ITS

WORKING PAPER
ITS-WP-99-1

# Understanding Travel Behaviour: Some Appealing Research Directions

by

David A. Hensher

January, 1999

ISSN 1440-3501

*Established and supported under the Australian Research Council's Key Centre Program.*

**INSTITUTE OF TRANSPORT STUDIES**

The Australian Key Centre in Transport Management

The University of Sydney and Monash University

| | |
|---|---|
| **NUMBER:** | Working Paper ITS-WP-99-1 |
| **TITLE:** | Understanding Travel Behaviour: Some Appealing Research Directions |

**ABSTRACT:**

This paper presents one researchers perception of selective emphases in the body of travel behaviour research which have had and/or may in the future have a non-marginal impact on the way that research activity is undertaken. Some of the contributions are well established and have moved from state of the art to state of practice; other efforts are relatively new and maturing in their role as paradigms of thought. The contributions can broadly be grouped into four classes of research: decision paradigms, in particular the interpretation of the choice process within a broad activity framework, and the recognition that agents making decisions do not always operate in a perfectly competitive market; releasing the analytical formalism of the choice/decision process from the restrictive IIA paradigm of the great majority of applied travel choice modelling - moving to nested structures, free variance and correlation among alternatives, random taste weights, accommodating unobserved heterogeneity and mixed 'logits'; combining sources of preference and choice data, including joint analysis of market and experimental choice data, interfaces between attitudinal and behavioural data, and generalising valuation to valuation functions; and advances in the study of the dynamics of traveller behaviour, especially the timing of change and its importance in establishing hurdle dates for forecasting traffic and revenue for infrastructure projects.

**AUTHORS:** 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:** January, 1999

# Background

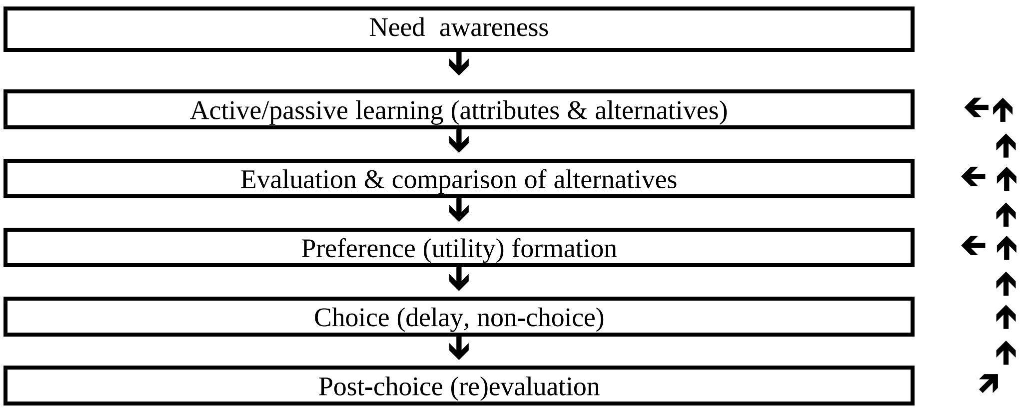
Travel behaviour is unquestionably complex, and challenging to understand and predict. Research on traveller behaviour is essential to an informed debate on how we might manage better the transport and traffic system in the interests of broad goals of efficiency, equity and environmental sustainability. There is a very large and growing body of literature drawing on a broadening disciplinary base (engineering, planning, economics, statistics, operations research, psychology, sociology, geography and environmental sciences), which offers innumerable hypotheses on how agents - individuals, households, organisation (public and private) and interest groups - make decisions and respond to changing opportunities and constraints. Qualitative and quantitative paradigms are offered as frameworks within which we seek an understanding of behavioural response to the multitude of transport and non-transport policy instruments continuously 'tested' by real and experimental markets. Advanced intelligence has become the backbone of research into travel behaviour; sometimes accompanied by accumulated wisdom from the market and/or from the extant body of research. Much of the research is repetitive, adding useful value at the margin and reinforcing an established paradigm; and from time to time we have significant breakthroughs in the way we think about travel behaviour.

This paper attempts to highlight one researchers perception of selective emphases in the body of travel behaviour research which have had and/or may in the future have a non-marginal impact on the way that research activity is undertaken. Some of the contributions are well established and have moved from state of the art to state of practice; other efforts are relatively new and maturing in their role as paradigms of thought. The contributions can broadly be grouped into four classes of research: C1 - decision paradigms, in particular the interpretation of the choice process within a broad activity framework, and the recognition that agents making decisions do not always operate in a perfectly competitive market; C2 - releasing the analytical formalism of the choice/decision process from the restrictive IIA paradigm of the great majority of applied travel choice modelling - moving to nested structures, free variance and correlation among alternatives, random taste weights, accommodating unobserved heterogeneity and mixed 'logits'; C3 - combining sources of preference and choice data, including joint analysis of market and experimental choice data, interfaces between attitudinal and behavioural data, and generalising valuation to valuation functions; and C3 - advances in the study of the dynamics of traveller behaviour, especially the timing of change and its importance in establishing hurdle dates for forecasting traffic and revenue for infrastructure projects.

In presenting a view of progress in travel behaviour research, the opportunity to highlight some practical policy implications of specific behavioural paradigms is exploited. For example, the usefulness of point estimates of valuation of travel attributes in contrast to a distribution of values; the worrying implications of using stand-alone stated choice methods in prediction and forecasting; and how elasticities from simple choice models are increasingly becoming misleading.

# The Evolution of More Realistic Decision Paradigms

Decisions taken by agents typically evolve out of a decision process as summarised in Figure 1 (Hensher et al., in progress). An agent first becomes aware that s/he has needs and/or problems to solve, which is followed by a period of information search in which s/he learns about services and products that can satisfy the need(s) or solve the problems. During search and learning, agents form beliefs about which services/products are available (if any) to satisfy the need(s) or solve the problem(s), product attributes germane to a choice and attribute values offered by services/products, as well as any associated uncertainties. Eventually agents become sufficiently informed about the service/product category to form a decision rule (or utility function) which involves valuing and trading off product attributes that matter in the decision. Given a set of beliefs or priors about attributes possessed by service/product alternatives, agents develop a preference ordering for such services/products; and depending upon budget and/or other constraints/considerations make decisions about whether to choose an alternative or reject all offerings. Given they decide to choose an alternative, agents finally must choose one or more alternatives, in certain quantities and with particular acquisition timings.




While travel behaviour research gives a lot of credit to the importance of all six stages, it devotes itself essentially to stages 4 and 5, as suggested in Figure 2. These represent the last decision stages at which agents form utilities or values and begin to compare services/products to form overall (holistic) preferences for an available set of alternatives. There is a maintained assumption of ‘equilibrium’ in the decision process ‘derived’ from stages 1 to 3, which denies the testing of the drivers which influence the establishment of the formed set of preferences which define the environment within which revealed and/or stated choices are made. Many of these drivers are ignored in simplistic travel choice and demand models such as the multinomial logit; while an increasing number are accommodated in various degrees of richness through more complex specifications of the explanatory variables and the unobserved or random error components (see Section 4).

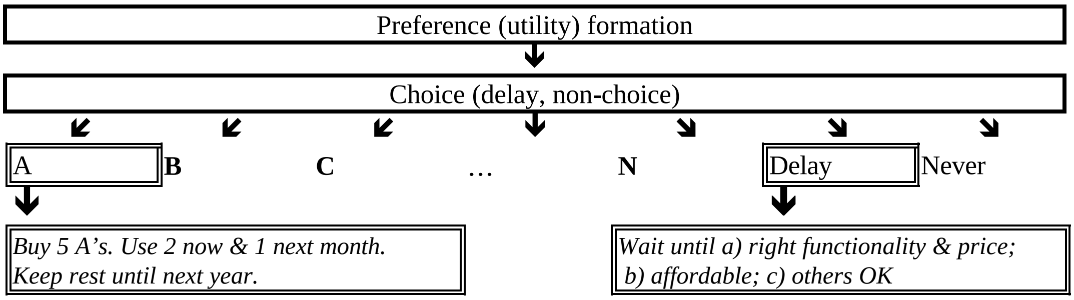



**Figure 2. Complex decision making and the choice process**

Figure 3 formalises the 6-stage process as a series of interrelated processes, linking each process to a stage in the decision making process and describes the general area of research connected to that area in contributing disciplines to travel behaviour research such as marketing, psychology, economics and econometrics. The conceptual framework outlined in Figures 1 to 3 is consistent with economic theory, accommodates random utility type choice and decision processes; and most importantly, allows one to "mix & match" measures from various levels in the process, assuming such measures are logically or theoretically consistent with the framework and each other. The advantage of the latter integration is that it allows explanation of the choice behaviour in terms of

1. physically observable and measurable (engineering) variables,
2. psychophysical variables (beliefs/product positions),
3. part-worth utility measures, or
4. holistic measures of each alternative's utility.

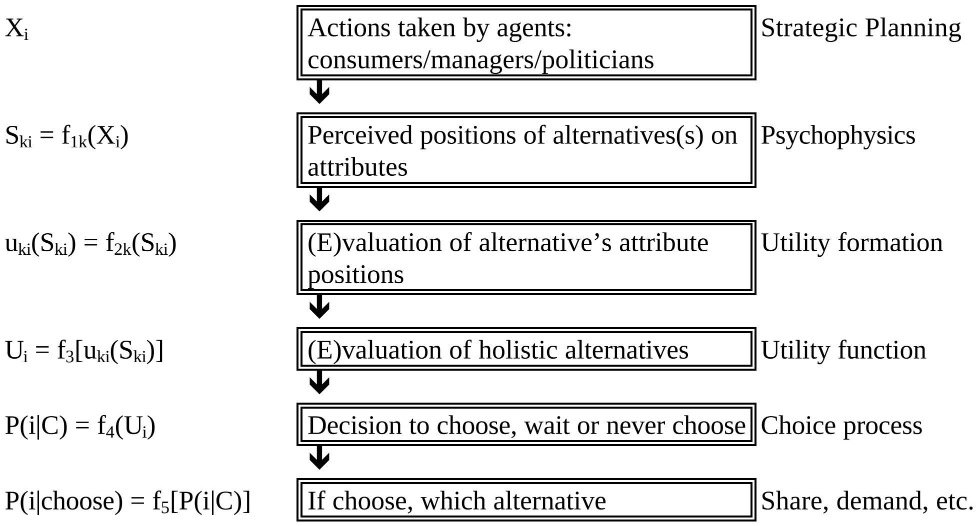



**Figure 3. Functional relationships implied by the framework**

Depending on one's research and/or analytical objectives, explanatory variables at one level can serve as instruments or "proxy" variables for measures at other levels. The advantage of the latter is that these instruments can be used to reduce specification errors and/or improve estimation efficiency. Equally importantly, the conceptual framework suggests that stated preference methods (which are very popular in travel behaviour studies) and measures used to model intermediate stages in the decision making process are consistent with parallel revealed preference (RP) methods and models. For example, the framework permits choices to be explained by direct observation and measurement of physical product attributes and/or managerial actions like advertising expenditures, but also suggests that such direct estimation may obscure important intermediate processes, and overlooks the potential role of intermediate models and measures in an overall behavioural framework that explains consumer choices.

## *Travel positioned within the content framework of an agent's choice*

The functional relationships offered in Figure 3 are driven by attributes and alternatives, conditioned on the characteristics of the agent, which when combined define activities undertaken by agents. We can adopt this framework to traveller behaviour, recognising that travel is a derived activity contributing to the ways in which people and commodities are (re)positioned in time and space. The underlying attributes driving this repositioning of agents and alternatives are extensive. They might best be viewed as a multi-dimensional vector of *contents*, whose dimensions may simultaneously include information content (IC), physical content (PC), social content (SC), material content (MC), and aesthetic content (AC). An activity is then defined by its contents (Hensher and Golob 1998) and occurs within *communication space*. This provides an attractive context within which to study the broadening base of travel behaviour research. Communication space, more specifically, might be thought of as a global construct capturing all forms of *contents' exchange*. This definition is sufficiently general to incorporate near or local distance such as an in-home activity, and non-local distance (Zumkeller 1997) and the role of telecommunications as a tele-access substitute and/or complement to travel.

Within this behavioural framework, an activity becomes the overarching unit of analysis, differentiated by the *mix of contents*. For example, the activity 'shopping' might be described by an internet order placed at work, followed by a car trip to the shop, a business meeting at the shopping centre, the collection of the ordered shopping and a car trip home. The *social content* is high, the *information content* may not be as high as ordering on site where the merchandise can be felt (but this may depend on the nature of the goods being purchased and one's overt experience with them). The *physical content* in terms of physical exertion is high compared to a home-delivery; the *aesthetic content* may be not applicable, and the *material content* is high because of the travel component and the greater ability to do comparative shopping.

Alternative ways of shopping are particular specifications of the contents mix. For example, the internet order may generate a home-delivery instead; however the social content of the activity would have been zero, and the physical content high in the form of a commercial urban goods movement rather than a person-shopping trip. Here we have substitution between a trip by a household agent and a trip by a firm's agent.

The idea of content mix can be applied to a simple activity such as home-walk-home and a complex activity (or chain) such as home-walk-restaurant-walk-home or home-teleshop-delivery to home-eat, all of which have physical, material, social, aesthetic and information content. Content mix also helps with the problem of simultaneous activities e.g. driving to work while using a mobile phone to plan social and business activities accompanied by a colleague with whom one is discussing strategy for an immediate meeting. It is the content mix which defines the sources of relative utility associated with simple and complex activities.

### *A conceptual framework to capture the content-mix of an activity*

The communication of agents requires them to act *as if* they are maximising utility through the way that activities are constructed. Conceptually we might define the demand for communications by agent *q* as a function of the contents exchange demanded by the agent located at spatial point *i*, the generalised goods content located at spatial point *j*, and the generalised cost of exchanging contents between *i* and *j*. A specific *ij* pair can be defined very finely as the same location (eg an office and kitchen at home) or a different location (eg office *i* at home and kitchen *j* at home, office *i* and restaurant *j* in the same building). The activity choice set embodies alternative amounts of content exchange. The presence of physical and material content involves travel.

The choice of an alternative activity is a function of, amongst other influences, the nature of exchange content, the degree to which this content can be obtained remotely ('at a distance' without travel) or through travel, and the (dis)utility associated with the sources of benefits and costs (eg the trade-off between the utility of information content with the utility of social content). The indirect utility function associated with an agent *q* and an activity *a* might be defined as:

$V_{qa}$ = f($IC_{qa}$, $PC_{qa}$, $MC_{qa}$, $Sc_{qa}$, $Ac_{qa}$, other benefits, travel time/stress, telecom cost, telecom time, other costs, socioeconomic characteristics$_q$, constraints….)

In recognition that the communication choices made by individuals are influenced by other members of the household and/or other non-household agents such as an employer, the indirect utility function may be generalised further to recognise 'interactive' agency effects - $V_{q'a}$ (detailed further in Section 3):

$V_{qa}$ = g($V_{q'a}$, $IC_{qa}$, $PC_{qa}$, $MC_{qa}$, $Sc_{qa}$, $Ac_{qa}$, other benefits, travel time/stress, telecom cost, telecom time, other costs, socioeconomic characteristics$_q$, constraints)

The framework can be generalised to incorporate dynamics in recognition of the importance of awareness, ability to adopt and experience with particular communication possibilities in an activity chain.

## Not all Agents Live in Perfectly Competitive Markets

How often do we discuss our plans with our household partners, our business associates and friends? How often do decisions involve negotiation, bargaining and sometimes non-cooperation wherein a *individual's* personal utility maximising outcome is not permissible? For example, employees and employers may have different views and preference functions for alternative work practices such as telecommuting (Brewer and Hensher 1997), compressed work weeks; shippers negotiate with freight forwarders and accept or reject the best offer from a limited set of competing forwarders. Two workers in a household negotiate on workplace and working hours to accommodate the transport needs of children.

All of these examples represent cases of preference formation and choices which are not necessarily independent of the equivalent functions of other agents. Interactive agency decision making is little understood and studied in the quest to improve our understanding of travel behaviour. It opens up challenges in handling contemporaneous and temporal sequential and simultaneous interactions between agents, unobserved heterogeneity within and between agents, and the need for more general choice models such as the multi-period multinomial probit (see Section 4 on analytical perspectives).

Any situation involving interaction between two or more individuals has elements of cooperation and non-cooperation. The choice outcome matters to each of them and depends on the actions of both or all of the players. At the outset of a negotiation, each person perceives the extent to which the other party will be cooperative as a commitment to choose a joint plan of action. This does not imply that either party sacrifices their interests for the sake of the other, although it may; only that each communicates and coordinates with a view to furthering their own unchanged interests by so doing. The central position here is, that, in neo-classical economic terms, private decision-making leading to everyone's good (or agent-independent utility maximisation) depends critically on assuming a regime of perfect competition with numerous participants. In the context of negotiations between employees and employers, and between household members, this may be an unrealistic assumption.

Game theory provides compelling support for the application of a cooperative game in which two agents such as an employee and the employer or two household members attempt to cooperate. Cooperation assumes compliance with two tests: (1) for *both* agents it cannot be bettered by some agreement, and (2) for *either* agent it cannot be bettered by one participant going their own way. Importantly, however, whether an agent will end up acting as a unified agent (i.e. cooperation), depends on decisions made entirely non-cooperatively by each party. The dynamics of game play is noticeably absent in the literature on traveller behaviour research is general and stated choice experiments in particular.

To illustrate how bargaining in a game context works, assume that the alternatives in the trade are telecommuting 2 days a week, 1 day a week, and non-telecommuting (see Brewer and Hensher 1997 for further details). The employees first strategy ($ee_1$) might be to opt for telecommuting 2 days a week; the second strategy ($ee_2$) to telecommute 1 day a week; and the third option ($ee_3$) not to telecommute. The employers strategies ($er_1$, $er_2$ and $er_3$) might consist in offering the options in the reciprocal order. If they do not agree to one of these exchanges, they will remain under current work practice (the status quo). The payoffs might be as given in Table 1. For example, in (3,-1), 3 represents the payoff to the employee and -1 is the payoff to the employer. If the employee chooses to telecommute 1 day a week ($ee_2$), and the employer chooses not to telecommute ($er_1$), the employee receives a payoff of 1.5 and the employer a payoff of 0.5.

