental values. *Ecological Economics*, 18, 243-253.
- Boyd, J. & Krupnick, A. (2009). The definition and choice of environmental commodities for nonmarket valuation. Resources for the Future Discussion Paper 09-35. Washington, DC: Resources for the Future.
- Boyle, K. J., Johnson, F. R., McCollum, D. W., Desvousges, W. H., Dunford, R. W. & Hudson, S. P. (1996). Valuing public goods: Discrete versus continuous contingent-valuation responses. *Land Economics*, 72, 381-396.
- Bunch, D. S., Louviere, J. J. & Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. Working paper. University of California-Davis.
- Cameron, T. A. (2010). Euthanizing the value of a statistical life. *Review of Environmental Economics and Policy*, 4, 161-178.
- Cameron, T. A. & DeShazo, J. R. (2013). Demand for health risk reductions. *Journal of Environmental Economics and Management*, 65, 87-109.
- Carlsson, F., Frykblom, P. & Liljenstolpe, C. (2003). Valuing wetland attributes: An application of choice experiments. *Ecological Economics*, 47, 95-103.
- Carlsson, F. & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health Economics*, 12, 281-294.
- Carson, R. T. & Groves, T. (2007). Incentive and information properties of preference questions. *Environmental and Resource Economics*, 37, 181-210.
- Carson, R. T. & Groves, T. (2011). Incentive and information properties of preference questions: Commentary and extensions. In J. Bennett (Ed.), *The International handbook of non-market environmental valuation* (pp. 300-321). Northampton, MA: Edward Elgar.
- Carson, R. T. & Louviere, J. J. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*, 49, 539-559.
- Cattin, P. & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. *Journal of Marketing*, 46 (3), 44-53.
- Chen, H. Z. & Cosslett, S. R. (1998). Environmental quality preference and benefit estimation in multinomial probit models: A simulation approach. *American Journal of Agricultural Economics*, 80, 512-520.
- Cooper, J. C. (1994). A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. *Land Economics*, 70, 111-122.
- Court, A. T. (1939). Hedonic price indexes with automotive examples. In *The dynamics of automobile demand* (pp. 99-117). New York: General Motors.
- Cummings, R. & Taylor, L. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review*, 89, 649-665.
- Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R. & Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management*, 63, 73-91.
- Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.
- Ferrini, S. & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, 53, 342-363.
- Gan, C. E. C. & Luzar, E. J. (1993). A conjoint analysis of waterfowl hunting in Louisiana. *Journal of Agricultural and Applied Economics*, 25, 36-45.
- Green, P. E. & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8, 355-363.
- Green, P. E. & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5, 103-123.
- Green, P. E. & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, 53, 107-117.
- Greene, W. H. (2002). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

- Griliches, Z. (1971). Hedonic price indexes for automobiles: An econometric analysis of quality change. In Z. Griliches (Ed.), *Price indexes and quality change: Studies in new methods of measurement* (pp. 55-77). Cambridge, MA: Harvard University Press.
- Gupta, S. & Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 31, 128-136.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255-262.
- Hanemann, W. M. (1999). Welfare analysis with discrete choice models. In J. A. Herges and C. L. Kling (Eds.), *Valuing recreation and the environment: Revealed preference methods in theory and practice* (pp. 33-64). Northampton MA: Edward Elgar.
- Hausman, J. & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52, 1219-1240.
- Hensher, D. A., Rose, J. M. & Greene, W. H. (2005). *Applied choice analysis: A primer*. New York: Cambridge University Press.
- Hess, S. & Rose, J. M. (2009). Should reference alternatives in pivot design SC surveys be treated differently? *Environmental and Resource Economics*, 42, 297-317.
- Holmes, T. P. & Boyle, K. J. (2005). Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. *Land Economics*, 81, 114-126.
- Horne, P., Boxall, P. C. & Adamowicz, W. L. (2005). Multiple-use management of forest recreation sites: A spatially explicit choice experiment. *Forest Ecology and Management*, 207, 189-199.
- Huber, J. & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33, 307-317.
