WORKING PAPER
ITS-WP-97-8

Combining Sources of Preference Data: The Case of the Lurking λ's

by

David A. Hensher
Jordan Louviere
Joffre Swait

April, 1997

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-97-8
TITLE:	Combining Sources of Preference Data: The Case of the Lurking λ's

ABSTRACT:

This paper brings together several research streams and concepts that have been evolving in random utility choice theory: first, it reviews the literature on stated preference (SP) elicitation methods and introduces the concept of testing data generation process invariance across SP and revealed preference (RP) choice data sources; second, it proposes a general data generation process an useful framework for viewing this data combination process; third, it describes the evolution of discrete choice models within the random utility family, where progressively more behavioural realism is being achieved by relaxing strong assumptions on the role of the variance structure (specifically heteroscedasticity) of the unobserved effects. This latter topic is central to the issue of combining multiple data sources. Particular choice model formulations incorporating heteroscedastic effects are presented, discussed and applied to data. The rich insights possible from modeling heteroscedasticity in choice processes is illustrated in each of the empirical applications, which examine its relevance to issues of data combination and taste heterogeneity.

AUTHORS:

David A. Hensher
Institute of Transport Studies, The University of Sydney

Jordan Louviere
Faculty of Economics, The University of Sydney

Joffre Swait
College of Business Administration, University of Florida

CONTACT:

Institute of Transport Studies (Sydney & Monash)
The Australian Key Centre in Transport Management
Graduate School of Business 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: April, 1997

1. Introduction

Economists have primarily focused their interests on the actual market behaviour of economic agents. Despite well-developed economic theory (see, e.g., Lancaster 1996, McFadden 1981) for dealing with real market choice data (generally termed *revealed preference*, or RP, data in this paper), there are a number of compelling reasons why economists should be interested in *stated preference* (SP) data, which involve choice responses from the same economic agents, but evoked in virtual (or hypothetical) markets:

1. There are serious and compelling needs to model demand for new products with new attributes or features for which there is no RP history, for which one cannot safely forecast by analogy to existing products.

2. Even when there is an RP history, not infrequently explanatory variables exhibit little variability or high colinearity in the marketplace of interest. Such cases frequently arise as a consequence of two interrelated forces: *market technology*, which drives correlations among some variables or attributes of products and limits the ranges over which they can vary; and *market forces*, which often drives competitors to offer exactly the same levels of attributes, or at least restrict products and services to compete within very confined ranges on certain attributes. Colinearity also arises due to market forces which tend to create negative correlations between attributes for any particular efficient frontier. In many cases, variables may be perfectly correlated with one another and/or with linear combinations of other variables, which occasions serious identification problems.

3. There may be an RP history, but as markets for products evolve and change over time, it is often the case that new attributes are introduced that now explain choices, but which previously were not present. For example, in the current PC market, many new features now exist that were unavailable on computers as recently as five years ago, and these new features not only drive current choices but may appeal to quite different segments of the market.

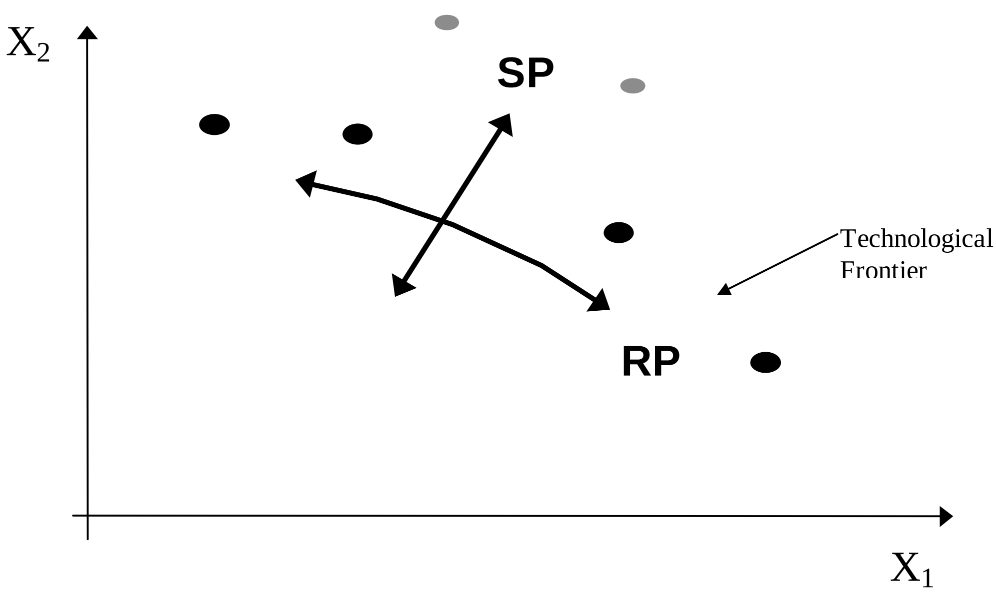
4. Even when RP data exist, they often do not satisfy model assumptions and/or contain statistical nuances which lurk in real data. Despite the development of ever more powerful statistical tools and models, one imposes maintained assumptions in the analysis of any set of data. Failure to satisfy assumptions often leads to problems of bias, many of which can be ameliorated by an SP data collection strategy.

5. Even with an RP market history, RP data often are time consuming and expensive to obtain. Thus, it is often more cost effective and quicker to collect SP data.

6. Finally, many problems of interest to marketers and resource economists involve products which are not traded in real markets, such as environmental goods. By definition, there is no RP data available for such classes of problems, although in some instances it is possible to make inferences indirectly via RP data for impacted behaviours (e.g. the "travel cost" method used to assess environmental impacts, see Englin and Cameron 1996). Hanemann and Kanninen (1997) recently reviewed the extensive and growing literature on discrete-choice contingent valuation data applications in environmental economics.

Many of the above comments can be understood via Figure 1. By definition, RP data generally are limited to helping us understand preference behaviour within an existing market and technology structure. On the other hand, while SP data certainly are useful in this same realm, they come into their own for problems which involve shifts in technological frontiers.

Figure 1: The Technological Frontier and the Roles of RP and SP Data



The latter type of application is central to many problems of marketing interest: e.g., new product introductions, line extensions, etc., are common concerns for many firms. Forecasts of likely demand, cannibalisation, appropriate target markets, segments, and the like are often needed to develop appropriate corporate and marketing strategy. Thus, a need for appropriate and valid models to reduce uncertainties associated with such decisions helped to spur development of various SP methods and models which we later briefly discuss. In particular, RP and SP choice data sources have different strengths and weaknesses, suggesting the possibility that one's weaknesses might be complemented by another's strengths. Indeed, it is this motivation that has seen significantly increased interest over the past decade into combining RP and SP choice (and more generally, preference) data in transportation, marketing and resource economics.

The primary focus of this paper, therefore, is to report on developments and advances in this stream of research. In particular, we first propose a conceptual framework within which the relationship of RP and SP methods and models can be understood. This serves as an informal introduction to a more general model of the data generation process which takes into account various sources of error variability, and allows us to characterise previous research. We then develop a number of different model forms that incorporate different error variability assumptions. We also discuss econometric issues involved in estimating and testing these models. We next provide a number of empirical examples of research in this paradigm. We end the paper with some conclusions based on this stream of research, as well as a set of research implications arising from this work.

2. Decision Making And Choice Behaviour

2.1 Conceptual Framework

Figure 2 suggests an overall ordering of stages in a consumer's decision process. S/he 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 products that can satisfy the need(s). During search and learning, the consumer forms beliefs about what products will or won't do, attributes that should be considered and the values of attributes possessed by products, as well as any associated uncertainties. At some point the consumer has sufficient information about the product category to form a decision rule or utility function which involves valuing and trading off various product attributes that matter in the decision. Given a set of beliefs or priors about attributes possessed by product alternatives, the consumer develops a preference ordering for products, and depending upon budget and/or other constraints and considerations makes a decision about whether to purchase, and given a decision to purchase, which one or more alternatives, and in what quantities.