**Table 1. A Bi-Matrix of Payoffs in an Interactive Telecommuting Preference Game**

| | | TC2 employer | TC1 employer | SQ employer |
|---|---|---|---|---|
| | | $er_1$ | $er_2$ | $er_3$ |
| TC2 employee | $ee_1$ | (3,-1) | (1.5, 0.5) | (1.5, 0.5) |
| TC1 employee | $ee_2$ | (1.5, 0.5) | (2.5, 1) | (1.5, 0.5) |
| SQ employee | $ee_3$ | (1.5, 0.5) | (1.5, 0.5) | (1,2) |

*Note:* the scale and origin of individual utility functions are arbitrary and so *afortiori*, the utility indices are not directly comparable, SQ = status quo

The off-diagonal payoffs show the payoffs if their demands are not acceded to (on the assumption of indifference to the alternative non-cooperative outcomes). It makes no difference which demand is refused. If they fail to agree the outcome is always the same - current work practices. The attainable region R in payoff space and the status quo point ($ee_s$, $er_s$) are shown in Figure 4.

The arbitration associated with the bargaining game is defined by Nash (1950) as follows: For any point (Ua, Ub) in R, consider the quantity (Ua-Sa)(Ub-Sb), the product of the employee and the employers utility increment from the status quo. Now find (Ua, Ub) in R that maximises this product subject to the constraints that Ua ³ Sa (ie Ua - Sa ³ *0*), Ub ³ Sb. This bargaining solution in outcome space, represents the basket of attributes which are sources of expected utility. The outcome of cooperative games, the pairs of baskets or attribute mixes, define the feasible set of distributions in outcome space. The search for the feasible sets can be implemented through (interactive) choice experiments. The choice probabilities from the choice experiment provide the information to construct the expected utility matrix, an input into interactive agency utility maximisation.

The mapping between payoff and expected utility is not exact. The off-diagonal expected utilities are likely to be different in each cell and thus the validity of the Nash solution of equal payoff does not translate into identical expected utilities in the off-diagonal cells. Indeed game theory sets out to describe *not* behaviour but non-cooperative and cooperative modes of choice. This is why the off-diagonals can be equivalent. In the study of behavioural responses in preference space this need not be so. The sum of the joint probabilities in the three diagonal cells define the cooperative probability set. The choice probabilities from a discrete choice model might be as given in Table 2. These are illustrative and bear no relationship to Table 1. *Cooperation is achieved up to a probability of 0.402*, comprising the current outcome (0.0667), telecommuting 1 day a week (0.2324) and telecommuting 2 days a week (0.1027).These cooperative probabilities can be identified at each stage (or pass between the two agents) in the sequential-move interactive agency experiment (see Brewer and Hensher 1997). Given an interest in expected utility, the embodiment of game theory is in spirit only.

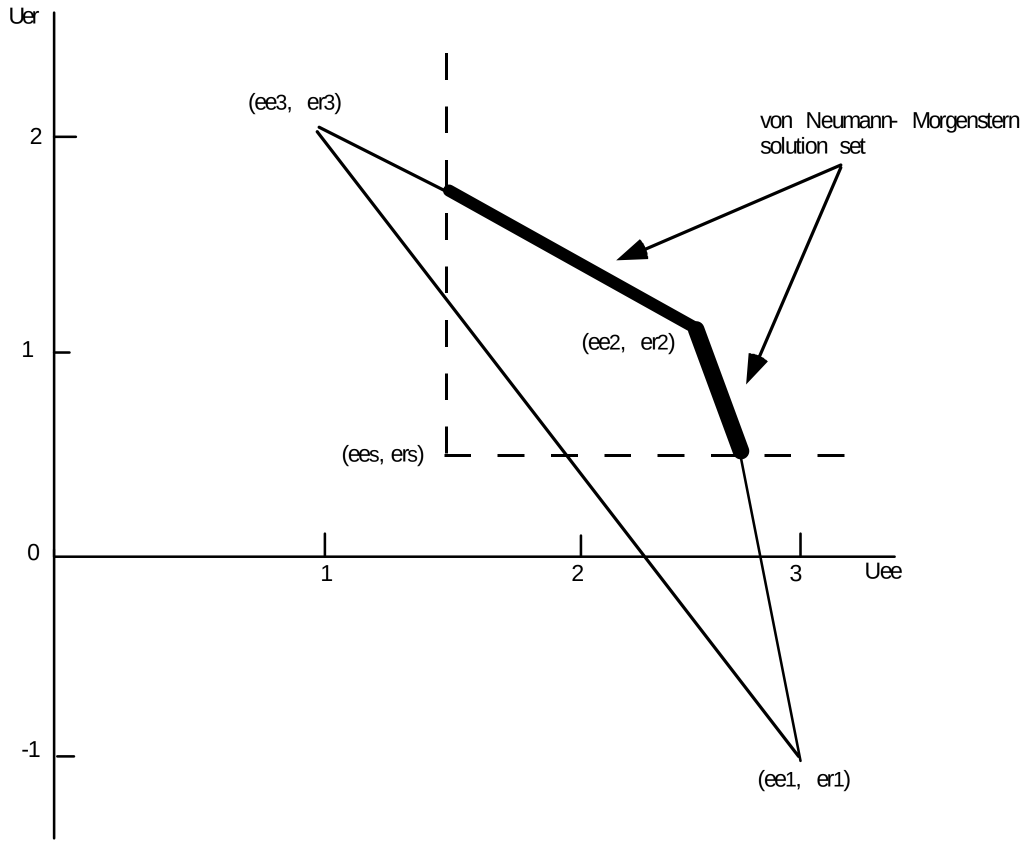



**Figure 4. Telecommuting Preferences**

**Table 2. A Bi-Matrix of Choice Outcomes in an Interactive Telecommuting Preference Game**

| | | employer | employer | employer |
|---|---|---|---|---|
| | | $er_1$ | $er_2$ | $er_3$ |
| employee | $ee_1$ | *(.208,.250)* | (.208,.583) | (.208,.168) |
| employee | $ee_2$ | (.375,.250) | *(.375,.583)* | (.375,.168) |
| employee | $ee_3$ | (.417,.250) | (.417,.583) | *(.417,.168)* |

## *The empirical paradigm - interactive stated choice experiments*

A set of attributes and alternatives provides the basis for designing an interactive stated choice experiment. The analysis of each pass in an interactive agency choice experiment might be represented as a recursive discrete choice paradigm in which the prior agent's choice conditions the subsequent agent's choice. The recursive structure embodies the agents 'flip-flopping' as the prior and subsequent agent in each round of the ICE. Sequential estimation of each agent's choice process at each pass in the sequential negotiation process will enable tracking of the choices made and their revisions up to the point of cooperation or experiment termination if there is no agreement after a predetermined number of rounds. In the bargaining literature, this is described as a dynamic sequence of concessions.

*Sequential* interactive choice experiments can be administered to a sample of agents using a randomised ordering of a series of showcards. For example, in *Round 1*, agent type A (ATA) is first selected. ATA completes the first round of the choice experiment involving the evaluation of a number of alternative scenarios, repeated a number of times. In *Round 2*, agent type B (ATB) is asked to make an offer to ATA in the context of the same choice experiment but under two information scenarios - with and without knowledge of ATA's preferred choice on each of three replications.

In *Round 3*, a set of first round ATB responses are fed back to ATA who assesses the 'preferred offer(s)' of ATB and then repeats the same experiment as in Round 1, revising or staying with their preferred first round offer. Reasons for maintaining or revising Round 1 preferences can be identified. In *Round 4*, the outcome is then fed back to ATB who re-evaluates their position in the face of ATA's second-round response, this time in full knowledge of ATA's, original or revised, preferred response. ATB is now supplied with *and* second round 'preferred' offers. For each of the experiment replications, ATB then makes a further offer which may maintain the first round offer or lead to a revision. Reasons can be sought for maintenance or revision of round 1 preferences. If the offer from ATB is accepted, that is the end of the process - a cooperative solution has been produced. If s/he rejects the offer, a stalemate is the outcome. The experiment may or may not extend into the negotiation space of a new option set. The experiment discussed above is sequential. Simultaneous interactions can also be investigated.

## *Empirical analysis strategy*

To implement the interactive agency choice, a sample of paired agents ATA and ATB have to be selected. The notion of a pass needs definition. A pass is defined as a sequence of ATA and ATB offers. Pass 1 involves round 1 and round 2; pass 2 involves round 3 ATA review given round 2 ATB response. If there was agreement in pass 1 on a replication of the *n* ICE's then the experiment finishes for that replication.

A series of choice models can be estimated to evaluate potential influences on ATA's and ATB's preference for each alternative. The following steps are involved in a sequential recursive paradigm:

- *Step 1 ($ATA_1$)* : First sequential move offer of ATA - *n* replications per ATA. As the first experiment there is no involvement of ATB.
- *Step 2 ($ATB_1$):* First sequential move offer of ATB - the same *n* replications as per ATA. The knowledge of the ATA's offer is revealed to half of the ATB's only. If ATA and ATB agree on the offer in pass 1, it can be identified by a 'pass agreement' dummy variable (=1 if agree and 0 otherwise). An important hypothesis is that the agents are more likely to agree where the ATB has knowledge of the ATA's preferred offer, and where the offer of both parties is to stay with the current situation. We also expect that the agreement of agents in an *earlier* pass will arise when the status quo offer is preferred; and that the agreement on the status quo will weaken as the preferred outcome through the pass negotiations, increasing the probability that a non-status quo offer is more likely. If this is true, then we have identified the limiting value of an empirical strategy which fails to reveal the views of both parties in negotiation, denying the opportunity to relax constraints where the status quo is not preferred by both parties.
- *Step 3 ($ATA_1$, $ATB_1$)*: Evaluate the influences on the pass agreement (1,0) outcome for the first sequential move offers. These influences include design attributes, individual characteristics and the individual's perception of the opportunities and constraints associated with the alternatives in the choice set.
- *Step 4 ($ATA_1$, $ATB_1$)*: Calculate the expected utility matrix for the agents and identify the cooperation probability for each alternative. The non-cooperation probabilities for each off-diagonal pair of alternatives are also identified.

The 4-step process is repeated subject to the number of steps required to achieve a cooperative outcome and the limits on sample size for model estimation.

An alternative specification involves identification of the equilibrium pass in which each pair of agents cooperate. An interesting research task is to compare the cooperative probabilities and sources of influence at each round of negotiation with the influences on the equilibrium pass alone. Each agent pair exhibits an equilibrium at different passes. If the gain in information from the dynamic 'negotiation' process adds value to our understanding of incentives and constraints required to establish a higher incidence of cooperative equilibrium, then the sequential process should be favoured over the analysis only of the 'final pass' strategy. The explanation for the number of passes required to reach cooperative equilibrium is also of great interest. Hensher, Boersch-Supan and Brewer (in progress) are investigating the alternative strategies.

## Analytical Perspectives

Discrete choice models belonging to the family of random utility models remain the most popular specification of travel demand models. Their appeal is largely due to their justification within a microeconomic framework which embodies the idea of random utility maximisation, wherein the importance of unobserved influences on choices from the analyst's perspective are explicitly handled in the behavioural definition of the choice problem for a sampled population. Much of the research in this field has concentrated on the search for parsimonious yet behaviourally realistic econometric models capable of assisting us to understand travel behaviour and to predict the consequences of change on such behaviour.

The literature is extensive (see Ben-Akiva and Lerman 1985, Hensher et al in progress, Bhat 1997, Koppelman and Wen 1997), with numerous applications of multinomial (MNL) and nested (NL) logit models in academic and practitioner studies. The strengths and weaknesses of such models are well documented, especially the strong (testable) assumptions imposed on the distribution and correlation structure of the variances and covariances of the unobserved (ie random) components of the indirect utility expressions associated with each alternative in a choice set. The importance of this research for policy is highlighted in Sections 5 and 6 when one compares estimates of elasticities and marginal rates of substitution (eg behavioural value of travel time savings) derived from simple and more complex models.

Much of the advanced research in the field of travel choice modelling has concentrated on relaxing these strong assumptions in ways which are behaviourally enriching, computationally tractable and practical. In part this is a response to criticism that the simpler models are a noticeable contributing source of error in forecasts (the external validity challenge) within the confines of the specific application; and partly a recognition that travel decision frameworks increasingly involve more complex trades between component choices leading to a joint outcome, which are often exogenised and/or confounded in simplistic applications.

To gain an appreciation of the progress made in relaxing some of the very strict assumptions of the multinomial and nested logit models (the closed form work horses of most transport researchers and practitioners), the essential generality of interest can be presented through the specification of the indirect utility expression $U_{it}$ associated with the i*th* mutually exclusive alternative in a choice set at time period *t*, and the diversity structure of the random component(s) (equation 1).

$$U_{it} = \alpha_{it} + \psi_{i,t-1} \frac{\lambda_{it}}{\lambda_{i,t-1}} \text{Choice}_{i,t-1} + \lambda_{it} \beta_{ikt} X_{ikt} + \lambda_{it}\gamma_{qt} + \varepsilon_{it} \qquad (1)$$

where

$\alpha_{it}$ = alternative-specific constant representing the mean of the distribution of the unobserved effects in the random component $\varepsilon_{it}$ associated with alternative i at time period t.

$\psi_{i,t-1}$ = the taste parameter associated with the lagged choice response from period t-1, which takes the value 1 if the chosen alternative in period t is the same as the chosen in period t-1 (ie, it = i,t-1) (see Hensher et al 1992).

$\lambda_{it}$ = the scale (or precision) parameter, which in the family of extreme-value random utility models is an inverse function of the standard deviation of the unobserved effects for alternative *i* at time period *t*. This parameter can be set to 1 across all alternatives when the standard deviations are identically distributed. $\lambda_{it}$ may vary between data sets such as a stated choice and market choice (RP) data drawn from the same or different samples of individuals in a closed population.

$\beta_{ikt}$ = the taste weights which represents the relative level of satisfaction or saliency associated with the kth attribute associated with alternative *i* in time period *t*.

$X_{ikt}$ = the kth exogenous attribute associated with alternative *i* and time period *t*.

$\gamma_{qt}$ = individual-specific effect or unobserved heterogeneity across the sampled population, for each individual *q* in time period *t*. This parameter may be a fixed effect (ie a unique estimate per individual) or a random effect (ie a set of values randomly assigned to each individual drawn from an assumed distribution). As a random effect, this unobserved term is part of an error components structure, assumed to be independent of other unobserved effects but permissible to be correlated across alternatives.

$\varepsilon_{it}$ = the unobserved random component comprising a variance and a set of covariances linking itself to the other alternatives. The full variance-covariance matrix across the choices set permits J variances and J*(J-1)/2 covariances; with at least 1 variance normalised to 1.0 and at least one row of covariances set to zero for identification (noting that the model is estimated as a series of differences between the chosen and each non-chosen alternative). By separating the unobserved heterogeneity across the sample, $\gamma_{qt}$, from $\varepsilon_{it}$ we have a components form of the random source of indirect utility. Any suppression of other sources of unobserved influences not included in equation (1), such as errors-in-variables (ie measurement error of the observed attributes), is confounded with the residual sources of random utility.

The taste weights and the scale parameters may themselves be a function of a set of exogenous characteristics which may or may not define the attributes of alternatives. This can include socio-economic characteristics of the sample and contextual effects such as task complexity, fatigue, data collection method, interviewer etc. The functional form can be of any estimable specification:

$$\lambda_{it} = \delta_{oi} + \delta_{1it}Income_{it} + \delta_{2it}TaskCompexity_{2it} \tag{2a}$$

$$\beta_{it} = \varphi_{oi} + \varphi_{1it}Age_{it} + \varphi_{2it}HouseholdSize_{2it} \tag{2b}$$

As one permits complex structures for the unobserved effects through introducing variation and covariation attributable to contemporaneous patterns between alternatives, and temporal patterns between alternatives (eg autoregressive structures), there exist complex and often ‘deep’ parameters associated with the covariance matrix which necessitate some simplification to achieve any measure of estimability of a model with application capability. The set of models presented below have the potential to be practically useful and to enrich our understanding of behaviour and behavioural response.

The majority of discrete choice models assume:

- A single cross-section and thus no lagged structure;
- Non-separation of taste and other component ‘weights’ defining the role of explanatory variables in each indirect utility expression (due to a confoundment with scale);
- Scale parameters that are constant across the alternatives (ie constant variance assumption);
- Random components that are not serially correlated (See Morikawa 1994, Hensher, Boersch-Supan and Brewer in progress);
- Fixed taste weights; and
- No unobserved heterogeneity.

The *error term* in discrete choice models has often been treated as intrinsic randomness in behaviour arising from unspecified sources. This is misguided. It is important to interpret the error term in the light of behavioural models. To illustrate the kind of misunderstanding that can arise from failure to do so, consider a simple random utility model, in which there are heterogeneous preferences for observed and unobserved labelled attributes:

$$U_{qjt} = a_{qj} + P_{qjt}g_{q} + X_{qjt}b_{q} + e_{qjt} \tag{3}$$

Here, $U_{qjt}$ is the utility that agent *q* receives given choice of alternative *j* on occasion *t*. In travel data, *t* could index time, destination visits, or travel occasions. In a stated choice experiment, *t* would index choice tasks. $P_{qjt}$ denotes price, and $Xqjt$ denotes an observed attribute of *j* (which for complete generality we allow to vary over agent and choice occasions). The $\alpha_{qj}$ denotes the agent specific intercept for alternative *j*, which can be interpreted as arising from q's preferences for unobserved attributes *j*. The $\gamma_q$ and $\beta_q$ are agent specific utility parameters which are intrinsic to the agent and hence invariant over choice occasions. The $\in_{qjt}$ can be interpreted as occasion specific shocks to q's tastes, which for convenience are assumed to be independent over choice occasions, alternatives and agents.

