I T S

Combining Sources of Preference Data

By

Jordan J. Louviere
David A. Hensher

July, 2000

INSTITUTE OF TRANSPORT STUDIES

The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University

NUMBER:	Working Paper ITS-WP-00-16

TITLE:	Combining Sources of Preference Data

ABSTRACT:	Mixtures of revealed preference and stated preference data are recognised by transportation researchers as offering a richness of behavioural input to travel choice modelling that is often absent from the isolated use of each data source. Accumulating evidence from various literatures, especially in marketing, psychology and transportation, provides support for the desirability of combining sources of preference data as a way of transferring increasing power of understanding of travel behaviour from the econometrics of a model to the underlying data inputs. Together with advances in the specification and estimation of discrete choice models, we are beginning to see that the simpler choice models such as multinomial logit (MNL) deliver an amazing amount of behavioural power providing that the underlying data specification is given a statistically rigorous treatment. There still remains however a growing role for more general choice models as a way of establishing the relevance of the simpler MNL model. This paper reviews the major contributions to the literature on combining sources of preference data and suggests new directions for fruitful research.

AUTHORS:	Jordan J. Louviere David A. Hensher,

CONTACT:	Institute of Transport Studies (Sydney & Monash) The Australian Key Centre in Transport Management C37, The University of Sydney NSW 2006 Australia

	Telephone:	+61 2 9351 0071
	Facsimile:	+61 2 9351 0088
	E-mail:	itsinfo@its.usyd.edu.au
	Internet:	http://www.its.usyd.edu.au

DATE:	July, 2000

Introduction

Stated Preference (SP) and Revealed Preference (RP) data are the outcomes of research design processes that involve decisions made about one or more preference elicitation procedures (PEP's). In the case of RP data, decisions are made about what travel choices and behaviours will be observed and recorded, in what ways, and from what populations. Similar decisions are made about the design and collection of SP data, the major difference being that RP data typically relate to actual choices and/or behaviours in real transport environments, while SP data typically relate to travel choices and/or behaviours in hypothetical transport environments.

The purpose of this paper is to discuss pooling of preference data sources. Our focus is on what sources of data can be used for what purposes, and how to combine them to enrich our understanding of travel choices and behaviours and develop more accurate and policy-relevant models. To accomplish this we develop a conceptual overview of the consumer's decision process. We also focus on the design and analysis of stated preference experiments and pooling SP data with revealed preference data and discuss combining mixtures of SP data as well as combining mixtures of RP data. The challenging issues of design complexity, survey length and number of profiles are discussed. We conclude the paper with a list of questions that can usefully position the ongoing research agenda.

To begin the discussion, consider two hypothetical but realistic consumer cases that serve to introduce our conceptual discussion:

1. John is a 30 year old single male currently employed at location X, residing at location Y. X and Y are about 12 kms apart, and are well-served by bus and commuter train as well as the road network. John typically commutes to and from work by train, which takes about 5-7 minutes to walk to the origin station, 12-17 minutes of in-vehicle time and another 2-3 minutes walk to the workplace from the destination station. About once per week John takes a taxi to a shopping centre near location Y to shop, drop clothes off at cleaners, etc. after work, and also eats out before walking home. John has been offered a new job at a substantial increase in pay and responsibility. The new position is in an outer suburb some 15kms from location Y that is poorly served by public transport. John decides to take the new job.

2. Sue is a female single parent with a 13 year-old son who lives at location Q and works part-time at location R that are about 3 kms apart. Sue's 13 year-old son, Jack, travels by bus to school at location S, 4kms from location Q and 5 kms from location R. Jack is a promising athlete who is being recruited by a high school located at V, 7kms from Q and 7kms from U. Sue is completing an MBA degree part-time two evenings per week at location U about 7kms from Q and 8 Kms from S. Sue finishes her MBA in two months, and has decided to do a PhD at U.

John's new job motivates him to buy a car to commute to/from his new workplace, and after two weeks of driving in heavy traffic for 45 minutes each way, he decides to find a new residence closer to X. Sue begins work on her PhD at U and Jack starts high school at V; after several months of driving Jack to V and picking him up after football practice, Sue decides that she has to move to deal with the demands of her new situation, and she

chooses a new location T that has good public bus service to both U and V. Note that changes to John's and Sue's personal circumstances trigger decisions that involve place of work, place of residence, mode of travel, and activity patterns. Some of these decisions are relatively immediate, like choice of residence, but others arise out of new experiences. In both cases, there also are likely to be on-going travel and behavioural consequences that are not observed in the brief scenarios above.

Most past research *combining* SP and RP data has focused on single outcomes such as choice of mode, choice of shopping destination or choice of residence (for example, Ben-Akiva and Morikawa 1990, Hensher and Bradley 1993, Bradley and Daly 1997, Hensher 1998, 1999, Brownstone and Train 1999). But some of these choices co-occur, and/or give rise to second, third and higher order choices and consequences. For example, now that John has a car, his shopping activities increase from a multi-purpose, one time per week activity to a number of single and multi-purpose activities per week, and his recreational choices patterns also change. Sue's departure times change because of class schedules and somewhat greater flexibility outside of class and meeting times at university; her choice of mode also changes as does Jacks's (from auto-passenger to bus). Her new location significantly reduces her travel activities and travel times, which in turn allows more leisure time during the day and on weekends, so she increases the number of single and multi-purpose trips for shopping and entertainment. Like John, Sue's travel choices and behaviours also are likely to change and evolve over time as a consequence of her recent choices. These interactions are well-recognised in the broader transportation literature and have spawned an interest in activity modelling (as reviewed by Bhat 1997, Bhat and Koppelman 2000, Golob 2000). However, to date the literature on pooling preference data has been isolated from the activity paradigm.

The two cases presented above illustrate limitations of conventional stand-alone SP-RP studies. Both SP and RP projects tend to be cross-sectional snapshots at one point in time (or a few weeks), and there is little longitudinal or panel data available using either preference elicitation approach (the commitment to RP panels in the 1980's seems to have waned, even if the research enthusiasm is still alive – see Golob et al. 1997). The behaviours and outcomes illustrated in the two cases suggest that much behaviour of interest is conditional on antecedent conditions and/or behaviours, and it is difficult to gain much insight into these types of processes without longitudinal observations. Moreover, external to the two cases environmental changes occur at different time scales, such as the rapid changes to communications, human interactions and commerce provided by the Internet and other interactive channels, or changes to urban environments like large-scale transport and housing infrastructure projects under way in many cities. Thus, these cases help us understand and appreciate some of the shortcomings of present practice.

