



London School of Economics and University of Bath London School of Economics

Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity

Author(s): Jordan J. Louviere

Source: *Journal of Transport Economics and Policy*, Vol. 22, No. 1, Stated Preference Methods in Transport Research (Jan., 1988), pp. 93-119

Published by: [University of Bath](#) and [London School of Economics and Political Science](#)

Stable URL: <http://www.jstor.org/stable/20052837>

Accessed: 02-11-2015 22:39 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of Bath, London School of Economics and University of Bath, London School of Economics and London School of Economics and Political Science are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Transport Economics and Policy*.

<http://www.jstor.org>

CONJOINT ANALYSIS MODELLING OF STATED PREFERENCES

A Review of Theory, Methods, Recent Developments and External Validity

By Jordan J. Louviere*

1. INTRODUCTION

This paper reviews three aspects of work in conjoint analysis:

- (1) the pros and cons of theory and methods that are available to design, implement and analyse conjoint judgment and choice studies, using the modes of ranking, rating, discrete choice and resource allocation;
- (2) some recent developments in conjoint judgment and choice analysis and their potential applications to analysis of travel choice;
- (3) evidence of the external validity of conjoint analysis methods applied to predictions of actual behaviour.

The intention of this review is to provide a frame of reference for so-called “stated preference” analysis in transport.

The term “conjoint analysis” means decomposition into part-worth utilities or values of a set of individual evaluations of, or discrete choices from, a designed set of multiattribute alternatives. A number of different paradigms are now available to researchers interested in applying conjoint analysis to study consumers’ travel decisions. These paradigms have in common the use of experimental or quasi-experimental design techniques to construct sets of multiattribute alternatives (for example, travel modes described by different combinations of values or levels of attributes such as travel time, travel cost, etc.). Despite a common reliance on experimental design techniques to construct combinations of attributes, the various conjoint analysis paradigms differ in (a) response modes used to obtain information from subjects, (b) methods of analysis, and (c) inferences that can be made about judgment or choice behaviour.

Traditionally, “conjoint analysis” has been a collective term, covering both the theory and methods of a variety of different paradigms that can be used to design,

* Professor of Marketing and Economic Analysis, University of Alberta, Canada.

implement and analyse *judgment data* experiments (see Green and Srinivasan, 1978). We define judgment data as evaluative rankings or ratings of a set of multi-attribute alternatives obtained from individuals. These data are assumed by analysts to be at least ordinal in measurement level.

Recently, theory and methods have been developed to allow one to design and analyse conjoint *choice data* experiments (Louviere and Hensher, 1982; Louviere and Woodworth, 1983). We define choice data as responses that identify one and only one of a set of alternatives as the "highest", "best", etc. We also regard as choice data responses which involve the allocation of fixed sets of resources such as trips, time or money to a set of alternatives, so that the individual allocations sum to the total fixed amount.

Differences between judgment and choice data are important, because judgment data may not contain information about choice behaviour and may not satisfy various assumptions necessary to forecast choice behaviour. By definition, choice data contain information about individuals' choice behaviour, but one must make assumptions about which of several possible choice processes underlies the data. External validity is an issue for both types of conjoint data. These distinctions, unfortunately, have been confused in both marketing and transport research. For example, Green (1974) in discussing considerations in the design of conjoint *judgment* experiments, refers to them as "choice" experiments, which they are not.

The distinction between judgment and choice data continues to be blurred, as is revealed in recent papers by Green and Srinivasan (1978) and Cattin and Wittink (1982, 1986), among others. In transport research, conjoint analysis has been variously referred to as "direct utility assessment" (Lerman and Louviere, 1978), "functional analysis" (Benjamin and Sen, 1983), and "stated preference analysis" (Ministry of Transport and Public Works, the Netherlands, 1985). To avoid semantic and other confusion, it is important for transport researchers to agree on a nomenclature for the study of utility and preference formation by individual travellers and/or aggregates of travellers. We therefore recommend that this be henceforth referred to as "conjoint analysis", the term by which it is known in marketing and other social science disciplines.

But perhaps the nomenclature used to describe what transport researchers do in "stated preference" analysis is less important than the recognition that different theoretical and methodological considerations apply to different types of conjoint response data. These considerations are intimately associated with particular paradigms for the design and analysis of conjoint studies, and account for differences among various conjoint paradigms. Writers on marketing research, in particular, have tended to concentrate on commonalities in conjoint paradigms (see, for example, Green and Srinivasan, 1978); but it is equally important to understand and appreciate their differences.

In the next section of this paper we provide a brief overview of several of the most commonly applied conjoint paradigms, emphasising the "differences" rather than the commonalities among them. After that brief review, we discuss several recent developments in conjoint capabilities that have applications potential in transport research. Then we review external validity evidence and issues. Finally, we discuss some unresolved issues that need further research.

2. A BRIEF OVERVIEW OF SEVERAL CONJOINT PARADIGMS

2.1 Rank-order judgment methods

Conjoint analysis was popularised as a tool for the practical analysis of rank-order consumer judgment data by Green and Rao (1971) and Green and Wind (1973). The theory that underlies the design and analysis of rank-order judgment experiments was developed by several writers (for example, Luce and Tukey, 1964; Kruskal, 1965; Tversky, 1967), and is summarised in Krantz *et al.*, (1971). Unfortunately, the theory and practice of rank-order judgment analysis are somewhat unrelated, since the methods of analysis (a) are not based on the theory, (b) do not have a statistical error theory, and therefore (c) cannot be used to test the adequacy of the theory.

The axiomatic theory of rank-order conjoint analysis is called "Conjoint Measurement" (Krantz *et al.*, 1971). This theory requires real ranking data to satisfy a large number of ordinal conditions before one can conclude that a particular utility specification is appropriate for scaling (that is, estimating) part-worth utilities from an individual's judgment data. Most individuals are not perfectly consistent in their rankings, and therefore there is error in their data. However, Conjoint Measurement has no error theory on which to base statistical tests of part-worth parameters or competing utility specifications.

Consequently, most practical and academic researchers who analyse ranking judgments assume that individuals' rankings are generated by a strictly additive (no non-additivities or interactions) function of the unknown part-worth utility measures. Part-worth utilities are estimated by least-squares procedures (for example, MONANOVA) that optimise the fit between observed and predicted rankings, *assuming* that an additive utility specification is correct. "Badness-of-fit" statistics known as "stress" measures are used as an index of how well additive or other specifications fit the observed rankings.

Unfortunately, "stress" measures are closely related to the quantity $(1 - R^2)$; and it is well known that R^2 or any monotone transformation of R^2 is an unreliable measure of the adequacy of conjoint models (not to mention other models). Dawes and Corrigan (1974), Wainer (1976) and Anderson and Shanteau (1977) are among those who have demonstrated that (a) conjoint experiments ensure high goodness-of-fit or low badness-of-fit measures, (b) many possible specifications can produce approximately equivalent fit measures, and (c) wrong specifications can produce "better" fit measures than "right" specifications in real, fallible data.

In particular, factorial-type experiments guarantee that main effects or other simple specifications will account for most of the variance in judgment data, even when wrong. This happens because "true" utility functions are conditionally monotone in each attribute, and the joint combination rule can be well approximated by functions that predict "higher overall utility corresponds to more high part-worths" and "lower overall utility corresponds to more low part-worths". Conjoint models mimic these conditions very well.

It is more important that in rank-order analysis methods like MONANOVA (Kruskal, 1965; Green and Wind, 1973), as in general linear models, the omission

of important variables can bias the parameter estimates of attributes that are included. Most of the practical, rank-order, conjoint studies use experimental designs known as “main effects” designs. Main effects designs permit one to estimate part-worth utility parameters in an unbiased manner if, and only if, the true underlying utility specification is additive. For example, if there are interactions or non-additivities in the utility surface, estimates of part-worth utilities derived from main effects designs will be biased. Many conjoint studies of travel judgments in a variety of travel contexts (for example, Louviere *et al.*, 1973; Norman and Louviere, 1974; Lerman and Louviere, 1978; Norman, 1977; Louviere and Meyer, 1981) have demonstrated that utility specifications are not additive; therefore one should question the adequacy of additive assumptions.

More recently, Chapman and Staelin (1982), Chapman (1984) and Hensher and Louviere (1983) have discussed methods for developing choice models and choice forecasts from ranking data. These methods require the following assumptions to be satisfied by individuals’ rankings: (a) the multinomial logit choice (MNL) model is a good approximation to the unobserved choices implied by the rankings, (b) the individual is perfectly transitive in the unobserved choice sets implied by the rankings, and (c) the individual is perfectly consistent in his/her ranking behaviour in the unobserved choice sets implied by the rankings.

These assumptions cannot be tested, because real choices are not observed; instead, models are estimated from simulated choices based on an explosion of the rankings (Chapman and Staelin, 1982; Chapman, 1984), or based on deterministic expressions for the marginal frequencies of choices of each alternative — the frequencies which result from applying an individual’s ranking to all possible choice sets (Hensher and Louviere, 1983). In view of the well known violations of the assumptions required to translate rankings into choices (for example, Tversky, 1972), caution should be used in applying these methods to estimate the parameters of MNL models.