Figure 2: Overview of the consumerís choice process

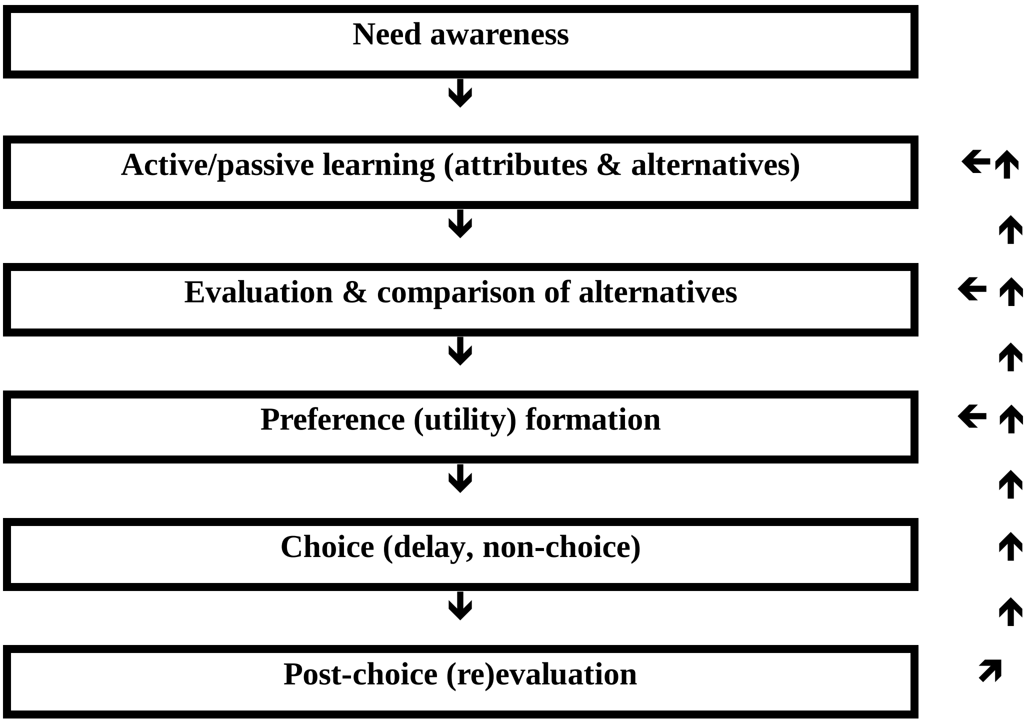


Figure 3 isolates this last decision and focuses on the stage at which the consumer forms utilities or values and begins to compare products to form overall (holistic) preferences for an available set of alternatives. Figure 4 formalises this process as a series of interrelated processes, links each process to a formal stage in the decision making process and describes the general area of research connected to that area in marketing, psychology and/or economics/econometrics.

Figure 3: Complex decision making and the choice process

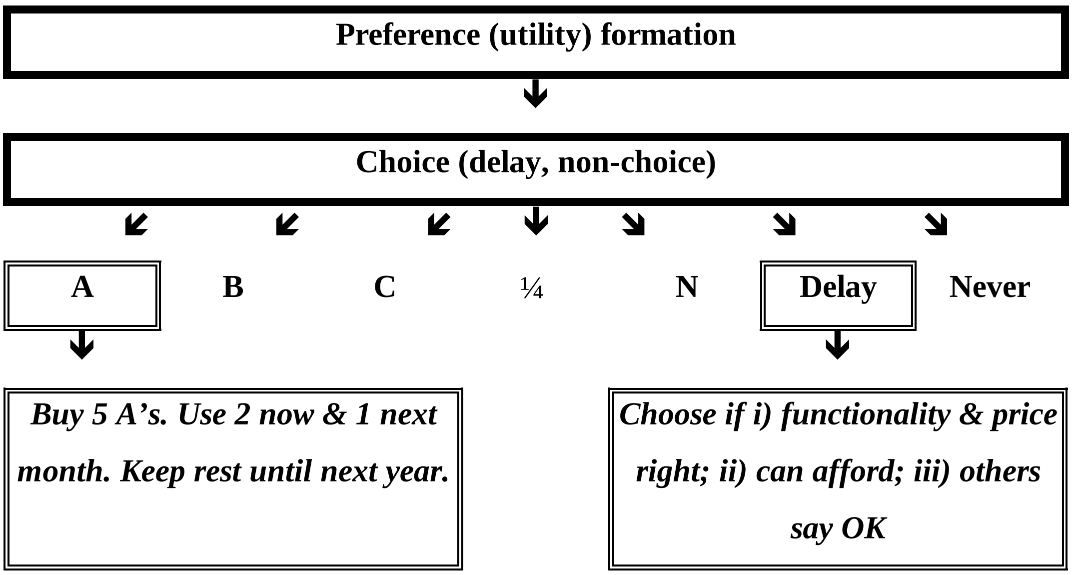


Figure 4: Functional relationships implied by the framework

X_i	Actions taken by managers/politicians ↓	Planning strategy
$S_{ki} = f_{1k}(X_i)$	Perceived positions of alternatives(s) ↓	Psychophysics
$u_{ki}(S_{ki}) = f_{2k}(S_{ki})$	(E)valuations of alternative positions ↓	Utility formation
$U_i = f_3[u_{ki}(S_{ki})]$	(E)valuations of holistic alternatives ↓	Utility function
$P(i \vert C) = f_4(U_i)$	Decision to choose, wait or never choose ↓	Choice process
$P(i \vert choose) = f_5[P(i \vert C)]$	If choose, which alternative	Share, demand, etc.

The key point made by Figures 2 to 4 is that the conceptual framework 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, provided such measures are logically or theoretically consistent with the framework and each other. The advantage of the latter integration is that it allows the marketer or econometrician to explain 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.

Thus, 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 overall framework indicates that SP methods and measures used to model intermediate stages in the decision making process are consistent with parallel 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 obscures important intermediate processes, and overlooks the potential role of intermediate models and measures in an overall behavioural, explanatory framework.

Because econometricians and readers of this journal are likely to be significantly less familiar with SP methods and models developed to study, explain and model these intermediate processes, we now turn our attention to a brief overview of SP methods and models. We particularly emphasise their application in marketing, and more particularly, concentrate on recent research involving combinations of sources of preference data that enhance our ability to explain and model the links implied in Figures 2 to 4.

2.2 A Brief Overview of SP Methods and Models

The literature on SP methods and models is vast and can be found in most social science disciplines, although psychology and marketing arguably contain most of the literature of primary interest to economics and econometricians. Of particular interest is the literature related to random utility theory (RUT) based choice models, and we concentrate our brief overview on that research stream. RUT was developed by Thurstone (1927) in his classic paper on modeling choices (comparisons) of pairs of alternatives. The problem of extending Thurstone's ideas to the multiple choice case was eventually solved by McFadden (1974), although psychologists and others had developed fixed (as opposed to random) utility multiple choice models some time earlier (e.g., Luce 1959; Restle 1961; Rushton 1969).

Interest in SP models in marketing arises from two dominant research streams: 1) modeling consumer tradeoffs, exemplified by the large literature in conjoint analysis (CA) and similar techniques used to develop quantitative descriptions of consumer preferences (e.g., Green and Wind 1971; Green and

Srinivasan 1978; 1990; Louviere 1974; 1988a; 1994; 1995), and 2) understanding and developing more behaviourally realistic models of preference and decision processes, exemplified by the Behavioural Decision Theory (BDT) literature in marketing, psychology and economics (e.g., Meyer, et al. 1984; Kahneman and Tversky 1979, 1984; Keller 1993). Generally speaking, the second research stream is often incompatible with RUT, and has evidenced little communality with parallel research on modelling preferences in econometrics and economics. Thus, we confine our attention to the first research stream in this overview.

The literature on conjoint analysis and related techniques designed to measure and model consumer tradeoffs, preferences and choices has roots in axiomatic work in utility and preference formation, such as axiomatic conjoint measurement, utility theory and risky decision making (e.g., Krantz and Tversky 1971; Krantz, et al., 1971; Machina 1987). Work in marketing has tended to emphasise practical application rather than understanding of process (although, see Lynch 1985), and has largely dealt with the riskless decision making case (e.g., Louviere 1988a). Little behavioural theoretical work is evident in the conjoint analysis literature in marketing, outside of statistical theory. Indeed, it is fair to say that until recently virtually all the conjoint analysis literature in marketing was statistically driven, but conjoint analysis has been placed more directly in the family of RUT models and links with RUT-based choice models established (Louviere 1988b; 1994; 1995). This latter stream of recent research forms the basis of our overview.

