Marketing Letters 5:4, (1994): 351–368

Experimental Analysis of Choice

RICHARD T. CARSON
Department of Economics, University of California, San Diego, La Jolla, CA 92093, (619) 534-6319

JORDAN J. LOUVIERE
University of Utah

DONALD A. ANDERSON
University of Wyoming

PHIPPS ARABIE
Rutgers University

DAVID S. BUNCH
University of California, Davis

DAVID A. HENSHER
University of Sydney

RICHARD M. JOHNSON
Sawtooth Software

WARREN F. KUHFELD
SAS Institute

DAN STEINBERG
San Diego State University

JOFFRE SWAIT
Decision Research

HARRY TIMMERMANS
Eindhoven University of Technology

JAMES B. WILEY
University of Alberta

Key words: stated preference data, discrete choice models, external validity

Abstract

Experimental choice analysis continues to attract academic and applied attention. We review what is known about the design, conduct, analysis, and use of data from choice experiments, and indicate gaps in current knowledge that should be addressed in future research. Design strategies consistent with probabilistic models of choice process and the parallels between choice experiments and real markets are considered. Additionally, we address the issues of reliability and validity. Progress has been made in accounting for differences in reliability, but more research is needed to determine which experiments and response procedures will consistently produce more reliable data for various problems.

1. Introduction

Consumers in real markets make decisions among competing alternatives. Much marketing research is directed toward predicting how those consumers will react to changes in available choice sets, typically in the context of adding new products or modifying existing ones. Survey methods that ask consumers to make choices from "experimental choice sets" enable researchers to learn about consumer preferences for products and attributes that do not yet exist in real markets.

Choice sets are termed "experimental" because some aspects of their composition are under the control of the researcher. The objective is to simulate real choice situations to determine how consumers will react when faced with particular choice situations. We focus on (1) how to model the choice process of interest, (2) how to take accounting of respondent ability to provide information and answer questions concerning choice sets, (3) how to estimate how the model parameters of interest can be estimated in a cost-effective manner, and (4) how to validate the estimated model. We conclude with a brief overview of techniques and practices that various authors have found useful in the course of their applied work.

2. Models of the choice process

At a micro level, consumers make three basic types of choices. The first and most common is to select the most preferred among a set of discrete alternatives. The second is, conditional on a decision to buy, how many. The third is how long to wait between purchases or, equivalently, how long to undertake an activity. The first type of choice lends itself to analysis with statistical models for discrete choice data (Ben-Akiva and Lerman, 1985); the second lends itself to analysis with statistical models for count data (Cameron and Trevedi, 1986); and the third lends itself to analysis with statistical models for duration data (Nelson, 1982). The latter two models can generally be expressed using a less parsimonious (but also less restrictive) discrete choice model (Sueyoshi, 1992) by discretizing the observed dependent variable. Joint discrete/continuous models (Hanemann, 1984; Barnard and Hensher, 1992) are also possible. Although space constraints allow us to consider only models for discrete choice data, much of the discussion is applicable to surveys that elicit other forms of choice data.

The simplest discrete choice models are those with only two choices, I and II, such as whether or not to purchase the product in question. Such models are common where there are very limited data on actual choices rather than in choice experiments. Binary choice models are also common in related contexts, such as voting in a referendum or choosing to undertake a medical operation, and can serve to illustrate a number of the basic features of choice experiments. In marketing, there are typically more than two possible choices, which we will denote with increasing Roman numerals.

Each product has attributes, such as price, and attributes potentially have different levels that a researcher may want to vary. Products potentially have large numbers of attributes, A_k ($k = 1, 2, \ldots, K$), and each attribute potentially has a large number of levels, A_{kl} ($l = 1, 2, \ldots, L$). Often, "brand" is an attribute used in choice experiments that tends to encompass a larger set of attributes and levels. Effectively, brand becomes a common attribute of the products taking on a different level (i.e., specific brand) for different products; it is frequently implemented as a choice specific constant in statistical models. One product may also have an attribute not possessed by another. For instance, transportation researchers (Hensher, in press) frequently want to predict the choice between car or bus, and interest may center on how the percentage of the respondents choosing bus will vary with changes in frequency of bus service. Respondent characteristics, such as distance from home to bus stop, may also influence this choice.

Consumers are assumed to choose the alternative which gives them the most utility. Inherently an unobservable or latent variable, utility provides a convenient framework for relating observed choices to attribute levels. The utility of choosing alternative I can be represented by $U_I = V(A_{1l}, A_{2l}, \ldots, A_{kl}) + e_I$, where $V(\bullet)$ is the systematic part of utility and e_I is the random component that has been interpreted variously in the literature. The utility of other choices can be represented similarly.

In the case of two choices, the model is relatively simple: some function of the probability of choosing alternative I (rather than II) is regressed on some function of the differences in the utility levels of the two choices. With more than two choices one typically must normalize on some aspect of the choice problem and consider how to account for possible correlations between alternatives. Choice models can be estimated from a sufficiently large number of answers by a single respondent to different experimental choice sets or can be estimated with multiple respondents providing answers to multiple choice sets, which is more typical of applied work. Multiple respondents providing answers to multiple choice sets also allows one to estimate models with both product-attribute levels and respondent characteristics as predictors.