In particular, individuals can (and do) use elimination and nesting strategies which violate the IIA assumptions of MNL models. In addition, it is unrealistic to assume that individuals’ rankings will perfectly correspond to the choices that they would make if faced with subsets of the ranked alternatives varying in size and composition. Individuals’ rankings of a single set of alternatives are thus unreliable estimates of their choices in subsets of the full choice set.

Rank-order conjoint experiments are more commonly used to estimate the parameters (part-worth utilities) of additive utility models for each individual in a sample of interest. The vectors of estimated part-worth utility parameters (often, in recent studies, dummy variable O.L.S. regression coefficients — see, for example, Cattin and Wittink, 1982, 1986) are used to predict each individual’s overall utility values for certain sets of multiattribute alternatives. A computer algorithm identifies the “highest” utility alternative in each set; and one assumes the individual would choose that alternative if asked to make a choice from the choice set in question.

The use of estimated individual utility parameters to simulate the expected choices of a sample is problematic also because one must assume that (a) individuals are perfectly transitive and consistent, (b) there is no bias in the estimated parameters, (c) individuals are aware of all the alternatives and possess

perfect information about their attributes, and (d) there are no income, time or other constraints to cause individuals to choose something other than the alternative with the highest predicted utility. Moreover, it is logically inconsistent that stochastic models are used to derive estimates of part-worth utility parameters (for example, linear regression models) but deterministic rules are used to apply the parameters to predict choice.

This logical inconsistency may be avoided by using the Chapman and Staelin (1982) or Chapman (1984) approach, or the Hensher and Louviere (1983) approach, or by making assumptions about the statistical properties of the estimated utilities. For example, if a subject's estimated utilities satisfy the scale assumptions of the MNL model, and the standard error of the regression model used to estimate the utilities satisfies the variance assumptions in the error distributions used to derive the MNL (or other) choice models, one can use the predicted utility values to forecast choice *probabilities*.

For example, Louviere and Woodworth (1983) and Louviere (1987a) compare parameters of MNL models estimated from *simulated choice data*, using two different assumptions: (a) "highest estimated overall utility equals first choice", and (b) predicted utilities satisfy the scale assumptions of the MNL model. Louviere (1987a), in a study of the future migration destination choices of a random sample of Iowans aged 55–65, found little difference in MNL model parameters estimated from choice data simulated by the two assumptions.

Transport researchers should understand that these choice forecasting techniques are *ad hoc*, with the possible exception of the Chapman and Staelin (1982) and Chapman (1984) approaches. However, these latter approaches require one to make untestable assumptions about individuals' choices in subsets of the full choice set on the basis of their rankings of the full choice set; only then can one use the MNL model to estimate part-worth utilities. The preceding approaches predict the probability of choosing an alternative, given that an individual will make a choice; but in real markets individuals often choose not to choose, or to delay a choice. Furthermore, awareness of alternatives influences the formation of a choice set, and it is not easy to incorporate measures of future awareness into choice models if one has no information except on rank order. Thus one assumes that all individuals have identical choice sets, or one must develop a way to forecast what a choice set will include.

2.2 Rating scale judgment methods

Rating scale conjoint judgment data are most closely associated with Information Integration Theory (Anderson, 1981, 1982) and Social Judgment Theory (Brunswick, 1956; Slovic and Lichtenstein, 1971; Adelman *et al.*, 1974; Hammond *et al.*, 1977). Only Information Integration Theory (IIT) has a theory of errors for rating judgments; therefore only IIT has both measurement models and the means to falsify them. The use of IIT models to estimate part-worth utilities from ratings judgments in conjoint tasks is called "Functional Measurement". If the Functional Measurement axioms are satisfied, analysis of variance or multiple linear regression analysis can be used as an error theory; this makes it possible to test or diagnose alternative conjoint model specifications.

Social Judgment Theory (SJT) is based on Brunswick's (1956) Lens Model, which is a descriptive theory about the relationship between real stimuli, perceived stimuli and individuals' responses to each. The estimation of individuals' part-worth utilities by the use of the Lens Model framework is referred to as "Judgment Policy Capturing". SJT has no error theory, so one must assume that multiple linear regression models, the primary analytical tools of this paradigm, are adequate descriptors of rating scale judgments of multiattribute alternatives.

Conjoint rating scale paradigms require one to assume that ratings data satisfy assumptions on interval (cardinal) level measurement if one is to diagnose and test models and estimate part-worth utilities. There has been much argument about the satisfaction of interval scale assumptions by ratings data, particularly in the debate between Anderson (1971) and Krantz and Tversky (1971a). However, the debate has largely died down, and it now seems to be more acceptable to assume that ratings data can and do satisfy cardinal measurement properties under appropriate experimental and task conditions. In fact, several former critics have attempted to axiomatise IIT, partly as a result of the compelling empirical support for ratings scales amassed by Anderson and others since the early 1960s (see, for example, McClelland, 1980; Luce, 1981; Anderson, 1981, 1982).

Thus both IIT and SJT use relatively well known statistical techniques to diagnose and test model forms and estimate part-worth utility parameters. But, as with rank-order conjoint techniques, to forecast choices from conjoint ratings data one must assume that either (a) "highest predicted ratings equals first choice", or (b) predicted ratings values satisfy MNL or other choice model scale properties (for example, Louviere, 1987a). Thus, assumptions required to forecast choice behaviour from ranking data by computerised choice simulators also apply to ratings data.

The main advantage of ranking data is that one does not have to make cardinal measurement assumptions. Unfortunately, the statistical behaviour of ranks observed in response to multiple attributes has not been formally axiomatised; so the properties of parameter estimates are unknown, and competing specifications cannot be tested. On the other hand, if rating data satisfy cardinal measurement assumptions, powerful tests for model form and adequacy are available; but it may be difficult to satisfy cardinal measurement assumptions in applications in the field. Both problems can be avoided by using discrete choice or resource allocation responses.

2.3 Discrete data conjoint methods

Recently, Louviere and Woodworth (1983), Louviere (1984) and Louviere and Hensher (1982) have formalised the necessary and sufficient conditions which experimental designs must meet to satisfy the statistical requirements of MNL choice models. Louviere and Woodworth (1983) note that their approach to the design and analysis of choice experiments can also be used with other choice models, and Louviere (1986) describes how to develop experimental designs that allow one to generalise MNL models to account for violations of IIA.

Thus, if one can assume that experimental conjoint choice data satisfy the

error assumptions of the MNL model, the unknown utilities can be estimated in a statistically efficient manner from various experimental designs discussed in Louviere and Woodworth (1983, 1986) and Louviere (1984a, b; 1986). Moreover, as with IIT, the MNL model serves as an error theory to diagnose or test various specifications for the utility function, if the choice experiment is designed in such a way as to accommodate the required tests. Unlike rank-order or ratings, this method does not require any assumptions to be made about order or cardinality of measurement; one need only assume that the response data are discrete (nominal or classificatory in level of measurement).

A more important advantage of discrete choice conjoint methods is that one can estimate choice models directly from choice data, and thus avoid the potentially unrealistic *ad hoc* assumptions about choice behaviour that would be implied by the rating or ranking of a single choice set. Instead, one can observe and model choices directly as the outcome of a conjoint choice experiment.

Unfortunately, however, in practical applications one rarely has enough choice data to estimate choice models at the individual level, though one often can obtain models for segments. Unlike observational, econometric choice data, multiple choice observations can usually be obtained from each individual. This makes it possible to cluster individuals on the basis of similarities in their *choice behaviour* rather than of similarities in individual characteristics that may be unrelated to choice behaviour. More detail about discrete data experiments is given in the next section.

We now turn to some recent developments in the three types of conjoint methods that have potential application in travel choice analysis.

3. SOME RECENT DEVELOPMENTS IN CONJOINT ANALYSIS

The previous section reviewed traditional conjoint paradigms, variants of which have been applied in many transport studies. Traditional conjoint methods played an important role in developing the capabilities of stated preference modelling in transport research, but there are limitations and problems which make the research community reluctant to accept and apply them. For example, the objective of most stated preference studies is to forecast the choice behaviour of travellers. Traditional conjoint methods require us to make untested, and in many cases untestable, assumptions about the relationship of stated preference data and choices. We need tasks that more closely mimic real travel choice environments, that are easier for respondents to complete, and that have more external validity. The next section reviews a number of recent developments which expand the design and analytic and task capabilities of more traditional methods.

3.1 Discrete or allocation response conjoint methods

Discrete choice or resource allocation responses have a number of important advantages. In particular, one can design choice or allocation experiments to mimic real choice environments closely. This is important because individuals in

real environments probably do not rank or rate travel alternatives; they choose one of them, or they choose not to choose any alternative. Moreover, discrete response tasks impose no order or metric assumptions on response data. Finally, choice experiments discussed by Louviere and Hensher (1982) and Louviere and Woodworth (1983) allow travel choice researchers to estimate choice models that are consistent with transport planning and forecasting practice.