Suppose we estimate a multinomial choice model, falsely assuming that the intercept and slope parameters are homogeneous in the population. The error term in this model will be

$$w_{qjt} = \hat{a}_q + P_{qjt}\hat{g}_i + X_{qjt}\hat{b}_i + e_{qjt} \tag{3a}$$

where ^ denotes the individual specific deviation from the population mean. Observe that (from the analyst's perspective) the variance of this error term for agent *q* on choice occasion *t* is

$$\mathrm{Var}(w_{qjt}) = s_a^2 + P_{qjt}^2 s_g^2 + X_{qjt}^2 s_b^2 + s_{\in}^2 \tag{3b}$$

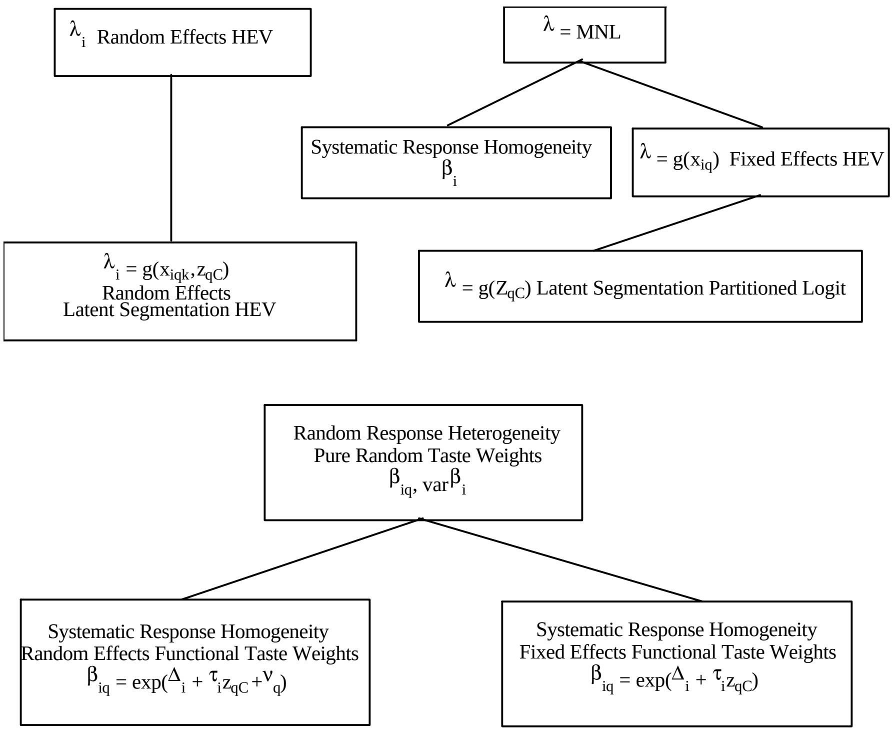
and the covariance between choice occasions *t* and *t*–1 is

$$\mathrm{Cov}(w_{qjt}, w_{qjt,t-1}) = s_a^2 + P_{qjt} P_{qj,t-1} s_g^2 + X_{qjt} X_{qj,t-1} s_b^2 \tag{3c}$$

Equations (3b) and (3c) reveal two interesting consequences of ignoring heterogeneity in preferences (Keane 1997). First, the error variance will differ across choice occasions as the price *P* and attribute *X* are varied. If one estimates an MNL model with a constant error variance, this will show up as variation in the intercept and slope parameters across choice occasions. In a stated choice experiment context, this could lead to a false conclusion that there are order effects in the process generating responses.

Second, equation (3c) shows how preference heterogeneity leads to serially correlated errors. That heterogeneity is a special type of serial correlation is apparently not well understood in the transportation literature. To obtain efficient estimates of choice model parameters one should include a specification of the heterogeneity structure in the model. But more importantly, if preference heterogeneity is present it is not merely a statistical nuisance requiring correction. Rather, one must model the heterogeneity in order to obtain accurate choice model predictions, because the presence of heterogeneity will alter cross-price, lead to IIA violations, etc. This is just one example of how paying attention to the behavioural source of the error terms in a choice model leads to new insights into how the model should be estimated, interpreted and applied. Much of the criticism of models is often based on the application of poorly specified models accompanied by inadequate data which fail to capture the real behavioural processes underlying travel decisions.

A hierarchy of models are evolving in the literature, relaxing progressively some of the testable assumptions imposed on the restrictive MNL form (Figure 5). We will concentrate on a number of the incremental improvements in the behavioural structure of choice models recognising that the λ's are more than noise parameters - they are behavioural opportunities to capture real behavioural processes.




$z_{qC}$ = A vector of agent-specific characteristics and/or contextual variables
$x_{iq}$ = A vector of alternative-specific attributes which vary across the agents

Multi-period Multinomial Probit
with IID or non-IID
autoregressive errors
unobserved heterogeneity (random effects)
inter-alternative correlation

**Figure 5. Taxonomy of Behaviourally Progressive Models (leading to the general Multiperiod Multinomial Probit)**

We have selected three more general choice models for discussion - the Heteroskedastic Extreme Value (HEV) model in its random and fixed effects form, the Random Parameter (or Mixed Logit ) (RP/ML) model and the Multi-Period Multinomial Probit (MPMNP) model. The latter is a generalisation of all of the other model structures.

### *Heteroskedastic extreme value model - random effects HEV*

In reality there may be noticeable unobserved effects with relatively high or low variance associated with one or more alternatives, in contrast to other alternatives such that the constant variance assumption can over or under estimate the indirect utility from the observed effects of the former alternatives relative to the latter.

The HEV model removes the condition of *identically distributed* random components associated with the MNL model while maintaining zero inter-alternative correlation. Thus the $\lambda$'s can vary across the alternatives. This model can also be treated as a probit specification with heteroskedastic normal distribution (HND) on the standard deviations of the random components. Allenby and Glinter (1995), Bhat (1995, 1997a) and Hensher (1997, 1998a) have implemented the HEV model; Hensher (unpublished) has implemented the HND model. The probability density function and the cumulative distribution function of the random error term for the *i*th alternative under an unrestricted variances form with scale parameter $\lambda_i$ for the HEV unobserved effects are given as Equation (4)

$$f(e_i) = \frac{1}{l_i} e^{-\frac{t_i}{l_i}} e^{-e^{-\frac{\epsilon_i}{l_i}}} \text{ and } F_i(z) = \int_{e_i=-\infty}^{e_i=z} f(e_i)\, de_i = e^{-e^{-\frac{z}{l_i}}} \qquad (4)$$

The probability that an individual will choose alternative *i* ($P_i$) from the set *C* of available alternatives is given in equation (5).

$$\begin{aligned} P_i &= \text{Prob}\,(U_i > U_j),\ \text{for all}\, j \neq i,\ j \in C \\ &= \text{Prob}\,(e_j \leq V_i - V_j + e_i), \text{for all}\, j \neq i,\ j \in C \\ &= \int_{e_i=-\infty}^{e_i=+\infty} \prod_{j \in C,\, j \neq i} F\left[\frac{V_i - V_j + e_i}{l_j}\right] \frac{1}{l_i} f\left(\frac{e_i}{l_i}\right) \mathrm{d}e_i, \end{aligned} \tag{5}$$

where $f(\ .\ )$ and F( . ) are the probability density function and cumulative distribution function of the standard type 1 extreme value distribution, respectively. If the scale parameters (ie the) of the random components of all alternatives are equal, then the probability expression in eqn (5) collapses to that of the multinomial logit. The HEV model avoids the pitfalls of the IID property of the multinomial logit model by allowing different scale parameters across alternatives.

The HEV model is flexible enough to allow differential substitution among all pairs of alternatives. When the observed utility of some alternative *l* changes, this affects the observed utility differential between another alternative *i* and the alternative *l*. However, this change in the observed utility differential is tempered by the unobserved random component of alternative *i*. The larger the scale parameter (or equivalently, the standard deviation) of the random error component for alternative *i*, the more tempered is the effect of the change in the observed utility differential and smaller is the elasticity effect on the probability of choosing alternative *i*. We illustrate the implications of elasticities in a Section 8 when we compare the MNL and HEV models for mixtures of preference data.

### *Heterogenous agent- and context-specific segmentation nested logit: fixed effects HEV and latent segmentation partitioned logit*

Traveller behaviour research focuses extensively on segmenting potential and actual choosers of each alternative in the offered choice set. There are two primary segmentation strategies - by benefit segment (ie utility ranges) and by agents' characteristics. Sources of unobserved variance are candidates for identification through some functional mapping with characteristics of agents as well as data-specific effects (eg task complexity, collection method). A few studies have implemented a latent class segmentation model within the framework of a set of partitioned MNL models (Swait 1994), or a nested logit framework within which the scale parameter ($\lambda_{it}$) varies between branches of the partitioned model but is invariant within a branch between alternatives ie $\lambda_i = \lambda_j\ \forall\ j \in J$ (McFadden 1981). A typical nested logit model with an upper (U) and a lower (L) level is summarised in equations (6)-(9).

$$P_{ul} = \frac{\exp[\lambda_u(\underline{v}_u + \underline{v}_{u*})]}{\sum_{u' \in U} \exp[\lambda_u(\underline{v}_{u'} + \underline{v}_{u'*})]} \cdot \frac{\exp(\lambda_l \underline{v}_{ul})}{\sum_{l' \in L} \exp(\lambda \underline{v}_{ul'})} \tag{6}$$

$$\underline{v}_{u*} = \frac{1}{\lambda_u} \log \sum_{l' \in L} \exp(\lambda_l \bar{v}_{ul'}) \tag{7}$$

$$\lambda_u = \frac{\pi}{\sqrt{6\sigma_U^{+}}} \tag{8}$$

$$= \frac{\pi}{\sqrt{6}} \left[ \sigma_U^2 + \frac{\pi^2}{6\lambda_l^2} \right]^{-\frac{1}{2}} \tag{9}$$

By normalising the value of lambda at the lower level to unity, the latent segmentation model specifies the (fixed) variance of the unobserved effects at the upper level, $\lambda_u$, as a function of the characteristics of each agent (in principle one could include the attributes of the alternatives) with data collection variables included to 'cleanse' the segments of bias due to noise in information gathering (see equation 2). This is equivalent to a fixed effects HEV model and is referred to as a *latent segmentation logit* model (Figure 5 and Swait 1994 ). Swait and Adamowicz (1996) propose an information theoretic or entropy indicator to represent task complexity and accumulated task complexity in stated preference repeated measures surveys and include it in the latent segmentation function.

### *Random parameter logit (RPL) or mixed logit model*

The utility expression is the same as that for a standard MNL model except that the analyst may nominate one or more taste weights (including alternative-specific constants) to be treated as random parameters with the variance estimated together with the mean. The layering of selected random parameters can take a number of predefined functional forms, typically assumed to be normally or lognormally distributed. The normal form is $\beta_{qk} \sim N(\beta_k + \nu_{qk})$ where $\beta_k$ is the mean response sensitivity across all observations for attribute *k*, and $\nu_{qk}$ represents random taste variation of individual q around the mean. The lognormal form is often used if the response parameter needs to be a specific sign: $\beta_{qk} \sim \pm \exp(\beta_k + \nu_{qk})$.

This form has important behavioural implications. The presence of $\nu_{qk}$ terms as a representation of random tastes of individual q invariant across the choice set, can induce a correlation among the utility of different alternatives (Bhat 1997, McFadden and Train 1996). It is the mixture of an EV1 distribution for the overall utility expression and embedded normality for the distribution of the taste weights across a sample which has led to the phrase 'mixed logit' (Train, 1997). Specifically, by treating the deviation around the mean taste weight as a component of the random component such that we have $\nu_{qk}x + \varepsilon_i$, the RPL model has been interpreted as an error-components model, where the first component can take on any distributional assumption and the second component is assumed to be EV1. One can also choose to treat the random effects, $\nu_{qk}x$, as different across the alternatives but independent (ie different standard deviations); or as different across alternatives and inter-alternative correlated.

This engenders a relatively free utility structure such that IIA is relaxed despite the presence of the IID assumption for the random components, $\varepsilon_i$, of the alternatives. That is, the RPL model disentangles IIA from IID and enables the analyst to estimate models which account for cross-correlation among the alternatives. When the random taste weights are all zero, the exact MNL model is produced. Applications of the RPL/mixed logit model are given in Bhat (1996), Revelt and Train (1996), Brownstone et al 1997 and McFadden and Train (1996). Bhat (1997) has superimposed random response heterogeneity over the systematic response heterogeneity by including parameterised covariates ($Z_{qk}$) in the function: $\beta_{qk} \sim \pm \exp(\beta_k + \gamma_k Z_{qk} + \nu_{qk})$.

The comparison of the Random Effects HEV and the RPL model is quite informative. To make them directly comparable, we can specify an RPL model in which only the alternative specific constants ($\alpha_j$) are random. The form of the two models is given in (10) and (11).

$$\text{HEV:}\quad U_j = \alpha_j + \boldsymbol{\beta}'\mathbf{x_j} + \varepsilon_j \text{ with } \varepsilon_j = \text{EV1 distribution}$$
$$E(\varepsilon_j) = 0,\ \text{Var}(\varepsilon_j) = \pi^2 \sigma_i^2 / 6;\ \sigma_J^2 = 1. \tag{10}$$

$$\text{RPL:}\quad U_j = \mu_j + \boldsymbol{\beta}'\mathbf{x_j} + \varepsilon_j,\ \varepsilon_j = \text{EV1 distribution} \tag{11}$$

same as HEV, but $\sigma_j = 1$ for all j. (IID)
$\mu_j = \alpha_j + u_j$, where $\mu_j$ = a constant and $u_j = N[0,\sigma_j^2]$

Inserting the expression for $\alpha_j$ in the utility function produces a model which is identical to HEV except that the distributions of the error terms are very different: HEV has EV1 ($\sigma_j$) while RPL has $\varepsilon_j + u_j$.

### *Multi-period multinomial probit*

The most general model in the sense of the specification of the variance-covariance matrix of random effect components is the ***M**ulti**P**eriod-**M**ulti**N**omial **P**robit* (*MPMNP*) model. Special cases of this model accommodate all variations of assumptions about the autoregressive structure, the correlation of unobserved effects between alternatives and time periods, explicit treatment of unobserved heterogeneity across the sampled individuals, and differential variances across alternatives. The models presented in Sections 4.1-4.3 are special applications of the more general model. The generalised utility expression can be written out as equation (12).

$$i_{qt}=\text{argmax}_{j=1...I}(u_{jqt}=X_{jqt}\alpha + y_{qt}\beta_j + \varepsilon_{jqt}) \quad (12)$$

where $i_{qt}$: observed discrete *choice* (or *alternative*) by agent q in time t, i=1....I, t=1....$T_i$
$u_{jqt}$: latent *utility* of alternative j as perceived by agent q in time t
$X_{jqt}$: alternative-specific *attributes* of alternative j as perceived by agent q in time t
$y_{qt}$: agent-specific *characteristics* of agent q in time t
$\varepsilon_{jqt}$: multinormal error with cov($e_q$)=$\Omega$ ( $e_q = (e_{jqt})_{j=1...I,\ t=1...T}i$ )
where $\Omega$ is I×$T_i$, permitting interalternative and intertemporal correlation between $e_{jqt}$ and $e_{kqs}$ for the same agent q, and
$\alpha$, $\beta$j, and $\Omega$: to be estimated

A range of covariance structures evolve from this general formulation:

- Contemporaneous Correlations and Heteroscedasticity of $e_{qt} = (e_{jqt})_{j=1...I}$
  MNP: General deviations from IIA
- Intertemporal Correlations between $e_q = (e_{jqt})_{j=1...I,\ t=1...T}i$
  RAN: Random effects, specific to alternatives;
  AR1: First-order autoregressive errors, specific to alternatives

Combinations of these error processes yield the models summarised in Table 3. To illustrate the behavioural implications of alternative assumptions, we estimated a series of models using a 1997 stated choice switching data set for high speed rail in the car non-business market for the Sydney-Canberra corridor. Each sampled car traveller was asked to review four alternative high speed rail options defined by fare class (first, full economy, discount economy and off-peak), frequency (every 30 minutes, hourly and two hourly), and parking cost ($2 - $20 per day); and asked to select one of them or stay with the car for the current trip. This was repeated up to 4 times. All 355 individuals completed up to 2 profiles and 81 completed up to 4 profiles.

To illustrate the behavioural implications of alternative assumptions, we have derived the mean behavioural values of non-business travel time savings, reported in Table 3 together with the log-likelihood at convergence. The variation in the VTTS is substantial, ranging from a low of $4.63/adult person hours for the most restrictive model, up to $8.37/adult person hour for a less restrictive specification. This is nearly a doubling of the VTTS which has major implications for transport investments, given the important role played by time benefits in most transport project appraisal. Close inspection of Table 3 suggests that the failure to account for correlation between the alternatives and free variance are the major contributing influence on the downward biased MNL mean estimate of VTTS. Allowing for unobserved heterogeneity (through random effects) and serial correlation through a first order autoregressive structure, contributes far less to an increase in the mean VTTS (relative to the MNL estimate).