Experimentally varying the levels of an attribute allows the researcher to estimate how the frequency that a particular alternative is chosen varies with changes in the level of that attribute. Lack of sufficient variation in the levels that an attribute takes in actual market data is one of the principal reasons for using choice experiments since that variation is now under the researcher's control. The effect of changing the level of a single attribute is known as a main effect. Judicious design of the experimental choice sets will also typically permit the estimation of main effects as well as key interactions, thus allowing one to predict the effect of simultaneously changing two or more attribute levels. These effects can be changes in the levels of two or more attributes of one product or a change in the level of one or more attributes in both products. For example, the effect of increasing both the price and warranty period of product I in the presence and absence of a similar change in those attribute levels for product II.

Statistical models used with choice experiments differ according to (1) the specific functional form for the probability that a particular alternative is chosen, (2) the specific functional form that links the predictor variables to (1), and (3) the nature of the random component assumed for the difference of the utilities of the two choices. The mathematical expression for probability that any particular alternative is chosen is generally chosen to ensure that predicted probabilities lie within [0, 1]. Models are also typically chosen for computational tractability and generalizability to cases involving more than two alternatives. The most popular choices for (1) are logit and probit functions, while linear, quadratic, or log response functions for (2) are often chosen. Large data sets and increased computational power are now encouraging the use of more general functional forms.

At present, various generalizations of the binary logit model are generally used in applied work. Such models provide advantages in terms of computation and interpretability, at the expense of more restrictive assumptions concerning the relationship between alternative choices. Simple logit models (e.g., multinomial and conditional) embody what is known as the independence from irrelevant alternatives assumption (IIA). A consequence of their independent and identically distributed (IID) extreme value error assumption, IIA implies that the ratio of the choice probability for any two alternatives is unaffected by addition or deletion of alternatives. An equivalent way of expressing this assumption is that the random components of utility in logit model are uncorrelated between choices and have the same variance. As might be expected, this assumption may not hold in real data, and there are several ways to avoid this problem and remain in a logit framework. One is the "mother logit" specification (McFadden, Tye, and Train, 1977), which allows the attributes of one alternative to influence the utility of another alternative, or a frequently used special case, the nested logit (NL) specification, in which an inclusive value coefficient (similar in some respects to a correlation coefficient) is estimated between different branches of a nesting structure within which the IIA property holds. A third approach is to parameterize the model so that the IIA property holds, which may simply require a more complete specification of the systematic utility component.

A more general correlation structure can be accounted for in a natural fashion with a multinomial probit model (MNP), which allows for estimation of covariances between different alternatives. Unfortunately, such models tend to be computationally difficult for more than three alternatives and burdensome for large numbers of attributes and respondents. There is currently a growing literature on simulation methods for MNP models (see Joel L. Horowitz, this volume), which effectively increases the number of alternatives allowable in the estimation; we expect such models to become more common in the near future.

3. Respondent limits and choice experiments

Choice modeling is predicated on the integrity of the data collected from respondents who generally face some limits in their ability to process information. If tasks are too long or too difficult or lack sufficient realism and credibility, data quality will suffer in the sense of not containing the information sought. Unfortunately, respondents generally answer the questions asked and seldom go out of their way to point out problems with tasks posed.

To some extent questionnaire design is an art; consequently it is difficult to list rules that guarantee generation of good data. However, some general rules can be applied: survey instruments always should be worded in as simple and straightforward a manner as possible to help ensure respondent comprehension; choice tasks should be designed to be realistic and natural, approximating as closely as possible the actual choice context; and the choices offered should be credible.

We have had success in administering choice experiments over a wide range of numbers of choice sets, numbers of choices, numbers of attributes, and numbers of levels. Experience suggests that an "average" questionnaire deals with about seven attributes, contains four choice sets, and has about four alternatives per choice set. However, this "average" masks a large amount of variability across studies and should not be taken to constitute a best practice recommendation. For example, we have been involved in studies with numbers of attributes ranging from two to thirty, numbers of choice sets ranging from one to thirty-two and numbers of alternatives per set ranging from two to twenty-eight. It is worth noting, however, that when one of these factors increases, others are generally reduced accordingly.

Several things can be done to help respondents with tasks. First, one should ensure that respondents understand the different product attributes and levels. This may be accomplished by using the attributes and levels in questions prior to the choice task, such as having respondents pick attribute levels that best describe their "usual" products. Use of glossaries, detailed explanatory text, and/or visual representations may be helpful. Some recent studies (e.g., Anderson, Louviere, Daniel, and Orland, 1993) have used videotaped representations of attribute combinations in choice sets.

Second, it may be helpful to provide "warm-up" choice tasks to ensure that respondents understand tasks. In some instances, it may be desirable for those choice sets to present alternatives from product categories other than the one of primary interest. Without "warm-up" sets, the quality of responses to the first few choice sets may suffer. Recently Swait (1993) examined the issue of data quality related to choice set order and suggests it may be necessary to discard or discount information from early choice sets in some instances. Similarly, boredom and fatigue effects may be observed if respondents are presented large numbers of complex choice tasks. Such effects may manifest themselves in respondents simplifying choice tasks by focusing on a small set of key indicator attributes like brand and price or simply answering questions randomly. The presence of such effects can be tested by comparing individuals' marginal choice frequencies for choice tasks in different parts of the sequence. At some point there is undoubtedly a quantity versus quality trade-off in terms of the data collected, and it would be useful to have research that directly addressed these issues to provide more informed guidance for practice.

