Stated preference analysis of travel choices: the state of practice

DAVID A. HENSHER
Institute of Transport Studies, Graduate School of Business, The University Sydney, NSW 2006, Australia

Abstract. Stated preference (SP) methods are widely used in travel behaviour research and practice to identify behavioural responses to choice situations which are not revealed in the market, and where the attribute levels offered by existing choices are modified to such an extent that the reliability of revealed preference models as predictors of response is brought into question. This paper reviews recent developments in the application of SP models which add to their growing relevance in demand modelling and prediction. The main themes addressed include a comparative assessment of choice models and preference models, the importance of scaling when pooling different types of data, especially the appeal of SP data as an enriching strategy in the context of revealed preference models, hierarchical designs when the number of attributes make single experiments too complex for the respondent, and ways of accommodating dynamics (i.e. serial correlation and state dependence) in SP modelling.

Introduction

It is twenty years since the seminal papers by Davidson (1973) and Louviere et al. (1973) in transportation were published which alerted us to the appeal of methods for evaluating an individual's response to *combinations* of levels of attributes of modes of transport which are not observed in the market, but which represent achievable levels of service. Widespread interest in this "new" approach to travel behaviour modelling, however, was slow in developing, in part due to the high agenda interest in the development of discrete-choice models and activity approaches to the study of the continuous sequences of human actions over a period of time (see Hensher and Stopher 1979).[1] Indeed, until the early eighties, the transport contributions were dominated by

publications from Louviere and his colleagues (see Louviere 1979 for a summary) with an almost universal application to the study of mode choice (Meyer et al. 1978).[2]

Although it is always difficult to pinpoint the major events which heralded the beginning of a widespread interest in SP methods, the motivation seems to have evolved from a number of applications in which the behavioural response involved an alternative which was either not currently available (e.g. Louviere and Hensher 1983, Hensher 1982) or where there was difficulty in assessing substantially different attribute mixes associated with existing alternatives to those observed (e.g. Kocur et al. 1982, Hensher and Louviere 1983, Bradley and Bovy 1984, Louviere and Kocur 1983). An important paper by Lerman and Louviere (1978) demonstrated the theoretical links between revealed preference and stated preference models.

Prior to the paper by Louviere and Hensher (1983), the emphasis had been on judgemental tasks in which a respondent was asked to rate or rank a number of attribute mixes associated with a particular choice context. The modelling of this data using standard regression-based estimation procedures required simulation of choice environments in order to predict market share. Louviere and Hensher showed how a preference experiment (i.e. a number of alternative mixes of attributes) could be extended to incorporate choice experiments in which an individual chooses from among fixed or varying choice sets, enabling estimation of a discrete-choice model and hence direct prediction of market share. Stated *choice* experiments are now the most popular form of SP method in transportation and are growing in popularity in other areas such as marketing, geography, regional science and tourism. The papers by Louviere and Hensher (1982) and Louviere and Woodworth (1983) have become the historical reference sources for stated choice modelling in transportation.

The introduction of stated choice modelling using the set of established discrete-choice modelling tools routinely applied with revealed preference data widened the interest in SP-methods. For the first time travel behaviour researchers could see the benefit of stated-preference data in enhancing their travel choice methods. This I would argue was *the* major watershed which after 10 years has resulted in widespread acceptance of SP methods in practice in transportation. A number of monographs and special issues of journals are now available which capture the major contributions up to the late eighties (Pearmain et al. 1991, Louviere 1988, Bates 1988, and Louviere 1992). Louviere, Hensher and Shocker (1992) run an annual short course, covering all aspects of stated-preference modelling (i.e. relevance, design, estimation, and application). PTRC also run a London-based annual short course. Batsell and Louviere (1991) and Louviere (1993) have recently reviewed the state of the art in experimental analysis of choice experiments. Green and Srinivasan

(1978, 1990) are the recognised review sources in marketing. Louviere and Timmermans (1990) provide an overview in the context of tourism.

With this brief historical perspective behind us, this paper concentrates on some of the important developments in recent years which crystallise the state of practice in stated preference modelling. In particular, we evaluate the pros and cons of alternative response metrics (namely ranks, rates and choice), the major considerations in the design of an experiment (i.e. attribute selection, attribute levels, main and interaction effects, hierarchical designs and making the exercise comprehensive and comprehendible), approaches to model estimation (especially individual models, and individual choice models based on a sample of individuals where the data is maintained at a disaggregate level or aggregated within each observation to choice proportions), and the scaling of data with different metric dimensions to enable data aggregation and enrichment. We also refer to the growing software capability for experimental design, model estimation and market share prediction. Throughout the paper the emphasis is on the practice of SP analysis.

Defining the response dimension

There are two broad categories of stated response of interest in travel behaviour research: (i) An individual is asked to indicate his *preferences* among a set of combinations of attributes which define services or products. This judgemental task, usually seeks a response on one of two metric scales – a rank ordering or a rating scale. (ii) An individual is asked to *choose* one of the combinations of attributes. Information is not sought on the ordering or rating of each of the non-chosen combinations. This is often called a first-preference choice task.

In both stated preference and stated choice experiments, each combination of attributes can be defined as an *alternative* in the sense of representing a product of service specification which may not be observed in the market. The attributes can include not only well-defined sources of (indirect) utility such as travel times and travel costs, but also aggregators such as name of product (e.g. car, train) which represent the respondent's perception of the attributes of the alternatives which are not represented by the explicitly defined attributes. In both preference and choice experiments it is feasible to vary both the combinations of attributes and levels as well as the subsets of mixes to be evaluated. This can be achieved by either designing varying numbers of combinations or asking the respondent to a priori eliminate any combinations which are not applicable before responding (soliciting criteria for non-applicability – see Louviere and Hensher 1983).

In practice, it is common in preference experiments to hold the number

of alternative attribute mixes constant and only vary the attribute levels. However, in choice experiments, it is common to vary the number of alternatives, while either holding the attribute levels associated with each alternative constant, or varying them, producing *varying choice sets* (e.g. Hensher et al. 1989). Fixed choice set designs are also widely used (e.g. Louviere and Hensher 1983, Gunn et al. 1992).

The decision on which type of response strategy to pursue must be addressed at the beginning of an SP study, because it will define the available outputs.[3] A major consideration is the need for predictions of behavioural response, especially market shares. Rank order and ratings "predictions" must be transformed to accommodate useful predictive outputs (except where the interest centres on the image of, or attitude towards, a service or product – see Hensher 1991). Choice responses are directly translated into predictions, through the application of discrete-choice models such as multinomial logit (MNL), and are also relatively easier for the respondent. However, the advantage of the direct translation comes at the expense of information loss. In a first-preference choice experiment, no information is available on the ordering of all of the alternatives in contrast to ranking and even rating. In recognition of this information loss, a number of studies have investigated ways of maximising the information content of a response metric while both maintaining the ability of the respondent to handle a more difficult task and have the capability of estimating a model which can provide useful predictive outputs in the form of market shares (and attribute elasticities) (e.g. Elrod et al. 1992, Ben-Akiva et al. 1992).

Rank-order data

Rank order (non-metric) data is popular with analysts who subscribe to the view that individuals are more capable of ordering alternatives that reporting, by a rating task, their degrees of preferences. A choice experiment is a first-order ranking task. A procedure proposed by Chapman and Staelin (1982) for translating rank order data into choice responses, referred to as 'rank explosion', enables one to translate the full depth of R ranks into R-1 choice observations. Each choice set in a sequence excludes the alternative(s) ranked above each level in the rank as we redefine each rank level as the "chosen" from the set below the rank of the "previously chosen." For example, if we have four alternatives and each is ranked 1 to 4, the reconfigured sequential choice sets are the chosen as rank = 1 and the remaining 3 alternatives, the chosen as rank = 2 and the alternatives ranked 3 and 4, and the chosen as rank = 3 and the alternative ranked 4. Automatic explosion and estimation as a multinomial logit model can be executed in the LIMDEP package (Econometric Software, 1992) and ALOGIT (distributed by Hague Consulting Group).