All of the models can be described by a likelihood function which is a product of the choice probabilities (equation 13) across the sample of q=1,...,Q individuals, i=1,...,I alternatives and t=1,....T time periods.

$$L(b,M) = \prod_{q=1}^{Q} P(\{i_{tq}\} \mid \{X_{itq}\}; b, M \tag{13}$$

**Table 3. Alternative Error Processes in Discrete Choice Models**

| Mode l | Error Processes | RAN | AR1 | MNP | VTTS* | LogL |
|---|---|---|---|---|---|---|
| 1 | iid across periods, iid across alternatives | 0 | 0 | 0 | **4.63** | -1067.9 |
| 2 | iid across periods, correlated across alternatives | 0 | 0 | 1 | **6.46** | -1050.71 |
| 3 | random effects, iid across alternatives | 1 | 0 | 0 | **5.22** | --765.01 |
| 4 | random effects, correlated across alternatives | 1 | 0 | 1 | **6.88** | --759.57 |
| 5 | AR1 errors, iid across alternatives | 0 | 1 | 0 | **4.98** | -811.46 |
| 6 | AR1 errors, correlated across alternatives | 0 | 1 | 1 | **7.87** | -770,38 |
| 7 | random effects + AR1 errors, iid across alt's | 1 | 1 | 0 | **5.40** | -775.68 |
| 8 | free variance, random effects, iid across alts | 1 | 0 | 1 | **8.37** | -759.71 |
| 9 | free variance and iid across periods | 0 | 0 | 1 | **7.64** | -1040.25 |
| 10 | free variance, iid across periods, correlated across alts | 0 | 0 | 1 | **8.06** | --1044.3 |
| 11 | free variance, random effects, AR1 errors, correlated across alt's | 1 | 1 | 1 | **7.09** | -759.04 |

* Dollars per adult person hour, **-** = not able to identify an appropriate model

The cumulative distribution function is assumed to be multivariate normal or extreme value type 1 and characterised by the covariance matrix *M*. Estimating the parameters in (13) is a complex task when we move beyond the simple MNL and NL models. In the most general case we need to evaluate E = (I-1)*T dimensional integral for each agent and each iteration in the maximisation of the (log) likelihood function. What makes this particularly complex is the inter-alternative correlation on one or more of the error components. Numerical integration is not computionally feasible since the number of operations increases with the power of E, which dimensions the covariance matrix. Simulation of the choice probabilities is now the preferred method of estimating all parameters, by drawing pseudo-random realisations from the underlying error process (Boersch-Supan and Hajivassiliou 1990). The popular method is one initially introduced by Geweke (and improved by Keane, McFadden, Boersch-Supan and Hajivassiliou - see Geweke et al 1994, McFadden and Ruud 1994) of computing random variates from a multivariate truncated normal distribution. Although it fails to deliver unbiased multivariate truncated normal variates (as initially suggested by Ruud and detailed by Boersch-Supan and Hajivassiliou (1990), it does produce unbiased estimates of the choice probabilities. The approach is quick and generated draws and simulated probabilities depend continuously on the parameters β and M. This latter dependence enables one to use conventional numerical methods such as quadratic hillclimbing to solve the first order conditions for maximising the simulated likelihood function (equation 14)- hence the term simulated maximum likelihood (SML) (Stern 1997).

$$\bar{L}(b, M) = \prod_{r=1}^{R} \prod_{q=1}^{Q} \bar{P}_r(\{i_q\}) \tag{14}$$

Boersch-Supan and Hajivassiliou (1990) have shown that the choice probabilities are well approximated by the formula (15), even for a small number of replications. Our experience suggests that 100 replications is sufficient for a typical problem involving 5 alternatives, 1000 observations and up to 10 attributes. Such runs with appropriate software and a fast pentium (200 Mhz, 64 RAM) should take about 5-15 minutes to converge. All models in Table 3 except Model 1 were estimated using SML with 100 replications in the simulation estimator.

$$\bar{P}(\{i_q\}) = \frac{1}{R}\sum_{r=1}^{R} \bar{P}_r(\{i_{qn}\}) \tag{15}$$

## *The Way Forward*

Travel behaviour researchers have much still to learn about the nature of gains (if any) in moving to more complex travel choice models. Parsimony must remain a major objective, but it must be justified through creative use of more advanced ways of studying the complexities of behaviour and behavioural response. Through better quality data we can emphasise a richer set of explicit explanatory variables and reduce the dependence on complex error variance-covariance structures. For the time being we must recognise the benefits of potentially more behaviourally realistic (albeit statistically complex) models and continue to ask fundamental questions about the suitability of simpler models as sources of information on predictive response, elasticities and marginal rates of substitution between attributes. We have already raised the issue of differences in marginal rates of substitution associated with behavioural values of travel time savings; we now turn to the issue of the suitability of elasticities from simple models and mixtures of preference data.

## Tread Carefully Beyond the Current Market

A major growth industry in travel behaviour research and practice is the design and application of stated preference experiments (see Hensher 1994 for a review and Hensher, Louviere and Swait (forthcoming) for state of the art and practice applications). So popular are these methods that the great majority of 'respectable' travel choice and demand modellers have included an SP experiment (to vary degrees of detail, rigour and reliability) in their empirical analysis of current and future travel markets. Indeed this popularity has become so strong that we see many studies discarding market data (revealed preference data) in favour of 'stand-alone SP' models to explain and predict travel demand.
The reason for combining data sources is simple - to take advantage of the extended richness that a range of data sources may offer the empirical analyst, leading to a more robust set of taste weights for understanding and predicting behavioural responses. Since the standard deviations of the unobserved effects are not distinguishable from the taste weights, except under the unit-normalisation assumption of the simple MNL model, and that the former may be different between data sets due to the differential influence of the unobserved effects, we need to take this into account in the specification of models combining multiple data sources. Within the general class of random utility models, the link between $\lambda$ and the standard deviation of the unobserved effects (see Section 4) provides a powerful mechanism for combining data from multiple sources (Morikawa 1989, Bradley and Daly 1992, 1994) and revealing the profile of differential error variance between data sets. The *'case of the lurking lambda's'* has become synonymous with the literature on combining sources of preference data.

As we have come to learn more about the strengths and limitations of the SP paradigm, it is increasingly apparent that such stand-alone approaches are very unreliable as procedures for predicting behavioural response (ie elasticities) and hence predicting travel demand. Their greatest strength lies in their contribution to the enrichment of the taste weights (and hence valuation) associated with the attributes of alternatives in the choice set of interest. Since the attribute space for SP models is richer and not subject to measurement error compared to RP models, the valuations are less subject to constraints associated with real markets which distort the full preference space in which 'willingness to pay' should be identified.

The preferred paradigm involves the merging (or fusing) of market and SP preference data, interpreting SP experiments as the opportunity to enrich the utility space within which attributes describing alternatives can be processed. The attribute space represented in market (or revealed preference) data is limited to the attribute mix offered by existing technology (eg the limits on travel times of each mode) (Figure 6).

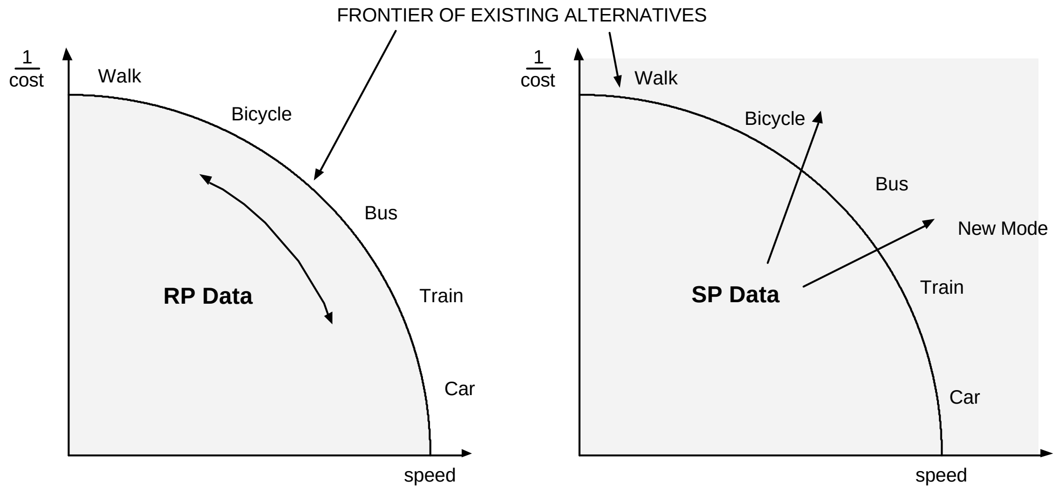



**Figure 6. The attribute space in revealed and stated markets**

Table 4 from Hensher (1998a) illustrates the implications of deriving elasticities from simple MNL vs HEV models using stand-alone SP, RP and combined RP-SP models. Hensher and Louviere (1998) undertook a similar comparison using a different data set for MNL, nested logit and HEV for mixtures of SP and RP data, producing conclusions that are consistent with the behavioural implications summarised below. The set of fare elasticities associated with a joint SP-RP model are based on the use of the SP parameter estimates for fare and cost, rescaled into the RP model, which provides the choice probabilities and fare (or car cost) attribute levels. We report the direct and cross elasticities from the SP partition of the joint SP-RP HEV model, the joint SP-RP MNL model and the stand-alone SP-MNL and RP-MNL model. The reported cross elasticities for a joint and stand-alone MNL model are, however, uninformative because of the imposed iid condition.

Each column provides one direct share elasticity and 6 cross share elasticities. For example, the column headed TW tells us that a 1% increase in the train weekly fare leads to a 0.093% reduction in the proportion of daily one-way trips by train on a weekly fare. In addition, this 1% single fare increase leads to a 0.001% higher proportion of one-way trips on a train travel pass and .004% increase in one-way trips on a bus travel ten ticket.

## Table 4. Direct and Cross Share Elasticities (Hensher, 1998a)

Note: Elasticities relate to the price per one-way trip. The RP elasticity precedes the SP elasticity in any pair. SP direct and cross elasticities from the HEV model are in parenthesis ( ). The direct elasticities from the stand alone RP and SP MNL models are in square brackets [ ]. Cross-elasticities for the stand-alone SP MNL model and the stand-alone RP MNL model are given in [ ]. The MNL RP and SP direct and cross elasticities are in brackets { } from the joint SP-RP MNL model. The interpretation for a specific fare class is obtained under each column heading.

| | TS | TW | TP | BS | BT | BP | Car |
|---|---|---|---|---|---|---|---|
| Train single | **-.218** (-.702) | .001 (.289) | .001 (.149) | .057 (.012) | .005 (.015) | .005 (.009) | .196 (.194) |
| (TS) | [-.161 -.517] | [.146,.110] | [.031,.067] | [.052,.035] | [.025,.041] | [.021,.024] | [.427,.601] |
| | {**-.057**,-.317} | {.134, .073} | {.004,.039} | {.048,.023} | {.012,.029} | {.018,.018} | {.134,.199} |
| Train weekly | .001 (.213) | **-.093** (-.635) | .001 (.358) | .001 (.025) | .001 (.024) | .006 (.019) | .092 (.229) |
| (TW) | [.062 , .087] | [-.057,-.313] | [.031,.067] | [.052,.035] | [.025,.041] | [.021,.024] | [.427,.601] |
| | {.054, .053} | {**-.018**,-.197} | {.004,.039} | {.048,.023} | {.012,.029} | {.018,.018} | {.134,.199} |
| Train travel | .001 (.210) | .001 (.653) | **-.196** (-1.23) | .001 (.023) | .012 (.022) | .001 (.017) | .335 (.218) |
| pass (TP) | [.062 , .087] | [.146,.110] | [-.111,-.597] | [.052,.035] | [.025,.041] | [.021,.024] | [.427,.601] |
| | {.054, .053} | {.134, .073} | {**-.002**,-.368} | {.048,.023} | {.012,.029} | {.018,.018} | {.134,.199} |
| Bus single | .067 (.023) | .001 (.053) | .001 (.031) | **-.357** (-.914) | .001 (.248) | .001 (.286) | .116 (.096) |
| (BS) | [.062 , .087] | [.146,.110] | [.031,.067] | [-.217,-.418] | [.025,.041] | [.021,.024] | [.427,.601] |
| | {.054, .053} | {.134, .073} | {.004,.039} | {**-.141**,-.239} | {.012,.029} | {.018,.018} | {.134,.199} |
| Bus travel | .020 (.020) | .004 (.037) | .002 (.023) | .001 (.206) | **-.160** (-.462) | .001 (.163) | .121 (.090) |
| ten (BT) | [.062 , .087] | [.146,.110] | [.031,.067] | [.052,.035] | [-.083,-.268] | [.021,.024] | [.427,.601] |
| | {.054, .053} | {.134, .073} | {.004,.039} | {.048,.023} | {**-.017**,-.159} | {.018,.018} | {.134,.199} |
| Bus travel | .007 (.025) | .036 (.063) | .001 (.034) | .001 (.395) | .001 (.290) | **-.098** (-.700) | .020 (.103) |
| pass (BP) | [.062 , .087] | [.146,.110] | [.031,.067] | [.052,.035] | [.025,.041] | [-.072,-.293] | [.427,.601] |
| | {.054, .053} | {.134, .073} | {.004,.039} | {.048,.023} | {.012,.029} | {**-.005**,-.154} | {.134,.199} |
| Car (C1) | .053 (.014) | .042 (.023) | .003 (.013) | .066 (.009) | .016 (.011) | .003 (.006) | **-.197** (-.138) |
| | [.062 , .087] | [.146,.110] | [.031,.067] | [.052,.035] | [.025,.041] | [.021,.024] | [-.130,-.200] |
| | {.054, .053} | {.134, .073} | {.004,.039} | {.048,.023} | {.012,.029} | {.018,.018} | {**-.265**,-.361} |

The results offer many implications. The differences in direct elasticities between the SP and RP choice sets reflects the different probabilities of choice. As is well known, although often ignored, studies which derive elasticities from stand-alone SP models produce different switching propensities to the RP estimates because the SP experiment is often searching in a more expansive utility space of choice opportunities, producing a different probability profile than an RP model. It is necessary to 'return' the parameter power of an SP model back to the RP space regardless of whether new alternatives are introduced to the market or existing alternatives removed. Since an elasticity calculation uses three inputs - a predicted choice probability, a taste weight (and a scale parameter in an HEV model) and an attribute level, the appropriate probabilities for predicting switching behaviour in the current market must come from a base or enhanced RP model.

For HEV direct elasticities, sensitivity within the commuter rail and bus markets decreases as we move from a single ticket through to multiple-trip tickets with the exception of train travel pass. For the MNL direct elasticities, the trend downwards in sensitivity is consistent across both train and bus markets. This has interesting implications for a fares policy - increasing the price of a multi-use ticket, especially in the bus market, offers higher revenue growth prospects for small losses of patronage than is the case for single tickets. The HEV cross elasticities suggest that there is more movement between modes for a given fare class than between fare classes within modes.

A comparison of the HEV and MNL revealed preference elasticities shows a systematically lower set of direct elasticity estimates for all public transport alternatives in the MNL model (and vice versa for car); thus we might conclude that an SP model tends to produce lower elasticities than its RP counterpart where the SP choice probabilities are higher than the RP probabilities (which is the situation herein). The MNL direct elasticity estimates for public transport alternatives tend to be lower than their HEV counterparts in both RP and SP models (and vice versa for car). The implication, *if generalisable (given the observation that the less chosen modes in an RP setting are chosen more often in an SP setting)*, is that all previous studies which have used an MNL framework and/or a stand-alone SP model specification have made sizeable errors in their estimation of direct share elasticities. Since the majority of travel choice studies have adopted this MNL framework, the findings are quite troublesome for the extant literature. Hensher (1998a) provides more policy interpretation.

## Valuation Functions Instead of Point Estimates for Travel Time Savings Valuation

The importance of the behavioural valuation of travel time savings (VTTS) has been recognised for many years (typically yielding over 60% of total user benefits); however in recent years, the introduction of tollroads has highlighted the critical nature of time savings in return for a toll and hence the implied valuation of such a trade. Forecasting traffic levels and revenue is inextricably tied up with the way that time savings are traded for a toll. With private sector investment so dependent on the reliability of toll revenue forecasts, behaviourally accurate VTTS's have become even more important than they have been in the past for publicly provided roads.

It is typically assumed that the marginal rates of substitution between any two attributes vary by a number of 'market segments' such as mode, trip purpose, trip length and personal income. As the heterogeneity of traveller preferences increases, the challenge to identify appropriate segments to capture this heterogeneity in a managed way increases. Indeed in the limit we might treat each individual in the sampled population as a segment with unique empirical valuation properties. Currently, within each highly aggregated market segment, point estimates of VTTS are routinely derived and applied as mean estimates of an unknown (but assumed) distribution. In addition, a strictly linear interpretation of the marginal rate of substitution between travel time and travel cost is assumed within the segment.

Little research has been undertaken to understand the *true* profile of VTTS within each segment or the consequent implications on traffic forecasts of ignoring the distribution of values. Some researchers (eg Horowitz, 1998) have suggested that the parametric models used to derive point estimates or distributions of values may also be inappropriate and require a non- or semi-parametric treatment (ie do we really have confidence in the EV1 or multivariate normal distribution? Maybe the true distribution is bi-modal?). In the few serious empirical studies, the set of segment-specific values are at best treated as threshold values with limited appreciation of the 'best' set of cut-off points if segmented single values are required. For example, the VTTS's for short, medium and long trips are preconditioned on an exogenous definition of each trip length. Bradley and Gunn (1990), for example, select a series of travel speeds with arbitrary ranges. Whether these arbitrary cut-off points are appropriate is not considered.

Valuation functions enable us to seek out a continuous distribution of values and to partition this distribution in a more useful way if a set of point estimates are required for practical analysis. Alternatively the richness of the continuous distribution can be preserved and implemented (see Hensher and Truong 1984, Bradley and Gunn 1990, Ben-Akiva et al 1993). So important are small movements in an average point estimate of VTTS on revenue (given the implied time-toll trades) that much more research is required into the implications of alternative specifications of the full distribution of VTTS throughout the heterogeneous population.