From the standpoint of econometrics, the primary problem with the CA and BDT literature is that there is little relationship between formal theory in economics, real marketing behaviour and the models and empirical studies available in these two streams of literature. In stark contrast, however, is the literature on designed choice experiments and preference elicitation procedures consistent with RUT (e.g., Louviere and Hensher 1983; Louviere and Woodworth 1983; Hensher and Louviere 1983; Louviere 1988a,b; Louviere 1994, 1995; Hensher 1994; Swait and Louviere 1993; Swait, et al. 1993; Adamowicz, et al. 1994). Beginning in the early 1980's researchers in marketing and transportation have steadily forged links between traditional revealed preference (RP) models and measures and corresponding SP models and measures. Indeed, it is now fair to say that these links are quite well-understood, and that (with some exceptions) it is possible to design SP studies which resemble corresponding RP studies in great detail

Thus, research into RUT-based SP methods and models has reached the point where the primary difference between more traditional econometric methods and models and the more recent SP methods and models is simply the data collected and analysed. In particular, the SP literature has reached the point where the theory and methods of analysis are identical to those developed in econometrics for cross-sectional and longitudinal analysis of choice or other data consistent with RUT (e.g., ordered dependent variables, complete rankings, joint continuous-discrete and related selection problems, etc.). Recognition of the isomorphism between the methods and models used to obtain and analyse the two types of data (RP/SP) has spawned a new and fast-growing interest in combining sources of preference data (e.g., Morikawa 1989; Ben-Akiva and Morikawa 1991; Swait and Louviere 1993; Swait, et al. 1993; Swait, Louviere and Williams 1994; Hensher and Bradley 1993; Louviere 1994, 1996 to name only a few). This more recent research paradigm is the focus of this paper, which is discussed in greater detail in Section 3.

The isomorphism between RP and SP data has been made possible primarily by advances in RUT-based SP choice modelling theory and methods. Although it has been well-known for many years that RUT-based discrete choice SP models can be estimated from paired comparison experiments (e.g., Thurstone 1927), only recently has the experimental basis for such models become well-established in marketing and other fields. In particular, Louviere and Woodworth (1983) in marketing and Hensher and Louviere (1983) in transportation provided the basic framework for the design and analysis of discrete choice experiments. Since that pioneering work, progress has been made on the design of SP choice experiments (e.g., Hensher and Barnard 1990; Anderson and Wiley 1992; Bunch, et al., 1996; Huber and Zwerina 1996), but much more remains to be done.

In the case of designing SP choice experiments, the key issues are a) identification, b) precision, and c) realism and cognitive limits. Identification refers to the specification of the utility function, and the objective often is to maximise the possible forms that can be identified (tested) from a particular choice experiment. Precision refers to the efficiency of the parameters estimated from the design, and the objective typically is to minimise the standard errors. Realism and cognitive limits refer to the degree to which the experiment mirrors the real choice faced by consumers in the marketplace, while cognitive limits refers to the complexity and demands of the task as far as limits to human processing ability are concerned. Typically, one wants to maximise realism while minimising complexity. A choice experiment developed in an applied setting usually involves some compromise among these objectives.

A choice experiment consists of sets of mutually exclusive alternatives, from which a consumer must choose one or more. Choice experiments are typically based on an experimental design which takes identification and precision into account. Choice experiments can be designed sequentially, by first designing one or more sets of alternatives with desirable identification properties (often treated as a "conjoint analysis" design problem, Louviere 1988, 1994), and then assigning the designed alternatives to sets to satisfy the properties of particular discrete choice models as well as provide desired levels of precision. For example, one can design three statistically equivalent sets of N attribute combinations that describe hotel options, place the designed sets of N options separately into three "urns", and randomly sample one combination from each urn without replacement to create N choice sets of size three (i.e., each set consists of three alternatives). If there are a sufficiently large number of sets, one can satisfy the asymptotic properties of particular discrete choice models, as well as produce sets that have desirable statistical properties taken as an entire set of sets.

More commonly, however, a choice experiment is developed simultaneously. That is, if there are N alternatives, and each has M attributes with exactly L levels, one can design experiments that are consistent with the most general member of the logistic regression family (e.g., "Mother Logit", McFadden, Tye and Train 1976). This is accomplished by treating all attributes of all alternatives as a collective factorial (i.e., L^{MN}), and selecting the smallest orthogonal main-effects design from the total factorial. This design approach insures that the marginals (i.e., alternative-specific and/or generic parameters) for each alternative are orthogonal within and between alternatives, which satisfies the necessary and sufficient conditions for estimating logistic regression models. The conditions for estimating the parameters of the Mother Logit model are satisfied because one can independently estimate all parameters which represent violations of the IID property of RUT choice models (i.e., "cross-effects" of the attributes of one alternative on the utility of a second).

There are other ways to design choice experiments (e.g., using 2^N designs, Louviere and Woodworth 1983), but the foregoing represent a reasonable description of much of the state-of-the-art. Designs differ in their ability to capture complexity of utility specifications (i.e., degrees of non-linearity and/or non-additivity), numbers of alternatives and numbers of attributes per alternative (Louviere and Woodworth 1983; Louviere 1988a,b; Louviere 1994; 1995; 1996; Carson, et al., 1994). In order to combine and compare RP and SP models, it is necessary that the units of measurement of the attribute levels be identical in both the choice experiment (SP) and the RP data. Typically, the latter can be accomplished by careful attention to detail in research leading up to the design of combined RP and SP data collection.

3. The Theory Of Choice Data Combination

3.1 The Fundamental Identification Problem: Inseparability of Taste and Scale

It has been well-known for some time that a fundamental link exists between the scale of the estimated parameters and the magnitude of the random component in all choice models based on Random Utility Theory (RUT) (see, e.g., Hensher and Johnson 1981; Ben-Akiva and Lerman 1985; Anderson et al. 1992). Let

$$U_{iq} = V_{iq} + e_{iq}, \qquad (1)$$

where U_{iq} is the unobserved, latent utility consumer q associates with alternative i; V_{iq} is the systematic, quantifiable proportion of utility which can be expressed in terms of observables of alternatives and consumers; and the e_{iq}'s are the random or unobservable effects associated with the utility of alternative i and consumer q. All RUT-based choice models are derived by making some assumptions about the distribution of the random effects. These distributions place in all models an imbedded scale parameter, inversely related to the variance of the random component and unidentifiable separately from the taste parameters.

For example, to derive the Multinomial Logit (MNL) choice model from (1), we assume that the e_{iq}'s are IID Type I Extreme Value (or Gumbel) distributed. The scale parameter $l \geq 0$ of the Gumbel distribution is inversely proportional to the standard deviation of the error component, thus, $s_{iq}^2 = p^2 / 6\, l^2$. As discussed in Ben-Akiva and Lerman (1985, chapter 5), the fundamental identification problem of RUT-based choice models reveals itself in the MNL model in that the vector of parameters actually estimated from any given source of RUT-conformable preference data is actually ($l\, b$), where b is the actual vector of taste parameters. This can be seen clearly in the full expression of the MNL choice probability:

$$P_{iq} = \frac{\exp(\lambda V_{iq})}{\sum_{j \in C_n} \exp(\lambda V_{jq})} = \frac{\exp(\lambda \beta X_{iq})}{\sum_{j \in C_q} \exp(\lambda \beta X_{jq})}, \qquad (2)$$

where P_{iq} is the choice probability of alternative i for consumer q, and the systematic utility $V_{iq} = \beta X_{iq}$. Because a given set of data is characterized by some value of λ, this constant is normalized to some value (say, one), and analysis proceeds as if $(\lambda\beta)$ were the taste parameters.[1]

Similar scaling and taste parameter identification problems arise with other members of the GEV (Generalized Extreme Value) family of choice models (i.e., Nested MNL, Tree Extreme Value and Ordered GEV; see McFadden 1981, Small 1987, 1994), with binary logit and probit models, and with the increasingly popular Multinomial Probit (MNP) choice model (e.g., Hausman and Wise 1978, Daganzo 1980, Keane 1995; see Bunch 1991 for a discussion of the scaling issue in MNP models). The reason for the pervasiveness of this identification problem is that choice models specify a structural relationship between a categorical response and a latent variable (i.e. utility). As in structural equation models involving latent variables (e.g. LISREL models), it is necessary to specify both origin *and* variance (read "scale") for the latent variable(s) to permit identification of utility function parameters.