Third, choice tasks sometimes can be made more realistic by specifying choice contexts explicitly, such as, "Suppose you've been driving for four hours and decide to stop for fuel; you have the choice of these service stations." Indeed, the more concrete the choice task context, the more reliable and interpretable the results are likely to be. However, care must be taken in constructing the desired context because there may be some situations that respondents find implausible.

Fourth, there may be a range outside of which attribute levels are implausible and thus not treated seriously by respondents. Hence, attention should be paid to the possibility that some attribute levels may stretch credibility, especially if they vary from set to set. Between-subject designs may be more appropriate in such cases. It also may be difficult to vary attribute levels of branded alternatives, with which respondents are familiar outside fairly narrow ranges. For example, suppose we want to estimate market share for a new type of vehicle offered by three different manufacturers. Respondents may not be able to imagine a high-end luxury Yugo or a low-end Mercedes.

Finally, it should be noted that it is possible to make the respondent's task too easy, for example, by including alternatives that "dominate" the others in the sense that they are "better" on every benefit and cost attribute than the other alternatives in the choice set. In that case, the respondent choices do not reveal information about trade-offs between the levels of different attributes. This is most likely to happen with quantitative attribute levels instances and it may be desirable to generate designs that do not allow for such dominance (Wiley, 1978; Krieger and Green, 1991).

3.1. Preference elicitation procedures

Stated (and revealed) choice studies usually involve characterization of preference through observations of consumers' first choices. However, other methods of obtaining indicators of latent utility are also possible in stated choice experiments. These include (1) ranking alternatives in order of preference, (2) series of paired comparisons, (3) pick best and worst alternative, (4) bundle selection (i.e., pick j best alternatives), (5) consider/not consider in making decision, (6) judgmental ratings, and (7) fixed-sum allocation tasks. Each method has associated with it one or more data collection procedures and model specifications, most of which are discussed in Louviere (1994a).

Such methods attempt to obtain more information from respondents than is revealed by simply picking the best among the set of alternatives provided. At one level, all methods should produce similar information about respondent preferences if equally valid. However, tests of procedural invariance with respect to the systematic component of utility must account for the fact that different elicitation procedures may influence the random component. This requires taking into account differences in the scale or variance of the estimates to compare estimates from different data types. For example, Swait and Louviere (1993) compare the cross-task validity of yes/no consideration and stated first choice data. In their empirical application, both tasks elicited the same preference function from respondents up to a multiplicative scale factor. They also show that once recognition is given to the different variance levels, pooling both data sources yields more information upon which to estimate the underlying utility function. Similarly, Ben-Akiva, Morikawa, and Shiroishi (1991) consider pooling of revealed first choice and stated ranking data by scaling utility functions to account for different levels of variability in the data types. Louviere, Fox, and Moore (1993) also consider cross-task validity of alternative stated preference elicitation methods and show that several methods can be rescaled and pooled to produce a common utility function.

One popular method that appears particularly troublesome is exploded logit (Chapman and Staelin, 1982) where ranked data is converted into a series of first choice data. This type of data seems to be particularly noisy and cannot be pooled without extensive corrections for differences in scale (Hausman and Ruud, 1987; Ben-Akiva, Morikawa, and Shiroishi, 1991).

Future academic research should compare the reliability of different preference elicitation methods to define a set of "best practices" for the academic and market research communities. Comparative reliability tests should not aim at suggesting "optimal" elicitation procedures, but should address broader issues. For example, it would be useful to (1) characterize contexts in which each data type does or does not deliver the desired information, (2) examine under what conditions and for what objectives data from multiple elicitation procedures should be combined, and (3) develop more sophisticated forms of data pooling that better account for heterogeneity differences.

3.2. Inclusion of constant alternatives in choice tasks

In some experimental choice situations it may make sense to include alternatives whose attributes are held constant from choice set to choice set. These situations tend to fall into one of three categories. The first is a specified alternative, which is appropriate when there are some explicitly mentioned (and known) brands that compete against each other, plus one or more store brands or generics. It is important that such specified categories be given an explicit description. This makes the task less confusing for the respondent and delimits more precisely the possible interpretations of modeling results.

The second is a "your current brand" alternative, which may be appropriate when research interest focuses on what attribute levels new products must have to induce switching in well-established consumer purchase patterns, or which changes in existing brands will induce consumers of a particular competing brand to change. In this instance it may be desirable to study current consumers of a competing brand and to offer that alternative explicitly.