Much more research is required to identify the distribution of VTTS derived from the non-linearity inherent in the attributes of alternatives (such as travel time and travel cost). Such research emphasises a *value function* defined as a functional relationship between VTTS and levels of time and cost both in respect of higher-order and interaction effects. To begin, assume that the theoretical parameter $\kappa_i$ from a theoretical indirect utility function of the linear additive form (Truong and Hensher 1985, Jara Diaz in press):

$$V_i = \alpha_i - \lambda C_i - \kappa_i T_i \tag{16}$$

is a function of $C_i$ and $T_i$:

$$\kappa_i = \kappa\ (T_i, C_i) \tag{17}$$

A Taylor series expansion of (17) around the mean levels $\overline{T}$ and $\overline{C}$ for each alternative *i* (neglecting second order terms) results in equation (18).

$$\kappa_i = \overline{\kappa} + (\partial\kappa/\partial T)_i\ (T_i - \overline{T}) + (\partial\kappa/\partial C)_i\ (C_i - \overline{C}) \tag{18}$$

Substitution of (18) into (16) and some rearrangement of terms gives:

$$V = \alpha_i - \lambda C_i - \overline{\kappa} T_i + (\beta T_i^2 + \gamma C_i T_i + \omega) \tag{19}$$

where

$\beta = (\partial\kappa/\partial T)_i$, $\gamma = (\partial\kappa/\partial C)_i$, and $\omega = -\beta\overline{T} - \gamma\overline{C}$.

By neglecting second-order terms in (18) we implied that $\partial\kappa/\partial T$ and $\partial\kappa/\partial C$ are constants, independent of alternative i. Equivalently, the parameters $\omega$, $\beta$ and $\gamma$ are unsubscripted. The VTTS can now be derived from (19) as follows (Hensher and Truong 1985, Hensher 1995):

$$\begin{aligned} \text{VTTS} &= \frac{\partial V/\partial T_i}{\partial V/\partial C_i}\bigg| V_i = \text{constant} \\ &= \frac{-\overline{\kappa} + \gamma C_i + 2\beta T_i}{-\lambda + \gamma T_i} \end{aligned} \tag{20}$$

Thus VTTS is dependent on the levels of travel time and cost. This formula can be generalised to account for the disaggregation of travel time. We can introduce interactions between each travel time and between travel time and other attributes of alternatives. The ability to enrich the valuation function to test for a richer specification is conditioned by the quality of data.

Revealed preference data is usually somewhat limiting in its ability to offer sufficient richness in both variability and correlation structure to enable each potential influence to be included without producing confoundment. This is particularly true when accounting for non-linearity and when 'new' alternatives are assessed for market share. Data derived from a stated choice experiment however increases the opportunity to account for the independent (ie additive) contribution of each source of variability in the valuation function in an expanded choice set. Hensher (1998b) has estimated a series of mode choice models and derived a distribution of VTTS (see Figure 7 for two modes - drive alone and ride share), and illustrated yet again the dependency of the VTTS on the specific assumptions underlying the random component (Table 5). A richer forecasting exercise should replace the mean estimates with a distribution of estimates using either a sample enumeration procedure of synthetic households with accompanying population weights; or some multi-way classification of the population by rich segments such as trip length by trip purpose by time of day by direction of travel by income class.

**Table 5. Mean Estimates of Commuter Values of Travel Time Savings ($/person hour)**

| Main mode | BL | HEVL | BnlL * | HEVnlL * |
|---|---|---|---|---|
| Drive Alone | 6.50 | 6.69 | 7.21 | 7.12 |
| Ride Share | 6.50 | 6.69 | 7.21 | 7.12 |
| Bus | 7.51 | 3.44 | 6.37 | 7.54 |
| Train | 7.51 | 3.44 | 6.37 | 7.54 |
| Light Rail | 7.51 | 3.44 | 6.37 | 7.54 |
| Busway | 7.51 | 3.44 | 6.37 | 7.54 |

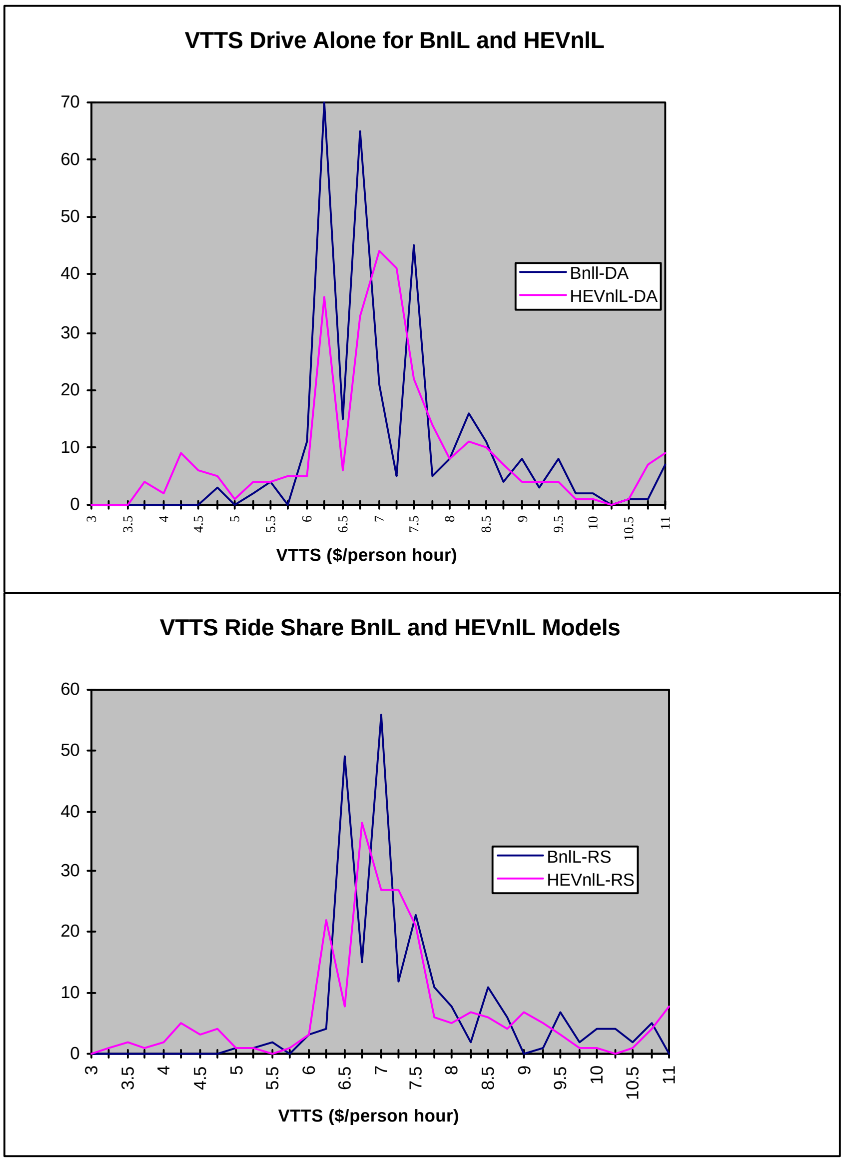



**Figure 7. Distribution of Value of Travel Time Savings for Drive Alone and Ride Share Commuters**

## Mapping what Agents Think with Policies which gain Political Favour - Attitudes Still Matter

Understanding travel behaviour can be broadly interpreted to include an understanding of the role of diverse sources of information on the decisions taken by politicians and their advisers which impact on travel behaviour. Great store should be placed on the importance of establishing a mapping between the views on specific potential policy and strategic issues and the stakeholder domain from which various degrees of support and opposition might evolve. Government agencies can use this information in positioning specific strategies and developing marketing plans to ensure that stakeholder support is maximised. Such a formula is likely to be attractive to the political process.

Hensher and Golob (1998) have extended the basis of traveller behaviour research into the domain of stakeholder-political matching of travel behaviour potentials in the context of the opinions of commercial freight operators concerning the priorities of various policies for transport planning and management. The survey instrument measures attitude in terms of overall opinions about the worthiness of each of a series of infrastructure investment priorities and policy options for the management of freight and commercial vehicle travel. The attitudes are measured in terms of a five-point scale, with the scale point descriptors being (1) “very bad idea,” (2) “bad idea,” (3) “neither good nor bad idea,” (4) “good idea” and (5) “very good idea.” The balance of this section draws on Hensher and Golob (1998a).

Three interrelated problems common to attitude surveys require special attention: First, attitudes can only be measured on scales that are ordinal, not cardinal. That is, favour or disfavour is monotonically related to the scale value, but it should not be presumed that the intervals between adjacent scale points are equal. Consequently, linear statistical analyses applied to the raw data (such as product-moment correlations, linear regression, and principal components factor analysis) will not necessarily yield accurate conclusions about relationships in the data because such methods assume equal intervals on the measurement scales.

Second, where interest is in evaluating a large number of infrastructure investment priorities and policy options, respondents are likely to judge many of them as being similarly good or bad ideas, and they may not have formed attitudes towards many of the initiatives. Thus, we can expect high levels of association among groups of attitudes, which need to be summarised by identifying patterns in attitudes. Third, we wish to determine how similarities in attitudes are related to the industry type represented by each respondent, since the mapping between attitudes and industry type is the foundation for the marketing of policy.

Given the objective of finding the best explanation of patterns in attitudes as a function of industry type, we have a nonlinear canonical correlation analysis (CCA) problem with an explanatory variable matrix defined by a single nominal (industry type) variable and a dependent variable matrix defined by a series of ordinal attitude scales. The linear combination on the explanatory variable side is undefined, because we have no metric to quantify the categories of each nominal variable. The linear combination of the variables on the dependent side is also undefined, because the categories of each variable can be re-scaled by any nonlinear function that preserves monotonicity. Thus, we need to optimally scale or quantify the variables while simultaneously solving the traditional linear CCA problem of finding weights for each explanatory variable.

An elegant solution to the nonlinear CCA problem was first proposed by researchers at the Department of Data Theory of Leiden University in the Netherlands. They developed a method for conducting canonical correlation analysis with variables of mixed scale types: nominal, ordinal, and interval (Gifi 1990). The method simultaneously determines both (1) optimal re-scalings of the nominal and ordinal variables and (2) explanatory variable weights, such that the linear combination of the weighted re-scaled variables in one set has the maximum possible correlation with the linear combination of weighted re-scaled variables in the second set. Both the variable weights and optimal category scores are determined by minimising a loss function derived from the concept of "meet" in lattice theory.

A nonlinear CCA solution involves, for each canonical variate, weights for all the variables, optimal category scores for all ordinal and nominal variables, and a canonical correlation. Graphical representations are very important in interpreting this plethora of results. Several authors have argued that graphical representations are even crucial in understanding the results of linear multivariate methods, particularly linear CCA, because patterns in the data can best be detected by the eye (Cailliez and Pagès, 1976; Ter Braak, 1990).

### *Interpreting the CCA solution in mapping attitudes and industry stakeholders*

To interpret the results of a nonlinear CCA solution for data with $p$ canonical variates, it is useful to generate a $p$-dimensional plot of the weights of the optimally scaled attitude variables and the weights of the nominal industry-type variable quantified for each canonical variate. The upper bound on $p$, the number of canonical variates, is the minimum of the number of attitude variables and the number of industry types (categories of the nominal variable). Analysts generally aim for a two-dimensional canonical solution ($p = 2$) due to the convenience of two-dimensional plots (Gifi, 1990); solutions in higher dimensions generally require multiple pair-wise plots. Optimal dimensionality of a CCA solution is determined by comparing canonical correlations and by further criteria detailed in Gittins (1985).

A second plot or series of category score plots provides the remainder of the information required to interpret a nonlinear CCA solution. Multiple treatment of the industry type variable results in different category scores on each canonical variate for this nominal explanatory variable, so a plot of the category scores in the space of the canonical variates allows us to visualise which industry or industries are associated with high or low values of each canonical variate. By comparing the component loadings and category scores plots we can then relate industries directly to attitudes towards policy initiatives.

### *Results*

The empirical investigation divided the infrastructure investment priorities and policy options into four classes - existing road infrastructure, new road infrastructure, other proposed infrastructure, and broad-based policy initiatives. A total of twenty initiatives, listed in Table 6, were evaluated. The location of specific initiatives is summarised in the footnote to Table 6. The optimal scaling method was implemented separately in each of the four classes of policy initiatives. We report the findings for the first category - existing infrastructure initiatives (see Hensher and Golob 1998a for the full set of results).

## Road infrastructure initiatives

Differences in attitudes towards the potential road infrastructure changes among the five categories of business sectors were identified by a two-dimensional nonlinear generalized canonical analysis yielded canonical correlations of 0.394 for the first dimension and 0.280 for the second. The first canonical dimension explains 70% of the variance of its object scores, while the second dimension explains approximately 64% of the variance in its object scores. These statistics indicate that a two-dimensional canonical solution provides fairly strong relationships between the two sets of variables, the optimally scaled ordinal attitude scales on one hand and the quantified five-category business sector variable on the other (Gittins, 1985). A three-dimensional solution was rejected, as the canonical correlation for the third dimension drops to 0.198. [1]

The key results from the CCA are graphed in Figures 8 and 9. The component loadings in Figure 8 reveal that attitudes towards the five potential operational changes align along two approximately orthogonal dimensions through the origin. The first dimension, rotated about fifteen degrees from the first canonical variate, passes between "freight vehicle-only lanes" and "roundabouts with wider lanes" in its negative domain and close to "freight vehicles on bus lanes during peak periods" in its positive domain. This shows that optimally scaled attitudes towards freight vehicle-only lanes and roundabouts with wider lanes are strongly positively correlated, and attitudes towards both are strongly negatively correlated with optimally scaled attitudes towards freight vehicles on bus lanes during peak periods.

[1] A comparative analysis conducted with the five attitude scales treated as numerical (linear), rather than ordinal, scales yielded canonical correlations of only 0.252 and 0.206. This improvement in canonical correlations demonstrates that treating the attitudinal scales as ordinal substantially improves the explanation of differences in attitudes among the five business sectors.

## Table 6. Proposed policy initiatives tested in the survey

| Policy Initiatives (Scale: 1=very bad idea, .... 5=very good idea) |
|---|
| **Existing Road Infrastructure Initiatives:** |
| parking restrictions on major roads from 6am - 9pm |
| B-double access to local road network |
| freight vehicles allowed on bus lanes during peak periods |
| freight vehicle only lanes |
| roundabouts with wider lanes |
| **New Road Infrastructure Initiatives:** |
| an orbital road around the Sydney CBD about 30 kms out |
| an orbital road around the Sydney CBD about 40 kms out |
| extension of the M5 east to Port Botany and Kingsford Smith Airport |
| Eastern Distributor |
| **Other Proposed Infrastructure:** |
| railheads and inland ports |
| location of Sydney's third airport at Badgery Creek |
| location of Sydney's third airport at Holdsworthy |
| proposed rail interchange terminal at Chullora |
| proposed rail interchange terminal at Bathurst |
| current rail interchange terminal at Blaney |
| common user terminal at Port Botany |
| **Policy Changes:** |
| plan transport for 24-hr. needs of people and freight rather than peak period demand |
| regulatory changes to allow collection and distribution centres to be open 24 hrs. |
| improved education of car drivers to improve attitudes towards trucks |
| priority to intermodal linkages, especially rail |

Notes: The M5 East Extension is a major freeway in the South West connecting into the M5 - a private tolled road. Badgery Creek and Holdsworthy (near Liverpool) are locations in Sydney's West. Chullora is near Enfield approximately 10 km from the Sydney CBD; Blaney and Bathurst are over the Blue Mountains at least 2 hours from Sydney CBD.

Referring to Figure 9, the first canonical variate separates contract carriers from retail, wholesale and distribution firms and, to a lesser degree, manufacturing and extraction companies. Thus, contract carriers are more in favour of operating freight vehicles on bus lanes during peak periods, while freight vehicle only lanes and, to a lesser degree, roundabouts with wider lanes are favoured by retail, wholesale, distribution, manufacturing and extraction firms.