The usefulness of preference ranking data has recently been questioned by Ben-Akiva et al. (1992). They found that response data from different ranking depths are *unequally* reliable, and that different ranks produce statistically significantly different estimates of the (indirect) utilities. To conform with the underlying properties of discrete-choice models, that is consistency with random utility maximisation and the well known properties of MNL models, the estimated indirect utilities from the full choice set should be proportional to the utilities estimated from any other choice set of another ranking depth. This requirement is rejected by Ben-Akiva et al. (1992) for several of the depths of comparisons. Although this evidence is based on only one empirical study, there is a growing view that rank order data provides limited information, at least below rank 4 (Hensher and Louviere 1983), a position corroborated by Bradley and Daly (1992).[4] It suggests the potential value of confining choice analysis to the first preference choice.

Further research is required to decide on the fate of ranking data as a basis of translation of a full or part profile of ranks into choice responses, to enable direct prediction of market shares. The usefulness of rank order responses *analysed as ranks* however is not under question per se, except to the extent of the reliability of the lower-order ranks. Hensher and Louviere (1983) proposed a way of transforming ranking responses into expected choice frequencies for analysis in the random utility framework. The translation produces choice proportions representing the responses of a sampled individual when faced with an alternative in every possible choice set. Iterative weighted least squares regression can be used to obtain parameter estimates. To my knowledge, the method has not been used by anyone else.

Ratings data

Ratings are, prima facie, the richest response metric, giving both order (including ties) and degree of preference. Analysts typically select a 5 or 10 point scale (and occasionally 100 points), to represent an underlying (i.e. latent) continuous distribution of interval scaled rates. A rating task is also the most demanding on a respondent, since the magnitude of the response associated with each attribute mix can vary across the entire rating scale. Ratings data are often assumed to have a monotonic translation into a utility scale, and after model estimation using techniques such as generalised least squares regression, the parameter estimates are applied via a logit transformation to obtain choice probabilities. The validity of this transformation is questionable, at least because of the discrete-nature of the ordered sets of rations available to the respondent. There are also different distributional assumptions for the error components of GLS and MNL.

A preferable approach to utilising ratings data in the derivation of choice

probabilities is to treat the observed ratings as a non-linear rating scale in an ordered responses model which defines points on the *observed rating scale* as thresholds (Henry 1982, Winship and Mare 1984, Crask and Fox 1987). Empirical rating scales are best viewed as discrete realisations of unmeasured continuous variables. The ordered probit or ordered logit model allows one to include ordinal dependent variables into the preference model in a way that explicitly recognises their ordinality and avoids arbitrary assumptions about their scale (Johnson 1990).[5] The essence of the approach is an assumed probability distribution of the continuous variable that underlies the observed ordinal dependent variable. Ordered probit or logit also takes into account the ceiling and floor restrictions on models which include ordinal variables, whereas a linear regression does not.

In specifying an appropriate preference model, we assume that the observed rating scale is a nonstrict monotonic transformation of an unobserved interval variable. Thus one or more values of an interval-level variable are mapped into the same value of a transformed ordinal variable. An underlying continuous variable is mapped into categories that are ordered but are separated by unknown distances. We cannot, for example, say that the difference between ratings 5 and 4 is identical to the difference between ratings 44 and 3, or 3 and 2. This method has been implemented in Hensher (1991), and Ortuzar and Garrido (1993). One possible practical limitation of the ordered probit model is that it does not have a closed-form solution. This means that each change in an attribute level must be evaluated through integration of the open-form choice probability model. LIMDEP fortunately can perform this task with ease.

Choice data

The attraction of choice responses in part evolves from the discussion of rank and rating data. Ultimately, the majority of travel behaviour practictioners want predictions of the demand or market share for a service or product. Individuals in reality make decisions by comparing a set of alternatives and selecting one. With this simple requirement in mind and the commentary above, the appeal of a first preference choice modelling approach is clear.

An appealing feature of stated choice (SC) data is the ability to view the experiment as the stated response counterpart to revealed preference (RP) data, the mainstay of econometric modelling. In addition to the capability of stated choice experiments to extend evaluation beyond observed attribute levels, the essential difference is one of scale. The recognition of the relative strengths and weaknesses of both types of data suggest that the joint utilisation of both data should enrich the modelling activity and further our understanding of choice behaviour. In particular SP data can be used effectively to enrich the

predictive capability of a base RP model, especially where the market share for a new alternative is being evaluated.

Whereas RP data describes actual choices in terms of a set of market-based measurements of attributes of alternatives (which by definition are restricted to the currently available feasible set), the SC data describe potential choices in terms of a set of constructed measures of combinatorial mixes of attributes of real and/or hypothetical alternatives. The opportunity to position an SC data set relative to an RP data set within the one empirical analysis on the common choice problem enables the modeller to extend and infill the relationship between variations in choice response and levels of the attributes of alternatives in a choice set, and hence increase the explanatory power of the RP choice model.

The mixing of sources of data however is not a matter of "naive" pooling. It requires careful consideration of the unit of the (indirect) utility scale. For example, the utility scale in an MNL model is inversely related to the variance of the unobserved influences, summarised as the random error term; hence the parameter estimates of two identical indirect utility specifications obtained from two data sources with different variances will necessarily differ in magnitude, even if the choice process that generated the indirect utilities is identical. The notion of scaling is not new. Horowitz (1981), for example, alluded to it. However prior to the contribution of Morikawa (1989), the scaling discussion was not specifically directed to the opportunity to enrich RP data with SC data. Some recent applications using mixed data are Morikawa (1989), Bradley and Daly (1991, 1992), Hensher and Bradley (1993) and Swait and Louviere (1993). Given the importance of this enrichment strategy, it is discussed in more detail in section four.

Experimental choice design

The *engine* of stated preference analysis is a controlled experiment, out of which comes a series of survey questions eliciting a response to alternative combinations of levels of attributes. A good experiment is one which has a sufficiently rich set of attributes and choice contexts, together with enough variation in the attribute levels necessary to produce meaningful behavioural responses in the context of the strategies under study.

There is a logical sequence of tasks required to design a choice experiment. This should be distinguished from the issue of statistical design complexity. The latter may require specialist support; whereas the overall process can be executed almost mechanically once the options within each task are known and understood. Importantly, the tasks must be undertaken with some broad awareness of the downstream implications of a decision taken at

each stage. Hence a considerable amount of revision is likely to occur before finalising an experiment. The key steps are summarised below, using commuter mode choice as the example context.

Task 1 involves the *identification of the set of attributes* which need not to be considered as sources of influence on mode choice. There may well be a large number of these attributes, requiring an early decision on which attributes to include in the experimental design and which to exclude, treating the latter as *contextual* or *covariate* effects. For example, we might exclude the number of transfers associated with the use of a public transport mode, because it is unlikely to vary within a mode. It may be better accommodated by describing a fixed level for a mode in the context-setting statement accompanying the experiment.

One way of preserving a large number of design attributes is to partition the attributes into generic groups, with each group defined by elemental attributes, and to design a number of linked hierarchical experiments (e.g. Hensher 1991, Louviere and Gaeth 1987, Hague Consulting Group 1988, Kroes and Sheldon 1988, Timmermans 1988).[6] The hierarchical approach assumes that individuals classify attributes into a set of generic decision constructs (e.g. comfort, convenience, cost and time). They then choose among the alternatives based on the generic attributes. A separate experiment for each generic attribute involves the respondent rating each of the elemental attributes associated with each mode, in order to give some substantive interpretation to the sources of utility underlying the generic constructs. Choosing a mode is not so meaningful for a subset of sources of influence on choice.