Recognition of the role of the scale parameter in the estimation and interpretation of choice models has come somewhat recently, but was fostered by the desire to combine RP and SP data. The paradigm shift involving efforts to combine sources of preference data was inspired by Morikawa's (1989) insight that if data generation processes underlying SP and RP data are the same, model parameters should differ only by a constant of proportionality. Morikawa (1989) noted that the fundamental identification problem was confined to a single preference data source, and that the ratio of λ's in two or more sources of data could be identified. Morikawa's dissertation (1989), and subsequent work (e.g., Ben-Akiva and Morikawa 1991) demonstrated that the ratio(s) of λ's could be estimated both sequentially (Swait and Louviere 1993) and simultaneously (Morikawa 1989; Hensher and Bradley 1993; Bhat 1995).

As we later discuss, the estimation problem amounts to placing an equality restriction on the taste

[1]Note that the MNL model predicts random choice when $\lambda \to 0$, and approximates a step function for the alternative with maximal utility as $\lambda \to \infty$ (see Ben-Akiva and Lerman 1985). This general behaviour applies to all choice model specifications.

parameters of K preference data sources to be combined (i.e. $b_1=...=b_K=b$) and estimating K additional scale parameters ($l_1,..., l_K$). One of these scale parameters must be fixed, say $l_1=1$. The remaining scale parameters can then be interpreted as inverse variance ratios with respect to the reference data source. The corresponding unrestricted model frees the taste parameters and the scale factors for the K data sources by estimating ($l_k\ b_k$), k=1,...,K. The null hypothesis of interest is that of taste invariance across data sources, after permitting variance/reliability differences.[2] In the present situation, such a hypothesis can be tested using a likelihood ratio statistic.

Greater understanding of the role of the scale parameter has spawned several related research streams, most notably a "data fusion" stream, primarily associated with travel demand modeling (e.g., Morikawa 1989; Ben-Akiva and Morikawa 1991; Hensher and Bradley 1993; Swait, Louviere and Williams 1995), and a more general stream concerned with comparing and testing models estimated from any sources of preference data consistent with RUT (e.g., Swait and Louviere 1993; Louviere, Fox and Moore 1993; Bradley and Daly 1994; Louviere 1994; Swait, Louviere and Williams 1995; Louviere 1995).

The latter paradigm represents a more general view of combining data sources than the former. First, it views the scale factor as an integral component of a real behavioural process, in contrast to the view that the scale is a nuisance parameter that must be accounted for to permit measurement of the true quantities of interest (i.e. taste parameters). Second, it encompasses a wider scope of data combinations, involving RP with RP and SP with SP, as well as the RP with SP combinations which are the sole interest of the "data fusion" stream.

3.2 A General Preference Data Generation Process

The ongoing challenge in capturing real behavioural processes through statistical modelling requires a modeling framework sufficiently rich to accommodate the structure of all observed and unobserved influences on choice responses. Such a framework should ideally be capable of allowing for both the real influences on choices as processed by the agents of choice (i.e. individuals, households, firms, groups) as well as for variation in response opportunities associated with the means used by analysts to acquire information from agents. The latter includes an extensive range of data acquisition paradigms

[2]Suppose each data source has L taste parameters. Note that under the null hypothesis the K parameter vectors can be represented in R^L as vectors superimposed upon one another, but with different lengths dictated by differences in reliability, i.e. in l's: more reliable measurements will correspond to longer vectors. The practical importance of this observation is that the plot of the L components of two coefficient vectors that differ only in terms of reliability, or

such as (1) revealed preference (RP) and stated preference (SP) methods, (2) the complexity of the task imposed on agents, such as the number of replications of an SP task, (3) reporting of attributes of non-chosen alternatives in an RP task, and (4) the actual method of data collection (e.g. telephone, mailout-mailback, face to face, etc. interview methods).

Choice modeling in marketing focuses on opportunities to capture real behavioural processes provided by deeper understanding of the behavioural meaning of traditional statistical constructs such as variance and covariance, scale and taste weights. For example, variance profiles associated with unobserved influences on the relative (indirect) utility of each alternative in a choice set have become a research focus because they offer opportunities to relax a basic limitation of traditional MNL models (i.e., the assumption that unobserved random effects are identically distributed) as well as use the variance properties to satisfy a common set of statistical assumptions when combining data from different sources (Morikawa 1989, Hensher and Bradley 1993, Swait and Louviere 1993). In addition, recognition that the data collection process itself may be a source of variability in behavioural choice response, and which, if not isolated, may confound the real behavioural role of the observed and unobserved influences on choice, can be handled by an appropriate functional specification of the structure of variance of the unobserved effects. For example, the variance associated with an alternative can be a function of task complexity (Swait and Adamowicz 1996) or some respondent characteristics which serve as proxies for ability to comprehend the survey task (Bhat 1996, Hensher 1996a).

Developments in refining the specification of the indirect utility expression associated with a mutually exclusive alternative in a choice set can be summarised by equation (3):

$$\tilde{U}_{iqt} = a_{it} + b_{qt} X_{iqt} + q_{qt} Y_{iqt-1} + d_{qt} \tilde{U}_{iqt-1} + \tilde{e}_{iqt} \tag{3}$$

where

$\tilde{U}_{iqt}$ = indirect utility associated by person *q* with alternative *i* at time *t*

a_{it} = alternative-specific constant represents the mean of the distribution of unobserved effects in the random component associated with alternative *i* at time *t*

b_{qt} = taste weights for person *q*, representing the relative level of or saliency associated with the attributes at time *t*

variance, should produce a graph that contains a scattering of points that essentially lie along a straight line going through the origin, with positive slope given by the ratio of the respective scale parameters.

X_{iqt} = the attribute vector faced by individual *q* for alternative *i*, time period *t*

Y_{iqt-1} = choice indicator (0=not chosen,1=chosen) for person *q*, alternative *i*, time *t-1*

q_{qt}, d_{qt} = utility weight associated with *state dependence* and *habit persistence* (see Heckman 1981), respectively, for person *q*, time period *t*

$\tilde{e}_{iqt}$ = error term

Assume that the scale of $\tilde{e}_{iqt}$ (hence of $\tilde{U}_{iqt}$) is l_{iqt}. Define $U_{iqt} = l_{iqt}\tilde{U}_{iqt}$ and $e_{iqt} = l_{iqt}\tilde{e}_{iqt}$. Multiply both sides of (3) by l_{iqt} to obtain

$$U_{iqt} = l_{iqt}a_{it} + l_{iqt}\, b_{qt} X_{iqt} + l_{iqt} q_{qt} Y_{iqt-1} + l_{iqt}\, d_{qt}\tilde{U}_{iqt-1} + e_{iqt} \quad . \tag{4}$$

The effect of this multiplication is that the error terms e_{iqt} have unit scale. Note, however, that the habit persistence term in (4) involves the previous period's utility $\tilde{U}_{iqt-1}$, which has associated scale l_{iqt-1}. Thus, (4) can be rewritten to involve $U_{iqt-1} = l_{iqt-1}\tilde{U}_{iqt-1}$ as follows:

$$U_{iqt} = l_{iqt}a_{it} + l_{iqt}\, b_{qt} X_{iqt} + l_{iqt} q_{qt} Y_{iqt-1} + \left(\frac{l_{iqt}}{l_{iqt-1}}\right) d_{qt} U_{iqt-1} + e_{iqt} \quad . \tag{5}$$

Recognize now that $d_{qt} U_{iqt-1}$ is a person-specific, time-dependent unobserved heterogeneity effect which we shall subsume into a random variable g_{qt}. Thus, the final form of the utility function that we shall consider is:

$$U_{iqt} = l_{iqt}a_{it} + \left(\frac{l_{iqt}}{l_{iqt-1}}\right) g_{qt} + l_{iqt}\, b_{qt} X_{iqt} + l_{iqt} q_{qt} Y_{iqt-1} + e_{iqt} \quad . \tag{6}$$

This form permits incorporation of (1) brand-specific temporal heterogeneity, (2) person-specific heterogeneity (which captures habit persistence, among other things), (3) state dependence, (4) temporal (for RP) or repeated measures (for SP) dependence in the error structure, and (5) temporal, person-specific and product-specific heteroscedasticity.