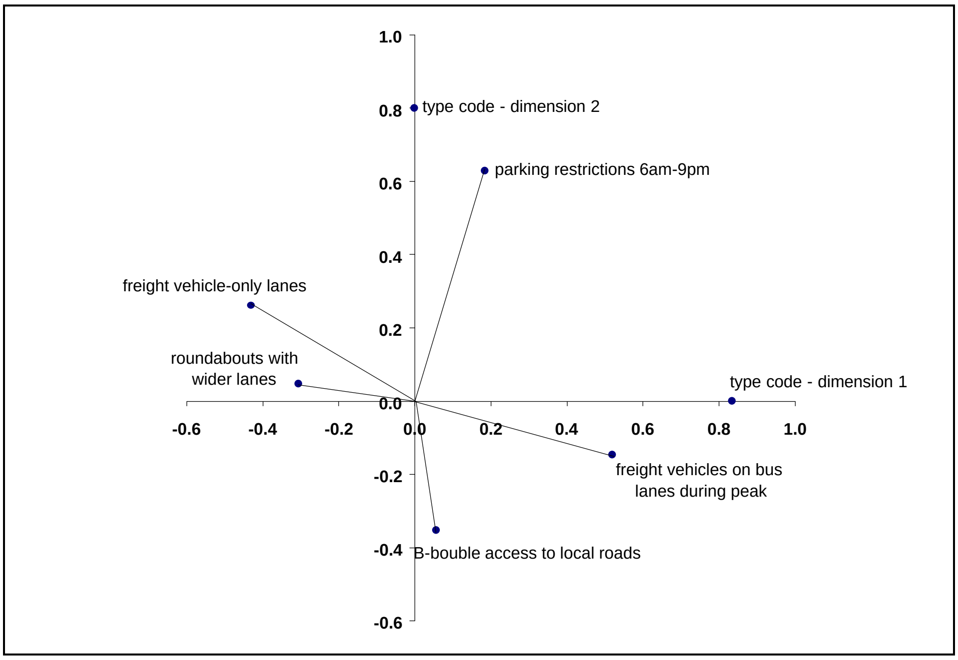



**Figure 8. Non-linear canonical analysis of attitudes towards road infrastructure initiatives versus business sector: Component loadings for the optimally quantified attitude scales**

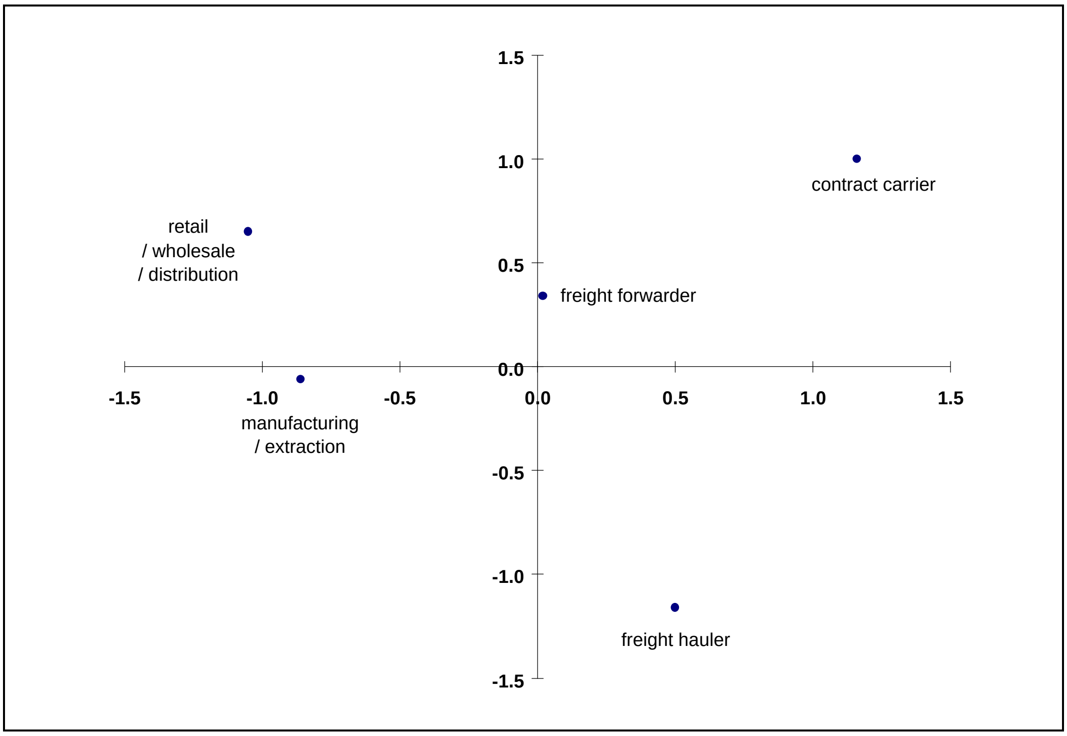



**Figure 9. Non-linear canonical analysis of attitudes towards road infrastructure initiatives versus business sector: Category scores for the business sector variable**

The second canonical variate is closely aligned with a dimension that distinguishes two negatively correlated policy initiatives: daytime parking restrictions, on the positive side of the dimension, and B-double access to local roads on the negative side (Figure 8). Freight hauliers, as indicated by their negative category score on the second dimension, tend to be more in favour of B-double access, while contract carriers, and to a lesser extent, retail, wholesale and distribution firms, prefer daytime parking restrictions (Figure 9). This in intuitively plausible given the dominant amount of urban goods movement on arterial roads by contract carriers. Of the five categories of firms, freight forwarders exhibit the least strong opinions about these five road infrastructure initiatives, as indicated by the position of this category near the origin of the category scores plot in Figure 9.

The method used to map attitudes about policy initiatives into the stakeholder domain provides an important framework for targeting (and hence marketing) specific policies in the market place. This marketing can be as much to reinforce the value of support for a specific policy or set of policies as it can be to more fully inform specific stakeholders about the benefits of specific policies where there is limited support. The implications in a political market of this process initially driven by an appreciation of stakeholder behavioural intent is very clear.

## The Timing of Forecasts: A Major Political Challenge

With an increasing presence of the private sector in transportation as owners or long term franchisees of transport infrastructure such as toll roads, tunnels, traditional railways, light rail, high speed rail, airports and ports, there is very close scrutiny of the forecasts of revenue reported in prospectus provided to potential investors. A significant component of revenue is attributable to the traffic using the facility. This is also the most risky element of the calculation of the financial viability of a major infrastructure project.

There is one thing known for certain about traffic forecasts - they will be wrong. But to what degree, why and when? The ‘when’ is extremely important and is often neglected in the marketing of the forecasts. Any major new infrastructure initiative such as a toll road will produce diverted and induced traffic throughout its life; however what is of major importance is establishing the hurdle date at which one can conclude with a high degree of certainty that the traffic levels have stabilised under the current service regime. To expect that this will happen almost immediately that new infrastructure is in service is foolhardy, yet it is the ‘food’ of the media and lobby groups opposed to specific infrastructures.

We need analytical support tools which can assist in identifying the hurdle date(s) at which one should pronounce a comparison between forecasts and actual traffic, and hence revenue flows. An understanding of ways of establishing such a capability should be a major agenda item for travel behaviour researchers. One way forward is through the incorporation of a longitudinal perspective rich with data on transitions between states, durations in states and variables capable of explaining such transitions and durations. In this section, we review the method of continuous time event history as an appealing framework provided there is a commitment to the collection of data capable of representing travel behaviour through time. Such data is very rare indeed.

Structurally, the data needed to model an event history perspective is illustrated by the example provided in Figure 10. In this example, five individuals are sampled to obtain information on commitment to a new vacation package. Information on the package choices of these individuals is collected over some period of time until the survey is terminated at time C. At time C, there will likely be a group of individuals (e.g. individuals represented by person 2) that either; a) will never try the new package, or b) will eventually try the new package, but just have not done so up to time C. The duration spells of these individuals will be censored since they are not observed trying the new vacation package. This type of censoring is referred to as right-censoring.

Another type of censoring could arise if the survey was begun some time after the new package was introduced. In Figure 10, for example, if the survey was started at time B, it may be difficult to determine when an individual (such as individual 3) was first exposed to the new vacation option. Being unable to determine when duration's begin is referred to as left-censoring. Left-censoring poses the additional problem of not knowing the value of the determinants of duration (e.g. income, household size, attitudes) at the beginning of the duration period. Left-censoring can be avoided, in this case, by beginning the survey when the vacation package is first introduced (time A). This will ensure full knowledge of the lengths of durations as well as possible determinants of durations. The duration model in its statistical form is referred to as a hazard function.

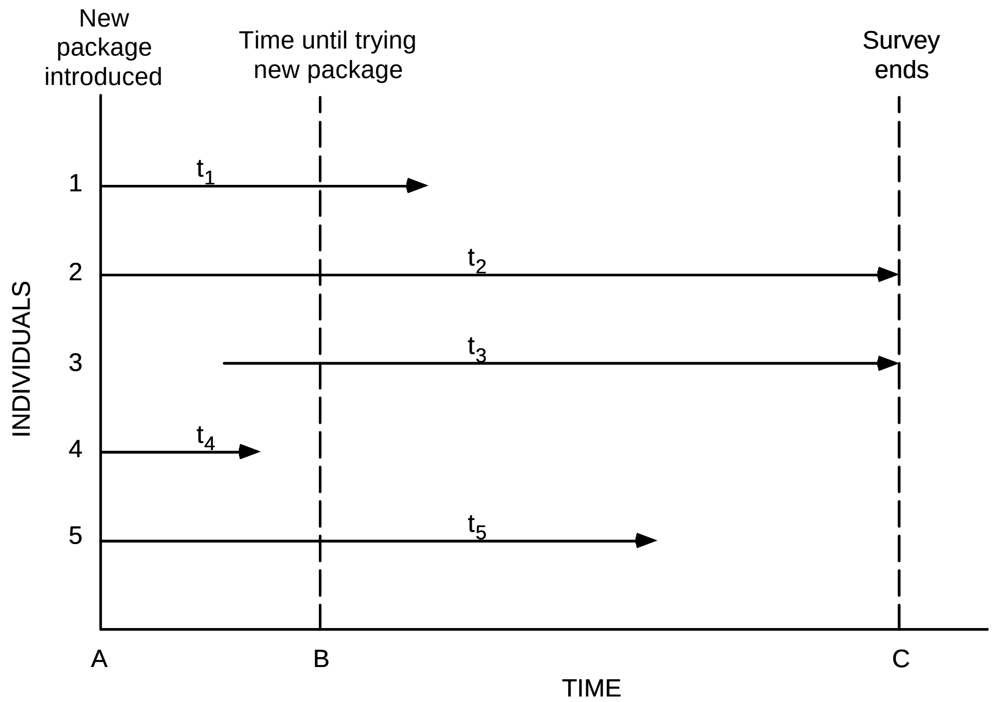



**Figure 10. Example of Event History Data**

### *The fundamental equations for a duration model of activity timing*

Formally, the hazard function can be expressed in terms of a cumulative distribution function, F(t), and a corresponding density function, f(t). The cumulative distribution is written as,

$$F(t) = \text{Prob}[T < t] \tag{21}$$

where Prob denotes the probability, T is a random continuous time variable, and t is some specified time. Equation (21) for example, identifies the probability of replacing a vehicle before some transpired time (assuming no left censoring) or switching travel to a tollroad after a particular date. The corresponding density function is

$$f(t) = dF(t)/dt \tag{22}$$

and the hazard function is,

$$h(t) = f(t)/[1 - F(t)] \tag{23}$$

where h(t) is the conditional probability that an event will occur between time t and t+dt given that the event has not occurred up to time t:

$$\text{Prob}\ (T_o \geq t+1 |\ T_o \geq t) \tag{24}$$

Information relating to duration dependence, as derived from the first derivative of the hazard function with respect to time (ie. its slope) provides insights into the duration process being modelled. Plotting the hazard function against time gives important empirical information for the parameterisation of the baseline hazard (Hensher and Mannering 1994, Bhat 1997). The probability of ending a duration or spell in a particular state may be dependent on the length of the duration. There may also be important determinants of duration (eg. socioeconomic characteristics) that should be included in the modelling approach. These covariates are included in hazard-based models using alternative methods such as *proportional hazards* and *accelerated lifetime*. Establishing the length of time before someone changes state (eg switching from a free route to a new toll road) and an explanation for the duration between states is the motivation for an interest in these methods. The challenge is to seek out behaviourally appropriate functional forms to explain

state duration.
Proportional hazards models operate on the assumption that covariates act multiplicatively on some underlying or baseline hazard function. The proportionality is due to the decomposition of the hazard rate into one term dependent upon time, and another dependent only on the covariates (Prentice and Gloeckler 1978). To accommodate time varying covariates we can assume that they are well approximated by their mean over the interval. This gives a clue to the interval size for continuous time, given the particular application (Hensher 1997a). A relatively general form of the hazard is specified as:

$$h_o(t) = \lambda_b(t) \exp (z_o(t)\beta) \tag{25}$$

where $\lambda_b(t)$ is an arbitrary baseline hazard and $\exp (z_o(t)\beta)$ is the parametric component including time varying covariates associated with an origin state *o*. A discrete set of time intervals are observed, which can be very small to capture the essence of continuous time. The conditional probability rule in (24) translates into the following function, given (26):

$$\text{Prob } (T_o \geq t+1 | T_o \geq t) = \exp ( -\exp (\gamma (t) + z_o(t) \beta)) \tag{26}$$

where

$$\gamma (t) = 1n \left(\int_t^{t+1} \lambda_b(u)du\right) \tag{27}$$

u is any function in terms of time. The model allows for a continuous 'failure' time $T_o$ and (right) censoring $c_o$, but with observation taking place only at $t_o$, t = 0,1,2,..., *J*-1, or in the final interval (*J*, ∞). If the baseline hazard is assumed to be well approximated by its mean over the time interval, it is completely captured by the single term $\gamma$ (t). Left censoring may exist if an event was well under way when the panel commenced. Right censoring exists since the endpoint of the last episode of an individual cannot be observed.

### *Competing risks and multispell models*

The dominating emphasis in empirical analysis of event history data, particularly in transportation, but also in economics (eg. Lancaster 1979) and marketing (eg. DuWors and Haines, 1990)) involves the study of a single initial or origin state, a single final or destination state, and a single period of time between successive events, often referred to as a single episode or spell. An example of the singular dimensionality would be studying the time before a traveller switches from a free public route to a tolled private route (Hensher

1997a). Multistate (or competing risks) and multispell situations are common in transportation, but they impose substantial complexity on the estimation of models. The combination of complexity and the general absence of packaged software for multistate and multispell models has limited applications, despite the realism (Hensher 1998, Bhat 1997)

Transport researchers usually assume that a competing risks model with *n* possible outcomes, has a likelihood function that could be separated into *n* distinct pieces. Under such an assumption, estimation could proceed by estimating separate hazard models for each of *n* possible outcomes. Gilbert (1992), for example, introduced a competing risks specification and separate estimation for three transitions in a study of automobile ownership duration. Treating competing risks independently is analogous to assuming recursivity in more traditional simultaneous equations problems, which can be solved using three-stage-least squares and similar methods (Hensher and Mannering 1994, Bhat 1997).

Some researchers also regard the various spells as being analysed as independent events, and apply the methods developed to handle single spells. This is problematic if the populations are heterogenous which would result in a mixing that may lead to a time dependency and incorrect inferences. Since transport applications are characterised by high levels of interdependency between variables, the homogeneity assumption is quite improbable. Incorporating observed and unobserved heterogeneity is necessary, or at least should be tested. Segmentation by socioeconomic characteristics is partially useful - it is however unable to handle the sources of unobserved heterogeneity (and its probable correlation with duration dependence). The importance of introducing time varying covariates and unobserved heterogeneity into a proportional hazards (PH) model is appreciated when it is understood that the PH model, in the presence of time invariant covariates, assumes that the ratio of the hazard for any two sampled members of a population should be constant throughout the observation (i.e. it is independent of time).

Two important issues in the study of event histories are (i) ways of capturing the unobserved heterogeneity in the sampled population and (ii) the dependency of duration and states over time (Diamond and Hausman 1984, Han and Hausman 1990, Sueyoshi, 1992, Meyer, 1986,1990, Bhat, 1996 and Hensher 1998). These phenomenon accommodate elements of the dynamics of event histories which influence the nature of transitions. To introduce these ideas, it is useful to define the information requirements of an event history, and then introduce the essential formulae required to parameterise a competing risks multispell duration model as extensions of equations (25) - (27).

An event history of a sampled household over some observed time period requires information on (i) the initial state (ii) the number of spells in the observation period (iii) the points in time at which some state transition has occurred or a specific event has taken place (iv) state occupancies corresponding to the above points in time (v) an indicator that identifies whether a particular spell is censored and (vi) the set of covariates, measured at the beginning of each spell. Covariates take on three forms: time invariant (e.g. sex), time dependent (e.g. age), and time varying (e.g. lifecycle stage).

Heterogeneity in general is handled by a mixing distribution over separate (but jointly estimated) hazard functions. A popular way of incorporating heterogeneity is as a random multiplicative factor that shifts the baseline hazard:

$$h_o(t) = \exp(\theta_o \lambda_b(t)) \exp(z_o(t)\beta) \tag{28}$$

where $\theta_o$ is a random variable associated with initial state *o* with a distribution defined by the analyst, representing the distribution of the unobserved heterogeneity within the population of sampled households. The random variable must be limited to positive values (given that the hazard rate is not negative). If we set $E(\theta) = 1$, then on average one obtains $\lambda_b(t)$. Parametric specifications have been investigated, especially the gamma, normal, and logistic mixing distributions. Heckman and Singer (1984) proposed a non parametric finite mixture model to accommodate the highly sensitive nature of the parameter estimates associated with the covariates to alternative distributional assumptions. Unlike the parametric specification of $\theta_o$, a fraction of the population can have a zero hazard rate Trussell and Richards (1985).

The hazard model for a competing risk model can be defined as (Hensher 1998):

$$h(t_{od}|\mathbf{z},\theta) = \exp\{\delta_{odb} + z_o(t_{od} + \tau)\beta_{od} + \sum \delta_{odk} f_k(t_{od}) + c_{od}\theta\} \tag{29}$$

and $$f_k(t) = \sum_{k=1}^{K} (t_{od}^{\lambda_k} - 1)/\lambda_k; \quad \lambda_k = \mu_{odk} \tag{30}$$

where $\delta_{odb}$ is the baseline hazard for a multi-state model, $z_o(t_{od})$ defines time-varying covariates, $z_o\tau$ defines time dependent covariates, and $f_k(t_{od})$ is defined by equation (30) as a Box-Cox transformation over time to capture general duration dependence. Setting K = 1 and hence $\lambda_1 = 0$ gives a Weibull distribution; setting K= 1 and $\lambda_1 = 1$ gives a Gompertz distribution. Other functional forms are possible. For example, setting K = 2 with $\lambda_1 = 1$ and $\lambda_2 = 2$ produces a quadratic duration dependence. Lillard (1993) chose a piecewise linear spline to represent the dependence of hazards on calendar time. $c_{od}\theta$ is a weighted unobserved heterogeneity index, where $\theta$ is common across all transitions o to d, and the weight, $c_{od}$ conditions the unobservable scalar to have a differentiating role in different transitions or different spells. Equation (29) is a very general specification of a hazard function allowing for time varying covariates, unobserved heterogeneity and duration dependence. Setting $\beta_{od} = c_{od} = f_{k\,(t)} = 0$ gives an exponential form for the hazard function. Parametric or non-parametric assumptions can be imposed on $\theta$ as discussed above.

Equation (29) is the kernel of the specification of a multistate multispell model with allowance for time varying covariates, unobserved heterogeneity and duration dependence.

### *Empirical Illustration I - Automobile transactions (from Hensher 1998)*

A sample of 200 households from the Sydney automobile panel (Hensher et al. 1992) who provided complete information over a 12 year period on a limited number of socioeconomic and vehicle characteristics (Table 7) were used in the empirical application.