It is recognised that longitudinal panels are very capital and resource intensive, and require special attention to issues like panel attrition (see Golob et al 1997). Additionally, it has proven difficult in the past to get transport researchers to agree on basic transport and travel measures to be used and/or observed in SP and RP projects, much less agree on measuring them the same way to ensure maximum comparability across space and time (eg, Louviere 1988b). We now turn our attention to some key conceptual considerations that bear on data pooling.

Conceptual Considerations as a Flag for New Mixtures of Data

The introduction points us in the direction of a conceptual framework, although the framework we outline below is necessarily incomplete and should be viewed as a tentative first sketch. As has been long recognised, the demand for transportation is typically a derived demand, consequent to satisfying some other need or purpose. We propose that a number of key events act as "triggers" to increase the probability that a consumer will make particular types of travel and related choices. Some of these triggers include changes in the household life cycle, such as births, deaths, separations, marriages, etc., while others include events such as job offers, accumulated or unanticipated gains or losses in income or wealth, gentrification of suburbs, changes in overall housing or travel prices, etc. It should be possible to develop a reasonably comprehensive and exhaustive list of the major triggers or drivers, and begin to observe these in future travel data collection efforts. What may be more difficult is modelling the state in which any particular traveller might be at a particular time, where a "state" refers to some cross-classification of initial household conditions based on the above types of drivers. Nonetheless, it should be possible to observe each traveller's initial state reasonably easily once a classification is developed, which would permit one to model choices and behaviour conditional on that; these data could be updated reasonably often if need be without great expense (eg, phone surveys).

Travel behaviour researchers continue to recognise the need for a much more comprehensive classification of travel choices and behaviours than simple mode or destination choices. Through integrated land use and travel model systems we strive to understand what other choices are linked to these decisions and follow from them (Hensher 2000, Waddell 2000), and we seek to identify and begin to observe second, third and high-order choices that follow from initial choices like choice of residential or school locations. Some of these choices and behaviours will be immediate, whereas others will be longer term. For example, if we provide new LRT facilities and link them to purpose-built new urban developments, we can observe how choices of LRT depend on the way in which we design and configure the system and associated services, we can observe choices of the new residential development and we can observe the joint choice of the residential development and LRT. However, a choice to live in the new development will give rise to a variety of other linked choices, such as possible changes in jobs, shopping locations, entertainment patterns, frequencies of visiting friends and relatives, etc. If we want to plan large-scale infrastructure projects many years into the future and forecast the associated behavioural changes we have to understand how these changes depend on one another and on the choices and behaviour of other actors such as developers, transport planners, local and higher-level political bodies, etc. While none of this is new, transportation planners continue to search for ways of capturing even a minimal amount of the response matrix.

Cross-sectional data sources provide little insight into such dynamics, and generally speaking we lack a comprehensive behavioural framework to guide data collection. If it could be developed, such a framework should provide better and richer insights than are presently possible. However, longitudinal data collection is very expensive and resource intense, hence this is likely to be less feasible in the short-run, despite the fact that it

offers insights into the dynamics of behavioural, household and environmental changes (Golob et al 1997).

One possible way to proceed would be to try to develop SP approaches that can simulate some of the dynamic changes, and combine the outputs of these studies with available RP data. It has yet to be established whether SP methods can simulate household changes, such as asking consumers to imagine changes in their personal and household circumstances like job changes, income changes, family composition changes and the like, and it may be that SP cannot be used to simulate these types of changes successfully. However, SP approaches do appear to be able to simulate environmental changes, such as changes to urban structure, transport systems, and policies. In turn, this suggests that a fruitful area of research would be to identify ways to pool cross-sectional and longitudinal RP data that can provide insights into changes in household circumstances with SP data that can provide insights into the likely effects of certain types of environmental changes. It also may be the case that we can develop SP approaches to simulate the likely choices that would be made by other actors in the system, such as developers, planners, etc., and/or develop ways to design and implement interactive agent experiments to better understand and predict how choices made by key actors in the system depend on the choices made by other key actors and consumers (eg, Brewer and Hensher 1999). We now turn our attention to a discussion of issues related to the design of SP experiments that can provide such insights and how the data resulting from such experiments can be pooled with RP data to provide new insights and potentially more useful models.

Design of SP Experiments

Despite early leadership in the design and analysis of SP experiments and surveys (see Louviere 1988b and Hensher 1994), transport researchers and practitioners seem to lag behind their colleagues in other fields, most notably marketing and environmental and resource economics. Like the conclusions reached by the Arrow-Solow committee about the state of valuation methods in environmental and resource economics that followed the Exxon Valdez oil spill, SP work in transport appears to have made relatively limited real progress in recent years (despite a proliferation of empirical studies), and much academic and practical SP research seems to be overly influenced by opinion instead of empirical evidence, especially the design and implementation of SP choice experiments. For example, much "conventional wisdom" in transport suggests analysts cannot and/or should not undertake complicated or lengthy SP exercises (Stopher and Hensher 2000). Unfortunately, there is little agreement on what "complicated" or "lengthy" means, and worse yet, there is little empirical evidence to support this conventional wisdom. For example, Louviere, et al (1993) undertook an extensive literature review to determine how survey length and complexity impacted decision-making and choice behaviour, and found surprisingly little real empirical evidence on the issue. Indeed, subsequent research has suggested that:

1. consumers will evaluate many more *choice sets* and/or scenarios than previously believed,
2. consumers will evaluate many more *attributes* and choice *alternatives* than previously believed; and most importantly,

3. increasing length and/or survey complexity primarily appear to impact the variance of the random component, not the mean utilities (Brazell and Louviere 1995, Uldry, et al 1998, Bunch, Brazell and Louviere 1998, Deshazo and Fermo 1999, Johnson and Orme 1996).

Hensher (2000a) has recently shown that increasing the number of treatments in blocks of four from 4 through to 8,12 and 16 in the context of three alternatives and six attributes had no systematic influence on the mean value of travel time savings.