Task 2 involves *selecting the measurement unit for each attribute.* In most cases the metric for an attribute is unambiguous; however there are situations where this requires consideration of alternative metrics. This is particularly true for generic attributes such as comfort. For example, one could define an ordinal scale of high, medium and low (which may be problematic if the analyst does not describe precisely what each level represents). Alternatively one could endogenise the construction of the metric scale by asking each respondent to first place values on each of the generic attributes, possibly on a satisfaction rating scale, to define one of the levels as the current level (not necessarily the medium level), and then the analyst can construct the other two levels as variations from the reported level. This is not so clear for a new alternative; however one way is to use a very clear description of the new mode as the surrogate for current awareness/experience. It may be as good as reliance made on *experience* with available alternatives which have never been used.

Well-defined attributes such as wait time can be treated in a number of ways, although in practice the selection is described primarily by the number of levels. For example, a two-level attribute could be "zero" and "greater than

zero"; a three-level attribute could be "zero", 5 and 10 minutes. There is more information in the latter, although the complexity of the experiment and/or the accuracy required on this attribute and/or the inability of a respondent to perceive any noticeable difference between levels may have a bearing on the selection.

Task 3 involves the *specification of the number and magnitudes of attribute levels*. As a rule of thumb, one should be extremely cautious about choosing attribute levels which are well outside the range of both current experience *and* believability. For existing alternatives one should construct a range which contains the level currently faced by an individual, no matter how the attribute is measured, and define it as one of the levels in the design. Consideration of the magnitude of an attribute to be evaluated in an application is also crucial to ensure that the design can assess behavioural response in the new attribute level regime. When new alternatives are being evaluated, making the attribute levels believable (and deliverable) becomes a primary consideration.

The number of levels for each attribute will be decided by the overall complexity of the design. This involves consideration of the combinations of attribute levels generated, the manner in which they are exposed to a respondent (i.e. partitioned or in their entirety), the need to investigate non-linearity (which is not possible with only two levels), and the extent to which interaction effects between pairs of attributes may be important. The final selection and format of implementation must be decided by the criterion of being comprehendible to the respondent.

Task 4, statistical design, is where the attribute levels are combined into an experiment. A combination of attribute levels describes an alternative, referred to in the literature as a *profile* or *treatment*. The alternative can be abstract in the sense of being an attribute mix which is not defined for a particular mode (e.g. a travel time and cost); or it can be mode-specific (e.g. travel time and cost by car). The former is often referred to as an *unranked alternative* and the latter as a *ranked alternative*.

Alternatives are generated with the aid of statistical design theory. In a statistical experiment each attribute has levels, and it is these levels that are the input data required to construct a factorial design (i.e. combinations of attribute levels for all attributes in the design). A *full factorial design* contains descriptions of all possible alternatives, enabling one to independently estimate the statistical effects of each attribute on the choice response. In practice the full number of combinations is impracticable to evaluate, and so a *fractional factorial design* is constructed. The price one pays for making the experiment manageable is that some statistical efficiency is lost. In designing a fractional factorial experiment, the analyst has to assume that certain interaction effects among the attributes are not statistically significant. This is a

very reasonable non-testable assumption for a large number of possible interactions, especially interactions of more than two attributes (e.g. three-way interactions), and indeed for many two-way interactions.[7] If interactions are statistically significant, their effects in a fractional factorial design will be loaded onto the individual main effects, giving erroneous results. This is referred to as *confounding* main effects with interaction effects. The analyst has to be creative in selecting a limited number of two-way interactions which enable one to include up to that number of interactions to test for statistical significance. It is important to note that the two-way interactions can be any two attributes, up to the maximum allowed for statistical independence.

The most common fractional factorial design is a *main effects* plan. The majority of previous applications in transportation are of this structure. A main effects plan does not, in a statistical design sense, provide a sufficient number of alternatives to be able to detect unobserved but possibly significant interactions effects, preventing determination of whether the estimated main effects are statistically biased. Main effects plans assume that individuals process information in a strictly additive way, such that there are no significant interactions between attributes.[8] A main effects plan does enable the analyst to define a linear and high-order dimensions (e.g. quadratic) for each attribute. Higher-order effects are important where the marginal rate of substitution between two attributes (e.g. the value of travel time savings) is a function of the magnitude of a design attribute (e.g. the value of travel time savings is a function of the level of travel time).

The fractional factorial designs are given in standard experimental design tables (e.g. Hahn and Shapiro 1966), which indicate to the user (i) if all the main effects are independent of two-way interactions, (ii) the number of independent interactions permissable for each fraction, (iii) the residual degrees of freedom, and (iv) the actual combinations of levels of attributes. For example, 3 attributes at 3 levels gives a full factorial of 27 combinations. This can be reduced to a one-third fraction of 9 combinations in which all the main effects are independent of two-way interactions, but there are no independent two-way interactions.[9]

There are three popular choice designs which highlight the nature of stated choice experiments:

- *A Varying Choice Set Double Conditional Design* (e.g. Hensher et al. 1989) with the first stage a fractional design from a Z^n full factorial where Z is the number of levels of each attribute and n the number of attributes. For example, 3 attributes at 3 levels yields 9 attribute combinations (a one-third fraction). Choice sets are created using a 2^J design where J defines the number of alternatives and '2' indicates their presence or absence in each choice set. A 2^9 design yields a 16 choice set fraction.

- *A Fixed Choice Set Double Conditional Design* (e.g. Louviere and Hensher 1983). The first stage defines alternatives in terms of combinations of levels of attributes. For example, 3 attributes at 2 levels is a 2^3 design giving 8 alternatives. These 8 alternatives can be configured in a fixed choice set with 2 price levels attached to each alternative. In a 2^8 design this gives 12 combinations of low and high prices across the 8 alternatives.

- *A Fixed Choice Set Design* (e.g. Gunn et al. 1992) based on combinations of alternatives and attributes. A fractional design is derived from the number of alternatives and the number of attributes. For example we might have 4 product names and 3 attributes giving 12 combinations. We can assign 2 levels to each attribute (high, low) and determine a fraction from a 2^{12} factorial. 16 choice sets is one outcome, each with 4 alternatives defined in terms of low or high levels of the 3 attributes.

One of the important issues in statistical design is *orthogonality*, which ensures that the attributes presented to individuals are varied independently from one another. This property of zero-correlation between attributes enables the analyst to undertake tests of the statistical contribution of main effects and interactions, and is promoted as a major appeal of SP data compared to RP data. There is a view that although this is a desirable property, it is not a necessary condition for useful SP modelling. RP modellers have had to live with some amount of correlation, and have suitable tests for multicollinearity to identify when correlation is a problem. Mason and Perreault (1991) show in a cross-sectional context that fears about the harmful effects of collinear attributes often are exaggerated. Indeed the major benefit of SP methods is the ability to capture the response to diverse attribute combinations which are not observed in the market. One suspects that this is the dominating reason for the popularity of SP methods in transportation.

Hensher and Bernard (1990) have made a distinction between design-data orthogonality (DDO) and estimation-data orthogonality (EDO) in order to highlight that DDO is not always preserved in model estimation. This is very important for the most common procedure in travel behaviour modelling of estimating an MNL model with three or more alternatives on the individual response data, namely pooling all data (i.e. number of individuals in the sample by number of stated choice replications per individual) across the sampled population, but *not* aggregating the response data within a sampled individual. Estimation orthogonality using individual data and discrete choice models requires that the *differences* in attribute levels be orthogonal, not the absolute levels. Techniques such as MNL estimated on individual data require the differencing on the attributes to be *the chosen minus each and every non-chosen*. Since the chosen alternative is not known prior to design development, it is not possible to design an experiment which has DDO, and which also satisfies EDO (Hensher and Barnard 1990).