The purpose is to identify the influences on the probability that a sampled household will undertake a particular type of transaction over the period 1974-85 given the observation of one of three states in each time interval. The three states are no change, replace a used vehicle with a used vehicle, and replace a used vehicle with a new vehicle. Out of 2400 observations across 200 households and 12 years, we have 2011 (83.8%) states of no change, 235 (9.8%) replacements with a used vehicle and 154 (6.4%) replacements with a new vehicle. They represent 886 spells.

There are 9 possible transitions with the following sample sizes:

1. State 1 to state 2 (no change to replace used vehicle with a used vehicle) = 197,
2. State 1 to state 3 (no change to replace use vehicle with a new vehicle)= 137,
3. State 2 to state 1 (replace a used vehicle with a used vehicle to no change= 212,
4. State 3 to state 1 (replace used vehicle with a new vehicle to no change)= 140,
5. State 1 to state 1 (no change to no change) = 163,
6. State 2 to state 2 (replace a used vehicle with a used vehicle in both states) = 23,
7. State 3 to state 3 (replace a used vehicle with a new vehicle in both states= 14,
8. State 2 to state 3 = 0 (replace a used vehicle with a used vehicle and then a used vehicle with a new vehicle), and
9. State 3 to state 2 = 0 (replace a used vehicle with a new vehicle and then a used vehicle with a used vehicle).

We limit the empirical assessment to joint estimation of transitions 1 to 2 and 1 to 3. The average duration of the transition from no change to replace with a used vehicle is 3.90 years; the equivalent mean for a replacement with a new vehicle is 4.41 years. In the context of the household’s timing and duration of automobile transactions, four empirical model specifications are investigated:

M1: parametric baseline hazard, time varying covariates, no unobserved heterogeneity, duration dependence
M2: parametric baseline hazard, time varying covariates, unobserved heterogeneity, duration dependence
M3: parametric baseline hazard, no time varying covariates, no unobserved heterogeneity, duration dependence
M4: parametric baseline hazard, no time varying covariates, unobserved heterogeneity, duration dependence

In models M2 and M4 we investigate one parametric distribution - log normal - and a non-parametric finite mixture model for unobserved heterogeneity. Duration dependence is evaluated under Weibull, and Gompertz distributions.

**Table 7. The Data Set Used in Model Estimation**

| No. | Acronym | Definition | Mean (sd) |
|---|---|---|---|
| 1 | END | End of case identifier (1,0) | |
| 2 | YR | Year (74,75,76,....,85) | |
| 3 | STATE | State (1 = no change, 2 = replace used with used vehicle, 3 = replaced used with new vehicle) | |
| 4 | HSIZE | Household size | 2.96 (1.44) |
| 5 | NHINC | Number of Income earners in household | 1.67 (0.64) |
| 6 | LIFA | lifecycle A (1,0) young adults (<35), no children | 0.053 |
| 7 | LIFBCD | lifecycle BCD (1,0) two heads, children up to 12 years old | 0.196 |
| 8 | LIFEF | lifecycle EF (1,0) one or two heads, children over 16 years | 0.191 |
| 9 | LIFG | lifecycle G (1,0) older adults, no children | 0.228 |
| 10 | LIFH | lifecycle H (1,0) retired persons over 65 years old | 0.226 |
| 11 | LIFIJ | lifecycle IJ (1,0) single head, | 0.107 |
| 12 | RGHH | 1 or more vehs. are private registered (1,0) | 0.705 |
| 13 | REGHS | 1 or more vehs are household business registered (1,0) | 0.171 |
| 14 | REGOT | 1 or more vehs are other company registered (1,0) | 0.127 |
| 15 | LOCAL | Prime county of manufacture (1 = local, 0 = other) | 0.491 |

There are 5 time varying covariates - household size (HSIZE), number of income earners (NHINC), household stage in the lifecycle (LIF..), number of vehicles in each registration category (REG..) and the prime country of vehicle manufacture (LOCAL). Within the limits of the data a number of broad issues are worthy of investigation. In particular we want to evaluate the role that changing household life cycle and vehicle registration status plays in the households automobile replacement decision. To what extent are households loyal to the used car market or are willing to trade up to new vehicles? Automobile manufacturers are particularly interested in this question as might be proponents of alternative fuelled vehicles in the early formative years. Since there is almost certainly likely to be some important missing covariates, allowance for unobserved heterogeneity will be important to the results.

The set of models estimated under different assumptions on the form of duration dependence and unobserved heterogeneity for a given set of significant time varying covariates are summarised in Table 8. We have limited Table 8 to a sufficiently broad range of situations to illustrate the diversity of results.

## Table 8. Illustrative Model Results for Alternative Specifications (Hensher 1998)

| | Variables | DD=Weibull 1 to 2 | DD=Weibull 1 to 3 | DD=Gompertz 1 to 2 | DD=Gompertz 1 to 3 |
|---|---|---|---|---|---|
| UH=0 | constant | -3.139 (-9.33) | -4.301 (-10.4) | -2.820 (-8.65) | -3.424 (-9.71) |
| | gamma | 1.069 (9.22) | 1.723 ( 8.2) | 0.314 (9.50) | 0.399 (7.94) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | nhinc | -0.117 (-.77) | -0.109 (-.71) | -0.105 (-.72) | -0.082 (-.53) |
| | lifa | 0.624 (1.46) | -0.021 (-.04) | 0.598 (1.31) | -0.083 (-.13) |
| | lifbcd | 0.125 (0.45) | 0.128 (0.46) | 0.083 (0.30) | 0.098 (0.35) |
| | lifef | -0.045 (-.17) | -0.373 (-1.3) | -0.085 (-.33) | -0.038 (-1.31) |
| | lifg | 0.112 (0.44) | 0.215 (0.75) | 0.095 (0.37) | 0.191 (0.67) |
| | lifij | 0.101 (0.33) | -0.497 (-1.37) | 0.103 (0.36) | -0.483 (-1.34) |
| | regot | -0.195 (-.67) | -0.011 (-.03) | -0.126 (-.42) | 0.052 (0.16) |
| | reghs | -0.392 (-1.7) | -0.129 (-.56) | -0.414 (-1.80) | -0.153 (-.64) |
| | LL (0) | -1029.30 | | -1057.74 | |
| | LL (C) | -1021.46 | | -1052.30 | |
| UH=0 | constant | -3.278 (-18.5) | -4.462 (-13.95) | -2.963 (19.09) | -3.582 (-16.04) |
| | gamma | 1.025 (9.49) | 1.652 (8.52) | 0.298 (9.54) | 0.379 (8.27) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | LL (0) | -1069.32 | | -1061.8 | |
| | LL (C) | -1030.86 | | -1061.3 | |
| UH =lognormal | constant | -3.03 (-7.31) | -4.086 (-8.47) | -2.878 (-2.46) | -3.535 (-2.51) |
| | gamma | 1.112 (8.82) | 1.947 (8.10) | 0.325 (7.24) | 0.403 (5.40) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | nhinc | -0.127 (-.81) | -0.150 (-.85) | -0.117 (-.78) | -0.085 (-.53) |
| | lifa | 0.638 (1.42) | -0.008 (-.01) | 0.609 (1.32) | -0.060 (-.09) |
| | lifbcd | 0.156 (0.55) | 0.214 (0.70) | 0.085 (0.31) | 0.103 (0.36) |
| | lifef | -0.037 (-.14) | -0.370 (-1.15) | -0.094 (-.36) | -0.377 (-1.26) |
| | lifg | 0.128 (0.50) | 0.251 (0.82) | 0.093 (0.36) | 0.191 (0.65) |
| | lifij | 0.090 (0.29) | -0.547 (-1.44) | 0.122 (0.43) | -0.473 (-1.29) |
| | regot | -0.223 (-.75) | -0.085 (-.25) | -0.131 (-.44) | 0.045 (0.13) |
| | reghs | -0.375 (-1.49) | -0.052 (-.20) | -0.427 (-1.78) | -0.157 (-.66) |
| | factor loading | -0.103 (-.61) | -0.341 (-1.24) | 0.035 (0.05) | 0.070 (.09) |
| | LL (0) | -2132.55 | | -1047.72 | |
| | LL (C) | -1016.11 | | -1047.45 | |

***Table 8 continued***

| | | | | | |
|---|---|---|---|---|---|
| UH= lognormal | Constant | -3.187 (-10.4) | -4.293 (-9.4) | -2.806 (-16.3) | -3.357 (-12.2) |
| | gamma | 1.057 (9.36) | 1.778 (8.03) | 0.403 (11.1) | 0.649 (10.7) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | factor loading | -0.084 (-.45) | -0.239 (-.78) | -0.278 (-3.09) | -0.785 (-3.91) |
| | LL (0) | -1218.63 | | -1060.13 | |
| | LL (C) | -1025.85 | | -1046.30 | |
| UH=non-parametric | constant | -2.594 (-3.34) | -2.484 (-1.84) | -1.796 (-3.50) | -0.180 (-.27) |
| | gamma | 1.144 (8.95) | 2.104 (8.56) | 0.422 (9.61) | 0.768 (12.3) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | nhinc | -0.135 (-.83) | -0.184 (-1.00) | -0.131 (-.81) | -0.193 (-.99) |
| | lifa | 0.651 (1.42) | 0.0126 (0.02) | 0.707 (1.52) | 0.148 (0.24) |
| | lifbcd | 0.183 (0.63) | 0.310 (0.95) | 0.238 (0.84) | 0.544 (1.70) |
| | lifef | -0.023 (-.09) | -0.334 (-1.02) | -0.017 (-.06) | -0.200 (-.65) |
| | lifg | 0.139 (0.54) | 0.293 (0.94) | 0.203 (0.76) | 0.519 (1.72) |
| | lifij | 0.092 (0.29) | -0.553 (-1.45) | 0.095 (0.32) | -0.459 (-1.22) |
| | regot | -0.244 (-.82) | -0.146 (-.43) | -0.278 (-.94) | -0.351 (-1.127) |
| | reghs | -0.381 (-1.44) | -0.033 (-.12) | -0.447 (-1.77) | -0.176 (0.67) |
| | factor loading | -1.053 (-.79) | -4.014 (-1.45) | -2.152 (-3.17) | -7.614 (-6.20) |
| | support point | 0.841 (4.58) | 0.841 (4.58) | 0.805 (17.99) | 0.805 (17.99) |
| | LL (0) | -1021.46 | | -1052.30 | |
| | LL (C) | -1020.33 | | -1033.38 | |
| UH= non-parametric | Constant | -3.276 (-.00) | -4.457 (-.016) | -1.923 (-5.21) | -0.681 (-1.20) |
| | gamma | 1.025 (8.85) | 1.652 (5.65) | 0.402 (10.3) | 0.682 (12.5) |
| | lambda | 0.00 | 0.00 | 1.00 | 1.00 |
| | factor loading | -0.002 (-.00) | -0.005 (-.01) | -2.187 (-3.17) | -6.705 (-6.10) |
| | support point | 0.004 (0.01) | 0.004 (0.01) | 0.819 (17.4) | 0.819 (17.4) |
| | LL (0) | -6724.37 | | -1065.36 | |
| | LL (C) | -1030.87 | | -1047.64 | |

The Weibull distribution is a generalised form of the exponential distribution. The Weibull distribution imposes the monotonicity restriction on the hazard. We are able to identify whether loyalty to the used car market is time-dependent or time-independent. The Gompertz distribution, derived from the extreme-value distribution, is truncated at zero so that no negative values are possible. Unobserved heterogeneity is evaluated as a parametric lognormal distribution and as a non-parametric mixture specification. We have assumed 10 intervals on each side of the mean to approximate the lognormal distribution.

The hazard of replacing a vehicle with a used vehicle (transition 1 to 2) or with a new vehicle (transition 1 to 3) varies quite noticeably between the transition types and the distributional assumptions on duration dependence and unobserved heterogeneity. Beginning with no unobserved heterogeneity, the shape parameter (gamma) for duration dependence for both distributions is significantly positive in all models across both transitions suggesting that for both distributions the hazard is an increasing function of time. When we control for unobserved heterogeneity the shape parameter has a stronger influence on the hazard, increasing the expected time in a state, ceteris paribus.

The only three covariates approaching acceptable statistical significance are REGHS (household has at least one household-business registered vehicle) in transition 1 to 2 , and LIFBCD (households in lifecycle stage of two heads and children up to 12 years old) and LIFG (households with older adults and no children) in transition 1 to 3 for Gompertz duration dependence and non-parametric unobserved heterogeneity. The negative sign on REGHS suggests that the hazard of replacement with a used vehicle decreases, ceteris paribus, where households have access to a household-business registered vehicle relative to a privately registered vehicle. The life cycle effects are both positive implying that a household in either of these life cycle stages, ceteris paribus, has a higher hazard of replacement with a new vehicle.

A useful way of comparing the alternative specifications is to tabulate the hazard as a function of time. Given the statistical insignificance of the covariates we limit this to the models containing the scale, duration shape and unobserved heterogeneity parameters (Table 9). The predicted hazards in parenthesis relate to parametric unobserved heterogeneity. The Weibull and Gompertz specifications are monotonically increasing in duration implying that the longer a household goes without exiting a duration, the more likely it is to exit soon. The effect is stronger for transition 1 to 3 than transition 1 to 2.

The turnover is greater for used vehicles than new vehicles. For transition 1 to 2, the hazard is higher for the Weibull distribution for 2 to 6 years with the Gompertz producing a greater hazard for 7 to 10 years. For transition 1 to 3 the Gompetz has the higher hazard up to 2 years and after 7 years with the Weibull higher in the middle time durations. When allowance is made for unobserved heterogeneity we find some re-ordering of relativities and some significant adjustments in the hazard for transition 1 to 3: allowance for unobserved heterogeneity reduces the hazard with the gap increasing as duration increases. The difference for transition 1 to 2 is not noticeable at all. This leads one to conclude that failure to control for unobserved heterogeneity tends to lead to an over-estimate of the hazard for transitions involving replacement of a vehicle with a used vehicle, but its has no effect in the new car market.

Table 9 provides the important policy output for identifying the hurdle data for a settling down of a forecast. In the case of automobile acquisitions, we can see that on average, the probability of a household moving from a state of *no change* to a state of *replacing a used vehicle with another used vehicle* over 10 years varies from 0.396 to 0.843, a very large variance indeed. This illustrates the importance of research into the appropriate functional form of duration dependence. One should be able to take the probability profile and identify the weighted average probability of a trade up to a new vehicles given knowledge of the last time each household moved from State 1 to State 3 (or between any pair of states where the new state is state 3). When applied to forecast of tollroad traffic, for example, this methods enables one to seek out the time duration in which the probability of choosing the state 'tollroad' plateaus. The example of a tollroad is summarised below and presented in detail in Hensher (1997a)

**Table 9. Estimated Hazard Functions**

| Time (years) | DD=Weibull 1 to 2 | DD=Weibull 1 to 3 | DD=Gompertz 1 to 2 | DD=Gompertz 1 to 3 |
|---|---|---|---|---|
| 1 | 0.038 (0.037) | 0.012 (0.009) | 0.052 (0.051) | 0.028 (0.023) |
| 2 | 0.077 (0.076) | 0.036 (0.029) | 0.070 (0.079) | 0.041 (0.039) |
| 3 | 0.116 (0.115) | 0.071 (0.056) | 0.094 (0.102) | 0.059 (0.052) |
| 4 | 0.156 (0.155) | 0.114 (0.090) | 0.126 (0.129) | 0.087 (0.072) |
| 5 | 0.196 (0.195) | 0.165 (0.130) | 0.170 (0.173) | 0.127 (0.113) |
| 6 | 0.237 (0.235) | 0.223 (0.175) | 0.229 (0.230) | 0.185 (0.162) |
| 7 | 0.277 (0.275) | 0.287 (0.226) | 0.309 (0.311) | 0.270 (0.235) |
| 8 | 0.318 (0.315) | 0.358 (0.282) | 0.416 (0.417) | 0.395 (0.346) |
| 9 | 0.358 (0.355) | 0.435 (0.343) | 0.560 (0.558) | 0.577 (0.523) |
| 10 | 0.399 (0.396) | 0.518 (0.408) | 0.755 (0.753) | 0.843 (0.721) |

### *Empirical Illustration II - Forecasting the timing of traffic using a tollroad*

There is growing government interest in private sector supply and operation of new tolled motorways. The banking sector is keen to identify how long it takes for the traffic volume to reach a certain level and settle down, so they can obtain the best estimate of revenue required to make the investment financially attractive. Traditional forecasting procedures are not able to advise on this matter. It is an issue of the timing of change, and is well suited to duration modelling using event history data.

To illustrate the usefulness of duration models to forecasting tollroad traffic and revenue, we estimate a duration model to obtain the distribution of non-tolled route use time lengths and to identify the influence that a time-varying effect (i.e. travel time difference between the two routes) and a time-invariant effect (i.e. ownership status of the automobile - private or company car) might have on non-switching time length. The estimated hazard function can generate the distribution of probabilities of switching in or out of the tollroad state, and hence the traffic forecasts at each point in time.

The data is drawn from a tollroad which opened in Sydney in the early 90's. For each sampled individual we were able to identify the precise date of switching to the new tollroad after its date of opening. There is no left censoring since the tollroad state did not exist prior to the known commencement date. Right censoring exists since the endpoint of the last episode of an individual cannot be observed. We have allowed for right-censoring under the assumption that over the period of monitoring a number of individuals are still in the non-switching state.

A parametric duration model has been estimated in which the "survival time" is defined as the time from the commencement of the tolled route until the sampled individual switched to the tolled route. That is, the length of time until the user fails to continue with the free route.

The empirical results (see Hensher 1997a) suggest that the greater the time savings in using the tolled route, the less time an individual stays with the existing free route. That is the probability of failure increases. Likewise, individuals driving a company car are more likely to switch earlier than an individual driving a privately-registered automobile. The company car effect reduces the duration on non-switching. The distribution of times until switching (Table 11) suggest that at the sample means of the exogenous effects, that in the model assuming a homogeneous survival function, 95 percent of the sample remain in the state of non-switching 4.56 weeks after the commencement of the toll route, dropping to 75% after 10.34 weeks, 50% after 15.71 weeks and 25% after 21.83 weeks.