The foregoing is a non-trivial issue insofar as it bears on the ability of SP to simulate complex, multi-stage and conditional behaviours and choices. That is, simplistic SP experiments involving 3-4 attributes and two or three choices cannot possibly provide insights into long- and short-term choices made in order to cope with imposed transport policies. Thus, if we impose road tolls, congestion charges or the like, this may impact mode, route and destination choices in the short-run, but also may impact residential and job choices in the longer term, as well as trip timing and vehicle ownership decisions in the moderate term. Transport SP surveys typically focus on one type of choice outcome (eg, choice of mode), but there seems to be little a priori reason why a series of choices cannot be elicited in response to one or more scenarios. The authors have participated in the design and analysis of several SP surveys that elicited multiple choice outcomes per scenario and/or hierarchical or conditionally structured choices in both the USA and Australia (eg, Hensher 2000). Moreover, it is worth noting that many transport decisions and choices in fact *are* complex; hence it is unclear why one would simplify tasks in surveys when the real choices to be simulated are complex. Indeed, simplifications omit relevant variables and/or fail to capture the full range of choices or associated decisions. This is a very real challenge for transportation researchers who appear inclined to select applications where the simplifications needed to make an SP experiment realistic are achievable. While laudable, it limits the potential scope for the SP paradigm.

Much has been written about the design of SP experiments (eg, Louviere and Hensher 1982, Louviere and Woodworth 1983, Louviere 1988a,b, Louviere et al. 2000), so we will not revisit previous work. Instead, it suffices to note that choice experiments remain one of the most powerful and flexible forms of preference elicitation, and in general one typically can design choice experiments to simulate choice situations faced by consumers as closely as possible, including at the limit actually designing real choice situations in real markets. Naturally, the more closely one wants to simulate reality, the more time and resources typically are required, but this varies greatly depending on the research objectives and associated experimental requirements/conditions to satisfy them. At the present point in time we understand a great deal about model identification and the experimental design requirements to satisfy this property of choice experiments (Louviere et al. 2000), but we understand far less about the statistical efficiency of choice experiments in general or for individual applications in particular. We now consider the latter issue to show that statistical efficiency cannot be separated from respondent efficiency.

The statistical efficiency of choice experiments is as much a behavioural as a statistical issue. In particular, unlike the objects of analysis in many classical statistical experiments, and indeed in the mathematical and statistical theory that underlies design optimisation,

humans _interact_ with choice experiments in ways not previously considered by transport analysts.

The multidimensional nature of the random component

Consider the familiar classical expressions in random utility theory for the utility and choice probability of the i-th choice option (McFadden 1981, 1986):

$$U_i = V_i + \varepsilon_i \tag{1}$$

$$P(i|C) = P[(V_i + \varepsilon_i) > (V_j + \varepsilon_j)], \text{ for all j options in choice set C} \tag{2}$$

In much of the theoretical and empirical work undertaken in the random utility theory (RUT) paradigm the ε's are a unidimensional component associated with each choice option. This view of the ε's has obfuscated the fact that these random components of utility are multidimensional. That is, the ε's are better thought of as variance components that include variation within-subjects, between-subjects, variation due to measurement instruments, etc. Unless a given empirical study is designed to separate these components, the data and model outcomes will be confounded with the variability and cannot be separately identified. Of these components, there has been recognition of between-subjects variation especially in the form of preference parameter heterogeneity, but less attention has been paid to the fact that there also can be between-subjects variation in model forms (eg, see Kamakura et al 1996).

Indeed, research in the RUT paradigm in transport has paid little attention to within-subject, measurement instrument and other sources of variation (exceptions include Ben-Akiva and Morikawa 1998; Daniels and Hensher 2000; Morikawa 1994). There has been attention paid to design optimisation, particularly optimisation of parameter efficiency, but it has not been recognised that one cannot optimise choice experiments without understanding and incorporating the likely effects of the experiment itself on all components of variation that matter. For example, to the extent that task complexity is an issue, one well could argue that the issue is its likely impact on within- and between-subject variability. That is, as task complexity increases, all else equal, one expects within-subject variability to increase, as shown by Deshazo and Fermo (1999). Other aspects of choice experiments also can impact between-subject variability, and hence overall design efficiency, such as using wide or narrow ranges of levels of important numerical attributes like fares and travel times. That is, choice of levels can impact both within- and between-subject variability, as we now discuss.

The Range of Attribute Levels

The wider the range of levels of key numerical attributes like fares and times, the easier it is for subjects in SP experiments to find levels that can be easily identified as "high" or "low." Moreover, the wider the range of levels, the more likely it will be that more subjects agree that some levels are "high" whereas others are "low." Thus, the more easily subjects can identify extreme levels, the more likely they are to respond to them more consistently, which reduces within-subject variability. Similarly, if more subjects

agree that extreme levels are extremes, between-subject variability should decrease. However, the latter two variance outcomes also can be offset by the fact that more extreme levels may induce subjects to behave more extremely, thereby accentuating between-subject differences.

These effects are likely to be non-linear such that overall variability is low at the extremes and higher in the middle range. Likewise, if there are two or more ways to elicit preferences given a particular experiment, it is likely that some will differ in the amounts of within- and between-subject variability that they elicit. Thus, the key message is that choice or response variability is a *behavioural phenomenon*, and is an outcome of a choice experiment as much as observed choices and/or model preference parameters or specifications. Failure to recognise that experiments can produce different impacts on the behaviour of random components may lead to misinference, incorrect interpretations and/or possibly even biased results. Hence, it is not possible to optimise choice experiments a priori without also taking the behaviour of the random component into account as an outcome of the design and experiment.

Hensher (2000c) found in a study in New Zealand of long distance car travel that there is very strong evidence that the range of the travel time attribute (after controlling for time and cost heterogeneity) has a statistically significant influence on mean VTTS, increasing as the range narrows. By imposing a range restrictions varying from 10% to 100% of the existing range (ie a mean from 14.9 to 298.68 minutes) the mean VTTS varied from $20.41 to $6.92. At present, despite the evidence on sensitivity to attribute range, there appears to be no 'magic' formula to establish a behaviourally optimal attribute range (see Toner et al 1998, Frisch 1972, Saelesminde 1999). .