Using (6) as an overarching structure, we note that the majority of discrete choice models reported in the literature assume one or more of the following restrictions:

1. homoscedastic error terms (i.e. the scale parameters are constant across people, products and time periods, but see Hensher et al. 1992, Bhat 1995, Swait and Adamowicz 1996, Swait and Stacey 1996);

2. no temporal or repeated measures dependence in the error term (but see Morikawa 1994, Keane 1995, Swait and Naik 1996);

3. homogenous taste weights across the population (but see Keane 1995);

4. no unobserved individual heterogeneity (but see Morikawa 1994, Elrod and Keane 1995).

Overcoming these restrictions is a challenge when considering a single source of choice data. They are even greater challenges when combining multiple sources of data. The role played by heteroscedasticity in pooling such data sources (unrecognized until recently) is the main reason we refer to the "case of the lurking λ's", which has become synonymous with the literature on combining sources of preference data.

3.3 Alternative Forms of Handling Heteroscedasticity in Choice Models

A hierarchy of models is evolving in the literature, relaxing progressively some of the testable assumptions imposed on utility function (6). We concentrate on discussing some possible routes taken in the literature with respect to relaxing the homoscedasticity assumption with respect to $\tilde{e}_{iqt}$ when combining sources of preference data, though we also consider relaxing certain other restrictions usually imposed on choice models. We also restrict our presentation exemplars from the GEV family of models (mainly variants of the MNL model) for cross-sectional choice data. However, it should be noted that the general principles and concepts we discuss are equally valid for other choice models and data types.

3.3.1 Random Effects Heteroscedastic Extreme Value Model

One way to relax the constant variance assumption is by means of a more complex choice model called the heteroscedastic extreme value (HEV) model. Allenby and Ginter (1995), Bhat (1995) and Hensher

(1996a) recently implemented the HEV model on a single data source, whereas Hensher (1996b) applied the Heteroscedastic HEV model to joint estimation of SP and RP data.

With respect to utility function (6), we assume that the data are cross-sectional (hence no temporal effects), there is no state dependence or serial dependence and tastes are homogenous. Specifically,

$$U_{iq} = \lambda_{iq}\alpha_i + \lambda_{iq}\beta X_{iq} + \varepsilon_{iq} \quad . \tag{7}$$

Assume further that the λ_{iq} equal λ_i for all individuals *q*; in addition, assume they are independently, but not identically, distributed across alternatives according to the Type I Extreme Value density function f(t) = exp(-t)*exp(-exp(-t)) = -F(t)*log(F(t)), where F(.) is the corresponding cumulative distribution function. If the decision rule is maximal utility, then the choice probabilities are given by:

$$P_{iq} = \int_{-\infty}^{\infty} \prod_{j \neq i} F(\lambda_j)[V_{iq} - V_{jq} + \varepsilon_{iq}]\, \lambda_i f(\lambda_i \varepsilon_{iq}) d\varepsilon_{iq} \quad . \tag{8}$$

The probabilities must be evaluated numerically because there is no closed-form solution for this single dimensional integral. The integral can be approximated, for example, using Gauss-Laguerre quadrature (Press et al 1986) and computational experience has shown a 68 point approximation is sufficient to reproduce taste parameter estimates (Greene 1996).

The HEV model nests the restrictive MNL, and is flexible enough to allow differential cross-elasticities among all pairs of alternatives, while avoiding the *a priori* identification of mutually exclusive market partitions of a nested MNL structure. It is parsimonious compared to the MNP model, introducing only J-1 additional parameters in the covariance matrix as opposed to the [J(J-1)/2]-1 additional parameters in the more general model (J is the total number of alternatives in the universal choice set). In contrast to the MNP model, the computational burden is significantly reduced, requiring only the evaluation of a one dimensional integral (independent of the number of alternatives), whereas MNP requires the evaluation of a J-1 dimensional integral. Importantly, in contrast to MNP, the HEV model is easy to interpret and its behaviour intuitive (Bhat 1995).

Hensher (1996b) suggested that HEV is a useful device to identify an appropriate partitioning of the MNL model into a nested structure, replacing the search for structure in nested MNL partitions. The reason for specifying a nested form of MNL is to accommodate systematic dependencies among the unobserved effects (leading to violation of the independence of irrelevant alternatives (IIA) condition), which are not handled properly by the MNL model. HEV does not have a closed-form, hence its practical appeal is that a nested specification consistent with the HEV profile of λ_i will be easy to apply without the numerical integration required by expression (8).

The HEV model can be specified for multiple data sources, jointly estimated using a FIML specification to produce a set of alternative-specific λ's across both RP and SP choice sets, normalising on an arbitrarily selected alternative. An empirical example of HEV is presented in Section 4.

3.3.2 Fixed Effects Heteroscedastic MNL Model

Researchers may be able to formulate a theory to explain the heteroscedasticity structure present in preference data of the type described by (6). For example, depending upon its content, exposure to advertising might lead to more or less variability in the observed choice behaviour of a population. While this effect might be captured partially by a random effects specification of the scale parameters, the origin of the variability would not be explicit. The latter consequence led to the development of fixed effect heteroscedastic MNL (FEHMNL) models[3] (Swait and Adamowicz 1996; Swait and Stacey 1996), wherein the scale parameters are given by

$$\lambda_{iq} = \exp(\gamma Z_{iq}), \tag{9}$$

where γ is a parameter row-vector and Z_{iq} are covariates. FEHMNL choice probabilities are given by

$$P_{iq} = \frac{\exp(\lambda_{iq}\ \beta X_{iq})}{\sum_{j \in C_q} \exp(\lambda_{iq}\ \beta X_{jq})} \ . \tag{10}$$

The parameters to be estimated are β and γ. As with the HEV model described previously, the FEHMNL allows complex cross-substitution patterns among the alternatives.

[3]The derivation of the FEHMNL when the scale factor does not vary by alternative is different than when it does. If the scale factor is *not* alternative-specific, the model can be derived using a heteroscedasticity argument (see Swait and Adamowicz 1996); when the scale factor is alternative-specific, the model must be derived as a special case of the Tree

Swait and Adamowicz (1996) hypothesize that task complexity (for SP data) and choice environment (e.g. market structure for RP data) influence levels of variability found in preference data. They propose a specific measure to characterize complexity and/or environment, and test and find evidence of its impact in a number of SP data sets, as well as in an RP data source. Their measure of complexity does not vary across alternatives, consequently scale parameters in their model vary across individuals and SP replications, but not across alternatives. They also found that different degrees of complexity between preference data sources can impact the propriety of combining RP/SP data.

Swait and Stacey (1996) apply FEHMNL to scanner panel choice data, allowing the variance (i.e. scale) to vary by person, alternative and time period as a function of brand, socio-demographic characteristics, interpurchase time, state dependence, etc. They show that accounting for non-stationarity of the variance in terms of the explanatory variables Z_{iq} enhances insight about panel behaviour and greatly improves model fit with respect to standard choice models such as the MNL, nested MNL and MNP models with fixed covariance matrices.

3.3.3 Latent Class Heteroscedastic MNL Model

This third model introduces the additional complexity of taste heterogeneity along with the heteroscedasticity that is our central interest. In most econometric models which permit taste heterogeneity, a random coefficients approach is adopted. That is, β_q is assumed to be a draw from some joint density function (e.g. multivariate normal), and estimation recovers the parameters of the distribution. In marketing, latent class models were often used instead of random coefficient formulations (e.g. Dillon and Kumar 1993). Instead of a continuous joint distribution, latent class models assume that a discrete number of support points (say, S) are sufficient to describe the joint density function of the parameters.