- Itaoka, K., Saito, A., Krupnick, A., Adamowicz, W. & Taniguchi, T. (2006). The effect of risk characteristics on the willingness to pay for mortality risk reductions from electric power generation. *Environmental and Resource Economics*, 33, 371-398.
- Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B. & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health*, 16, 3-13.
- Kanninen, B. J. (2002). Optimal design for multinomial choice experiments. *Journal of Marketing Research*, 39, 214-227.
- Kessels, R., Jones, B., Goos, P. & Vandebroek, M. (2008). Recommendations on the use of Bayesian optimal design strategies for choice experiments. *Quality and Reliability Engineering International*, 24, 737-744.
- Kim, H. N. (2014). The demonstration and capture of an ecosystem service value: Three different methodological approaches. Ph.D. thesis. Edmonton: University of Alberta.
- Kling, C. L. (1991). Estimating the precision of welfare measures. *Journal of Environmental Economics and Management*, 21, 244-259.
- Krinsky, I. & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, 68, 715-719.
- Krupnick, A., Alberini, A., Cropper, M., Simon, N., O'Brien, B., Goeree, R. & Heintzelman, M. (2002). Age, health, and the willingness to pay for mortality risk reductions: A contingent valuation survey of Ontario residents. *Journal of Risk and Uncertainty*, 24, 161-186.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, 132-157.
- Lareau, T. J. & Rae, D. A. (1989). Valuing WTP for diesel odor reductions: An application of contingent ranking technique. *Southern Economic Journal*, 55, 728-742.
- Lindhjem, H. & Navrud, S. (2011). Using Internet in stated preference surveys: A review and comparison of survey modes. *International Review of Environmental and Resource Economics*, 5, 309-351.
- List, J. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sports cards. *American Economic Review*, 91, 1498-1507.
- Louviere, J. J. (1988a). *Analyzing decision making: Metric conjoint analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-067. Newbury Park, CA: Sage.

- Louviere, J. J. (1988b). Conjoint analysis modeling of stated preferences: a review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy*, 10, 93-119.
- Louviere, J. J. & Hensher, D. A. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, 10, 348-361.
- Louviere, J. J. & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research*, 20, 350-367.
- Louviere, J. J., Flynn, T. N. & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, 3 (3), 57-72.
- Louviere, J. J., Hensher, D. A. & Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge, United Kingdom: Cambridge University Press.
- Louviere, J. J., Islam, T., Wasi, N., Street, D. & Burgess, L. (2008). Designing discrete choice experiments: Do optimal designs come at a price? *Journal of Consumer Research*, 35, 360-375.
- Luce, R. D. (1959). *Individual choice behavior A Theoretical Analysis*. New York: Wiley.
- Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.
- Mackenzie, J. (1993). A comparison of contingent preference models. *American Journal of Agricultural Economics*, 75, 593-603.
- Manski, C. (1977). The structure of random utility models. *Theory and Decision*, 8, 229-254.
- Marschak, J. (1960). Binary choice constraints on random utility indicators. In K. Arrow (Ed.), *Mathematical methods in the social sciences, 1959: Proceedings of the first Sanford Symposium* (pp. 312-329). Stanford, CA: Stanford University Press.
- Mazzotta, M. J. & Opaluch, J. J. (1995). Decision making when choices are complex: A test of Heiner's hypothesis. *Land Economics*, 71, 500-515.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). New York: Academic Press.
- Morey, E. R. (1999). Two RUMs uncloaked: Nested logit models of site choice and nested logit models of participation and site choice. In J. A. Herriges and C. L. Kling (Eds.), *Valuing recreation and the environment: Revealed preference methods in theory and practice* (pp. 65-120). Northampton, MA: Edward Elgar.
- Naidoo, R. & Adamowicz, W. L. (2005). Biodiversity and nature-based tourism at forest reserves in Uganda. *Environment and Development Economics*, 10, 159-178.
- Phaneuf, D. J., Taylor, L. O. & Braden, J. B. (2013). Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. *Land Economics*, 89, 30-52.