Choice experiments can be used in place of rank-order and rating tasks: instead of asking subjects to rate or rank a single set of alternatives, one can show subjects different sets of alternatives and ask them to choose among them or allocate resources among them in each set (see, for example, Louviere and Woodworth, 1983). Choice experiments therefore allow one to study how choices vary as a function of both size and composition of choice sets. Thus, unlike traditional conjoint methods, they enable us to study and model many possible choice processes, and there is no need to make *ad hoc* and potentially incorrect assumptions in order to create computerised choice simulators.

On the other hand, choice experiments are more difficult to design, because they require two separate designs to be combined¹: one to create the choice alternatives (conjoint treatments and/or existing alternatives) and a second to place choice alternatives (treatments plus possibly other non-designed choice alternatives) into choice sets. Both designs must satisfy certain statistical properties to enable one to estimate parameters and conduct statistical tests efficiently (see Louviere and Woodworth, 1983). Thus, with traditional conjoint rating or ranking methods one trades simplicity of design for unreality of task, strong response measurement level and choice simulation assumptions. With discrete choice experiments one trades simplicity of response, better understanding of choice process and minimal measurement assumptions for increased complexity of design.

Not only discrete choice models but other discrete response models considerably expand the analytic capabilities of conjoint techniques. For example, instead of making cardinal measurement assumptions about rating data, one could use double truncated Tobit models to analyse ratings data. Similarly, one might apply ordered logit or probit models to both ranking and rating data. Those models have not yet been applied to analyse conjoint data; we suspect that this is because discrete analytic methods have not been widely available till recently. We mention the potential of these other discrete techniques only in passing, because experience of them is limited and more research is needed to show how far they can be applied. On the other hand, discrete choice analysis is familiar to many transport researchers, and can be applied to conjoint choice experiments, as we now explain.

3.1.1 Discrete choice or allocation conjoint experiments

Louviere (1981), Louviere and Hensher (1982) and Louviere and Woodworth (1983) discuss the theory and logic behind discrete choice experiments. Briefly, if one assumes that the choices of a (reasonably) homogeneous behavioural

¹ An example is provided in the paper by Hensher *et al.* in this issue.

segment of individuals can be closely approximated by some form of an MNL model, it is possible to design statistically efficient experiments to study choice behaviour. MNL models can also be generalised to handle violations of the IIA axiom (Luce, 1959), as suggested by McFadden (1981, 1987). Indeed, McFadden (1987, p. 293) suggests that "In a conjoint study it is possible to construct powerful direct tests of IIA by observing choices from a set C and subsets A , as described by Louviere and Woodworth (1983)." Before explaining how to handle violations of IIA, we first consider experiments for choice problems that satisfy IIA.

MNL models that satisfy IIA can be fully specified by estimating the marginal choice probabilities for each alternative. Thus a necessary and sufficient condition for efficiently estimating MNL models is to be able to estimate these marginals independently. Louviere and Woodworth (1983) show that many factorial and fractional factorial designs commonly applied to conjoint problems can be used to construct sets of choice sets that guarantee independence-of-marginal-probabilities by design. Louviere and Woodworth prove that a 2^N (N = the total number of choice alternatives) factorial design can be used to put N choice alternatives into choice sets such that the parameters of MNL models, estimated from choices made in response to the design, will be near-optimal in statistical efficiency. Other types of factorial designs also can be used to construct sets of choice sets that satisfy the independence-of-marginal-probabilities condition, but the efficiency properties of these designs have not been studied. Louviere and Woodworth (1983) conjecture that other classes of factorial designs also have excellent efficiency, but research is needed to verify this.

To construct conjoint choice experiments using 2^N designs, one first creates a set of multiattribute conjoint alternatives by using a factorial or other design to produce combinations of levels of attributes (that is, alternatives). One next treats each of the designed multiattribute alternatives as a two-level factor (present or absent in a choice set). A 2^N fractional factorial design is used to generate combinations of "present/absent". Orthogonal 2^N designs have the property that the presence/absence of each alternative is independent of the presence/absence of all other alternatives; thus, constructing choice sets by using 2^N designs to place the N designed alternatives into choice sets according to which of the N are "present" satisfies the independence-of-marginal-probabilities condition needed for the efficient estimation and testing of parameters of MNL models.

Hence, sets of choice sets produced by 2^N designs satisfy the necessary and sufficient condition for estimating MNL choice models if such models are good approximations to choice data. Choice experiments constructed from 2^N designs are shown to samples of subjects, who are asked to choose one and only one alternative from each set. Louviere and Woodworth (1983) recommend the use of a constant "base" alternative in each choice set, to set the origin of the utility scale. Technically, a constant "base" alternative is not needed, but it is useful because it preserves the design orthogonality of the attribute vectors of conjoint alternatives, and it often provides a meaningful and important origin, such as "not choosing any alternative", "choosing to delay travel plans", or the like.

Choice data derived from such experiments are aggregated over participating

subjects for analysis. Thus, there are multiple observations of choice for each choice set, and a variety of methods can be used to estimate the parameters of MNL models from these data, as explained by Louviere and Woodworth (1983). However, one must be cautious about standard errors estimated from the aggregated data because the repeated observations are not independent, though the parameter estimates are consistent. That is, one must be concerned about variations in model parameters across individuals. MNL models have only asymptotic properties; therefore one cannot estimate separate models for individuals, as is done often with conjoint rating or ranking data. If one could estimate individual models, one could efficiently estimate the standard errors and test the model parameters against their respective variances across individuals.

Because, for a number of choice sets the choices made by each subject can be observed, subjects' choice response vectors from these sets can be compared with one another, and subjects who behave in a similar manner can be grouped together. Separate MNL models can be developed for each of these behaviourally "homogeneous" segments. Travel choice surveys normally record only a single choice made by a subject in a single choice set; in contrast, designed choice experiments allow one to observe choices made by individuals in multiple choice sets. If individuals respond to the same choice sets, their choices can be compared statistically, and individuals with similar choice patterns can be identified by the use of discrete cluster analysis techniques (for example, Wilkinson, 1986).

Louviere and Woodworth (1983) and Louviere (1983, 1984b, 1986) discuss ways of constructing choice experiments from other types of factorial designs. For example, if one has M alternatives, each with A attributes having L levels, one can satisfy the independence-of-marginal-probabilities condition by treating each attribute of each alternative as a separate factor in an L^{MA} factorial design. This design generates combinations of levels of the attributes of the M alternatives so that the attributes of each alternative are independent of one another. Louviere (1984b, 1986) discusses less demanding ways to generate choice sets that satisfy MNL conditions for model estimation.

Certain 2^N or variants of L^{MA} designs can be used to generalise MNL models to account for violations of IIA. The requirement for generalising MNL models (or testing violations of IIA) is that certain cross-alternative effects must be estimable. For example, including bus fare in the utility function of train is a cross-effect of bus on train, or an interaction among the effects of the different alternatives. Louviere (1986) illustrates how one can estimate and test these cross-effects.

Space precludes further discussion of methods of designing discrete choice experiments; however, Louviere and Woodworth (1983) and Louviere (1988, forthcoming) discuss a wide range of versatile and statistically efficient ways to construct choice experiments. Because the responses to these tasks are discrete choices, there is no need to simulate choices, or to make assumptions about choice processes. Instead, one can test a variety of competing choice models directly by designing choice experiments to permit efficient estimation of diagnostic test statistics.

Moreover, discrete choice tasks are easy for subjects to do, and more closely mimic choice situations faced in real life than traditional conjoint tasks. Indeed, the object of interest in many studies of travel behaviour is variation in choice in

response to variation in the attributes of travel alternatives, not variation in rating or ranking behaviour. Evidence advanced by Levin, Louviere, Schepanski and Norman (1983), Louviere and Hensher (1982) and Louviere and Woodworth (1983), and reviewed later in this paper, supports the external validity of discrete choice experiments. Indeed, external validity has been found to be high, and several studies have demonstrated close correspondence between predicted and observed choice proportions in real markets.

3.1.2 Correlated Conjoint

Many decisions made by real decision makers involve comparisons of sets of correlated attributes, because vectors of attributes of real offerings are often correlated in real markets as a result of market forces, technological capabilities, etc. For example, prices of alternatives are often strongly positively correlated with more features, more quality designed into features, longer trip lengths, and so forth. Existing conjoint methods cannot easily or satisfactorily deal with problems of correlated attributes: these problems make it necessary to design comparison tasks in order to obtain statistically efficient estimates. Comparison tasks are similar to and may be described as choice tasks, so we discuss them in this section.

Louviere and Woodworth (1986, 1987) propose a method for dealing with correlated attributes on the basis of the MNL choice model. To understand the approach, we rewrite the MNL model as follows:

$$\ln[p(a/A)] = \ln[p(\text{base}/A)] + [U(a) - U(\text{base})] \quad (1)$$

where $p(a/A)$ and $p(\text{base}/A)$ are the probabilities of choosing alternative a and the base alternative, respectively. Given that a is a member of some choice set A ; $U(\text{base})$, $U(a)$ are, respectively, the unknown utilities of a base alternative and alternative a , both members of choice set A ; and \ln is a symbol for the natural logarithmic operator.