The third is the inclusion of a "none" or no-purchase alternative. Inclusion of this alternative allows respondents to indicate that under the circumstances described in the choice set they would prefer not to purchase any of the alternatives shown. Inclusion of this alternative should be considered for two reasons: (1) it may enhance task realism, by making the set of alternatives more akin to the "typical" marketplace decision problem, and (2) it may help estimate market penetration, making it mandatory to consider whether consumers purchase the product. The latter clearly can be seen by asking a random sample of respondents which of three brands of computers they prefer for home use. If a respondent is knowledgeable about computers from work experience but has no desire to have one at home, very strong preferences might be exhibited although that respondent will never be part of the potential market for home computers. The correct alternative for such a respondent would be none. In other instances, consumers may lack the financial resources to purchase any alternatives offered. The fraction of respondents choosing the no-purchase option is likely to be highly context specific. Clearly, omission of a no-purchase alternative when it is a realistic alternative will adversely affect both market share and volume estimates.

Olsen and Swait (1993) find that the behavioral processes captured by unconditional (i.e., no-purchase alternative available) and conditional share models are not equivalent. This is not surprising because it would be necessary for IIA to hold between the no-purchase option and the other alternatives for both tasks to produce similar results. Hence, inclusion of a no-purchase option may require the use of a nested-logit model or other more complex choice models.

The no-purchase alternative estimates market penetration if and only if respondents use it only when all other alternatives present in a set provide less utility than making no purchase. Unfortunately, respondents may use no-purchase as a means to avoid making difficult decisions. Such "satisficing" is likely a function of task difficulty, respondent fatigue, and respondent characteristics (such as education). Good survey design and incentives can help motivate respondents, thus minimizing this effect; but it can adversely influence model estimates (Huber and Pinnell, 1993). Characterization of situations where it is desirable to force choices between specific alternatives and to include no-purchase alternatives remains an open research issue.

Of course, there are some situations where no-purchase alternatives make no sense and should not be included. For example, mode choice for work trips, wherein the population of interest (i.e., workers) necessarily have to make work trips, and the question is by which mode the trip will be made. Choice sets involving almost universally purchased consumer products like toothpaste and toilet paper also should not contain no-purchase alternatives, although "go to another store" might be a reasonable alternative to include. No purchase may also be inappropriate if respondents have been prescreened to be regular purchasers of the product class of interest and one of the alternatives offered is their current choice.

4. Constructing experimental designs

During the past decade considerable progress has been made in understanding and modeling choice behavior. New estimation methods are rapidly becoming available for more sophisticated and complex choice models. Concomitantly, new strategies for constructing experimental designs were developed during the past decade that permit researchers to simulate choice situations faced by decision-makers in most real market environments. In many cases, these designs support the estimation of more sophisticated choice models like mother logit, NL and MNP, in addition to the workhorse multinomial logit (MNL) model. However, there is little work on the comparative statistical properties of various design strategies.

Successful implementation of choice experiments requires considerable up-front work dedicated to understanding the choice process and context(s) involved, identifying the attributes that influence choices, and selecting appropriate attribute levels. Careful consideration must also be paid to differences among individual decision makers and task complexities. Once these critical elements are thoroughly understood, then a design strategy for creating sets of choice alternatives must be selected that is consistent with these elements and simulates actual market conditions of interest as closely as possible.

Following this systematic approach to design selection, applied research should also address issues related to model identification and statistical efficiency before proceeding. Specifically, researchers should identify (1) choice processes or model types to be studied, (2) functional forms for utility to be estimated, and (3) measures of statistical efficiency used in evaluating design alternatives. In general, the more flexible the design is with respect to the first two items, the better. However, precision of estimates, or statistical efficiency, is conditional on the choice model and utility function, and generally there is a trade-off between flexibility in the model specification and statistical efficiency in selecting designs. This aspect of design selection and the issue of how statistical efficiency should be measured have not been extensively studied except for the binary choice model (Alberini and Carson, 1993).

There are likely to be other issues that should influence the design chosen. Among these are (1) the level of aggregation in the analysis, (2) the presence and treatment of preference heterogeneity, (3) the data collection method (e.g., personal interviews, computer assisted interviews, telephone surveys, mail surveys), (4) the need for multiple versions of the survey and the number of respondents to be assigned to each version, (5) the need for individual survey customization, commonly the case in transport applications (e.g., individual travel times for the journey to work vary widely, making it impractical and unrealistic to use the same range of levels of travel times for all subjects), and (6) whether repeated measures are used and taken into account.

Researchers familiar with traditional designs for linear models will find that background provides only a starting point for designing choice experiments. For example, consider a relatively simple problem involving four three-level attributes, none of which are brand names. There are 3^4, or eighty-one, total possible alternatives. The design problem for choice experiments involves the selection of combinations (choice sets) of these eighty-one alternatives to satisfy the objectives discussed earlier. Assuming that one wants only choice sets of size three, there are 85,320 possible triples. The design problem involves selection of the total number of triples to be used and exactly which ones. Suppose the number of choice sets is arbitrarily restricted to fifty-four. There are more ways of selecting fifty-four triples from this universe of possible triples than the fastest computer could list in our life time.