A Formal view of Behavioural Variance

Let us now develop a more formal view of this process. As far as we are aware, regardless of the distributional assumption one makes about the random component(s), the scale (denoted by λ) of the estimated model parameters will be inversely related to the amount of variability in the random component. Lambda scales the model parameter vector as follows:

$$V_i = \sum_k \lambda \beta_k X_{ki}, \tag{3}$$

where V_i is the systematic or explainable component of utility that can be observed from a properly designed and executed data collection or survey exercise, β_k is a generic k-element vector of preference parameters associated with each of the k elements of the design matrix for option i, X_{ki}. λ "scales" the β_k vector in the sense that the estimated values of β_k are the quantities $\lambda\beta_k$, but λ cannot be identified separately in any one source of preference data.

For example, in the case of the multinomial logit (MNL) model, λ is related to the variance of the random component as follows: $\lambda^2 = \pi^2 / (6\sigma^2)$, where π is the natural constant, and σ^2 is the variance of the random component. Thus, λ is inversely proportional to the variance of the random component; and as the variance increases, the magnitude of the resulting β_k estimates decreases, or as the variance decreases, the

magnitude of the β_k estimates increases. Thus, the variance of the random component plays a critical role in inference and model comparisons.

In the case of inference, λ, or equivalently the random component variance, impacts the magnitudes of the estimated parameters, and hence plays a direct role in the quality of inference because the more response variability, the smaller the estimates and hence the less likely one is to find significant model effects. In the case of model comparisons, the variance of the random component can differ between data sources, and hence one cannot assume that the constant variance assumption holds between data sources. Indeed, response variability can differ within data sets between alternatives and/or choice sets, and this variability may be larger than the variability between data sources (Hensher 1998). The latter matters because if two preference data sources to be compared do not have a common level of variability, one can mistakenly infer that β_k estimates differ when in fact random component variances differ. Thus, separating data source differences due to model parameter differences from differences due to random component variability differences is critical to be able to combine and properly compare sources of preference data.

In fact, the notion of preference data source comparisons was pioneered in transport. Morikawa (1989) proposed that if the only difference between an SP and an RP data source was the level of random component variability, one could pool the two sources of data and compare them by estimating a ratio of λ's, which imposes a restriction on the estimation that the parameters in the two data sources be proportional. In a comparison of inter-city mode choice RP and SP data sources in the Netherlands, Ben-Akiva and Morikawa (1990) demonstrated that one could not reject the parameter restriction; that is, the RP and SP parameter vectors were proportional. Since that time, there have been a number of such comparisons, and the empirical record suggests that model parameter proportionality appears to hold to a close first approximation in many cases (eg, see Louviere, et al. 1999).

Relatively few transport researchers (see Ben-Akiva 1987 as an example of an exception) have recognised that the same logic used by Morikawa (1989) to deduce the proportionality relationship between RP and SP model parameters also can be used to compare any sources of preference information, not just RP and SP sources. That is, the theoretical logic of data pooling and model comparison can be used to compare SP with SP sources, SP with RP and RP with RP. Swait and Louviere (1993) discuss this more general context and develop a simple way to compare and test for parameter proportionality in two preference data sources. Louviere et al. (2000) discuss conjoint and choice experiments in the broader context of preference elicitation methods generally, and show how to pool and compare preference data from many sources.

Data Pooling Rationale

The theoretical rationale for combining data or pooling can be summarised as follows:
  1. Each source of preference data contains elements that are common to other sources of preference data as well as elements that are unique to that source.

2. We denote common elements or attributes X_{ks}, and specific elements as Z_{ls}, where k indexes elements of the common parameter vector for data source s, and l indexes elements of the unique parameter vectors for data source s.

3. Each common and data source-specific attribute vector has associated model parameters, respectively β_{ks} and θ_{ls}.

4. The utility of the ith choice option can be expressed as a function of the common- and data source-specific elements using a linear-in-the-parameters specification involving common and data source-specific error components.

5. Preference invariance holds among sources of preference data if two or more sources of preference data exhibit proportional common utility parameters.

Louviere et al. (2000) provide a conceptual basis for pooling and comparing sources of preference information that also allows analysts to deal with the fact that some elements of data sources provide common information while other data source elements provide unique information about preferences and choices. In turn, this conceptual framework provides a number of important insights about modelling preferences and choices:

1. Because of the critical role played by the variance of the random component, it behooves all academics and practitioners to spend as much time as possible in advance of going into the field to collect data to understand the key drivers of choice, *the choice process itself* and anything else that may impact on behaviour associated with the phenomenon of interest. That is, investing time and resources to gain insight and understanding in advance serve to decrease the size of the variance of the random component, and hence benefit inference and model development. Similarly, simplistic tasks involving few choice options or choice attributes are likely to greatly increase random component variability due to omitted effects and/or model misspecification and/or overly inflate the role of alternative-specific constants.

2. Failure to take random component variance differences into account when comparing sources of preference data or models can lead to mis-inference and biased results. Of particular concern would be confounding differences in model parameters with differences in random component variance magnitudes.

3. *It is not possible to optimise choice or preference experiments in the absence of understanding how designs and tasks impact the variance of the random component.* For example, experiments that use comparisons of similar or comparable choice options (so called "utility balanced" experiments) are likely to significantly increase within-subject variability, and hence may be far less statistically efficient that other designs.

4. Many previously published results involving comparisons of model parameters that did not take random component variance differences into account should be reconsidered in light of the critical role played by the random component in model comparisons.

Response variability also can be associated with risk and uncertainty. For example, it is likely to be the case that in many markets, the probability that consumers will switch from their present choice options is significantly impacted by the hassle of switching and the uncertainty of whether they will be better off if they switch. In many decisions relevant to transport analysts and planners, switching hassles and option uncertainty are likely to act as significant deterrents to switching, and may be confused with habit persistence or serial correlation in choices. Thus, a consumer may not particularly like

where they live or the way they commute to work, but there are significant time and cost implications of investigating alternatives; and as is the case with many experience and credence goods, one often cannot know whether and if one is better off without many repeated experiences. Thus, low within-subject choice variability associated with repeated choices of the same option may reflect perceived or actual hassles and uncertainties associated with switching and not true unconstrained preference revelations.