The innovative method proposed by Louviere (1988) for overcoming EDO is not feasible where individual data are applied in estimation. The Louviere method defines a base alternative and derives all attribute combinations from a given *difference of attribute levels* satisfying an orthogonal-difference design. It is however suitable when the choice responses are aggregated within each individual's set of replications to derive choice proportions for each alternative. In this case, logit regression is a suitable estimation methods, which does not require any further differencing in estimation. Transportation modellers have tended to opt for the preservation of the individual discrete-choice responses, and hence (without realising it in most cases), accepting some amount of correlation. Indeed with computer-based interviewing, random variation around a realistic correlation structure works well.

There are a growing number of software packages which can be used to design fractional factorial experiments. The user defines each attribute by the number of levels, and then follows menu-driven screen instructions on how to select a particular fractional factorial which has the statistical properties the analyst requires. The most popular packages in transportation are SPEED/MINT (Hague Consulting Group, The Netherlands), CONSURV (Intelligent Marketing Systems – Canada and Econometric Software – Australia), and GAME GENERATOR (Steer, Davies Gleave – UK).

In *Task 5*, the experiment designed in task 4 has to be translated into a set of *questions and showcards for execution in the data collection phase*. The survey instrument can be designed for either a notebook computer or non-computerised administration. Whatever the preferred collection strategy, the design must be translated from a set of orthogonal or near-orthogonal design attribute levels into real information for respondents to comprehend and respond. Where feasible, it is suggested that a respondent be asked to both choose an alternative and either rank or rate the full set of alternatives (or a subset derived from a prior question on applicability or non-applicability of particular alternatives). The subset issue is particularly important where there are too many alternatives to rank or rate, although it may be of interest in a choice response context to ascertain some additional information on relevant sets. If the request for ranking or rating responses may jeopardise the cooperation across the replications of the experiment, it is more important to limit the task to the first preference choice.

Where there are a lot of replications, it is popular to block or randomise the experiment in such a way that subsets of respondents are asked to respond to either a fixed subset or a random subset in a way which ensures that all replications have equal representation. The only concern about this strategy is the extent of segmentation heterogeneity with respect to response profile, which could lead to a distortion of the population's response profile.

In *Task 6* the selection of an appropriate *estimation procedure* will be dependent on the metric of the response variable and the level of aggregation of the data for modelling. The main approaches are summarised in Table 1.

Table 1. Alternative model estimation methods.

Data response	Rating	Rank-order		First-preference
Dimension:		Unexploded	Exploded	
Modelling strategy: Aggregated sample choice proportions	B	D	F	
[Grouped data]				
Individual choices [Individual data]	A	C	E	G

Exploded ranks involves converting data derived from a preference ranking task into choice data for modelling. This procedure has been described above in section 2.1. Approaches B, D and F can use regression based estimation methods such as generalised least squares because the response variable is continuous. However, aggregating data derived from repeated observations on each individual introduces a number of statistical problems due to the non-independence of intro-individual observations (Louviere and Woodworth 1983). Approach A should be estimated by techniques such as ordered logit or ordered probit; GLS would produce biased parameter estimates.

Unexploded ranking data analysed as individual observations (approach C), can be utilised in a choice context by assuming the independence of irrelevant alternatives (IIA) property, and transitivity in the unobserved choice sets implied by the rankings (Hensher and Louviere 1983). A simple translation of rankings into choice proportions for each alternative is then possible. If there is doubt about the use of rules for ranking which violate the IIA assumption, then, the translation is dubious (Louviere 1988). The resulting choice proportions can be used in the estimation, by generalised least squares or ordered logit/probit, of an individual level choice model.

In approach G (or equivalently E) the data relate directly to discrete choice responses, and estimation takes place using the repeated observations on each individual. The MNL model has been used in the majority of the stated choice applications (e.g. Louviere and Hensher 1983, Bates et al. 1987, Wardman 1988, Bradley and Bovy 1984). All of the statistical procedures are available in LIMDEP (Econometric Software 1992). Specialised software is readily available for logit modelling such as ALOGIT (Hague Consulting Group,

the Netherlands), PCLOGIT which superseded BLOGIT (Institute of Transport Studies, University of Sydney), NTELOGIT (Intelligent Marketing Systems – Canada and Econometric Software – Australia) and HLOGIT (ITS, Sydney).

Task 7 uses the estimated parameters in a simulated choice context to obtain choice probabilities for each alternative for each sampled individual, which together with population weights can be used to obtain predictions of market shares or total demand. To obtain predictions of total demand, it is important to allow for a no-choice response in the experimental design (Louviere and Hensher 1983).

Scaling and Enrichment

The behavioural framework underlying discrete-choice models such as MNL is applicable for both RP and SP data. The definition of the observed and unobserved influences on the choice outcome however varies. First, the observed levels of the attributes of alternatives typically obtained in an RP study are sought directly from the decision maker or taken from exogenous data such as posted prices. The responses are reported perceived levels, which may vary from the "actual" levels. By contrast, the attribute levels associated with an SP study are fixed by the analyst, and are by definition "actual" levels. Thus we have at least one source of variation in the metric of the observed attributes of alternatives. Second, the choice outcome in the RP study is the known outcome, whereas for the SP study it is the potential outcome or the outcome with the highest likelihood of occurrence given the combination of attribute levels offered in an experimental replication. Third, the SP study elicits choice responses from a repeated measures experiment in which the attribute levels (and even the choice set) are varied, in contrast to the single response in an RP study. Thus there is a greater amount of information on decision maker response to a range of possible attribute profiles.

After recognising the likely sources of observed variation between RP and SP data, the remaining unobserved sources of indirect utility are most unlikely to display identical distribution profiles within the common sampled population. Hence the *"naive"* pooling of the two types of data cannot be treated as if they display identical unobserved effects. Given that the variance of the unobserved effects is an important piece of information used in the derivation of the functional form of a probabilistic discrete choice model (McFadden 1981), this variance deviation has to be recognised and accommodated. One solution proposed originally by Morikawa (1989) is to scale the variance of the unobserved effects associated with the SP data so that the equality of variances across the RP and SP components of a pooled model

is reinstated. The indirect utility expressions are defined as (Morikawa 1989):

$$V_{rp} = \alpha + \beta X_{rp} + \psi Y + \varepsilon_{rp}$$
$$V_{sp} = \delta + \beta X_{sp} + \gamma Z + \varepsilon_{sp} \quad ; \quad \theta^2 = var(\varepsilon_{rp})/var(\varepsilon_{sp})$$

where

X_{rp}, X_{sp} = a vector of observed variables common to rp and sp data

Y, Z = vectors of observed variables specific to one data set or the other

β, α, δ, ψ, γ = unknown parameters

ε_{rp}, ε_{sp} = the unobserved effects associated with the rp and sp data configurations

θ^2 = the scaling parameter, enabling the scaling of V_{sp}' equal to θV_{sp}, and hence joint estimation of the two data sets.

The probability of a decision maker selecting an alternative out of the available set of alternatives is defined as the probability that the observed and unobserved indirect utility of an alternative is greater than or equal to the observed and unobserved indirect utility of each and every other alternative in the choice set:

$$\text{Prob}_j = \text{Prob}\{(V_j + \varepsilon_j) \geq (V_{j'} + \varepsilon_{j'}); \ j \in J; \ j \neq j'\}$$

Particular assumptions on the distribution of the unobserved effects within the sampled population lead to a particular functional form of the discrete choice model (see below). A priori the relative magnitudes of the variances is unknown, due to the many sources of differences between the RP and SP contexts. The equality of variances is a permissable empirical outcome, but not one to be assumed ex ante.