Equation (1) indicates that the probability of choosing a particular alternative can be decomposed into two components: (i) a component associated with the base alternative, which is a constant for each choice set, and (ii) a difference-in-utilities component involving each non-base and base alternative in a particular choice set. Equation (1) suggests how one can develop Correlated Conjoint experiments. Because the choice of a base alternative is arbitrary, it is under the control of the experimenter. Thus, one can select the levels of the attributes of the base alternative arbitrarily so that the vectors of attribute levels possess whatever correlational structure one wishes.

To do this one constructs an orthogonal design to vary *differences in the levels of quantitative variables or contrasts in the levels of qualitative variables*. The vectors of orthogonal differences/contrasts are used to create combinations of absolute attribute levels of non-base alternatives from the arbitrary levels of the attribute vectors of the base alternative. That is, if bus fare is an attribute, one first arbitrarily assigns numerical values of fare (or travel cost) to the base mode in each choice set (for example, one can assign the levels of fare so that the fare vector is positively correlated with the travel time vector). Next one uses the

elements of the differences-in-bus-fare vector to calculate the levels of absolute fare (or travel cost) to be assigned to other alternatives in each choice set. The number of combinations of difference levels determines the number of choice sets.

The levels of the attribute difference vectors, however, are not arbitrary; rather they are column vectors in an orthogonal design, the elements of which are discrete levels of "differences in the attributes relative to the base" (for example, difference in fare or travel time). Thus, in a paired choice between bus and train, if train is the base, there may be one designed column difference vector for bus fare, a second designed column difference vector for bus travel time, and additional designed column difference vectors for other attributes. These vectors represent columns of an orthogonal array of difference and/or contrast column vectors. Additional alternatives may be included by designing additional difference vectors. For example, if the alternatives were auto, bus and train, one might have two designed difference column vectors for travel cost and two for travel time to represent the differences between the attribute levels of train (the base) and the attribute levels of auto and bus.

The difference vectors define the absolute attribute levels to be assigned to each treatment combination that describes a second, third, fourth, etc., alternative. For example, if the modes were bus, train and private car and the base was taxi, six vectors would be required to vary differences in levels of cost and travel time (bus, train and private car each require two difference vectors for time and cost). As indicated by equation (1), this is accomplished by simply subtracting (adding) the level of the bus cost difference vector in a particular choice set to the arbitrarily chosen level of cost of taxi in the same set to obtain the level for bus. Similarly, one adds the levels of train cost difference to the levels of taxi cost to obtain the absolute levels of cost for train. If there are additional quantitative attributes, additional orthogonal difference vectors are needed to construct the absolute attribute levels for each choice alternative.

The treatment of qualitative attributes is less obvious, because they have to be expressed as contrasts. A contrast is a difference in response between two or more different levels of an experimental variable. From a logical and statistical standpoint, an attribute cannot influence choice if all alternatives have the same level of that attribute (for example, footrests attached under bus sets = "yes"). Thus, qualitative contrasts of interest in choice experiments are contrasts between unequal levels. For example, if two choice alternatives (say, bus and train) are each described by a two-level qualitative attribute with levels "yes" or "no", there are four possible combinations of outcomes for bus and train on that particular attribute: (1) bus = "no", train = "no"; (2) bus = "no", train = "yes"; (3) bus = "yes", train = "no"; and (4) bus = "yes", train = "yes".

Outcomes 1 and 4 provide no statistical information for modelling choice behaviour; so the outcomes of interest are 2 and 3, which can be expressed as a two-level contrast, and this contrast can be used to determine the levels of both a base and a second alternative. Hence, defining qualitative contrasts in this way exactly fixes the levels of both a base and a second alternative, but the levels of qualitative attributes of the base are designed and not arbitrarily determined by the experimenter. One can arbitrarily define the qualitative levels if the con-

trast is defined as “equal” or “not equal”. This contrast is applied to a vector of arbitrarily assigned qualitative levels for a particular attribute of the base alternative to create the levels of a second alternative. One should define the same number of “yes” “no” contrasts as “no” “yes” contrasts within the level “not equal”. If one assigns “not equal” levels deliberately, one can create two contrast dummies as follows: (1 if “yes” “no”; -1 if “no” “yes”; 0 otherwise) and (1 if “no” “yes”; -1 if “yes” “no”; 0 otherwise).

Louviere and Woodworth (1986) indicate that contrasts of “not equal” levels are restricted to paired choice experiments. The more general contrast described in the preceding paragraph can be used to create multiple choice experiments. However, as one adds alternatives, one loses the freedom to assign levels to the base alternative arbitrarily, because one must satisfy the contrast conditions in a statistically efficient manner. Thus, Correlated Conjoint designs cannot be developed by “off the shelf” methods; some trial and error is needed to produce a design that is efficient and also has the correlational structure of interest. Responses to such designs can be analysed by generalised least-squares (see, for example, Louviere and Woodworth (1983) and Louviere and Hensher (1982)).

Correlated conjoint designs can be used with ratings tasks if restricted to paired comparison tasks. However, discrete choice and resource allocation response modes make more sense for such tasks; and, as indicated by equation (1), binary or multinomial logit models are theoretically consistent with these types of responses in these kinds of tasks. There are many potential applications of correlated tasks to problems of transport research: attributes of modes may be correlated for particular journeys; evaluations of freight shipping options may be correlated among corporate decision makers; some attributes of shopping destinations (for example, numbers of shops and parking spaces) may be correlated. Thus correlated conjoint designs can be applied to a variety of problems in transport research.

3.2 Rank order conjoint methods

We have already discussed the Chapman and Staelin (1982) and Chapman (1984) methods for estimating MNL models by exploding the information obtained in rankings of conjoint experiments. Louviere (1981) and Hensher and Louviere (1983) use the rankings to simulate the choices that individuals should make in all possible choice sets defined by a 2^N factorial expansion of the treatments in a conjoint experiment. This power set arises because each conjoint treatment is either “in” or “not in” a particular choice set, and a factorial of these two states yields all possible choice sets, as previously explained. The marginal choice frequencies of the i th treatment for an individual can be determined exactly by the expression:

$$F(i) = 2^{(N-1)} \{2^{-[r(i)-1]}\}, \quad (2)$$

where $F(i)$ is the total number of times that the i th treatment is chosen in all choice sets, N is the number of treatments ($i = 1, \dots, N$), and $r(i)$ is the ranking of the i th treatment (rank = 1 is the “highest” or “best”).

The $F(i)$ represent the total marginal choice frequencies for an individual that are expected for each treatment (or alternative) if an individual's choices in all possible choice sets are exactly consistent with his/her rankings. Louviere and Woodworth (1983) demonstrate that these frequencies contain all the information needed to estimate the parameters of an MNL model. Thus, if one can assume that the choice outcomes are perfectly (deterministically) determined by the ranks, one can estimate an MNL model for an individual or a segment of individuals. Segments can be identified by clustering individuals with similar rank-orderings of a common set of multiattribute alternatives.

If individual- or segment-level choice models can be obtained, the significance of a particular part-worth utility parameter can be determined by estimating its sample mean (or segment mean) and standard error. The ratio of the mean to its standard error is distributed as a t -statistic, which can be tested in the usual way. This test measures whether the parameter variability in the sample is statistically acceptable (that is, whether the confidence interval is small enough to be acceptable). If all individuals rank the same treatments, the resulting $F(i)$ are repeated measures, and the usual MNL statistical tests of the parameters do not apply. In any case, if one accepts the assumptions, one can use this approach to develop individual-level MNL models.

A second approach to developing MNL models is to calculate the number of first-rank responses given to each treatment for a sample or subsamples. The frequencies of first-ranks estimate the choice outcomes that would be expected if individuals were given a single choice set consisting of all the treatments and were asked to choose one. These frequencies can be used as empirical estimates of the probabilities, and one can estimate the parameters of an MNL model, as explained in Louviere and Woodworth (1983). The parameter estimates will be consistent but inefficient, because the choices are repeated measures observed on the same choice set, and the estimated standard errors will therefore be incorrect.

If one's interest is in forecasting, consistent estimates are probably sufficient. A crude approximation to the correct standard errors can be obtained if one clusters the individuals and estimates separate MNL models for each cluster. The mean parameters and their standard errors can be estimated (weighted by the sample sizes in each cluster) from the separate cluster results, and tested as explained above in the case of individual parameters.

Thus, if one is to use any methods for translating conjoint ranking responses into choices one must assume that individuals will (a) be perfectly transitive and consistent, (b) possess perfect information about all alternatives, (c) lack constraints and (d) choose one of the alternatives. Further, it is difficult to incorporate existing choice options and/or non-choice alternatives in conjoint ranking tasks, and external validity is therefore an issue. For example, one might ask individuals to rank their present travel mode, or several existing travel alternatives, together with a set of conjoint treatments describing alternative travel modes. Unfortunately, if one obtains such rankings, one cannot use common conjoint estimation techniques such as MONANOVA to analyse the data; one must resort instead to rank-order explosion procedures or to estimating the expected frequencies of choices in the power set of choice sets, as explained previously.