Decomposition of the Variance of the Random Components

To the extent that there is external information available, it also is possible to decompose the random component variance *within a data source*. For example, Swait and Adamowicz (1997) develop a number of measures of choice set and task complexity and show how one can decompose and compare choice set-to-choice set variability in responses by selecting one choice set as a reference and parameterising the variance ratios of all choice sets relative to the reference as a function of their complexity measures. Deshazo and Fermo (1999) also demonstrate that this can be done in two environmental valuation data sources from Costa Rica and Guatemala. More generally, the variance of the random component can be decomposed and parameterised by designing data collection efforts or SP experiments a priori to allow the components to be specified. For example Dellaert et al. (1999) demonstrate that one can design an experiment to manipulate absolute levels of price and levels of price differences as a framework within which an SP experiment is developed, and how to decompose the random component variance and parameterise it as a function of these factors. The key message from this is that one CAN identify, decompose and parameterise random component variance components within and between data sources by designing research that makes it possible. Alternatively, one also can attempt to test hypotheses and/or parameterise the variance of the random component a posteriori by developing and testing measures that are hypothesised to explain the differences as illustrated by Swait and Adamowicz (1997).

With some exceptions (eg Morikawa 1994, Kim 1998, Daniels and Hensher 2000, Owersloot and Rietveld 1996), much past SP modelling has ignored serial correlation (or repeated observation effects) in responses. Like the previous discussion of repeated choices that can exhibit habit persistence, SP choices can exhibit correlations over repeated scenarios. For example, in a study of decisions to re-screen for breast cancer (Gerard et al. 1999), serial correlation was found to play a very significant role and was confounded with effects hypothesised to impact the mean propensity to say "yes" (the model intercept). A single serial correlation effect accounted for more variation than 10-12 statistically significant demographic and personal characteristic measures, and once the serial correlation was taken into account, these measures were no longer significant. Indeed it may be easy to confuse serial correlation for model intercept effects in choice models because serially correlated choices impact the intercepts, and more work is needed to determine whether the insights into the propensity to choose options should be sacrificed in favour of simply estimating a serial correlation effect. This needs to also recognise the role of RP observations when jointly estimated with SP treatments. Daniels and Hensher (2000) also investigated the effects of serial correlation on the estimates of the value of travel time savings (VTTS) from SP models, and found that failure to take serial correlation into account resulted in significantly higher values of VTTS.

Of the components of variance previously discussed, between-subject response variability has been the most researched, at least that portion of the component due to unobserved heterogeneity (eg Bhat 1997, 1997a, Munizaga et al 1997, 1999, Louviere et al 2000). A number of approaches have been proposed and applied to deal with this problem, including random coefficients specifications (Revelt and Train 1996, Hensher 1999), mixed logit (Train 1997, Brownstone and Train 1999), latent segmentation (Swait 1994), hierarchical Bayes estimation (Wedel, et al 1999), multinomial probit (Bolduc 1992) and/or embedding hypothesised individual difference measures in model intercepts and parameters (Ben-Akiva and Lerman 1985, p.). As previously noted, in the case of SP surveys with multiple scenarios, serial correlation has been shown to account for some demographic effects interacted with model intercepts (Shanahan et al. 1999), so there is a clear need to begin to separate these effects in data sources and determine the proper antecedent links.

As McFadden (1986) noted, there is little reason to estimate separate effects for individuals in most applications because the distribution of preferences can be specified in other ways, such as imposing joint multivariate distributions on parameters and estimating the parameters of that distribution or specifying the distribution parameters to be a function of covariates. While the latter approach is appealing conceptually, little research is available on sample sizes and data properties needed to reliably and efficiently estimate the parameters of such choice models; and as noted in Louviere, et al. (1999), it is likely that as the number of choice options and attributes increase, sample size requirements increase exponentially, such that even moderately sized estimation problems may be infeasible with current technology. The latter considerations bear on data pooling and model comparisons in which one may wish to compare and test hypotheses about parameter proportionality for variance-*covariance* matrices as well as parameter vectors.

Optimal Designs

Optimal designs for SP experiments remain elusive, as previously discussed. Present practice focuses on identification, but there are few results for efficiency even in the case of paired comparisons (Street et al 1999). Apart from the lack of results for design efficiency, the issue of respondent efficiency previously discussed also bears on non-design aspects of SP experiments and surveys. That is, framing, formatting, survey length, survey complexity, types of response questions, and the like, can impact both response means and variances (Louviere et al 1999). Thus, we need to distinguish between differences in such task related factors associated with real response bias as opposed to random component variance. For example, the behavioural decision theory paradigm views consumer's decisions through a lens that suggests that consumers frequently resort to decision heuristics and often exhibit various forms of response bias (eg, Ben-Akiva et al 1999), leading to systematic violations of economic theory, or at least formal models of rational behaviour. This view can be challenged on many fronts, but as Louviere et al (1999) note, there is substantial evidence for model preference parameter stability in many areas, while at the same time there is now considerable and ever-increasing evidence that many effects previously reported in various literatures as being response mean effects, in fact are response variance effects.

Survey Length and Task Complexity

Many behavioural stories can be told about survey length effects, some of the most popular being that experimental subjects use early scenarios to learn tasks and develop decision rules, which they then try to apply consistently until some point at which the task becomes sufficiently lengthy that they grow bored or angry, become inattentive and so begin to produce biased responses. Brazell and Louviere (1999) examined several such "stories" in two separate studies in which both survey length and scenario order were systematically varied in between-subjects choice experiments involving canned soups and holiday tours to Mexican resorts. Survey length was manipulated by designing experiments involving 12, 24, 48 and 96 choice sets (soups) and 16, 32, 64 and 128 choice sets (tours), and using latin squares to vary presentation orders within the 12 and 16 choice set conditions. Subjects were randomly assigned to each of the choice set size conditions in both studies. Results show little effect on response rates due to survey length, although there is a significant negative trend. There were no significant differences in preference parameters estimated from the various conditions, but there were significant random component variance differences, but the random component variance differences were not systematically related to survey length. Similarly, there were no significant preference parameter differences due to choice set order (appearance in survey); there were random component variance differences between some orders, but there was no systematic relationship between order and variability. Louviere et al. (2000) review a number of such results and findings, so we state the key result, which is that in a large majority of cases investigated the primary impact of task-related factors appears to be on the variance of the random component, not on response means or model parameters.