The Econometric Specification

The distribution of the unobserved effect in an indirect utility expression has always been an important consideration in econometrics. Within the family of random utility models centred on discrete choices, the multinomial logit (MNL) form requires that the unobserved effects are independently and identically distributed (IID) across the alternatives in the choice set, according to the extreme value type I distribution (Hensher and Johnson 1991, Borsch-Supan 1986, Ben-Akiva and Lerman 1985). The violation of this constant variance condition (alternatively referred to as the independence of irrelevant alternatives property) resulted in the development of the nested (or hierarchical)

logit (NL) model, which permitted differential variance between levels and/or branches within a level of the nested structure but a common variance within a branch (Hensher 1986, 1991, Borsch-Supan 1986). The explicit accommodation of differential variance within a nested-logit model provides a means of identifying the scale parameter required to rescale parameter estimates associated with data derived from more than one source which are combined in a single empirical model.

The RP parameters to be estimated are the simple values of α, β and ψ, the SP parameters are $\theta\delta$, $\theta\beta$ and $\theta\gamma$. This scaling has no other effect on the distributional assumptions or on the conversion of the indirect utility expressions to choice probabilities. The scaling of $\theta\beta$ is the essential link between the two data models. The SP model, however, is non-linear. This estimation problem can be solved by available nested-logit software, by setting up an *artificial tree structure* as follows (Bradley and Daly 1991). The artificial nest is constructed to have *at least* twice as many alternatives as are observed in reality. One subset is labelled as RP alternatives, the other subset as SP alternatives. The SP subset can include additional new alternatives not in the RP subset. The indirect utility functions in each case are defined by the V_{rp} and V_{sp} expressions, defined above without theta. The RP alternatives are placed just below the "root" of the nest, whereas the SP alternatives are each placed in a single-alternative "nest". For the SP observations, the average indirect utility of each of the "dummy composite" alternatives (Figure 1 – after Bradley and Daly 1991) uses the theoretical basis of the inclusive value concept associated with linking levels in a nested logit model (McFadden 1981) to define

$$V^{comp} = \theta\log \sum_{j=1}^{J_{sp}} e^{V_{sp}^{j}}$$

in which the summation is taken over all alternatives in the next corresponding to the composite alternative. Because each nest contains only one SP alternative, V^{comp} reduces to θV_{sp}, the expression for a single SP alternative, with every parameter including the unobserved component associated with an SP alternative scaled by θ. It is because the approach operates as if we are estimating a traditional nested logit model and drawing on the empirical content of the inclusive value which links levels in a tree structure that we refer to the estimation of the scaling approach as an *artificial* nested (logit) model. The scale θ does not have to lie in the unit interval, the condition for consistency with random utility maximisation (Hensher and Johnson 1981, Ben-Akiva and Lerman 1985), because individuals are not modelled as choosing from the full set of RP + SP alternatives. The scale for SP relative to RP can be greater than one.
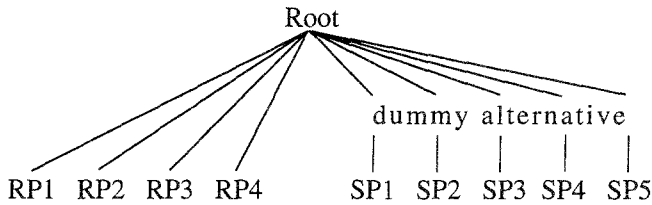
Fig. 1. The Estimation Structure.

The joint estimation using two types of data involves a choice outcome associated with the RP data and a number of choice outcomes associated with the SP data. This is not a typical discrete choice application where there is only one choice outcome in either a MNL or NL configuration. To allow for this multiple response the observations are *stacked* in such a way that for each RP observation there is a null choice set for the SP observation, and for each SP observation there is a null choice set for the equivalent RP observation. The 'hierarchical' structure, given in Figure 1, ensures that each of the parameter estimates associated with the SP data are scaled by the ratio of the variances. The different thetas on each dummy node are constrained to take the same value, a requirement for the scaling condition. Different theta's can be allowed for each *additional* type of SP data set.

The concentration on the unobserved effects is deliberate, given that the scaling of *all* parameters associated with one type of data is necessary to enable joint estimation of two or more types of data within an IID model framework. Any additional sources of variation between attributes can be accommodated by the inclusion of data-type specific dummy variables (i.e. scale factors) such as a fatigue effect dummy for SP data (measured for example by the sequence of replications) (see Bradley and Daly 1992). The scale parameter can be normalised to unity on either the RP or SP side.

To scale the variance of the unobserved effects in the SP component relative to the RP component, a sequential or a simultaneous scaling approach can be used. Simultaneous estimation of the "nested" structure using the method of full-information maximum likelihood (FIML) is the most efficient approach; although sequential estimation can also be used, both allowing us to normalise the variance of the one data source to unity and allowing the variance of the other data source to be empirically determined around unity. In sequential estimation, the calculated standard errors are not efficient and are likely to be underestimated leading to inflated t-statistics (Hensher 1986). Morikawa (1989) cites underestimates by a factor of 10 to 500%. Sequential estimation is also inefficient in the sense of loss of sample points if differential choice sets are permitted across the sample, since one encounters parts of the tree

structure without a chosen alternative or a single alternative (Hensher 1986). As an enrichment strategy for RP modelling, one anticipates a burgeoning literature of RP-SP applications.

Conclusions and new challenges

There are many challenges still to be faced in making the existing set of tools both more user-friendly and capable of assisting in the resolution of further issues emanating from state of the art research. In concluding this paper, we have selected two topics of particular importance in ongoing research to highlight the richness of stated preference modelling. Other important topics not discussed herein are transferability of models over time and location (e.g. Hensher and Battellino 1993), using SP experiments to value environmental effects (e.g. Adamowicz et al. 1992), defining consideration sets and an efficient number of alternatives (e.g. Bunch and Batsell 1989) and external validity (e.g. Horowitz and Louviere 1993).

Incorporating uncertainty in an experiment

It is widely acknowledged that the levels of many attributes actually offered in the marketplace have an element of uncertainty. This is especially the case for services where the variability in total demand affects the ability to deliver a certain level of service. There are many examples in transportation, such as the reliability of travel times due to traffic congestion, the breakdown of a bus, and an accident on the train system. In these circumstances, a fixed attribute level in a design experiment may be more realistically redefined to account for the expectation of variation (i.e. uncertainty). This issue has been addressed in a general way in the theoretical literature on risk and uncertainty, and only recently has there been a serious effort in designing uncertainty into a choice experiment.

Senna (1992) is one example of recent effects. He gave each respondent a set of five travel times for the same trip repeated five times, and allowed the levels to vary across the five trips. No variability implied certainty. The design had three levels of mean travel time, three levels of travel time variability and costs, with a modified rating scale of 5 levels. By treating uncertainty as an additional attribute with obvious links to another mean level attribute, the property of orthogonality becomes attractive in statistical estimation to enable identification of the role of uncertainty in choice response.

Introducing dynamics into stated preference model estimation

Some RP-SP choice models incorporate the endogenous choice variable from the RP choice set as an exogenous variable in the SP indirect utility expressions to represent inertia; which implies the presence of state dependence. Furthermore, it is well known in the econometric's literature that true state dependence is not transparent if (serial) correlation exists between the random components of the RP and SP indirect utility expressions. Morikawa et al. (1992) propose a method of handling stated dependence and serial correlation in a discrete choice modelling context and apply it in an intercity mode-choice context.