MONANOVA cannot be used to estimate part-worth utilities from rankings that include non-designed alternatives, because MONANOVA results depend upon the orthogonality of the design and the precision of measurement and identification of the attribute levels of each alternative. Adding non-designed alternatives to a MONANOVA analysis can destroy orthogonality of design, cause the attribute levels of some alternatives to vary from individual to individual, and call into question the reliability of the attribute measures of existing alternatives. Reliability of measurement is an issue because MONANOVA assumes the levels of the attributes to be fixed and measured without error.

Even if one accepts the assumptions of the techniques discussed above, one needs to be able to map existing or future alternatives into choices. One way to do this is to obtain individuals' estimates of the levels of the attributes possessed by existing alternatives in a manner consistent with the levels used in the conjoint experiment. For example, individuals can be asked to associate particular attribute levels with existing alternatives. This is known as "product positioning"; each individual indicates which level of each attribute best describes where he/she thinks particular real alternatives are "positioned" on each attribute. Our work has relied on a multiple choice test-like task (unlike most product positioning tasks common to marketing research studies) in which individuals identify one level from a set of levels that they feel best describes a particular attribute of a particular alternative. Thus these "beliefs" measure attribute levels of real alternatives in a way that corresponds to the attribute levels varied in conjoint tasks, and can be used as input to MNL models to predict how the choices of a population of interest are likely to change as one changes the attribute levels.

In order to forecast future alternatives using rank-order conjoint tasks, one must assume that the conjoint attributes contain all the information that travellers will use to choose among these alternatives when they are available. Furthermore, one must also assume that individuals are aware of all alternatives and their attributes, and include the alternatives in their choice sets. Thus, future choices of freight shippers, airline schedulers, transport planners, travel agents and the like can probably be forecast with rank-order conjoint techniques if the assumptions can be satisfied. However, the future mode, route and destination choices of consumers may be less amenable to prediction if experimental attributes do not provide a good description of alternatives, or if one cannot assume that all individuals know about the attributes, are aware of all the alternatives, and include the alternatives in their choice sets. On the other hand, choice of residence, business location and other decisions in which individuals are motivated to study carefully, acquire information, and analyse alternatives should be more amenable to prediction.

3.3 Rating scale conjoint methods

Rating scales in conjoint tasks have a number of advantages not possessed by ranking methods:

- (1) Individuals can rate conjoint treatments, existing travel alternatives, and/or a non-choice alternative on the same scale in the same task.

(2) If ratings responses satisfy the scale assumptions of the MNL model, one can predict the choice probabilities of existing alternatives competing with new alternatives described by different combinations of attribute levels, as well as the probability of no choice.

(3) One can use the ratings data as estimates of “high” and “low” utility values, avoiding the cardinal measurement assumptions needed to analyse (preference) ratings data with metric statistical procedures like analysis of variance or multiple linear regression.

Information about magnitudes of differences in utilities contained in rating responses makes them potentially more useful than ranking measures. An individual's ranking of a set of multiattribute options contains no information about the magnitude of “how likely” (“how much preferred”, “how desirable”, etc.) alternatives are to be chosen (including the choice of “none”). Instead, ranking responses provide no information about choice likelihoods or preferences except their order. For example, individuals may be unlikely to choose any alternative, or may have low levels of preference for all alternatives. In such cases, ranking tasks cannot distinguish among different levels of response to a set of alternatives; they only distinguish among differences in responses relative to the levels of the attributes of the alternatives in the set.

If one assumes that rating data are approximately cardinal in measurement level, one can identify subjects who dislike (or like) all or most alternatives. For example, Meyer, Levin and Louviere (1978) show that individual differences in the magnitude of conjoint ratings reflect important aspects of travel choice behaviour. Hence, relative magnitudes of responses to a collection of alternatives may provide important information about individual differences or biases toward travel alternatives.

Thus, one has to trade off the advantages of less restrictive ordinal assumptions, which sacrifice relative information on magnitude, for more restrictive cardinality assumptions that produce information on magnitude. With these caveats in mind, there are several new developments in conjoint ratings methods that can be applied to the analysis of transport-related decision making.

3.3.1 Hierarchical Conjoint

Louviere (1984a) and Louviere and Gaeth (1987) describe a method for the design and analysis of experiments in hierarchical information processing. Conceptually, these experiments view individuals as having learned to combine many separate attributes into a few composite concepts which are used as the basis for decision making. Thus, for many types of routine travel choices, such as choice of a supermarket or shopping centre as destination, it seems reasonable to assume that individuals have learned to combine a number of cues in order to form opinions or beliefs about various destinations, such as “how convenient”, “how expensive”, “how wide a selection”, “how much service”, “how much quality”, etc. Hierarchical Conjoint provides both theory and methods for modelling how composite impressions are formed and integrated (processed) to evaluate alternatives such as shopping destinations.

To design a Hierarchical Conjoint study, one must identify both composite concepts and the attributes that define them. One develops separate conjoint tasks for each composite concept in which attributes that define particular concepts are varied. (Some attributes may be part of the set of defining attributes of more than one concept.) In each concept task, individuals are asked to rate “how much” of each concept is produced by a particular combination of attribute levels. For example, in the case of shopping destinations, one might ask subjects to judge “how convenient”, “how expensive”, etc., are alternatives described by different combinations of levels of defining attributes.

One must also design an “integrative” conjoint task to combine the separate concept evaluations. In the integrative task the experimental factors are the concepts, and the levels of the concepts are individuals’ “pseudo ratings” of “how convenient”, “how expensive”, etc. are the combinations of attribute levels in the separate conjoint concept tasks. We refer to these levels as “pseudo ratings” because they are not the actual ratings given by individuals, but are combinations of possible ratings that individuals might have given to the different concepts.

For example, individuals might judge the “convenience” of shopping destinations described by different combinations of levels of number of parking spaces, width of aisles, cheque-cashing policy, opening hours, travel time from residence or work, etc. If subjects use a 0 to 10 numerical rating scale to judge supermarket “convenience” in a convenience concept task, one might use the pseudo ratings 1, 5 and 9 as levels of “convenience” in the integrative design, to study how individuals combine ratings of convenience with ratings of other concepts. Thus, in the integrative conjoint task, subjects judge alternatives that are combinations of pseudo ratings of concepts. The relationship between the concept tasks and the integrative task can be expressed as follows:

$$R_1 = f_1(A_{1i}) \quad (3a)$$

$$R_2 = f_2(A_{2j}) \quad (3b)$$

$$\vdots$$

$$R_M = f_M(A_{Mm}), \quad (3M)$$

$$V_{1, 2, \dots, M} = g(R_1, R_2, \dots, R_M), \quad (4)$$

where

R_1, R_2, \dots, R_M are the ratings made in response to each of the M subtasks involving the 1st, 2nd, \dots , M th concepts.

$V_{1, 2, \dots, M}$ is the overall evaluation of a combination of pseudo ratings of the M concepts measured as a rating on some scale, a choice between two concept combinations, etc.

$A_{1i}, A_{2j}, \dots, A_{Mm}$ are separate sets of attributes or explanatory variables that define the M concepts.

g, f_M are mappings defined respectively on the composite concept task and on the M separate concept tasks.

We assume that all of the M separate mappings in equations (3a) to (4) have additive disturbances, the expectations of which are zero. Therefore, we can concatenate the separate functions as follows:

$$V_{1, 2, \dots, M} = g[f_1(A_{1i}), f_2(A_{2j}), \dots, f_M(A_{Mm})], \quad (5)$$

where all terms are as previously defined.

Equation (5) indicates that one can integrate the effects of the separate attributes of each concept into a single closed-form model of the overall utility of the alternatives of interest. Hierarchical Conjoint allows one to incorporate a large number of variables in conjoint tasks by decomposing a task into separate tasks for concepts and an integrative task. Many travel choices of interest may involve a large number of attributes (for example, choice of destination for a vacation); Hierarchical Conjoint offers a logical way to study and model such decisions.

3.3.2 Organisational Conjoint

Organisational Conjoint, similar in concept to Hierarchical Conjoint, can be applied to decisions made by groups, organisations and decision-making units like families (Louviere and Larsen, 1987). The basic idea is to study decision makers separately and then study the influence that each has on the organisational decision. One must first identify the principal decision attributes that influence the decisions of the organisation. Then one develops a conjoint task to be administered to each individual taking part in the decision process. As in Hierarchical Conjoint, individuals evaluate the description of each attribute in this first conjoint task on a numerical rating scale.

A second conjoint task is developed in which each decision maker is treated as a separate experimental factor. For example, if the separate decision makers are a husband and wife, there would be two factors representing, respectively, the husband's and the wife's ratings of a set of alternatives. The levels of these factors are their "pseudo ratings" of the attribute combinations in the first conjoint task. Thus each treatment combination in the second task consists of a description of the rating that each individual decision maker might have given to an alternative in the first task. The group of organisational decision makers next judges the descriptions of the pseudo ratings of the separate group members, and makes an overall group decision about each description.