From a marketing point of view latent classes correspond to underlying market segments, each of which is characterized by unique tastes β_s, s=1,...,S. However, Swait (1994) pointed out that these classes also can be characterized by variance differences. He postulated that members of class s have taste β_s and scale λ_s. If the indirect utility function for members of that class is

$$U_{iq|s} = \lambda_s \alpha_{i|s} + \lambda_s \beta_s X_{iq} + \varepsilon_{iq|s}\,, \qquad (11)$$

Extreme Value model (Daly 1987, McFadden 1981, Hensher 1994, Swait and Stacey 1996). In either case, the final and the $e_{iq|s}$ are conditionally IID Type I Extreme Value within class, the choice probability for members of class *s* is

$$P_{iq|s} = \frac{\exp(l_s\ b_s X_{iq})}{\sum_{j \in C_q} \exp(l_s\ b_s X_{jq})}\ . \qquad (12)$$

The final specification of the choice model requires the development of a classification mechanism to predict an individual's membership in a class (see Swait 1994 for full details). If the probability of being in class *s* is given by W_{qs}, the unconditional probability of choosing alternative i is simply

$$P_{iq} = \sum_{s=1}^{S} P_{iq|s} W_{qs}\ . \qquad (13)$$

It is not possible to simultaneously identify scale factors and taste parameters in this model. Some possibilities for dealing with this issue are as follows: (1) let taste parameters vary across classes but constrain scale parameters to be equal (i.e. estimate b_1,..., b_S and set l_1=...= l_S=1); (2) force taste parameter homogeneity but let scale parameters vary across classes (i.e. estimate b_1=...= b_S= b and l_2,..., l_S, normalizing l_1=1); (3) restrict combinations of taste and scale parameters such that either l_s or b_s is estimated for any particular class *s*. Each of the foregoing possibilities represents different behavioural assumptions concerning taste heterogeneity *vis-à-vis* error term variance within latent classes.

Gopinath and Ben-Akiva (1995) and Swait and Sweeney (1996) proposed similar models to that of Swait (1994); differently from that earlier work, however, Gopinath and Ben-Akiva (1995) and Swait and Sweeney (1996) assume that the latent classes are ordered with respect to an additional underlying latent dimension (e.g. value of time, orientation towards value for money in a retail setting). Swait's (1994) model assumes no particular relationship holds between latent classes and the multiple latent dimensions permitted in his segmentation framework.

3.3.4 Summary

Several variants of basic MNL models were proposed and discussed above. All involve some form of

expression for the choice probability is given by expression (10).

relaxation of the constant variance assumption of basic MNL, either through random or fixed effects. We also discussed combining taste heterogeneity and heteroscedasticity into a choice model, and these approaches can be combined in different ways to deal with panel data, introduce dependence in error terms, etc. The contribution of the foregoing discussion was to illustrate the general principles of controlling for heteroscedasticity in choice models, which is the major thrust of this paper.

4. Applications

In this section we present empirical applications of particular models presented in Section 3.3. As in that section, our objective is not to be exhaustive but to illustrate the additional insight gained from choice models when the scale parameter is viewed as an integral part of the specification, rather than as a nuisance parameter to be modeled and subsequently discarded. The first case illustrates estimation of an HEV model in a combination of RP and SP data. The second case involves the use of a Fixed Effect Heteroscedastic MNL model to investigate certain market segment differences within a single SP data source. Finally, we illustrate the types of insight that can be gained by investigating the possibility that apparent taste parameter heterogeneity might be explained largely by variance differences.

4.1 Case 1: The HEV Model

The data used to illustrate the application of the HEV model are drawn from a pre-feasibility market study associated with the Very Fast Train (VFT) project in Australia. We extracted 197 surveys of non-business travel between Sydney, Canberra and Melbourne in 1986. The RP mode choice set comprised four modes (plane, car, coach and conventional train). The SP choice set included the four RP modes plus a new high-speed rail alternative. For their most recent intercity trip, each sampled decision maker provided details of the travel time components (access, linehaul and egress), the cost, and transfers (if public transport was used) of their chosen means of transport and each of the modal alternatives. Earlier, Hensher and Bradley (1993) used these same data to estimate a joint fixed effect RP/SP model with a constant scale factor ratio between the RP and SP sources.

A stated choice experiment was designed for the five modes, and each mode was described by three attributes (access plus egress time, in-vehicle time for the main mode (i.e. linehaul) and total cost; Hensher and Bradley 1993) with three levels each. Attribute levels were selected to be realistic variations around experience on each of the reported RP trips. There were a total of 27 possible combinations of attribute levels for each mode, and assuming a constant choice set size of five, there are a total of 27^5 possible sets, an unwieldy number. A fraction of the total 27^5 possible sets was created by creating a one-third fraction for each mode alternative. The resulting nine combinations for each mode were randomly allocated to create nine choice sets. The nine resulting choice sets were administered in different random orders, respondents evaluated four of the sets and ranked the five modes in order of choice. The first-preference rank was defined as the chosen mode in the current example application.

MNL models with constant variance were estimated from the RP and SP data, respectively (Table 1). Examination and comparison of these models shows that the sensitivity to mode attributes (travel time and cost) are of the same order of magnitude in both data sources. However, earlier in this paper, it was noted that direct comparisons of choice model coefficients from different data sources are not possible without controlling for variance differences because the estimated coefficients actually confound scale and taste.

Table 1 - Heteroscedastic Extreme Value Model Estimation Example

Parameter	**RP MNL Model Estimated Parameters** (t-stats in parentheses)	**SP MNL Model Estimated Parameters** (t-stats in parentheses)	**RP+SP HEV Model Estimated Parameters**[1] (t-stats in parentheses)
Utility Function (Taste) Parameters			
Alternative-Specific Constants			
Plane - RP	-0.228 (-0.5)		1.981 (2.8)
Train - RP	0.725 (3.0)		1.983 (3.1)
Coach - RP	-0.225 (-0.9)		1.897 (2.1)
Car - RP	-0-		-0-
Plane - SP		0.502 (0.4)	1.714 (0.7)
VFT - SP		2.014 (2.2)	2.019 (2.6)
Train - SP		-0.546 (-1.2)	1.705 (1.1)
Coach - SP		-0.736 (-2.2)	1.768 (0.4)
Car - SP		-0-	-0-
Generic Mode Attributes			
Cost ($)	-0.0259 (-3.9)	-0.0317 (-2.0)	-0.00184 (-2.0)
Door-to-Door Time (min)	-0.00396 (-5.3)	-0.0027 (-1.75)	-0.000187 (-2.1)
Scale Parameters			
Plane - RP	1.00		0.167 {8.37 (1.1)}
Train - RP	1.00		0.179 {7.77 (1.4)}
Coach - RP	1.00		0.152 {9.13 (1.2)}
Car - RP	1.00		1.465 {0.95 (2.8)}
Plane - SP		1.00	0.182 {7.65 (1.7)}
VFT - SP		1.00	0.154 {9.04 (1.9)}
Train - SP		1.00	0.192 {7.22 (1.7)}
Coach - SP		1.00	0.150 {9.26 (1.7)}
Car - SP		1.00	1.392 1.00

Goodness-of-fit			
Log Lik. @ convergence	-244.9	-220.4	-536.6
Rho-squared	0.096	0.301	0.373

Notes:

1. Parameters in {} are estimated standard deviations of error terms, and corresponding t-values.

Table 1 also presents the HEV model on the pooled RP/SP data, which controls for scale differences between RP and SP, in this case at the mode-specific level (the scale parameter for Car in the SP data set is normalized to one.). In contrast, most academic and commercial applications of data fusion have not controlled for scale differences at the alternative level (e.g. Adamowicz et al. 1996, Ben-Akiva and Morikawa 1991, Swait and Louviere 1993); instead, error terms have typically been assumed IID within data source, leading to a single variance ratio to be estimated.

The HEV model in Table 1 reveals some significant scale coefficients in the data, suggesting interesting comparisons within and between data sources. For example, note first the similarity between the Car scale parameters in the RP and SP data. Because the Car SP standard deviation is normalized to one, the referrent mode within both data sets has comparable levels of variation. This observation seems to hold for all common mode pairs across the two data sources, such that the ratio of the RP to SP scale factors for each mode is about unity. This suggests that the SP experimental choice task was well-designed in the sense that it captured error variability levels comparable to those in the RP data; intuitively, this should enhance our ability to pool these data.

Another interesting insight provided by the mode-specific scale factors can be seen by noting that the ratio of the scale factors of all modes in either data set to that of Car is quite small (on the order of eight). Hence, these other modes have error variances an order of magnitude *greater* than that of Car, both in the real market place and in the SP choice task. This further suggests that consumers as a group behave much less consistently (i.e. reliably) when evaluating non-Car modes than when evaluating the Car mode. These rich behavioural insights are indicative of the benefits of modeling heteroscedasticity in choice models, and are the basis for formulating potentially rewarding avenues of research to explain the behavioural foundations of the heteroscedastic errors.