   Hensher (1993) proposes a panel-data approach for handling these correlated sources of potential bias in parameter estimates in repeated measures SP data when ratings data is used; drawing on the analogy to the time series of cross-sections data profile common in econometrics (i.e. a panel). The panel approach is only applicable to SP data except where the RP alternatives have been rated. This approach recognises that there are unobserved effects which are constant within an individual between replications. These can be defined either as fixed (ω_i) or random (μ_i) individual-specific effects. With a large sample, it is likely that a random-effect which can be allowed for by the inclusion of an order-specific fixed (τ_r) or random (η_r) effects variable. The panel data estimator for a fixed effect specification is:

$$R_{ir} = \alpha_0 + \beta' x_{ir} + \omega_i + \tau_r + \varepsilon_{ir}$$

and for a random-effects specification is:

$$R_{ir} = \alpha + \beta' x_{ir} + \mu_i + \eta_r + \varepsilon_{ir}$$

where R_{ir} is the rating response for replication r and individual, i, x_{ir} is the set of attributes from the experimental design and contextual/covariate influence, ε_{ir} is the residual error effect,[10] and α and the β' matrix define unknown parameters, to be estimated. τ_r and η_r can capture sources of replication bias including order effects μ_i and ω_i can capture inertia effects as well as other person-specific effects such as fatigue.

A final word

Stated preference analysis has come a long way. It is now widely accepted as a logical approach to extending the behavioural response space for studies of traveller behaviour and travel demand. The papers in this special issue illustrate the progress which has been made.

126

Notes

1. During the 1970's the field of travel behaviour was dominated by economists and econo-metricians, and engineers with an interest in the economist's approach to the study of travel demand. Geographers were beginning to contribute in the activity modelling area, but there was a noticeable absence of input from psychologists, sociologists, and marketing researchers. Economists were sceptical about experimental data (often referred to as "wish" data), indeed many economists still are, although a sub-discipline of experimental economics is now attracting interest from mainstream economists as they wrestle with the inadequacy of revealed preference methods in contexts which are not observed in the marketplace (see Madden (1992) for an overview).

2. The widespread use of the phrase "state preference methods" has evolved out of a number of earlier literatures under the titles of *information integration theory* (Anderson 1981, 1982), *functional measurement, direct utility assessment, functional analysis* and *controlled experimental design*. The SP terminology has been popularised in the transportation literature for what is generically referred to as *conjoint analysis* in the marketing and psychology literature. The SP nomenclature can be confusing when one distinguishes preference and choice experiments, which led this author to promote the phrase *stated response methods*. However, given the common reference to SP methods in practice in transportation, we will adopt it in this paper, with the occasional reference to stated choice (SC).

3. Historically, the traditions of a disciplinary specialisation and the availability of computer software had a significant influence on the selection of the response metric. Most notable is the division between the traditional conjoint approach (as it is called in marketing) which relies on ranks or rates and which estimates *separate* models for each and every sampled observation using essentially ordinary least squares regression (simulating the market shares in benefit segments, the latter defined by grouping on the predicted utility values associated with each individual); and the discrete-choice approach which uses stated choice or aggregated choice proportions across the replicated choice experiments to obtain sample estimates of parameters in a logit or probit modelling framework. In this paper, all discussion is in the context of models estimated on a sample of respondents. There are a number of software packages available for both preference and choice design develop-ment. For example, Bretton-Clark's Conjoint Designer, Sawtooth's Adaptive Conjoint Analysis, Hague Consulting Group's SPEED and MINT, Peter Davidson's SP-Ask, Steer, Davies Gleave's GAME GENERATOR and Intelligent Marketing's CONSURV.

4. Marketing researchers sometimes claim value in the lower ranked information as useful in identifying possible *detractor* attributes which need to be given special consideration in any marketing campaign to ensure that their offered levels in the marketplace do not deteriorate. Such attributes may be sufficiently sensitive dimensions of choice such that a small deterioration would detract from a product or service's appeal and result in a lower ranking.

5. Formally, let R denote an unobserved continuous rating variable $(-\infty < R < +\infty)$, and $\omega_0, \omega_1, \ldots, \omega_{J-1}, \omega_J$ denote the cut-off or threshold points in the distribution of R, where $\omega_0 = -\infty$ and $\omega_J = +\infty$. Define R* to be an ordinal (observed ranting) variable such that $R* = j$ if, and only if, $\omega_{J-1} \leq R \leq \omega_J$ $(j = 1, 2, \ldots, 5)$. Since R is not observed (but R* is observed), its mean and variance are unknown. Statistical assumptions must be introduced such that R has a mean of zero and a variance of one. To operationalise the model, we define a relationship between R and R*. Consider the likelihood of obtaining a particular rating value of R and the probability that R* takes on a specific rating value. If R follows a probability distribution such as normal with density function $f(R)$ and a cumulative density function $F(R)$, then the probability that $R* = j$ is the area under the density curve between

ω_{j-1} and ω_j. Formally this is given by (Winship and Mare 1984):

$$P(R^* = j) = \int_{\omega_{j-1}}^{\omega_j} f(R)dR = F(\omega_j) - F(\omega_{j-1}).$$

where $F(\omega_j) = 1$ and $F(\omega_{j-1}) = 0$.

For a sample of individuals for whom R^* is observed, we can estimate the thresholds ω_j as $\omega_j = F^{-1}(p_j)$, where p_j is the proportion of observations for which $R^* < j$ and F^{-1} is the inverse of the cumulative density function of R. Given empirical estimates of ω_j, we can obtain estimates of the mean of R for observations within each interval of the rating scale. If R is a standard normal, its mean for the sample for which $R^* = j$ is:

$$R_{\omega_j, \ \omega_{j-1}} = \frac{\phi(\omega_{j-1}) - \phi(\omega_j)}{\Phi(\omega_j) - \Phi(\omega_{j-1})}$$

where ϕ is the standardised normal probability density function and Φ is the cumulative standardised normal distribution function. Explanatory variables obtained from both the experimental designs and contextual data can be readily incorporated into the ordered response model in the usual manner that they are incorporated into a regression equation. If we assume that the randomly distributed error term is uncorrelated with the observed explanatory variables and its probability distribution is normal, then the ordered response model is referred to as ordered probit.

The implication of this distributional assumption is that the probability that R^* takes on successively higher values increases or decreases slowly at small values of the explanatory variables, more rapidly for intermediate values, and more slowly again at large values. A linear model assumes the probability that R^* takes successively higher values increases or decreases a constant amount over the entire range of the explanatory variables. The advantages of ordered probit over a linear (probability regression) model are greatest when R^* is highly skewed.

The LIMDEP package (Econometric Software 1992) automates ordered probit and ordered logit, making it easy to obtain parameter estimates for the design and socioeconomic variables, as well as the threshold parameters *The latter set of parameters indicate the extent to which the categories of the rating scale are equally spaced in the probit scale.* All parameters are obtained by iterative maximum likelihood estimation, using the Davidson-Fletcher Powell optimisation method with the first derivatives used to define the variance matrix. Starting values are obtained by the ordinary least squares regression on a binary dependent variable. Sample cell frequencies on the observed ratings are used to initially divide up the real line in order to define the starting threshold values on a normalised scale. The lowest threshold is normalised to zero.