As in Hierarchical Conjoint, we assume that the ratings responses of individual decision makers can be approximated by a linear-in-the-parameters-and-variables function, for which the expected value of the error component is zero. Such a study supplies two different pieces of information (which can be concatenated as in equations 3a to 5): (1) the utility function of the individual decision makers, and (2) the effect of each individual on the group decisions.

An alternative way of studying organisational decision making would be to construct the group decision-making task so that each decision maker's *rating of each level of each decision attribute* is described. For example, if the decision

makers are a husband and wife, and if four attributes are varied in the first conjoint task, eight variables must be varied in the second, or “family”, decision task. The levels of the variables in the second task are “pseudo ratings” for each decision maker for each attribute. Thus, a “family” judges descriptions (that is, alternatives) of both spouses’ ratings of each attribute. The “family” responds to each description of the husband/wife attribute ratings; and one analyses these responses as a function of the levels of the separate pseudo ratings of each attribute for both spouses.

In this approach, the separate ratings of each decision maker of each attribute are predicted from their responses to the first conjoint task. One way to do this is to estimate the expected rating of each level of each attribute from their respective marginal means. The marginal mean rating of the level of a particular attribute is the best least-squares estimate of the effect of that level in an analysis of variance model. This estimate corresponds to the dummy variable effect for that level in the multiple linear regression model; so it is the theoretical parameter of interest in this task. For example, if a decision process is additive, one can use an individual’s marginal means to predict his/her expected ratings of combinations of levels of four attributes as follows:

$$R_{ijkl} = R_{i \dots} + R_{.j \dots} + R_{..k \dots} + R_{...l} - 3R_{\dots}, \quad (6)$$

where

R_{ijkl} is the individual’s expected overall rating of a combination, respectively, of the i th, j th, k th and l th levels of the four decision attributes.

$R_{i \dots}$, $R_{.j \dots}$, $R_{..k \dots}$ and $R_{...l}$ are the marginal means of the four decision attributes.

R_{\dots} is the grand or overall mean.

Marginal means are, of course, functions of the levels of the attributes to which they correspond. Hence, if one uses regression equations to describe individuals’ ratings of combinations of attribute levels (R_{ijkl}), one can also predict individuals’ marginal means from regression equations. Marginal means correspond to levels of pseudo ratings varied in the group decision-making task. So, if one assumes that the marginal means are functions of their corresponding attribute levels and that these functions have error expectations of zero, one can combine the individual and overall group equations to deduce which attributes were important to which individuals, and which individuals played which roles in the group decision.

This approach to studying organisational decision making necessitates the development of larger and more complicated experimental designs than the first approach. Furthermore, size and complexity increase with the number of decision makers. On the other hand, the second approach provides much more information about the effects of each individual’s attribute judgments on the group’s decisions.

3.3.3 Product Positioning Conjoint

Product Positioning Conjoint involves simultaneous estimation of both product positions and utilities from a conjoint task (Louviere and Johnson, 1987). “Product

positions" measure how alternatives stack up on each attribute relative to other alternatives in individuals' minds. Product Positioning Conjoint involves no new technology, but only a different approach to defining levels of attributes. For example, to study consumer preferences for travel modes for the journey to work, one identifies attributes that influence this choice, and makes the levels of these attributes the *names* of the travel modes of interest. Thus, if the attributes are travel time, travel cost, and walking distances from residence to mode(s) and from final mode to work, one might define the levels of the attributes as: (1) "like private automobile, van or pickup truck", (2) "like commuter train", (3) "like tram" and (4) "like city bus". One might then construct a 4⁴ factorial (four attributes) or some fraction thereof to produce attribute combinations that consist of mixtures of the four mode names.

Thus, in Product Positioning Conjoint, treatments consist of combinations of names of modes or other travel alternatives. Because subjects have previous experience with the various modes on the decision attributes, and have opinions or impressions of them, the responses reveal how each mode is positioned subjectively on each attribute. If new modes can be described as a combination of performance features of current modes, one can apply the results to make predictions about likely responses to the new modes. Product Positioning Conjoint can also be applied to choice of supermarket or shopping centre as destination, to freight shippers' choices among modes (or among competitors within a mode), and to many other uses.

The recent developments in conjoint described in this paper make it possible to construct more realistic tasks for subjects, and thus to simulate reality more closely. However, simulating reality in a task and being able to predict choices in reality are not the same. Thus the external validity of conjoint models is important; we discuss this issue in the next section.

4. THE EXTERNAL VALIDITY OF CONJOINT MODELS

Academic research has largely ignored the issue of external validity in conjoint studies. External validity has seldom been tested on real choice objects (see, for example, Wright and Kriewal, 1980; Davidson, 1973; Parker and Srinivasan, 1976; and Wittink and Montgomery, 1979). Unfortunately, in most academic studies, conjoint model predictions have been compared with responses to hold-out treatments or their equivalents, or measures of choice intentions, and not actual choices (for example, Green, Rao and DeSarbo, 1978; Jain *et al.*, 1979). Comparing model predictions with hold-out data measures test-retest reliability or prediction shrinkage (the loss of goodness-of-fit observed when a model calibrated on one subsample is used to predict observations in a second); it does not measure external validity.

The majority of the writers who have examined the ability of conjoint methods to predict the real choices of real people in real markets have conducted their studies in transport or related contexts. Much of the work on external validity up to 1982 is summarised in Levin *et al.* (1983), so we give only a brief summary of

this report. Levin *et al.* (1983) provide a table summarising more than a dozen studies; they conclude that there is considerable evidence of external validity.

Some studies of particular interest include the following: Louviere and Piccolo (1977) and Lerman and Louviere (1978) dealt with choice of residential location; Louviere and Piccolo (1977) studied choice of out-of-town shopping destination; Meyer, Levin and Louviere (1978) examined transport mode choice; Louviere and Meyer (1981) focused on choices of supermarket destinations (see also Louviere and Hensher, 1982); Louviere and Kocur (1983) studied transport mode choice; and Timmermans (1982) examined choice of shopping centre destination. Many of these studies found that the predictions of aggregate conjoint models developed from studies of individuals correlated well with the observed behaviour of aggregates of real people other than those who were studied.

There have also been studies of the external validity of conjoint choice experiments. Louviere and Woodworth (1983) report high correlations between predicted and observed choices for individuals asked to make choices among brands of pet foods, or asked to choose modes of transport between mainland Australia and Tasmania. Louviere (1986) reports a high correlation between the observed market shares of agricultural chemical products and the predictions of a choice model based on a random sample of farmers. Louviere and Ghosh (1986) report high correlations and relative parameter invariance between three different measures of shopping mall patronage (number of reported visits in past month; amount of money spent in past month; last mall visited) and the predictions of a choice model from a choice experiment.

Additional evidence of external validity is supplied by several other studies. Louviere (1974) studied the recommendations on fishing destination of public hatchery personnel in Iowa and found the recommendation predictions of a conjoint model to be strongly correlated with recommendations of actual destinations for trout fishing. Louviere (1977) reported a high correlation, a slope of approximately one and an intercept of approximately zero in a study of students' evaluations of courses at the University of Wyoming. The students in the sample used to estimate conjoint models were not members of the classes used to validate the models. Lieber (1976) studied the job choices of graduating seniors at the University of Iowa. Later, after identifying students offered more than one job, Lieber (1976) was able to predict 100 per cent of the job choices of students who actually had a choice. Kocur, Hyman and Aunet (1982) successfully calibrated conjoint model predictions to the mode, route and destination choices of a very large sample of Wisconsin residents. The calibrated conjoint/choice models were used to develop a state-wide system for forecasting travel choice for the Wisconsin Department of Transportation.

Louviere and Woodworth (1985) report that, on a model developed from an experiment involving choices among state and local parks, a graph of the predicted and observed park choices of the sample was consistent with a slope of unity and an intercept of zero. Similarly, Louviere *et al.* (1981) found that predictions of reported mode choices using conjoint parameters in the arguments of MNL models were not statistically inferior to predictions based on MNL parameters estimated from the data themselves. Finally, Koeppel and Neveu (1979) used rank-order conjoint analysis methods to predict changes in New York DOT

employees' work schedules from a fixed 8AM – 4:10PM workday to five alternative work schedules. Koepfel and Neveu (1979, p. 11) report that "The observable behaviour supports the use of trade-off analysis as a valid tool to predict the reaction of persons to major changes in their environment".

We therefore suggest that there is considerable evidence to support the conclusion that appropriately designed, implemented and analysed conjoint studies can predict the real behaviour of real individuals in real markets. The next section discusses some limitations of these conclusions and makes suggestions for further research.

5. DISCUSSION AND CONCLUSIONS

This paper has briefly reviewed differences in design, implementation and analysis among conjoint paradigms. It has also discussed some recent developments in various conjoint paradigms that have potential application to problems in transport research. In addition a brief overview has been given of external validity evidence for conjoint models. Conjoint paradigms differ primarily in their assumptions regarding the level of measurement of the response, and different assumptions lead to different analytical methods which have different pros and cons.