This suggests that research attention should focus on identifying combinations of task-related factors that lead to lower random component variance outcomes. In the area of task complexity, Swait and Adamowicz (1997) and Deshazo and Fermo (1999) have initiated research programs that bear on these issues, and as previously discussed can demonstrate how certain measures of task complexity impact the variance of the random component. This is a good start, but much more work is needed. Importantly, task-related factors bear on data source pooling and model comparisons because they can lead to potentially large differences in response variability, which in turn can affect inferences in data source comparisons. The importance of this issue can perhaps best be illustrated with reference to the very many studies in environmental and resource economics and transportation research that have tested hypotheses about differences in willingness to pay (WTP) and willingness to accept (WTA) responses.

By far the prevailing paradigm for these comparisons is the one advocated by the Arrow-Solow committee, namely closed-ended (discrete choice, or yes/no) Contingent Valuation (CV) tasks. At the risk of oversimplification, such CV tasks are based on an elaborate description of a single resource scenario to which all subjects are exposed. Following exposure to the scenario, a single value is drawn from a uniform distribution of values of a payment vehicle (eg, some form of a tax) and subjects essentially agree to pay the amount for the resource or not in WTP tasks (eg, to remediate a damage, improve a resource, etc.). In WTA tasks, subjects are offered a payment to accept the resource (eg, to sell their house under imminent domain, accept more pollution, etc.). The resulting models typically contain an intercept and a single payment vehicle effect.

Comparisons of WTP and WTA have almost always resulted in significantly different outcomes for welfare calculations; however, *these tasks confound mean outcomes with random component variance outcomes* because there is only a single effect, namely the effect of the payment vehicle. If we recall that the scale of each data source (WTA and WTP) can differ due to random component variance differences, a problem with such comparisons becomes obvious. That is, in the case of a single effect parameter, an infinite number of variance-scale ratios can satisfy the parameter proportionality condition imposed by Morikawa's (1989) hypothesis. Hence, in the case of a single parameter one cannot pool data sources and compare model parameters because model parameter (mean) differences are confounded with variability differences.

Likewise, many behavioural decision theory experiments rely on comparisons of response means between conditions. For example, in so-called "attraction effect" experiments, subjects are offered a choice of a bus that has a high fare but fast travel time or one that is slow and inexpensive (eg, Simonson 1989). After subjects make their choices, they are shown a second scenario that offers them a choice of the previous bus options and an additional third option that is dominated by one of the two. It is claimed that such experiments demonstrate violations of the regularity condition of random utility theory based choice models, namely that the choice probabilities of the two original buses can not increase up when a third bus is added.

Unfortunately, many empirical results in this paradigm rely on aggregate choices, raising immediate concerns about between-subject variability. However, it is easy to see that such experiments can produce different random component variances in each condition, which would lead to differences in choice probabilities unrelated to mean utility differences. Because of failure to control for other effects, and generally poor design practice, little really can be said about the results of such experiments because of such obvious confounds. Nonetheless, much has been made of the results of such experiments, including claims of preference reversals, violations of regularity and the like (eg, Ben-Akiva, et al 1999). The theory and conceptual discussion in this paper should make it clear that little can be inferred from such experiments except that aggregate choice proportions change between conditions.

The Role of RP Data

There are two final issues. The first deals with missing or unavailable RP data, and the second deals with ill-conditioned or "dirty" RP data. In many cases RP data are not available to be pooled with SP data for joint estimation. For example, in all cases in which products or services are new to the world or to a particular geographical location in the case of transport, there will be no previous RP data in which choices of the new option can be observed. The degree to which this is a problem depends on the degree to which the attribute effects are predominantly generic compared to alternative-specific. The more alternative-specific the attribute effects the greater the problem, such that in the limit there are no RP data available to pool. Another version of this problem is associated with the extent to which choices of each option are driven by unique as opposed to common attribute effects. The more unique effects, the less information is provided by the RP data for pooling purposes.

In the case of ill-conditioned or "dirty" RP data, the issues are identification and collinearity. That is, market forces, technology or sampling considerations may lead to perfect correlations between the attributes of some alternatives (eg, the same fares are charged for the same journey lengths on competing modes), some values may not be observed in the RP data and/or the predictor variables (attributes, covariates) may exhibit high or extreme levels of multicollinearity. Adamowicz et al. (1994) provide an example of all of these problems in the context of a recreational fishing destination choice application. In particular, some RP options could not be identified separately from others because their measures were identical in every respect, and some attributes were so highly correlated that reliable estimates could not be derived from the RP data set alone. However, when the data were pooled, it was possible to identify all effects, and the hypothesis of parameter proportionality between data sets could not be rejected.

Taking Stock of the Challenging Research Questions

The interest in stated preference methods continues unabated; however a growing number of issues can be raised as researchers acquire more experience with the design of SP experiments and analysis of SP data. Some of these issues have been addressed in previous sections, but it is useful to set out the most challenging issues as a research agenda for the immediate future. We offer the following set of questions as the kernel of the ongoing research agenda:

1. Which response metric(s) in SP experiments (ie choose, rank, rate, best-worst) provide appropriate information to understand and predict travel behaviour?
2. What evidence is there that there is a limit to the number of attributes and the number of alternatives that individuals can evaluate meaningfully in arriving at a response?
3. What evidence is there that the number of design treatments (or profiles) should be capped for whatever reason (eg fatigue, redundancy)? And if there should be a cap, what is the optimum number for particular applications?
4. What evidence is there that the actual levels and range of an attribute has an important influence of identification of statistical significance and associated outputs such as marginal valuations and elasticities?
5. What evidence is there that many design effects like complexity, ambiguity and data collection procedures influence variances and not means?
6. What evidence is there that simpler choice models like MNL confound a range of real influences on choice behaviour such as unobserved heterogeneity and serial correlation?
7. The great majority of the SP applications in travel behaviour research limit their estimation to linear additive main effects (see Ortuzar et al 1997 and as exceptions). To what extent are simple SP designs inadequate due to behaviourally limiting linear additive specification of utility expressions associated with each alternative in a choice set? If we move to more complex but potentially behaviourally richer non-linear specifications involving higher-order main effects and interactions, what are the design implications for particular applications (Stopher and Hensher 2000)?
8. What evidence is there to guide the selection on a data collection strategy as the complexity of an SP experiment increases based on the many examples of SP

data collection by mail, face to face (both paper and laptop), and telephone interview?