6. An example of a hierarchical design of generic and elemental attributes with a design linking facility is given below:

A parsimonious set of elemental and generic attributes in five stated response designs

Generic attribute	Elemental Attributes	Levels
Wait quality	Waiting time at stop	5, 15
	Punctuality	0, 5
	Bus shelter	1, 2
Vehicle quality	Modernity	1, 2
	Seat material	1, 2, 3, 4
	Step height	1, 2
	Interior cleanliness	1, 2
	Leg room	1, 2
Trip quality	Time to get a set	1, 2, 3
	Time to board bus	1, 2, 3
	Express Service	1, 2
Information quality	Timetable	1, 2
	Destination signs	1, 2
	Source of timetable	1, 2

Generic attributes	Levels	
Fare level	25% less, same, 25% more	
Wait quality	VS, Ave, VD	
Vehicle quality	VS, Ave, VD	
Information quality	VS, Ave, VD	
Modal interchange	No, two buses	

Note: VS = very satisfied, VD = very dissatisfied, Ave = average

7. Three- (or more) way interactions are extremely difficult to interpret in a behaviourally meaningful sense, and are typically excluded. It is often stated in the literature that the majority of the variability in behavioural response is explained by main effects and a few two-way interactions. Louviere cites 80% plus for main effects and up to an additional 6% for two-way interaction (Louviere 1988).
8. Although in a main effects only plan, interaction effects between design attributes may not be permissible in a statistical sense, it is valid to interact a design attribute from the main effects plan with a non-design variable. For example interaction of the design attribute *travel time* and *income*. In a mode choice model, the value of travel time savings then becomes a function of income.
9. To appreciate the importance of an awareness of possible statistical correlation between main effects and interactions, assume a situation where we have 5 attributes each of 2 levels. This is a 2^5 design giving 32 possible alternatives. To reduce the number of combinations to a more manageable size without loosing essential information, an orthogonal main effects design can be developed from only a 2^3 factorial if we make certain assumptions about interaction effects. By assuming that some interaction effects are not significant, we can use these "interaction effects" to represent the two main effects which would otherwise

be missing from the 2^3 design. For example the 8 combinations in the 2^3 design are as follows (using orthogonal coding to represent the levels of each of the two-level attributes):

invehicle time (IVT)	wait time (WT)	walk time (WK)	IVT*WT (Invehicle cost – IVC)	IVT*WK (Parking cost – PC)	IVT*IVC (= WK)	IV*PC (= WT)
−1	−1	−1	+1	+1	−1	−1
−1	−1	+1	+1	−1	+1	−1
−1	+1	−1	−1	+1	−1	+1
−1	+1	+1	−1	−1	+1	+1
+1	−1	−1	−1	−1	−1	−1
+1	−1	+1	−1	+1	+1	−1
+1	+1	−1	+1	−1	−1	+1
+1	+1	+1	+1	+1	+1	+1
0	0	0	0	0	0	0

To include the cost attributes in the fractional factorial (i.e. the 8/32 or one-quarter fraction), we can use the 4th and 5th columns which represent interactions between two of the time attributes to construct the extra main effects IVC and PC only if the interactions IVT*WT and IVT*WK are equal to zero. It is obvious why these two interactions effects must be excluded: the interactions of the cost main effects with IVT are perfectly correlated with the main effects WT and WK, as shown by comparing columns 3 and 6 and columns 2 and 7. This is the underlying rationale for fractional factorial designs. If some interaction effects may be significant, you will have to construct a larger design such as a 2^4 factorial. A repeat of the process outlined above will enable you to identify the number of two-way interactions which are independent of the main effects. All pairs of interest can be evaluated to identify the number of permissable independent two-way interaction effects. In this example, with the one additional main effects attribute, all two-way interactions are orthogonal with the main effects.

10. A set of SP replications have some specific characteristics which if not handled properly can cause misleading inference. Although such data is not long enough in time to produce the possibility of different stochastic processes applying to different replications (i.e. heteroscedasticity of unequal variances associated with the unobserved influences); the application of ordinary least squares (OSL) regression is not typically valid. OLS treats the data as if it were a pooled set of independent replications with the classical statistical properties for the error variance-covariance matrix of constant variance and zero covariance between all pairs of replications (i.e. homoscedasticity). There are a number of specifications for the structure of the random components variance-covariance matrix. Commencing with (i) the OLS assumption of homoscedasticity, we can allow (ii) the variance to vary across the sample observations (i.e. cross-sectional or individual-specific heteroscedasticity). In addition (iii) we can relax the entire error matrix set and allow for free correlation between the individuals for the first replication together with the individual-specific heteroscedasticity. The only assumption imposed in (i)–(iii) is that (iv) the observations are uncorrelated over replications. By allowing for one-lag autocorrelation which is either (v) invariant with each individual or (vi) allowed to vary across individuals, we are able to evaluate important sources of model misspecification. A final set of models would be the outcome of evaluating nine combinations of error variance-covariance and autocorrelation. The Lagrange multiplier (LM) test, asymptotically equivalent to the likelihood ratio test, can be used to test the null hypothesis of homoscedas-

ticity, using a chi-square critical value. Heteroscedastic models use a feasible generalised least squares estimator (Greene 1990).

References

Adamowicz W, Louviere JJ & Williams M (1992) Combining revealed and stated preference methods for valuing environmental amenities. Working Paper, Department of Rural Economy, University of Alberta, Edmonton (unpublished).

Anderson NH (1981) *Foundations of Information Integration Theory*. Academic Press, New York.

Anderson NH (1982) *Methods of Information Integration Theory*. Academic Press, New York.

Bates J (1988) Stated preference methods in transport research. *Journal of Transport Economics and Policy*. XXII(1): 1–137.

Bates JJ et al. (1987) *The Value of Travel Time Savings*. Policy Journals, United Kingdom.

Batsell RR & Louviere JJ (1991) Experimental analysis of choice. *Marketing Letters* 2(3): 199–214.

Beaton WP, Meghdir H & Carragher FJ (1992) Assessing the effectiveness of transportation control measures: Use of stated preference models to project mode split for work trips. *Transportation Research Record* 1346: 44–1346.

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

Ben-Akiva ME & Morikawa T (1990) Estimation of switching models from revealed preferences and stated intentions. *Transportation Research* 24A(6): 485–495.

Ben-Akiva M, Morikawa T & Shiroishi F (1992) Analysis of the reliability of preference ranking data. *Journal of Business Research* 24(2): 149–164.

Borsch-Supan A (1986) *Econometric Analysis of Discrete Choice*. Springer-Verlaag, Berlin.

Bradley MA & Bovy PHL (1984) A stated preference analysis of bicyclist route choice. PTRC Summer Annual Meeting, Sussex (pp 39–53).

Bradley MA & Daly AJ (1991) Estimation of logit choice models using mixed stated preference and revealed preference information. Paper presented to the *6th International Conference on Travel Behavior*, Quebec, May 22–24, 1991.

Bradley MA & Daly AJ (1992) Use of the logit scaling approach in stated preference analysis. Paper presented at the *7th World Conference on Transport Research*, Lyon, July.

Bradley MA & Daly AJ (1992) Stated preference surveys. In: Ampt ES, Richardson AJ & Meyburg AH (eds), *Selected Readings in Transport Survey Methodology*. Eucalyptus Press, Melbourne (pp 31–35).

Bunch DS & Batsell RR (1989) How many choices are enough? The effect of the number of observations on maximum likelihood estimator performance in the analysis of discrete choice repeated measures data sets with the multinomial logit model. Working Paper, Graduate School of Management, University of California, Davis (unpublished).

Chapman RG & Staelin, R (1982) Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* 19: 288–301.

Crask MR & Fox RJ (1981) An exploration of the interval properties of three commonly used marketing research scales: a magnitude estimation approach. *Journal of the Market Research Society* 29: 317–339.

Davidson JD (1973) Forecasting traffic on STOL. *Operations Research Quarterly* 24: 561–9.

Econometric Software (1992) *LIMDEP 6.0*, Econometric Software Inc., New York and Sydney.

Elrod T Louviere JJ & Davey KS (1992) An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing Research* XXIX(3): 368–377.