**Table 11. Results for the Weibull Single Risk Model of Toll Route Switching Allowing for Right Censoring and Null Switching (Endogenous Variable = ln(duration)).**

| Item | 25th percentile | 50th percentile | 75th percentile | 95th percentile |
|---|---|---|---|---|
| Survival | 0.25 | 0.50 | 0.75 | 0.95 |
| Time - Homogeneous | 21.83 | 15.71 | 10.34 | 4.56 |
| Time - Heterogeneity | 21.81 | 15.36 | 9.97 | 4.34 |

Average predicted failure probability for model (i) is 0.679 and for model (ii) it is 0.647.

When we allow for possible heterogeneity of the survival distribution across the sample, the survival periods for the four percentiles are shortened marginally respectively to 4.34, 9.97, 15.36 and 21.81 weeks. What we are observing is the timing of change and the role that the time savings and use of a company car have on the probability of staying in the state of no-switch. This is very important information for private financiers of major infrastructure where the revenue base is use-related. The importance of the particular application to the introduction of universal road pricing is also clear.

These illustrative applications of duration modelling motivate the importance of the timing of change. The most challenging features of this approach for ongoing research are the availability of high quality data in continuous time, handling a large number of discrete time periods, and the ability to forecast the historical relationships into the future.

## Future Directions - A Research Agenda for Practical Policy Initiatives

The contributions of traveller behaviour research presented in the previous sections are rich in detail and diversity, and demonstrate the progress that is being made as we strive to improve the frameworks and toolkits for investigating such behaviour. While there is much intellectual merit in pursuing new methods per se, the long term justification for a particular research path must be subject to the test of relevance in a policy formulating environment.

The research audit presented above, as selective as it may be, suggests a number of messages of value to the community of practitioners who strive to use these progressive frameworks and analytical methods in assisting the delivery of real world decisions on transport policy and planning. In concluding this lengthy paper, we offer the following interpretations as the basis of a research agenda for practical policy initiatives:

- There exist notable gains in our understanding of traveller behaviour from richer behavioural specifications of the ‘error’ structures of travel choice models. The implications for behavioural response (through elasticities), for valuation (such a value of travel time savings) and consequently prediction and forecasting are significant, as shown herein. This is not analytical sophistication for its own sake, but for the sake of more behavioural realism in policy advice and input into the planning process.

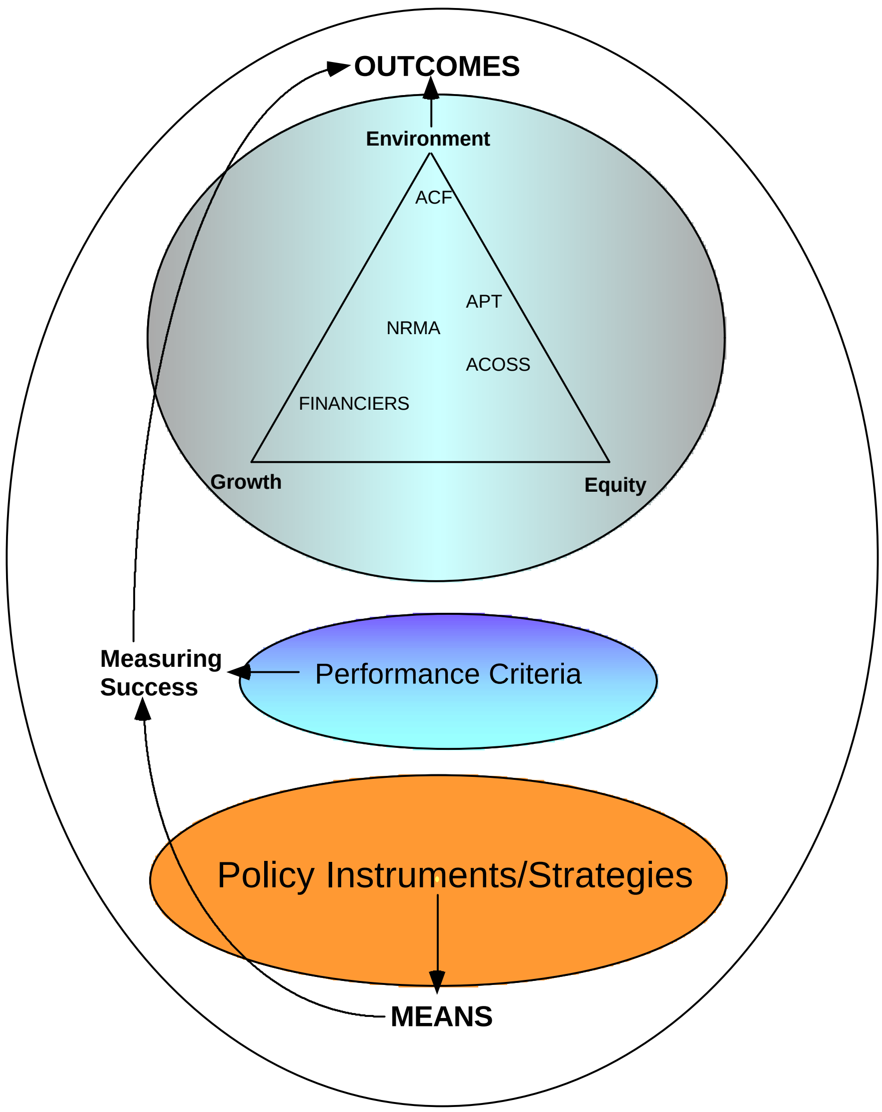
- Traveller behaviour research has shown little interest beyond exploratory research in the endogenous interactions between individuals, their household and the organisations they deal with on a regular basis (eg one’s work organisation). The potential contribution of interactive agency approaches to studying traveller behaviour, using game-theoretic and experimental choice frameworks appears high.

- There appear to be substantial benefits from combining sources of preference data, be it mixtures of revealed preference (market) data, mixtures of stated choice data or mixtures of RP and SC data. The evidence is accumulating to suggest that the application of stand-alone stated choice models in predicting/forecasting and in the derivation of elasticities is not to be encouraged. There are significant biases. Contrary, stand-alone stated choice models appear to provide suitable indicators of marginal rates of substitution such as the behavioural value of travel time savings.

- A great appeal of enriched data from multiple sources, especially when stated choice data is included, is the ability to develop valuation functions. Such rich extensions beyond point estimates enable planners and policy analysts to identify the appropriateness of mean estimates and to establish boundaries for segmented values.

- The timing of behavioural response remains one of the most challenging and important research topics. The reliability of forecasts and the pronouncement of the hurdle dates at which specific forecasts of traffic and revenue are deemed to be ‘stable’ is of major concern to investors in transport infrastructure, be they public or private. The risks attached to transport investment are high, and the largest is the revenue forecast produced by the traffic. 'When will the traffic settle down?' remains a key question.

- The set of advanced research tools reviewed above are useful contributions in themselves to the field of travel behaviour research; however their real worth will be forthcoming when we can integrate them into a framework capable of assessing the impact of the expansive set of potential policy instruments on levels of traffic, land use, the environment etc in accordance with a set of global performance measures such as improved accessibility, reduced traffic congestion, reduced global warming, increased air quality, and increased safety (Figure 14).




APT = Action for Public Transport, ACF = Australian Conservation Foundation, ACOSS = Australian Council of Social Security, NRMA = National Roads and Motorists Association

**Figure 14. Integrating Outcomes, Means and Measures of Success**

- Real progress in traveller behaviour research will be achieved when it can be translated into political gains. An important research task is the development of methods and data capable of providing the signposts to mapping policy and strategy which evolves from the study of traveller behaviour into the decision space of stakeholders who are influential in the political arena. Through the determination of a set of subjective weights, aligned with key stakeholders, which are attached to the importance of specific policy instruments which impact of travel behaviour, we can reveal the agenda necessary to promote and achieve the real gains that are offered by the outputs of traveller behaviour research. This link is currently extremely weak.

## Acknowledgments

Over the last 10 years I have been privileged to work with many researchers in a number of countries. While it is not possible to list all such individuals, I wish to especially thank Jordan Louviere, Joffre Swait, Tom Golob, Bill Greene, Chandra Bhat, Michael Beesley, Axel Boersch-Supan, Phil Goodwin, Joel Horowitz, Truong Truong, Ann Brewer, Jenny King, John Taplin, Mark Bradley, Andrew Daly, Tu Ton, Fred Mannering, Bill Waters and David Brownstone. The ideas on communication space and activity content evolved from a workshop in Texas in September 1997 - the intellectual contributions of Ira Salamon, Patricia Mokhtarian and Jackie Golob are acknowledged. The interactive agency choice experiments are the product of ongoing research with Ann Brewer. The research on the mapping of attitudes and stakeholders has been undertaken with Tom Golob. I dedicate this paper to the memory of Eric Pas.

# References

Allenby, G. and Ginter, J. (1995) The effects of in-store displays and feature advertising on consideration sets, *International Journal of Research in Marketing*, 12, 67-80.

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

Ben-Akiva, M., Bolduc, D. and Bradley, M. (1993) Estimation of travel choice models with randomly distributed values of time, *Transportation Research Record*, No. 1413, 88-97.

Bhat, C. (1995) A heteroscedastic extreme value model of intercity travel mode choice, *Transportation Research,* 29B (6), 471-483.

Bhat (1996) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling, forthcoming, *Transportation Research*

Bhat, C. (1997) 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 (1997a) Incorporating Observed and Unobserved Heterogeneity in Urban Mode Choice Models, under review, *Transportation Science*

Bhat (1997b) Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice, forthcoming, *Transportation Research*.

Boersch-Supan, A. and Hajvassiliou, V. (1990) Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models, *Journal of Econometrics*, 58 (3), 347-368.

Bradley, M. and Gunn, H. (1990) Stated preference analysis of values of travel time in the Netherlands, *Transportation Research Record*, No. 1285, 78-89.

Bradley, M. A. and Daly A. J. (1992) Uses of the logit scaling approach in stated preference analysis, paper presented at the *7th World Conference on Transport Research*, Lyon, July.

Bradley, M. and Daly, A. (1994) Use of the logit scaling approach to test rank-order and fatigue effects in stated preference data, *Transportation*, 21 (2), 167-184.

Brewer, A. and Hensher, D.A. (1997) Distributed Work and Travel Behaviour: The Dynamics of Interactive Agency Choices between Employers and Employees, paper presented at the *International Conference on Travel Behaviour Research*, Austin, Texas, September.

Brownstone, D., Bunch, D. and Train, K. (1997) Joint mixed logit models of stated and revealed preferences for alternative-fuelled vehicles, Conference Pre-prints, IATBR'97, *The 8th Meeting of the International Association of Travel Behaviour Research*, Austin, Texas, September.

Cailliez, F. and J.P. Pagès (1976). *Introdusction à l'Analyse des Données*. SMASH, Paris.

Diamond, P. and Hausman, J. (1984) The retirement and unemployment behavior of older men. In H. Aaron and G. Burtless (eds), *Retirement and Economic Behavior*, Brookings Institute, Washington, DC.

DuWors, R. and Haines, G.H. (1990) Event history analysis measures of brand loyalty, *Journal of Marketing Research,* XXVII (November), 485-493.

Geweke, J., Keane, M. and Runkle, D. (1994) Alternative computational approaches to inference in the multinomial probit model, *Review of Economics and Statistics*, LXXVI,(4), 609-632.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, Chichester.

Gilbert, C.C.M. (1992) A duration model of automobile ownership, *Transportation Research,* 26B(2), 97-114.

Gittins, R. (1985). *Canonical Analysis: A Review with Applications in Ecology*. Springer Verlag, Berlin.

Han, A. and Hausman, J. (1990) Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5, 1-28.

Heckman, J. and Singer, B. (1984) A method for minimising the impact of distributional assumptions in econometric models for duration data, *Econometrica*, 52, 271-320.

Hensher, D.A. (1994) Stated Preference Analysis of Travel Choices: The State of the Practice, *Transportation*, 21:107-133.

Hensher, D.A. (1997) A practical approach to identifying the market for high speed rail: a case study in the Sydney-Canberra corridor, *Transportation Research, 31A(6), 431-446.*

Hensher, D.A. (1997a) Timing of Change: Discrete and Continuous Time Panels in Transport Panels, In Golob, T., Kitamura, R. and Long, L. (eds), *Panels in Transport Planning*, Kluwer Scientific Publishers, Boston, 305-320.

Hensher, D.A. (1998) The Timing of Change for Automobile Transactions: A Competing Risk Multispell Specification in Ortuzar, J.D., Hensher, D.A. and Jara-Diaz, S. (eds) *Travel Behaviour*, Pergamon Press, Oxford.

Hensher, D.A. (1998a) Establishing a Fare Elasticity Regime for Urban Passenger Transport: Non-Concession Commuters *Journal of Transport Economics and Policy*, 32(2).

Hensher, D.A. (1998b) Extending Valuation to Controlled Value Functions and Non-Uniform Scaling with Generalised Unobserved Variances, in Garling, T., Laitilia, T. and Westin, K. (eds.) *Theoretical Foundations of Travel Choice Models*, Pergamon, Oxford.

Hensher, D.A, and Golob, J. (1998) Travel-telecommunications interactions, in Mahmassani, H. (forthcoming) *Recent Developments in Traveller Behaviour Research*, Pergamon Presss, Oxford.

Hensher, D.A. and Golob, T. F. (1998a) Searching for policy priorities in the formulation of a freight transport strategy: an analysis of freight industry attitudes towards policy initiatives, *Institute of Transport Studies*, University of Sydney, February.

Hensher, D.A. and Louviere, J.J. (1998) A Comparison of Elasticities Derived from Multinomial Logit, Nested Logit and Heteroscedastic Extreme Value SP-RP Discrete Choice Models, *paper presented at the 8th World Conference on Transport Research*, Antwerp, July.

Hensher, D.A and Truong, T.P. (1984) Valuation of travel time savings from a direct experimental approach, *Journal of Transport Economic and Policy*, XIX (3), 237-261.

Hensher, D.A. and Mannering, F.L. (1994) Hazard-Based Duration Models and Their Application to Transportation Analysis, *Transport Reviews*, 14 (1), 63-82.

Hensher, D.A., Smith, N.C., Milthorpe, F.M. and Barnard, P.O. (1992), *Dimensions of Automobile Demand: A Longitudinal Study of Automobile Ownership and Use* Amsterdam, North Holland.

Hensher, D.A., Louviere, J.J. and Swait, J. (forthcoming) Combining sources of preference data *Journal of Econometrics.*

Hensher, D.A., Louviere, J.J., and Swait, J. (in progress) *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation* (600pp, first draft).

Hensher, D.A., Boersch-Supan, A. and Brewer, A. (in progress*) Alternative behavioural perspectives on interactive agency choice*, Institute of Transport Studies University of Sydney and Department of Economics, Manheim University.

Horowitz, J. (1998) *Semi-Parametric Estimation Methods*, Springer-Verlag, Berlin.

Jara-Diaz, S. (in press) Time and income in discrete choice models, in Garling, T., Laitila, T. and Westin, K. (eds.) *Theoretical Foundations of Travel Choice Modelling*, Pergamon, Oxford.

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

Kitamura, R. (1987) A panel analysis of household car ownership and mobility, *Proceeding of Japan Society of Civil Engineers,* 338/IV-7 (Infrastructure Planning and Management), 13-37.

Kitamura, R. (1989) A causal analysis of car ownership and transit use, *Transportation*, 16(2): 155-173.

Koppelman, F.S. and Wen, C.H. (1997) Alternative nested logit models: structure, properties and estimation (mimeo).

Lancaster, T. (1979) Econometric methods for the duration of unemployment. *Econometrica*, 47(4), 939-956.

Lillard, L.A. (1993) Simultaneous equations for hazards: marriage duration and fertility timing, *Journal of Econometrics*, 56, 189-217.

Louviere, J.J. and Swait, J. (1996b) Searching For Regularities In Choice Processes, Or The Little Constant That Could, Working Paper, *Dept. of Marketing, Faculty of Economics*, University of Sydney, August.

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. and Ruud, P.A. (1994) Estimation by simulation, *Review of Economics and Statistics*, LXXVI,(4), 591-608.

McFadden, D. and Train, K. (1996) Mixed MNL models for discrete response, *Department of Economics,* University of California at Berkeley.

Meyer, B. (1986) Semi parametric estimation of hazard models, unpublished paper, Northwestern University, Evanston.

Meyer, B. (1990) Unemployment insurance and the distribution of unemployment, *Econometrica*, 58, 757-82.

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.

Nash, J.F. (1950) Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences* 36, 48-49.

Prentice, R.L. and Gloeckler, L.A. (1978)Regression analysis of grouped survival data with applications to breast cancer data, *Biometrics*, 34, 57-67.

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

Stern, S. (1997) Simulation-based estimation, *Journal of Economic Literature*, XXXV. December, 2006-2039.

Sueyoshi, G. (1992) Semi parametric proportional hazards estimation of competing risks models with time-varying covariates, *Journal of Econometrics*, 51, 25-58.

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

Swait, J., Louviere, J. and Williams, M. (1994) A sequential approach to exploiting the combined strengths of SP and RP data: application to freight shipper choice, *Transportation*, 21(2):135-152.

Ter Braak, C.J.F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, 55: 519-531.

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.

Truong, T.P. and Hensher, D.A. (1985) Measurement of Travel Times Values and Opportunity Cost from a Discrete-Choice Model, *Economic Journal*, 95 (378), 438-45.

Trussell, J. and Richards, T. (1985) Correcting for unobserved heterogeneity in hazard models: an application of the Heckman-Singer model to demographic data, *Sociological Methodology*, 242-276.

Zumkeller, D. (1997) Transport and telecommunications: first comprehensive simulation approaches, IATBR'97 Conference Pre-Prints, *Workshop on Telecommunications-Travel Interactions*, Austin, Texas, September 21-25.