4.2 Case 2: The Fixed Effects Heteroscedastic MNL Model

A sample of 100 undergraduates at a major North American university participated in a survey regarding preferences for shopping outlets for apparel purchases. The SP elicitation procedure used in the survey is called Best/Worst Conjoint (Louviere and Swait 1996a). In this procedure respondents evaluate a number of statistically designed product descriptions (in this case, retail outlets) one at a time, and indicate which aspect/attribute in the description is most/least attractive along some evaluation dimension. In this particular survey generic retail stores were described by six attributes that could take on one of two values (levels): merchandise assortment, travel time, frequency of sales, merchandise quality, sales personnel availability and return policy liberality. Besides the Best/Worst conjoint task, respondents also rated their general attitude towards apparel shopping on a nine category rating scale (the low end of the scale was labelled "I hate to shop," and the high end "I love to shop").

In the Best/Worst survey there are two categorical responses of interest: the most attractive attribute and least attractive attribute levels presented. Respondents may be more sure of their "best" level compared with "worst" level responses. Thus, it is postulated that:

> H1: The variance (scale) of the latent evaluation dimension is smaller (larger) for "best" responses compared to "worst" responses.

Moreover, it may be that respondents who like to shop will respond more consistently than those who dislike shopping. Accordingly, the sample was divided into two groups depending upon their stated attitude towards shopping (a "Love to Shop" group who responded 6-9 on the rating scale; and a "Hate to Shop" group who responded 1-5 on the scale). We therefore hypothesized the following:

> H2: The "Love to Shop" group should have lower (higher) variance (scale) in their responses compared to the "Hate to Shop" group.

Assuming that the attribute level selection frequencies can be modeled according to a FEHMNL model, we defined a scale function with the following variables to test the two preceding hypotheses:

$$Z_1 = \begin{cases} 1 & if \text{ "}Best\text{" } choice\,set \\ 0 & o.w. \end{cases}$$

$$Z_2 = \begin{cases} 1 & if \text{ "}Love\,to\,Shop\text{" } individual \\ 0 & o.w. \end{cases}$$

The scale function itself is given by

$$\ln\ l = y_{1}Z_1 + y_{2}Z_2 \quad . \qquad (14)$$

Under the hypotheses above, it is expected that both y_1 and y_2 will be significantly different from zero and positive in sign.

Table 2 presents the estimation results for this FEHMNL model and its restricted form (in which we impose the constraint $y_1 = y_2 = 0$, producing the MNL model). A likelihood ratio test for the joint statements made by H1 and H2 strongly supports both hypotheses, yielding a chi-squared statistic of 22.8 with 2 degrees of freedom, which can be compared with the table value of 5.99 at the 95% significance level. Individually, we cannot reject either hypothesis H1 or H2 at the 95% confidence level, using a one-sided asymptotic Z test. Thus, these results support the conclusion that this sample of subjects exhibits homogeneity of tastes but different variances in utility, depending upon (1) whether they were responding to best or worst elicitations and (2) their attitudes towards shopping.

Table 2 - Fixed Effect Heteroscedastic MNL Model Estimation Example

Parameter	MNL Model Estimated Parameters (t-stats in parentheses)	FEHMNL Estimated Parameters (t-stats in parentheses)
Utility Function (Taste) Parameters		
Intercepts		
Overall	-0.015 (-0.1)	0.197 (0.7)
Assortment	-0.095 (-0.6)	-0.121 (-0.8)
Travel Time	-0.580 (-3.6)	-0.493 (-3.0)
Promotions/Sales	0.797 (4.5)	0.665 (3.7)
Quality Merchandise	-0.232 (-1.5)	-0.276 (-1.7)
Personnel	0.292 (1.7)	0.230 (1.4)
Return Policy	-0-	-0-
Attribute Slopes		
Assortment (Wide=1,o.w.=0)	1.293 (13.5)	1.095 (8.2)
Travel Time (min)	-1.116 (-11.4)	-1.008 (-7.3)
Promotions/Sales (Occ.=1,o.w.=0)	-0.724 (-7.5)	-0.531 (-6.1)
Quality Merchandise (High=1,Low=0)	1.591 (17.7)	1.364 (8.7)
Personnel (Sales+Ch=1,o.w.=0)	0.703 (6.7)	0.583 (6.2)
Return Policy (No Questions=1,o.w.=0)	1.612 (11.6)	1.448 (6.9)
Scale Function Parameters		
Z_1 (Best=1,Worst=0)	-0-	0.207 (1.8)
Z_2 (Love=1,Hate=0)	-0-	0.419 (4.2)
Goodness-of-fit		
Log Lik. @ convergence	-1,965.7	-1,954.5
Rho-squared	0.202	0.206

The latter result deserves further elaboration. In particular, suppose we separate the sample into two groups, who either love or hate shopping for clothes, and estimate a choice model for each group. The results in Table 2 reveal that most analysts would have found a significant taste difference between the two groups, and more than likely, this result would have agreed with their priors. However, we found that the responses of individuals who dislike shopping are less consistent (i.e. have lower scale) than respondents who like to shop. From a marketing perspective, this latter finding contains significant strategic implications: if the two groups truly differ in their tastes one must (in principle!) design different stores for each. On the other hand, if tastes are the same but response variance differs, only one store type is necessary to serve both, but the factors which make certain people dislike shopping must somehow be overcome.

4.3 Case 3: Latent Class Heteroscedastic MNL Model

A public agency was interested in the tradeoffs and choices of accommodation and types of recreation destinations in a certain region of the US. Several of the accommodation types did not exist in the region of interest; hence, an SP choice experiment was designed that could accommodate the new options. The choice alternatives in the SP survey included two campground options, a lodge option, a cabin option and hotel/motel type accommodation. The descriptions of choice alternatives were based on an experimental design which systematically varied levels of the following types of attributes: type of setting, state/national park or forest, a variety of physical and other features that were present or absent, activities that could be enjoyed or not, fees, distances to nearby towns, typical weekend demand for sites or rooms, etc. A fraction of the complete factorial was designed by treating all attributes of all alternatives as a collective factorial, and developing the smallest orthogonal design that permitted all main effects to be estimated independently. This design was used to produce 64 different choice sets that consisted of descriptions of the accommodation and destination options. Eight versions of the survey, consisting of eight choice sets each, were created to insure that the demands on any one respondent were not too burdensome. Respondents were randomly assigned to one of the versions.

Respondents were sampled and recruited from lists of consumers who participate in and/or purchase outdoor oriented activities or products. The subjects were sampled by telephone and asked to participate in a follow-up mail survey containing the SP survey. All sampling and data collection were conducted by the Survey Research Center of a major US university, which resulted in a total sample of 622 useable respondents. The respondents to the survey had to decide whether to take an overnight recreational trip or not, and if so, to choose one of the available accommodation and destination options. Background characteristics of respondents and other information about recreational behaviour, preferences and experiences also were collected for use in modelling.

We postulated that latent classes existed in this population, defined on the basis of socio-demographic characteristics (education, income, marriage status, etc.), as well as attitudes towards recreation, as measured by a well-known scale called the Recreational Opportunity Spectrum (ROS) scale. The model described in Section 3.3.3 was estimated for different numbers of latent classes; Table 3 contains log likelihoods and numbers of parameters corresponding to 1, 2 and 3 latent classes. The last two of these models assume that tastes vary between classes but variances are uniform. In notation previously introduced, we estimated β_s, s=1,...,S, while constraining the scale parameters to be equal (i.e. $\lambda_1=...=\lambda_S=1$). It should be noted that the number of taste parameters is quite large (89 utility function parameters per class, including alternative-specific constants), due to the large number of attributes varied in the SP choice experiment.

Table 3 - Determination of the Number of Latent Classes

Number of Latent Classes	Log Likelihood	Number of Parameters	Akaike Information Criterion
1	-6,814.1	114	13,856.2
2	-6,490.0	188	13,356.0
3	-6,436.5	255	13,383.0
2 (restr.)	-6,573.4	118	13,382.8

To determine the likely value of S, a discrete parameter, we used the minimum AIC (Akaike Information Criterion), defined as $-2[LL_s+K_s]$, where LL_s is the log likelihood at convergence for the model with s segments, and K_s the corresponding number of parameters. The two-segment solution provided the best balance of explanatory power and model parsimony.