There is not yet a simple general theory that unambiguously yields the "best" design for any given situation; we question whether such a theory ever will be produced. Research to date has been primarily motivated by identification issues, using extensions of concepts and results from the general linear models literature. Useful reviews of this literature are provided by Louviere (1988) and the report of the Banff Conference (Batsell and Louviere, 1991). In general, these papers review designs that satisfy the statistical properties of the mother logit model or its nested forms such as MNL. A systematic theory of estimable utility functions for MNL models has yet to be developed, but the general principle of designing experiments consistent with mother logit models in combination with familiar results for linear models has led to the development of parsimonious designs that permit (1) a relatively wide range of utility specifications to be estimated and (2) the incorporation of useful tests of the IIA property of MNL models. For example, Louviere and Woodworth (1983) provide design strategies that support estimation of models ranging from the simple MNL case with alternative specific constants to mother logit models with alternative-specific utility functions that permit tests of IIA. As noted by Batsell and Louviere (1991), the statistical efficiency properties of these design strategies are not presently well understood.

4.1. Identification and inference

There is considerable interest in design strategies that allow one to estimate and test models that violate the IIA property of the MNL model. Most recent work uses a mother logit framework, which allows the direct capture of "context effects" on alternative I's choice probability from the presence of competing alternatives in the choice set. When variation in alternative II's attributes affects the probability of choosing I significantly more (or less) than would be predicted by IIA, the violation can be captured by the so-called attribute cross-effects. These are modeled by including terms for II's attributes in I's utility function (Batsell and Louviere, 1991).

Alternatively, if alternatives are treated primarily as "brands," the context effects are captured by "alternative cross-effects" in which the presence or absence of brands creates IIA violations. These are modeled through alternative specific constants in the utility function. Anderson and Wiley (1992) consider designs that allow estimation of these effects. Lazari (1991), Lazari and Anderson (1993), and Raghavarao and Wiley (1993) consider designs that allow alternative cross-effects along with a single attribute cross-effect. Anderson, Borgers, Ettema, and Timmermans (1992) develop a design for estimating both availability and attribute cross-effects for a transportation mode choice application.

Although the mother logit model is a computationally tractable and practical model on which to base choice experiments, it has theoretical drawbacks. Without the imposition of particular restrictions it is difficult to demonstrate that the mother logit model is consistent with utility maximization. More appealing models like NL and MNP allow one to account for IIA violations and remain within the utility maximizing paradigm. To date, however, there have been few applications (e.g., Bunch, Bradley, Golob, Kitamura, and Occhiuzzo, 1993), of these models to analyze experimental choices.

Efficiency

Efficiency issues have been largely unexplored. Maximizing the determinant of the Fisher information matrix for the estimated parameters (D-optimality) is a useful general approach for design selection in choice experiments. In some instances interest may center on the derivatives of the choice probabilities with respect to the covariates rather than on the actual model parameters estimated, and this should be taken into consideration in choosing a design. In other instances it may be reasonable to use a c-optimal approach to minimize the confidence interval around a particular statistic of interest. Entropy, the expected log density of the parameter estimates, may also be useful criteria, particularly in small samples.

Due to nonlinearities, design efficiency in most choice models depends on the (unknown) true values of the model parameters. One approach to this problem is to choose a "best guess" for parameters and derive a design based on that estimate. Probably a better approach is to look for designs that are relatively robust to changes in the true parameter values. Bunch, Louviere, and Anderson (1993) have performed such an exercise for MNL models using the D-optimality criteria. Their results highlight an important difference between choice models and standard linear models – namely, that choice model probabilities are based on utility differences and hence on differences among attribute levels rather than on absolute levels per se. They show that design efficiency is generally reduced by the number of zero attribute differences in the design matrix, which is related to the "information" obtained from the experiment.

Kuhfeld, Garratt, and Tobias (1993) and Sawtooth Software (1993) have proposed interesting schemes for the systematic generation of near-orthogonal designs using computerized search procedures that appear to have important advantages in constructing experimental designs. Their efficiency properties should be investigated further.

4.3. Repeated measures

At one extreme, each subject can be randomly assigned to a single choice set without replacement. At the other, an experiment can be designed that is sufficiently compact to be completed by all subjects. The former avoids correlations among responses both within and between subjects, the latter maximizes the chances such dependencies will occur. In practice, most researchers adopt design strategies between these two extremes. Several recent papers (Borjas and Sueyoshi, in press; Montopoli, 1992) have begun to provide the necessary statistical framework.

4.4. Hierarchical and sequential designs

Hierarchical choice models seek to address a problem that often arises with choice experiments – namely, very long lists of attributes. This new and emerging literature extends Louviere's (1984) hierarchical conjoint model. For instance, Oppewal, Louviere, and Timmermans (in press) recently used integrated choice experiments to break a large list of attributes into smaller choice experiments by grouping the attributes into logical subsets based on theory, empirical evidence, or commercial practice. They show that choice data from separate experiments can be combined into a single estimation by logical substitution. However, there is currently insufficient empirical experience with this approach to understand its limitations and advantages, and little attention has been given to alternative design strategies that would be consistent with the model.

Sequential designs are another area of possible future work. Work is just beginning on sequential analysis for choice experiments (Kanninen, 1993), but the idea has strong appeal because most optimal design strategies for probabilistic choice models require a priori knowledge of the true model parameters. Various updating procedures can be used to select new choice sets that maximize the expected informational value of additional responses. The sequential approach can incorporate information from previous work in a natural way and can potentially be implemented in an efficient manner using computer-assisted interviewing techniques.