- Qin, P., Carlsson, F. & Xu, J. (2011). Forest tenure reform in China: A choice experiment on farmers' property rights preferences. *Land Economics*, 87, 473-487.
- Rae, D. A. (1983). The value to visitors of improving visibility at Mesa Verde and Great Smoky National Parks. In R. D. Rowe and L. G. Chestnut (Eds.), *Managing air quality and scenic resources at national parks and wilderness areas*. Boulder, CO: Westview Press.
- Ready, R. C., Champ, P. A. & Lawton, J. L. (2010). Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. *Land Economics*, 86, 363-381.
- Roe, B., Boyle, K. J. & Teisl, M. F. (1996). Using conjoint analysis to derive estimates of compensating variation. *Journal of Environmental Economics and Management*, 31, 145-159.
- Rolfe, J., Bennett, J. & Louviere, J. (2000). Choice modelling and its potential application to tropical rainforest preservation. *Ecological Economics*, 35, 289-302.
- Rose, J. M. & Bliemer, M. C. J. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews: A Transnational Transdisciplinary Journal*, 29, 587-617.
- Sándor, Z. & Wedel, M. (2001). Designing conjoint choice experiments using manager's prior beliefs. *Journal of Marketing Research*, 38, 430-444.
- Scarpa, R. & Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, 52, 253-282.

- Scarpa, R. & Thiene, M. (2005). Destination choice models for rock climbing in the Northeastern Alps: A latent-class approach based on intensity of preferences. *Land Economics*, 81, 426-444.
- Ščasný, M. & Alberini, A. (2012). Valuation of mortality risk attributable to climate change: Investigating the effect of survey administration modes on a VSL. *International Journal of Environmental Research and Public Health*, 9, 4760-4781.
- Schultz, E.T., Johnston, R. J., Segerson, K. & Besedin, E. Y. (2012). Integrating ecology and economics for restoration: Using ecological indicators in valuation of ecosystem services. *Restoration Ecology*, 20, 304-310.
- Shonkwiler, J. S. & Shaw, W. D. (1997). Shaken, not stirred: A finite mixture approach to analyzing income effects in random utility models. Paper presented at the 1997 Annual Meeting of the American Agricultural Economics Association. August 2-4, Toronto, Ontario.
- Small, K. A. & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49, 105-130.
- Smith, V. K. & Desvousges, W. H. (1986). *Measuring water quality benefits*. Boston: Kluwer-Nijhoff.
- Street, D. J. & Burgess, L. (2007). *The construction of optimal stated choice experiments: Theory and methods*. Hoboken, NJ: Wiley-Interscience.
- Street, D. J., Burgess, L. & Louviere, J. J. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing*, 22, 459-470.
- Sur, D., Cook, J., Chatterjee, S., Deen, J. & Whittington, D. (2007). Increasing the transparency of stated choice studies for policy analysis: Designing experiments to produce raw response graphs. *Journal of Policy Analysis and Management*, 26, 189-199.
- Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional, revealed preference choice data. *Journal of Retailing and Consumer Services*, 1, 77-89.
- Swait, J. & Adamowicz, W. (2001a). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, 86, 141-167.
- Swait, J. & Adamowicz, W. (2001b). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, 28, 135-148.
- Swait, J. & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30, 305-314.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychology Review*, 34, 273-286.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Train, K. (2003). *Discrete choice methods with simulation*. New York: Cambridge University Press.
- Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics*, 74, 230-239.
- von Haefen, R. H. & Phaneuf, D. J. (2008). Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference. *Journal of Environmental Economics and Management*, 56, 19-32.
- Vossler, C. A. 2016. Chamberlin meets Ciriacy-Wantrup: Using insights from experimental economics to inform stated preference research. *Canadian Journal of Agricultural Economics*, 64, 33-48.
- Vossler, C. A., Doyon, M. & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, 4 (4), 145-171.
- Yellott, J. I., Jr. (1977). The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15, 109-144.
- Zhang, J. & Adamowicz, W. L. (2011). Unravelling the choice format effect: A context-dependent random utility model. *Land Economics*, 87, 730-743.