Green PE & Srinivasan V (1978) Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research* 5: 103–123.

Green PE & Srinivasan V (1990) Conjoint analysis in marketing research: New developments and directions. *Journal of Marketing* 54(4): 3–19.

Greene WH (1990) *Econometric Analysis*. MacMillan, New York.

Gunn HF, Bradley MA & Hensher DA (1992) High speed rail market projection: survey design and analysis. *Transportation* 19: 117–139.

Hague Consulting Group (1988) A stated preference analysis of public transport service, station and vehicle improvements in Stockholm. Report prepared for Stockholm Transport, October 1988 (mimeo)

Hahn GJ & Shapiro SS (1966) A catalogue and computer programme for design and analysis of orthogonal symmetric and asymmetric fractional experiments. *General Electric Research and Development Centre Report No 66-C-165*, Schenectady, New York.

Henry F (1982) Multivariate analysis and ordinal data. *American Sociological Review* 47: 299–304.

Hensher DA (1982) Functional measurement, individual preference and discrete-choice modelling: theory and application. *Journal of Economic Psychology* II(3): 323–335.

Hensher DA (1986) Sequential and full information maximum likelihood estimation of a nested logit model. *Review of Economics and Statistics* LXVIII(4): 657–667.

Hensher, DA (1991) Hierarchical stated response designs and estimation in the context of bus use preferences. *Logistics and Transportation Reviews* 26(4): 299–323.

Hensher DA (1993) Treating the replications of a stated preference experiment as a panel specification. Institute of Transport Studies, Graduate School of Business, The University of Syndey (in preparation).

Hensher DA & Barnard PO (1990) The orthogonality issue in stated choice designs. In: Fischer M, Nijkamp, P & Papageorgiou Y (eds), *Spatial Choices and Processes* (pp 365–278). North Holland, Amsterdam.

Hensher DA & Battellino HC (1993) The use of discrete choice models in the determination of community preferences towards sub-arterial traffic management devices. Proceeding of the 7th World Conference on Transport Research, Lyon, France.

Hensher DA & Louviere JJ (1983) Identifying individual preferences for international air fares. *Journal of Transport Economics and Policy* XVII(2): 225–245.

Hensher DA & Stopher PR (eds) (1979) *Behavioural Travel Modelling*. Croom Helm, London.

Hensher DA, Barnard P, Milthorpe F & Smith N (1989) Urban tollways and the valuation of travel time savings. *The Economic Record* 66(193): 146–156.

Horowitz JL (1981) Sampling specification and data errors in probabilistic discrete-choice models, Appendix C. In: Hensher DA & Johnson LW (eds), *Applied Discrete-Choice Modelling* (pp 417–435). Croom Helm (London) and Wiley (New York).

Horowitz JL & Louviere JJ (1993) Testing predicted probabilities against observed discrete choices in probabistic discrete choice models. *Marketing Science* (in press).

Johnson LW (1990) Discrete choice analysis with ordered alternatives. In: Fischer MM, Nijkamp P Papageorgiou YY (eds), *Spatial Choices and Presses* (pp 279–289). North-Holland, Amsterdam.

Kocur G, Adler T, Hyman W & Audet E (1982) *Guide to Forecasting Travel Demand with Direct Utility Measurement*. UMTA, USA Department of Transportation, Washington D.C.

Kroes E & Sheldon R (1988) Are there any limits to the amount consumers are prepared to pay for product improvements? Paper presented at the PTRC Annual Summer Meeting, July.

Lerman SR & Louviere JJ (1978) On the use of functional measurement to identify the functional form of the utility expression in travel demand models. *Transportation Research Record* No 673 (pp 78–86).

Louviere JJ (1979) Attitudes, attitudinal measurement and the relationship between attitudes and behaviour. In: Hensher DA Stopher PR (eds), *Behavioural Travel Modelling*, Croom Helm, London, 782–794.

Louviere JJ (1988) Analysing Decision Making: Metric Conjoint Analysis. Sage University Paper No. 67, Newbury Park, Beverly Hills.

Louviere JJ (1992) Special issue on experimental choice analysis. *Journal of Business Research* 24(2): 89–189.

Louviere JJ (1993) Conjoint analysis. In: R Bagozzi (ed), *Handbook of Marketing Research*. Academic Press, New York.

Louviere JJ & Gaeth GJ (1987) Decomposing the determinants of retail facility choice using the method of hierarchical information integration: a supermarket illustration. *Journal of Retailing* 63: 25–48.

Louviere, JJ & Hensher DA (1982) On the design and analysis of simulated or allocation experiments in travel choice modelling. *Transportation Research Record* 890: 11–17.

Louviere JJ & Hensher, DA (1983) Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, 10(3), 348–381.

Louviere JJ & Kocur G (1983) The magnitude of individual level variation in demand coefficients: a Xenia, Ohio, case example. *Transportation Research* 17A: 363–374.

Louviere JJ & Timmermans H. (1990) Stated preference and choice models applied to recreation research: a review. *Leisure Sciences* 12: 9–32.

Louviere JJ & Woodworth GG (1983) Design and analysis of simulated choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research* 20: 350–367.

Louviere JJ, Meyer R, Stetzer F & Beavers LL (1973) Theory, methodology and findings in mode choice behaviour. Working Paper No. 11, The Institute of Urban and Regional Research, The University of Iowa, Iowa City.

Louviere JJ, Hensher DA & Shocker A (1992) *Conjoint Measurement: A Short Course*. Course offered annually in Australia and the USA.

Madden GG (1992) Social experimentation in economics: an overview of the stated-preference experimental design method. Paper presented to the Fourth Conference of the Australian Centre for Experimental Economics, University of Adelaide, 19–21 June.

Mason CH & Perreault WD (1991) Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research* XXVIII: 268–280.

McFadden D (1981) Econometric models of probabilistic choice. In: Manski CF & McFadden D (eds), *Structural Analysis of Discrete Data* (pp 198–272). MIT Press, Cambridge.

Meyer RJ, Levin IP & Louviere JJ (1978) Functional analysis of mode choice. *Transportation Research Record* No. 673 (pp 1–7).

Morikawa T (1989) Incorporating Stated Preference Data in Travel Demand Analysis. PhD Dissertation, Department of Civil Engineering, M.I.T.

Morikawa T, McFadden D & Ben-Akiva ME (1990) Incorporating psychometric data in econometric travel demand models. Kyoto University Working Paper, Department of Civil Engineering, University of Kyoto, Japan.

Morikawa T, Ben-Akiva ME & Yamada K (1992) Estimation of mode choice models with serially correlated RP and SP data. Paper presented at the 7th World Conference on Transport Research, Lyon, July.

Ortuzar J de Dios & Garrido RA (1993) On the probabilistic interpretation of semantic scales in stated preference rating experiments. Department of Transport Engineering, Pontifica Universidad Catolica de Chile (unpublished).

Pearmain D, Swanson J, Kroes E & Bradley M (1991), *Stated Preference Techniques: A Guide to Practice*, 2nd edition. Steer Davies Gleave and Hague Consulting Group.

Senna L ADS (1992) Traveller's willingness to pay for reductions in travel time variability. Paper presented at the 7th World Conference on Transport Research, Lyon, July.

Swait J Louviere JJ (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* (in press).

Timmermans, HJP (1988) Hierarchical information integration applied to residential choice processes. Paper presented at the Annual Meeting of the American Association of Geographers, Phoenix, Arizona.

Wardman, M. (1988) A Comparison of revealed and stated preference models of travel behaviour. *Journal of Transport Economics and Policy* XXII(1): 71–92.

Winship C & Mare RD (1984) Regression models with ordinal variables, *American Sociological Review* 49: 512–525.