5. External validity

External validity results for choice models were discussed, and we concluded that the evidence, although encouraging, was incomplete. External validity was defined as evidence that the choice process and utility estimates obtained in a choice experiment are the same as the process and estimates that apply in the real market of interest. We eschewed issues like "how individuals really make decisions," regarding this as an example of process validity or "ultimate truth," of interest to but not centrally germane to our charge as a workshop. We believe consideration of context and other external factors is a key component of any study of external validity. There is a dearth of published studies related to external validity. Several members of the workshop provided anecdotal evidence of predictive validity, which, although encouraging, was deemed insufficient.

There are several empirical tests of predictive validity: (1) predicting the choice of a new product, and tracking the changes in choices of that product over time; (2) demonstrating spatial and temporal transferability of experimental choice model parameters; (3) predicting the real marketplace choices of separate but statistically equivalent samples of individuals; and (4) demonstrating that utilities from a model conditional on real market choices were the same as utilities from a choice experiment, up to the limits of sampling error and rescaling by a positive constant. In the case of the first and most rigorous test, favorable anecdotal evidence was reported by several participants of models tracking changes in choices over time in Australia and the United States. With respect to published work, Kocur and Louviere (1983) applied choice models to backcast to changes in choices of public transport options in Xenia, Ohio, following a disastrous tornado in the early 1970s and found the model predictions tracked the changes. Horowitz and Louviere (1990) estimated choice models from a random sample of students who took the ACT college exam in 1987. They used the model to predict the eventual choice of college of a subset of the students and could not reject the hypothesis that the utilities in the two samples were equal up to rescaling by a constant. Finn, Louviere, Timmermans, and Hutchinson (1992) examined differences in choice model parameters for choice of shopping centers in Edmonton, Eindhoven, Oslo and Orlando, and found that, except for Orlando, they could not reject the hypothesis of transferability. Mitchell and Carson (1989) report a number of comparisons where survey-based estimates of the value of environmental amenities are similar to those derived using techniques based on revealed preferences. Hensher, Barnard, Milthorpe, and Smith (1989) found their value of time estimates from revealed choice and stated choice travel data to be statistically equivalent. This will not always be the case. Hensher and Battellino (1993), for instance, showed that experimental utilities estimated at one point in time were unable to account for community acceptance of traffic management devices at a later point in time after installation. Thus, despite some encouraging progress, much more external validity research of a rigorous nature needs to be conducted before more definitive statements can be made.

As noted earlier, tests of the equivalence of utilities estimated from choice experiments and observations of real market choices require that differences in scale between two data sets be taken into account. There has been considerable progress since the Banff Conference. Work by Louviere and his associates, Hensher and his associates, and Ben-Akiva and the Hague Consulting Group have addressed this issue with encouraging results. Several recent sources support the general usefulness of rescaling, including Hensher and Bradley (1993), Bradley and Daly (in press), Swait and Louviere (1993), Swait, Erdem, Louviere, and Dubelaar (1993), Adamowicz, Louviere, and Williams (1993), Louviere, Fox and Moore (1993), and Swait and Louviere (in press).

Germane to issue of comparing revealed and stated choice data is the need to understand and predict differences in awareness and knowledge among consumers and to find ways to incorporate such differences directly in choice models along the lines begun by Anderson, Louviere, and Jenkins (1992). There is also related work in which choice experiments have been run in the context of laboratory test market and information acceleration studies, where the objective is to estimate trial and repeat rates for new products.

6. Additional topics

A variety of useful things have been learned by experience. These include techniques that aid in segment identification, exploratory analysis, and incorporating awareness and knowledge. With respect to segment identification, if the segments are not known a priori, one is forced to define them a posteriori using sample information. This can be done using either the discrete responses or the number of times a specific attribute such as brand was chosen as a basis for clustering the respondents. Work by Brossier (1990) and Gaul and Schader (1988) provide potentially useful approaches to this problem. Increasing attention is also being given to latent class or mixture model approaches (Cardell, 1989; Dillon and Kumar, in press; Swait, 1993). These models are special cases of a random coefficients model. Unfortunately, estimation of such models currently poses formidable operational and computational problems. These problems are rapidly being solved but no general purpose software is yet available.

If subjects receive different sets of choice sets, one must rely on other approaches. One obvious approach is to disaggregate the choice data and introduce individual difference measures to explain differences in tastes and preferences (e.g., Guadagni and Little, 1983; Louviere, Fox, and Moore, 1993; and Swait, Erdem, Louviere, and Dubelaar, 1993). While theoretically and empirically appealing, the number of possible terms potentially required in such a complex model may often be daunting. For example, the final college choice model estimated by Horowitz and Louviere (1990) contained well over 100 terms, which was only a small fraction of the number of possible terms.