We look forward to research directed towards these issues and to understanding the behaviour of random component variation in experiments and real market choices.

Acknowledgment

References

Adamowicz W., Louviere J., and Williams M. (1994) Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities, *Journal of Environmental Economics and Management*, 26, 271-292.

Ben Akiva, M.E. (1987) Methods to combine different data sources and estimate origin-destination matrices. In N.H. Gartner and N.H.M. Wilson (eds.) *Transportation and Traffic Theory*, Elsevier, New York.

Ben-Akiva M. E., and Morikawa T. (1990) Estimation of switching models from revealed preferences and stated intentions, *Transportation Research A*, 24A(6), 485-495.

Ben-Akiva, M.E. and Morikawa, T. (1998) Estimation of discrete choice models with correlated stated and revealed preferences, paper prepared for the Choice Symposium, HEC, France, July.

Ben-Akiva, M.E., McFadden, D., Garling, T., Gopinath, D., Bolduc, D., Borsch-Supan, A., Delquie, P., Larichev, O., Morikawa, T., Poly doropoulou, A and Rao, V. et al (1999) Extended framework for modelling choice behavior, *Marketing Letters*, 10 (3), 187-204.

Bhat, C.R. (1997) An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel, *Transportation Science*, 31, 34-48

Bhat, C. (1997a) Recent methodological advances relevant to activity and travel behavior analysis, Conference Pre-prints, IATBR'97, *The 8th Meeting of the International Association of Travel Behaviour Research*, Austin, Texas, September.

Bhat, C. (in press) A Multi-Level Cross-Classified Model for Discrete Response Variables, *Transportation Research*.

Bhat. C.R. and F.S. Koppelman (2000) Activity-Based Modeling of Travel Demand, in *The Handbook of Transportation Science*, Edited by Randy Hall, Kluwer Academic Publishers.

Bolduc D. (1992): Generalised autoregressive errors in the multinomial probit model, *Transportation Research*, 26B (2), 155-170.

Bradley, M. A. and Daly, A. J. (1997) Estimation of logit choice models using mixed stated preference and revealed preference information, in Stopher, P.R. and Lee-Gosselin, M. (Eds) *Understanding Travel Behaviour in an Era of Change*, Pergamon, Oxford, 209-232.

Brazell, J. and Louviere, J. (1995) Length Effects in Conjoint Choice Experiments and Surveys: An Explanation Based on Cumulative Cognitive Burden, Paper presented to INFORMS Marketing Science Conference, Sydney, Australia 1995.

Brewer, A. and Hensher, D. A. (1999) Distributed Work and Travel Behaviour: The Dynamics of Interactive Agency Choices Between Employers and Employees *Transportation* (earlier version presented at 8th IATBR Conference on Travel Behaviour Research, Austin, Texas, September 1997).

Brownstone, D. and Train, K. (1999) Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, 89 (1-2), 109-129.

Bunch, D., Brazell, J. and Louviere, J. (1998) "Implications of Choice Set Complexity Effects for Design and analysis of Discrete Choice Experiments," Paper presented to the INFORMS Marketing Science Conference, INSEAD, France (June).

Daniels, R. and Hensher, D.A. (in press) Valuation of Environmental Impacts of Transportation Projects: The Challenge of Self-Interest Proximity, *Journal of Transport Economics and Policy* edited by Ken Small

Deshazo, J.R. and Fermo, G. (1999) Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency, Unpublished working paper, School of Public Policy and Social Research, UCLA (February).

Frisch, R. (1972) Cooperation between politicians and econometricians on the formalization of political preferences. Reprint Series No. 90, Institute of Economics, University of Oslo.

Gerard, K., Shanahan, M. and J.J. Louviere (1999) Mammography Screening: Why do some women attend?, Unpublished working paper, *Centre for Health Economics Research and Evaluation (CHERE),* University of Sydney, November.

Golob, T. F. (2000) Resource paper, IATBR 2000, Gold Coast, Queensland, July2-7.

Golob, T., Kitamura, R. and Long, L. (eds), (1997) *Panels in Transport Planning*, Kluwer Scientific Publishers, Boston.

Hensher, D.A. (1994) Stated Preference Analysis of Travel Choices: The State of Practice, A special issue of *Transportation* on The Practice of Stated Preference Analysis, 21 (2), 106-134.

Hensher, D.A. (1998) Establishing a fare elasticity regime for urban passenger transport *Journal of Transport Economics and Policy*, 32 (2), 221-246.

Hensher, D.A. (1999) The Valuation of Time Savings for Urban Car Drivers in New Zealand: Evaluating Alternative Model Specifications, Institute of Transport Studies, The University of Sydney,December

Hensher, D.A. (2000) Towards an Integrated Strategic Urban Passenger Transport Modelling System, Institute of Transport Studies, The University of Sydney, January.

Hensher, D.A. (2000a) Measurement of the Valuation of Travel Time Savings, Special issue in memory of Michael Beesley of Journal of Transport Economics and Policy

Hensher, D.A. and Bradley, M. (1993) Using stated response data to enrich revealed preference discrete choice models, *Marketing Letters*, 4(2), 139-152.

Hensher, D.A., Louviere, J.J. and Swait, J. (1999) Combining sources of preference data, *Journal of Econometrics*, 89, 197-221.

Johnson, R.M. and Orme, B.K. (1996), How Many Questions Should You Ask In Choice-Based Conjoint Studies? Paper presented to the American Marketing Association's Advanced Research Techniques Forum, Beaver Creek, CO (June).

Kamakura, Wagner A., Byung-Do Kim and Jonathan Lee (1996), Modeling Consumer Choice Processes: Preference and Structural Heterogeneity, *Marketing Science,* 15 (2), 152-172.

Keane, M.P. (1997) Current issues in discrete choice modelling, *Marketing Letters*, 8(3), 307-322.

Kim, K.S. (1998) Analysing repeated measurement problems in SP data modelling, paper No 537, presented at the 8the World Conference on Transport Research, Antwerp, July.