These differences are rarely emphasised in applied research; it is rather the commonalities among conjoint paradigms that are highlighted: for instance, that all conjoint paradigms make use of some type of combinatorial experiment – and that, if the assumptions are satisfied, all provide methods for inferring individual or aggregate part-worth utilities. Our discussion focussed on the assumptions one must make before using the various major conjoint paradigms to estimate part-worth utilities and predict choices, and on the likelihood that the assumptions are satisfied in real data. Finally, we argued that the goal of most conjoint studies is to make inferences about *choice* behaviour, not rating or ranking behaviour; and we suggested that when possible transport researchers should consider the use of designed choice experiments to study behaviour in travel choice.

This last consideration is important because there may be a significant difference between the maximum utility alternative predicted from ranking or rating tasks and the alternative actually chosen by an individual. One way to avoid such errors in prediction is to use discrete choice tasks instead of ranking or rating tasks. Although more complicated to design and analyse than traditional conjoint tasks, choice tasks are easier for subjects and more closely mimic the behaviour they are designed to study. Also, rating or ranking a single set of alternatives tells nothing about the way in which choices can change in response to the size and composition of choice sets. Choice experiments have the great advantage of allowing one to observe how choices change as a function not only of changes in attributes of alternatives, but also of changes in the number and composition of competing alternatives.

A serious limitation in most previous applications of conjoint analysis methods

has been a lack of data on parallel product positioning for existing choices. One cannot assume, except in special circumstances (for example, in expert decision making), that consumers will perceive the levels of the attributes of real alternatives to be the same as the levels one can observe and (possibly) measure objectively. Thus, to forecast choice behaviour in most real markets, one needs product positioning data to serve as the baseline case. The product position measures must be in the same metric that one uses to describe the levels of the attributes in the conjoint experiment(s); otherwise the conjoint and positioning data will be incompatible. That is, it cannot be assumed (except in special circumstances) that individuals will think that the levels of attributes of transport alternatives are the same as those measured objectively by transport researchers.

Research is needed to discover how far the information provided by techniques such as Product Positioning Conjoint can be used in lieu of separate data on product positioning to predict current and future choices. Research is needed also to define transport decision attributes (such as times and costs) commonly used in many research studies in terms of the variables used by consumers to define these attributes. Likewise appropriate objective measures of travel decision variables should replace the present practice of defining objective or “engineering” measures on the basis of convenience or past practice or by other arbitrary or *ad hoc* means.

Designed choice experiments offer researchers the opportunity to test competing models of choice processes under controlled conditions. Existing designs for studying choice behaviour do satisfy the statistical properties of MNL choice models; but research is needed to identify optimal designs to estimate models other than MNL models. Similarly, research is needed to determine optimal designs for testing competing choice models. Because choice processes can now be studied under controlled conditions, and because there is now evidence to suggest that the results of such experiments are externally valid, time and money can be saved by using this method to study travel choice processes of interest.

Despite the growing evidence that conjoint methods can forecast choice behaviour, research is needed to determine the appropriate sphere of application of conjoint techniques: that is, in what types of problems are conjoint techniques most and least likely to exhibit external validity? – and are some conjoint techniques more likely than others to produce externally valid predictions? Practical research could benefit from a set of guidelines developed from rigorous research to define the scope and limitations of conjoint techniques. Previous attempts to define such guidelines (for example, Benjamin and Sen, 1983) have considered only ranking and rating conjoint methods, and have not reviewed rank-order explosion techniques or discrete choice techniques. Furthermore, no comparisons are available of product positioning measures with “engineering” measures. Hence, both academic and practical researchers could benefit from research to establish guidelines for application.

We hope this review of several aspects of conjoint analysis paradigms will give transport researchers a better understanding of the tradeoffs required in practical applications, and that it will encourage more academically inclined researchers to study the many unresolved problems and research issues we have discussed.

REFERENCES

- Adelman, L., T. R. Stewart and K. R. Hammond (1974): *A Case History of the Application of Social Judgment Theory to Policy Formation*. Report No. 173, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.
- Anderson, N. H. (1971): "An Exchange on Functional and Conjoint Measurement". *Psychological Review*, 78: 457–458.
- Anderson, N. H. (1981): *Foundations of Information Integration Theory*. New York: Academic Press.
- Anderson, N. H. (1982): *Methods of Information Integration Theory*. New York: Academic Press.
- Anderson, N. H., and J. Shanteau (1977): "Weak Inference with Linear Models". *Psychological Bulletin*, 84: 1155–1170.
- Benjamin, J., and L. Sen (1983): *An Empirical Test of the Utility of Attitudinal Measures for Short-term Public Transportation Planning*. Springfield, VA: National Technical Information Service.
- Brunswick, E. (1956): *Perception and the Representative Design of Experiments*. Berkeley, CA: University of California Press.
- Cattin, P., and D. R. Wittink (1982): "Commercial Use of Conjoint Analysis: A Survey". *Journal of Marketing*, 46: 44–53.
- Cattin, P. and D. R. Wittink (1986): *Commercial Use of Conjoint Analysis: An Update*. Paper presented to the ORSA/TIMS Marketing Science Meetings, Richardson, TX, March 12–15.
- Chapman, R. G. (1984): "An Approach to Estimating Logit Models of a Single Decision Maker's Choice Behavior". *Advances in Consumer Research*, 11:656–661.
- 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.
- Davidson, J. D. (1973): "Forecasting Traffic on STOL". *Operations Research Quarterly*, 24: 561–569.
- Dawes, R. M. and B. Corrigan (1974): "Linear Models in Decision Making". *Psychological Bulletin*, 81:95–106.
- Green, P. E. (1974): "On the Design of Choice Experiments Involving Multifactor Alternatives". *Journal of Consumer Research*, 1:61–68.
- Green, P. E., J. D. Carroll and S. M. Goldberg (1981): "A General Approach to Product Design Optimization via Conjoint Analysis". *Journal of Marketing*, 45:17–37.
- Green, P. E. and V. R. Rao (1971): "Conjoint Measurement for Quantifying Judgmental Data". *Journal of Marketing Research*, 8:355–363.
- Green, P. E., V. Rao and W. Desarbo (1978): "Incorporating Group Level Similarity Judgments in Conjoint Analysis". *Journal of Consumer Research*, 5:127–192.
- Green, P. E. and V. Srinivasan (1978): "Conjoint Analysis in Consumer Research: Issues and Outlook". *Journal of Consumer Research*, 5:103–123.
- Green, P. E. and Y. Wind (1973): *Multiatribute Decisions in Marketing: A Measurement Approach*. Hinsdale, IL: Dryden Press.
- Hammond, K. R., J. Rohrbaugh, J. Mumpower and J. Adelman (1977): "Social Judgment Theory: Applications in Policy Formulation". In Kaplan, M. F. and S. Schwartz (eds.): *Human Judgment and Decision Processes in Applied Settings*. New York: Academic Press.
- Hensher, D. A., and J. J. Louviere (1983): "Identifying Individual Preferences for International Air Fares". *Journal of Transport Economics and Policy*, 17: 225–245.
- Jain, A. K., F. Acito, N. K. Malhotra and V. Mahajan (1979): "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models". *Journal of Marketing Research*, 16:313–322.
- Kocur, G., W. Hyman and B. Aunet (1982): "Wisconsin Work Mode-Choice Models Based on Functional Measurement and Disaggregate Behavioral Data". *Transportation Research Record*, 895:24–31.
- Koeppel, P. K.-W., and A. J. Neveu (1979): *An Acid Test of the Trade-Off Method of Attitude Measurement*. Preliminary Research Report No. 160. Planning Research Unit, NY State Department of Transportation, Albany, NY.