Several exploratory research approaches have been useful in identifying potentially significant and meaningful individual difference effects. One is to cross-tabulate the choice data by alternative and individual difference variables to determine likely candidate effects that might shift the alternative specific constants. Similarly, one can cross-tabulate the choice data by alternative, attribute, and individual difference measures to identify differences in responses to attributes of different alternatives. These simple analyses "work" because all choice data, whether from real market observations or choice experiments, can be viewed as very large and sparse contingency tables. The usual caveats that apply to marginal analyses that do not take the effects of all other variables into account apply to this simple-minded approach. A slightly more sophisticated version of the foregoing exploratory approach is to apply multiple correspondence analysis (MCA) to the choice data, an appropriately coded and weighted design matrix and the individual difference data as proposed by Kaciak and Louviere (1990). A researcher using MCA for any purpose should be aware of the limitations, pointed out by Hubert and Arabie (1992), on its interpretation.

References

Adamowicz, W., J. J. Louviere, and M. Williams. (1993). "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities," Working paper, University of Alberta.

Alberini, A., and R. T. Carson. (1993). "Choice of Thresholds for Efficient Binary Discrete Choice Estimation," Discussion paper 90–34R, Department of Economics, University of California, San Diego, September.

Anderson, D. A., A. Borgers, D. Ettema, and H. Timmermans. (1992). "Estimating Availability Effects in Travel Choice Modeling: A State Choice Approach," *Transportation Research Record,* 1357, 51–65.

Anderson, D. A., J. J. Louviere, T. Daniel, and B. Orland. (1993). "Comparing Verbal and Visual Attribute Representations in Choice-Based Conjoint: Videotapes Versus Paper and Pencil," Working paper, Department of Marketing, University of Utah.

Anderson, D. A., J. J. Louviere, and W. S. Jenkins (1992). *Factors Affecting Users' Choices of National Forest Recreation Sites.* Report to the U.S. Forest Service.

Anderson, D. A., and J. B. Wiley. (1992). "Efficient Choice Set Designs for Estimating Cross-Effects Models," *Marketing Letters,* 3, 357–370.

Barnard, P. O., and D. A. Hensher. (1992). "The Spatial Distribution of Retail Expenditures," *Journal of Transport Economics and Policy,* 26, 299–312.

Batsell, R. R., and J. J. Louviere. (1991). "Experimental Choice Analysis," *Marketing Letters,* 2, 199–214.

Ben-Akiva, M., and S. R. Lerman. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press.

Ben-Akiva, M., T. Morikawa, and F. Shiroishi. (1991). "Analysis of the Reliability of Preference Ranking Data," *Journal of Business Research.* 24, 149–164.

Borjas, G., and G. Sueyoshi. (In press). "A Two-State Estimator for Probit Models with Structural Group Effects," *Journal of Econometrics*.

Bradley, M. A., and A. J. Daly. (In press). "Use of the Logit Scaling Approach in Stated Preference Analysis," *Transportation*.

Brossier, G. (1990). "Piecewise Hierarchial Clustering," *Journal of Classification*, 7, 197–216.

Bunch, D., M. Bradley, T. Golob, R. Kitamura, and G. Occhiuzzo. (1993). "Demand for Clean-fuel Vehicles in California: A Discrete Choice Stated Preference Pilot Project," *Transporation Research A*, 27A, 237–253.

Bunch, D., J. Louviere, and D. Anderson. (1993). "A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes," Working paper, University of California, Davis.

Cameron, A. C., and P. K. Trevedi. (1986). "Econometric Models Based on Count Data: Comparisons and Application of Some Estimators and Tests," *Journal of Applied Econometrics*, 1, 29–53.

Cardell, S. (1989). "The Hedonic Demand Model and Some Other Extensions of Multinomial Logit." Ph.D. dissertation, Harvard University.

Chapman, R. G., and R. Staelin (1982). "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model," *Journal of Marketing Research*, 19, 281–299.

Dillon, W., and A. Kumar. (In press). "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview." In R. Bagozzi (ed.), *Handbook of Marketing Research*. Oxford: Blackwell.

Finn, A., J. J. Louviere, H. Timmermans, and W. Hutchinson. (1992). "International Generalizability of Shopping Center Choice and Consideration Models," Paper presented at Marketing Science Conference, London, July.

Gaul, W., and M. Schader. (1988). "Clusterwise Aggregation of Relations," *Applied Stochastic Models and Data Analysis*, 4, 273–282.

Guadagni, P. M., and J. D. C. Little. (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

Hanemann, W. M. (1984). "Discrete Continuous Models of Consumer Demand," *Econometrica*, 52, 541–561.

Hausman, J. A., and P. Ruud. (1987). "Specifying and Testing Econometric Models for Rank-Ordered Data," *Journal of Econometrics*, 34, 83–104.

Hensher, D. A. (In press). "Stated Preference Analysis of Travel Choices: The State of Practice," *Transportation*.

Hensher, D. A., P. Barnard, F. Milthorpe, and N. Smith. (1989). "Urban Tollways and the Valuation of Travel Time Savings," *Economic Record*, 66, 146–156.

Hensher, D. A., and H.C. Battellino. (1993). "The Use of Discrete Choice Models in the Determination of Community Preferences Toward Sub-arterial Traffic Management Devices," Proceedings of the 7th World Conference on Transport Research, Lyon France.