Louviere, J.J. (1981) Some Comments on Premature Expectations Regarding Spatial, Temporal, and Cultural Transferability of Travel-Choice Models. In Stopher, P. R., Meyburg, A. H., and Brog, N. (eds.) New Horizons in Travel-Behavior Research. Lexington, Massachusetts: D. C. Heath and Co., 653-664.

Louviere J. (1988a) *Analyzing Decision Making: Metric Conjoint Analysis*, Newbury Park, CA:Sage Publications, Inc.

Louviere J. (1988b) Conjoint Analysis Modelling of Stated Preferences. A Review of Theory, Methods, Recent Developments and External Validity, *Journal of Transport Economics and Policy*, 20, 93-119.

Louviere J. (1994) Conjoint Analysis, in Bagozzi R. (ed.), *Advanced Methods of Marketing Research*, Cambridge, MA:Blackwell Publishers, 223-259.

Louviere J. (1995) Relating Stated Preference Measures and Models to Choices in Real markets: Calibration of CV Responses, in Bjornstad D.J. and Kahn J. R. (eds.), *The Contingent Valuation of Environmental Resources*, Brookfield: Edward Elgar, 167-188.

Louviere J., and Woodworth G. (1983) Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research,* 20, 350-367.

Louviere J., and Hensher D.A. (1983) Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event, *Journal of Consumer Research*, 10 (3), 348-361.

Louviere, J., Oppewal, H., Timmermans, H. and Thomas, T. (1993), Handling Large Numbers of Attributes in Conjoint Applications: Who Says Existing Techniques Can't Be Applied? But If You Want an Alternative, How About Hierarchal Choice Experiments? Paper presented to the American Marketing Association's Advanced Research Techniques Forum, Beaver Creek, CO (June).

Louviere, J.J., Meyer, R. J., Bunch, D.S., Carson, R., Dellaert, B., Hanemann, W.M., Hensher, D.A., and Irwin, J. (1999) Combining sources of preference data for modeling complex decision processes, *Marketing Letters*, 10 (3), 205-218.

Louviere, J.J., Hensher, D.A. and Swait, J. (in press 2000) *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*, Cambridge University Press, Cambridge, 697pp

Mazzotta, M.J. and Opaluch, J.J. (1995) Decision making when choices are complex: a test of Heiner´s hypothesis. *Land Economics 71*, 500-515.

McFadden, D. (1981) Econometric models of probabilistic choice, in Manski, C.F. and McFadden, D. (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press, 198-272.

McFadden D. (1986) The choice theory approach to marketing research, *Marketing Science*, 5 (4), 275-297.

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

Morikawa, T. (1994) Correcting state dependence and serial correlation in the RP/SP combined estimation method, *Transportation*, 21 (2), 153-166.

Munizaga, M.A., Heydecker, B.G. and de Dios Ortuzar, J. (1997) On the error structure of discrete choice models, *Traffic Engineering and Control*, 38 (11), 593-7.

Munizaga, M.A., Heydecker, B.G. and de Dios Ortuzar, J. (1999) Representing heteroscedasticity in discrete choice models, Department of Civil Engineering, Universidad de Chile, Santiago.

Ortuzar, J. de D., Roncagliolo, D.A. and Velarde, U.C. (1997) Interactions and independence in stated preference modelling. In *Proceedings 25th European Transport Forum*, Brunel University, England.

Ouwerslot, H. and Rietveld, P. (1996) Stated choice experiments with repeated observations. *Journal of Transport Economics and Policy XXX*, 203-212.

Revelt, D. and Train, K. (1996) Incentives for appliance efficiency: random parameters logit models for households; choices, *Department of Economics*, University of California, Berkeley.

Saelensminde, K. (1999) Inconsistent choices in stated choice data: use of the logit scaling approach to handling resulting variance increases. Working Paper, Institute of Transport Economics, Oslo.

Simonson, I. (1989) Choice Based on Reasons: The Case of Attraction and Compromise Effects, *Journal of Consumer Research*, 16, 158-174.

Stopher, P.R. and Hensher, D.A. (2000) Are More Profiles Better than Less? Searching for Parsimony and Relevance in Stated Choice Experiments. Paper presented at the Transportation Research Board Annual Conference, January (accepted for *Transportation Research Record*)

Street, D.J., Bunch, D.S. and Moore, B. (1999) Optimal Designs for 2^k Paired Comparison Experiments, Unpublished working paper, School of Mathematical Sciences, University of Technology, Sydney (November).

Swait, J. (1994) A Structural Equation Model of Latent Segmentation and Product Choice for Cross-Sectional Revealed Preference Choice Data, *Journal of Retailing and Consumer Services*, 1(2), 77-89.

Swait, J. and W. Adamowicz (1996) The Effect of Choice Environment and Task Demands on Consumer Behavior: Discriminating Between Contribution and Confusion, Unpublished Working Paper, Dept. of Rural Economy, University of Alberta, June, 39 pp.

Swait J., and Louviere J. (1993) The Role of the Scale Parameter in The Estimation and Use of Multinomial Logit Models, *Journal of Marketing Research*, 30, 305-314.

Toner, J.P., Clark, S.D., Grant-Muller, S.M. and Fowkes, A.S. (1998) Anything you can do, we can do better: a provocative introduction to a new approach to stated preference design. Paper 459, Topic Area D1, *World Conference on Transport Research*, Antwerp, Belgium.

Train, K. (1997) Mixed logit models for recreation demand, in Kling, C. and Herriges, J. (eds.) *Valuing the Environment Using Recreation Demand Models*, Elgar Press, New York.

Waddell, P. (2000) Land Use Models Resource Paper, IATBR 2000, Gold Coast, Queensland, July2-7

Wedel W., Kamakura W, Arora N, Bemmaor, A. Chiang, J., Elrod, T., Johnson, R., Lenk, P.. Neslin S, and C S Poulsen (1999) Discrete and Continuous Representation of Unobserved Heterogeneity in Choice Modelling, *Marketing Letters*,10 (3), 219-232.

Uldry, P., Yeoh, A., Louviere, J. and Burke, S. (1998) Does Missing Information Affect Means or Variances of Error Distributions, Paper presented to the INFORMS Marketing Science Conference, INSEAD, France (June).

20