- Krantz, D. H., and A. Tversky (1971a): "An Exchange on Functional and Conjoint Measurement". *Psychological Review*, 78:457–458.
- Krantz, D. H., and A. Tversky (1971b): "Conjoint Measurement Analysis of Composition Rules in Psychology". *Psychological Review*, 78:151–169.
- Krantz, D. H., R. D. Luce, P. Suppes and A. Tversky (1971): *Foundations of Measurement*. New York: Academic Press.
- Kruskal, J. B. (1965): "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data". *Journal of the Royal Statistical Society, Series B*, 27:251–263.
- Lerman, S. R., and J. J. Louviere (1978): "On the Use of Direct Utility Assessment to Identify the Functional Form of Utility and Destination Choice Models". *Transportation Research Record*, 673:78–86.
- Levin, I. P., J. J. Louviere, A. A. Schepanski and K. L. Norman (1983): "External Validity Tests of Laboratory Studies of Information Integration". *Organizational Behavior and Human Performance*, 31:173–193.
- Lieber, S. R. (1976): "A Comparison of Metric and Nonmetric Scaling Models in Preference Research". In Golledge, R. G., and G. Rushton (eds.): *Spatial Choice and Spatial Behavior*. Columbus, OH: Ohio State University Press.
- Louviere, J. J. (1974): "Predicting the Evaluation of Real Stimulus Objects from an Abstract Evaluation of their Attributes: the Case of Trout Streams". *Journal of Applied Psychology*, 59(5): 572–577.
- Louviere, J. J. (1977): *A Study of Teacher/Course Evaluations*. Unpublished Final Report to the Division of Basic Research, College of Arts and Sciences, University of Wyoming, Laramie.
- Louviere, J. J. (1979): "Modeling Individual Residential Preferences: a Totally Disaggregated Approach". *Transportation Research*, 13A: 374–384.
- Louviere, J. J. (1981): "On the Identification of the Functional Form of the Utility Expression and its Relationship to Discrete Choice". Appendix B in Hensher, D. A., and L. W. Johnson: *Applied Discrete Choice Modelling*. London: Croom-Helm.
- Louviere, J. J. (1983): "Integrating Conjoint and Functional Measurement with Discrete Choice Theory: An Experimental Design Approach". *Advances in Consumer Research*, 10: 151–156.
- Louviere, J. J. (1984a): "Using Discrete Choice Experiments and Multinomial Logit Choice Models to Forecast Trial in a Competitive Retail Environment: a Fast Food Restaurant Illustration". *Journal of Retailing*, 60:81–107.
- Louviere, J. J. (1984b): "Hierarchical Information Integration: A New Method for the Design and Analysis of Complex Multiattribute Judgment Problems". *Advances in Consumer Research*, 11:148–155.
- Louviere, J. J. (1986): "A Conjoint Model for Analyzing New Product Positions in a Differentiated Market with Price Competition". *Advances in Consumer Research*, 13:375–380.
- Louviere, J. J. (1987): "An Experimental Design Approach to the Development of Conjoint Based Choice Simulation Systems with an Application to Forecasting Future Retirement Choices". In: Golledge, R. G., and H. G. Timmermans (eds.): *Behavioral Modeling in Geography and Planning*. London: Croom-Helm. Forthcoming.
- Louviere, J. J. (1988): *Modeling Individual Decisions: Metric Conjoint Analysis – Theory, Methods and Practical Applications*. Sage University Papers Series. Sage Publications, Inc., Newbury Park, CA.
- Louviere, J. J., L. L. Beavers, K. L. Norman and F. Stetzer (1973): *Theory, Methodology and Findings in Mode Choice Behavior*. Working Paper No. 11, Institute of Urban and Regional Research, University of Iowa.
- Louviere, J. J. and G. A. Gaeth (1986): *Investigating Communications Content by Combining Management Judgment with Product Positioning and Choice Data*. Paper presented to the annual ORSA/TIMS international meetings, Miami, FL.
- Louviere, J. J. and G. A. Gaeth (1987): "Decomposing the Determinants of Supermarket Choice: A Hierarchical Information Integration Approach". *Journal of Retailing*, 63(1), 25–48.
- Louviere, J. J. and A. Ghosh (1986): *Deriving Measures of Mall Attractiveness*. Paper presented to the annual ORSA/TIMS international meetings, Miami, FL.

- Louviere, J. J., D. H. Henley, G. Woodworth, R. J. Meyer, I. P. Levin, J. W. Stoner, D. Curry and D. A. Anderson (1981): "Laboratory Simulation versus Revealed Preference Methods for Estimating Travel Demand Models: an Empirical Comparison". *Transportation Research Record*, 794: 42–51.
- Louviere, J. J. and D. A. Hensher (1982): "On the Design and Analysis of Simulated Choice or Allocation Experiments in Travel Choice Modeling". *Transportation Research Record*, 890:11–17.
- Louviere, J. J., and R. D. Johnson (1988): "Using Conjoint Analysis to Measure Retail Image". In: C. Ingene and A. Ghosh (eds.): *Spatial Aspects of Market Analysis*. In press.
- Louviere, J. J. and G. Kocur (1983): "The Magnitude of Individual Level Variations in Demand Coefficients: a Xenia, Ohio, Case Example". *Transportation Research*, 17A:363–374.
- Louviere, J. J. and L. Larsen (1987): *Organizational Conjoint: A New Method for Studying Group Decision Making*. Unpublished paper, Department of Marketing and Economic Analysis, University of Alberta, Edmonton.
- Louviere, J. J. and R. J. Meyer (1981): "A Composite Attitude-Behavior Model of Traveller Decision Making". *Transportation Research*, 15B:411–420.
- Louviere, J. J. and M. Piccolo (1977): "Information Integration Theory Applied to Real-World Choice Behavior: Validation Experiments Involving Out-of-Town Shopping and Residential Choice". *Great Plains/Rocky Mountains Geographical Journal*, special edition, 5–21.
- Louviere, J. J., H. Schroeder, G. G. Woodworth and C. H. Louviere (1987): "Do the Parameters of Choice Models Depend upon Mode of Stimulus Presentation? Visual vs. Verbal Presentation". *Advances in Consumer Research*, 14: forthcoming.
- Louviere, J. J. and G. G. Woodworth (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. and G. G. Woodworth (1985): *Models of Park Choice derived from Observational and Experimental Data: a Case Study in Johnson County, Iowa*. Unpublished Final Report to the US Forest Service, North Central Forest Experiment Station, Urban Forestry Project, Chicago, IL.
- Louviere, J. J. and G. G. Woodworth (1986): *You Can Have your Cake and Eat it too: Correlated Attributes Choice Designs that Permit Orthogonal Parameter Estimation*. Paper presented to the 1986 ORSA/TIMS Marketing Science Conference, Richardson, TX.
- Louviere, J. J., and G. G. Woodworth (1987): *On the Design and Analysis of Correlated Conjoint Experiments Using Difference Designs*. Paper presented to the annual meeting of the Association for Consumer Research, Boston, MA.
- Luce, R. D. (1959): *Individual Choice Behavior*. New York: John Wiley and Sons.
- Luce, R. D. (1981): "Axioms for the Averaging and Adding Representations of Functional Measurement". *Mathematical Social Sciences* 1:139–144.
- Luce, R. D. and J. W. Tukey (1964): "Simultaneous Conjoint Measurement: a New Type of Fundamental Measurement". *Journal of Mathematical Psychology*, 1:1–27.
- McClelland, G. H. (1980): *Axioms for the Weighted Linear Model*. Report No. 227, Institute of Behavioral Science, Center for Research on Judgment and Policy, University of Colorado, Boulder.
- McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior". In P. Zarembka (ed.): *Frontiers in Econometrics*. New York: Academic Press, 105–142.
- McFadden, D. (1981): "Econometric Models of Probabilistic Choice". In: Manski, C. and D. McFadden (eds.): *Structural Analysis of Discrete Data*. Cambridge, MA: MIT Press, 198–272.
- McFadden, D. (1987): "The Choice Theory Approach to Marketing Research". *Marketing Science*, 5:275–297.
- Meyer, R. J., I. P. Levin and J. J. Louviere (1978): "Functional Analysis of Mode Choice". *Transportation Research Record*, 673, 1–7.
- Ministry of Transport and Public Works, The Netherlands (1985): "Workshop 5: Stated Preference Methods". In: Ruhl, A. (ed.): *Behavioral Research for Transport Policy*, Vol. 2:

- Introductory Papers. Preprint of the Proceedings of the 1985 International Conference on Travel Behavior. Den Haag: The Ministry.
- Norman, K. L. (1977): "Attributes in Bus Transportation: Importance Depends on Trip Purpose". *Journal of Applied Psychology*, 62: 164–170.
- Norman, K. L. and J. J. Louviere (1974): "Integration of Attributes in Public Bus Transportation: Two Modelling Approaches". *Journal of Applied Psychology*, 59:753–758.
- Pampel, F. C., I. P. Levin, J. J. Louviere, R. J. Meyer and G. Rushton (1984): "Retirement Migration Decision Making". *Research on Aging*, 6:139–162.
- Parker, B. R. and V. Srinivasan (1976): "A Consumer Preference Approach to the Planning of Rural Health Care Facilities". *Operations Research*, 24:991–1025.
- Phelps, R. H., and J. Shanteau (1978): "Livestock Judges: How Much Information Can an Expert Use?" *Organizational Behavior and Human Performance*, 21:209–219.
- Slovic, P., and S. Lichtenstein (1971): "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment". *Organizational Behavior and Human Performance*, 6:649–744.
- Timmermans, H. G. (1982): "Consumer Choice of Shopping Centre: an Information Integration Approach". *Regional Studies*, 16(3):171–182.
- Tversky, A. (1967): "A General Theory of Polynomial Conjoint Measurement". *Journal of Mathematical Psychology*, 4:1–20.
- Tversky, A. (1972): "Elimination by Aspects: a Theory of Choice". *Psychological Review*, 79:281–299.
- Wainer, H. (1976): "Estimating Coefficients in Linear Models: It Don't Make No Nevermind". *Psychological Bulletin*, 83:213–217.
- Wilkinson, L. (1986): *SYSTAT: The System for Statistics*. Evanston, IL: SYSTAT, Inc.
- Wittink, D., and D. Montgomery (1979): "Predictive Validity of Trade-off Analysis for Alternate Segmentation Schemes". In *Proceedings, American Marketing Association Educators Conference*. Chicago: American Marketing Association.
- Wright, P. L., and M. A. Kriewall (1980): "State-of-Mind Effects on the Accuracy with which Utility Functions Predict Marketplace Choice". *Journal of Marketing Research*, 17: 277–293.