Hensher, D. A., and M. A. Bradley. (1993). "Using Stated Response Data to Enrich Revealed Preference Discrete Choice Models," *Marketing Letters*, 4, 139–152.

Horowitz, J. L., and J. J. Louviere. (1990). "The External Validity of Choice Models Based on Laboratory Experiments." In M. Fischer, P. Nijkamp, and Y. Papageorgiou (eds.), *Spatial Choices and Process*. Amsterdam: North-Holland.

Huber, J., and J. Pinnell. (1993). "Value and Process Thresholds Governing Acceptance of a Default Alternative in Multiple Choice Tasks," Paper presented at Duke Invitational Symposium on Choice Modeling and Behavior.

Hubert, L. J., and P. Arabie. (1992). "Correspondence Analysis and Optimal Structural Representations," *Psychometrika*, 56, 119–140.

Kaciak, E., and J. J. Louviere. (1990). "Multiple Correspondence Analysis of Multiple Choice Data," *Journal of Marketing Research*, 27, 455–465.

Kanninen, B. (1993). "Design of Sequential Experiments for Contingent Valuation Studies," *Journal of Environmental Economics and Management*, 25, 1–11.

Kocur, G., and J. J. Louviere. (1983). "The Magnitude of Individual Level Variations in Demand Coefficients: A Xenia, Ohio, Case Example," *Transportation Research,* 17, 363–374.

Krieger, A. B., and P. E. Green. (1991). "Designing Pareto Optimal Stimuli for Multiattribute Choice Experiments," *Marketing Letters,* 2, 337–348.

Kuhfeld, W. F., M. Garratt, and R. D. Tobias. (1993). "Nonorthogonal Experimental Design Theory with Marketing Research Applications." Paper presented at AMA Forum, Monterey, CA, June.

Lazari, A. (1991). "Designs for Discrete Choice Experiments Including Availability and Cross Effects." Ph.D. Dissertation, University of Wyoming.

Lazari, A., and D. A. Anderson. (1993). "Availability and Attribute Cross-Effects Model: Determinant Optimal Designs," Technical report, University of Wyoming.

Louviere, J. (1994a). "Conjoint Analysis." In R. Bagozzi (ed.), *Handbook of Marketing Research.* Oxford: Blackwell.

———. (1984b). "Hierarchical Information Integration: A New Method for the Design and Analysis of Complex Multiattribute Judgement Problems," *Advances in Consumer Research* 11, 148–155.

———. (1988). *Analyzing Decision Making: Metric Conjoint Analysis.* Beverly Hills: Sage.

Louviere, J. J., M. Fox, and W. Moore. (1993). "Cross-Task Validity Comparisons of Stated Preference Choice Models," *Marketing Letters,* 4, 205–213.

Louviere, J. J., and G. Woodworth. (1983). "Design and Analysis of Simulated Consumer Choice of Allocation Experiments," *Journal of Marketing Research,* 20, 350–367.

McFadden, D., W. Tye, and K. Train. (1977). "An Application of Diagnostic Test for the Irrelevant Alternatives, Property of the Multinomial Logit Model, *Transportation Research Record,* 637, 39–46.

Montopoli, G. (1992). "The Analysis of Discrete Choice Set Experiments with Correlated Error Structure and Related Logistic Topics." Ph.D. Dissertation, University of Wyoming.

Mitchell, R. C., and R. T. Carson. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Baltimore: Johns Hopkins University Press.

Nelson, W. (1982). *Applied Life Data Analysis.* New York: John Wiley.

Olsen, G. D., and J. Swait. (1993). "The Importance of Nothing." Working paper, University of Calgary.

Oppewal, H., J. J. Louviere, and H. Timmermans. (In press). "Modeling Hierarchical Information Integration Processes with Integrated Conjoint Choice Experiments," *Journal of Marketing Research.*

Raghavarao, D., and J. B. Wiley. (1993). "Experimental Designs for Availability Effects and Cross Effects with One Attribute." Working paper, Department of Marketing and Economic Analysis, University of Alberta.

Sawtooth Software. (1993). *The CBC System for Choice-Based Conjoint Analysis.* Sun Valley: Sawtooth Software.

Sueyoshi, G. (1992). "Semi-parametric Proportional Hazards Estimation of Competing Risks Models with Time Varying Covariates," *Journal of Econometrics,* 51, 25–58.

Swait, J. (1993). "A Structural Equation Model of Latent Segmentation and Product Choice for Cross-Sectional Revealed Preference Choice Data." Paper presented at the AMA Forum, Monterey, CA, June.

Swait, J., T. Erdem, J. J. Louviere, and C. Dubelaar. (1993). "The Equalization Price: A Measure of Consumer-Perceived Brand Equity," *International Journal of Research in Marketing.*

Swait, J., and J. J. Louviere. (1993). "The Role of the Scale Parameter in the Estimation and Use of Generalized Extreme Value Models," *Journal of Marketing Research,* 30, 305–314.

———. (In press). "A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice," *Transportation.*

Wiley, J. B. (1978). "Selecting Pareto Optimal Subsets from Multi-Attribute Alternatives," *Advances in Consumer Research,* 5, 171–174.