The question of most interest to this paper was whether the differences underlying the two classes uncovered in this sample could be explained by variance (scale) heterogeneity instead of taste heterogeneity. To answer this question, we estimated a model with two latent classes, with the restriction that all attribute parameters were constrained to be equal across segments, but alternative-specific constants and scale parameters were allowed to vary between segments. Hence, segments could have unequal average preferences for each alternative (i.e. alternative-specific constants were free) and unequal variances, but all attribute slopes (i.e. tastes for attributes) were forced to be homogenous. The scale of class 1 was normalized to 1.0, and the scale parameter of class 2 estimated relative to that of class 1.

Interestingly (Table 3), the taste-restricted, 2-segment model (with 118 parameters) has an AIC equal to that of the 3-segment solution (255 parameters) and a difference of only 26.8 AIC points with respect to the 2-segment unrestricted model with 188 parameters. Despite this performance, a formal likelihood ratio test between the restricted and unrestricted two-class models will reject the hypothesis of taste homogeneity but variance heterogeneity (the calculated chi-squared is 166.8 with 70 degrees of freedom, compared to the table value of 51.74 at a 95% confidence level).

The key point arising from this result is that restricting 71 taste parameters across the two classes and freeing *one scale parameter* (therefore yielding a total of 70 degrees of freedom) produced a parsimonious model which, for all practical purposes, performed statistically almost as well as the unrestricted two-class model. This suggests an intriguing hypothesis for further research, namely that judicious freeing of a very small number of taste parameters leads to intermediate specifications that explain observed behaviour as well as unrestricted models. These results suggest that it may be the case in a number of behavioural modelling applications that differences between classes principally are driven by variance heterogeneity, and only secondarily driven by true taste heterogeneity.

Such a behavioural possibility is of more than academic interest. As noted in the previous empirical example, if individual consumers or classes of consumers differ fundamentally in their tastes, then marketing strategy should be directed to develop differentiated products for each. On the other hand, if the main difference between consumers is in variance in response to options offered, marketing strategy should be directed towards addressing the causal factors behind variance differences (e.g. consumer information processing capabilities,[4] consumer uncertainty with respect to aspects of usage, misperceptions about brand positions, and the like.)

5. Summary and Conclusion

[4]de Palma et al. (1994) assume that consumers have different abilities to choose, such that an individual with lower ability to choose will make more errors in comparisons of marginal utilities. They then outline the implications of limited processing capability on choice and discover that heterogeneity over a population in the ability to choose produces widely different choices, *even if in all other aspects the individuals are identical* (including identical tastes).

Discrete choice models estimated from mixtures of data sources provide exciting opportunities for research into the demand for new products and services, as well as adjustments in the attribute space of existing products and services outside attribute domains currently available in markets. This paper reviewed several new analytical techniques being developed in marketing, resource economics and transportation to analyze, model and combine sources of preference data, and illustrated the value of applying a number of relatively new types of discrete choice models which explicitly allow for error heteroscedasticity.

Recognition that multiple data sources, especially mixtures of RP and SP data, can be used to estimate choice models within the well-developed random utility theoretic framework opens up many opportunities for further research into ways of taking into account both scale and taste, as well as capturing complex behavioural constructs such as latent segmentation, random taste variation, and dynamics, not to mention extend utility spaces well beyond the domains contained within today's actual markets. Central to the future enrichment of choice models is a greater respect for, and a richer representation of, the "lurking λ's" which impact in complex ways many elements of a choice model.

The implications of behavioural differences in both means and variances of the distribution of consumer choices are both theoretical and practical. Theoretically, the types of models discussed in this paper extend our ability to represent complex behavioural processes, and open the way for future research into the antecedents of differences in variances, as well as the previously well-trodden road of differences in means. To our knowledge, the suggestion that variance components can be specified in significant behavioural ways consistent with random utility theory is new. Similarly, this paper demonstrated that placing structure on variance components was not only possible, but practical; moreover, our empirical examples suggested significant gains in behavioural insights over more traditional models that focus largely on differences in means.

As discussed and demonstrated in the empirical examples in the paper, variance differences can have important practical implications for marketing (and other) policy. In particular, as the recreation destination/accommodation choice and the Best/Worst shopping example revealed, it may be that there are fewer differences in tastes than previously thought, but that even if tastes are relatively homogeneous, different segments may exhibit different variances in their preference evaluations. The latter finding has considerable practical import for marketing policy: in both these examples it suggests that resources should be devoted more to understanding the antecedents of variance and developing strategies aimed at overcoming and/or reducing same, rather than attempting to serve what are believed to be different taste segments. Indeed, a strategy based on beliefs in differences in means would be incorrect for a population displaying differences in variance, as the optimal strategy would be to direct efforts to reduce variance.

This paper provided theory and empirical evidence to suggest that future research directed towards models that can better capture differences in variances, as well as empirical work directed towards understanding sources of differences in variances and their role in choice, should be both interesting and fruitful. Presently, the authors individually and collectively are engaged in research programs consistent with these suggestions. For example, Louviere and Swait (1996b) have reviewed a number of published results, as well as provided new empirical results, that suggest that accounting for differences in variance will often account for much of the difference in taste parameters in studies of preference and choice. In particular, they show that context effects, previously shown to produce large differences in model results in consumer decision-making experiments, largely disappear when differences in variances between contexts are taken into account. This, in turn, suggests that previous conclusions regarding the effects of context on means may have to be revised to indicate that the primary effect appears to be on variance. Other research, notably de Palma et al. (1994) and Swait and Adamowicz (1996), is examining how complexity of decision tasks and decision environments impacts variance in preferences and choices. Preliminary results by Swait and Adamowicz (1996) support the conclusion that decision context complexity increases variance. Other research involves differences in effects due to types of product information, such as visual versus verbal information, as well as differences in the variance of new product trial and repeat choice rates over time as a function of a variety of competitive and environmental differences, as well as individual characteristics.

The foregoing constitute only a small example of the many new avenues for research suggested by the theory and empirical results reviewed in this paper. Taken as a whole, they would seem to constitute a compelling case that econometricians and marketing researchers have much to learn from each other to the mutual enrichment of both fields.

References

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

Adamowicz, W., Swait, J., Boxall, P., Louviere, J. and Williams, M. (1996) Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation, Journal of Environmental Economics and Management, forthcoming.

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.

Anderson, S., de Palma, A. and Thisse, J. (1992) Discrete Choice Theory of Product Differentiation, Cambridge: MIT Press.

Anderson, D.A. and Wiley, J.B. (1992) Efficient Choice Set Designs for Estimating Availability Cross-Effects Models, Marketing Letters, 3 (4), 357-370.

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

Ben-Akiva, M. and Morikawa, T. (1991) Estimation of travel demand models from multiple data sources, in M. Koshi (ed.) Transportation and Traffic Theory, Proceedings of the 11th ISTTT, Elsevier, Amsterdam, 461-476.

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

Bhat, C.R. (1996) An endogenous segmentation mode choice model with an application to intercity travel, Transportation Science (in press).

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.

Bunch, D. (1991) Estimability In The Multinomial Probit Model, Trans. Research, 25B (1):1-12.

Bunch, D.S., Louviere, J.J. and D.A. Anderson (1996) A Comparison of Experimental Design Strategies for Choice-Based Conjoint analysis with Generic-Attribute Multinomial Logit Models, Unpublished Working Paper, UC Davis Graduate School of Management, May.

Carson, R., Louviere, J., Anderson, D., Arabie, P., Bunch, D., Hensher, D., Johnson, R., Kuhfeld, W., Steinberg, D., Swait, J., Timmermans, H. and Wiley, J. (1994) Experimental Analysis of Choice Marketing Letters., 5(4):351-368.

Daganzo, C. (1980) Multinomial Probit, New York: Academic Press.

de Palma, A., Myers, G.M., Papageorgiou, Y.Y. (1994) Rational Choice Under an Imperfect Ability to Choose, American Economic Review, 84:419-440.

Dillon, W. and Kumar, A. (1993) Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview, in Handbook of Marketing Research, R. Bagozzi, Editor, London:Blackwell Publishers.

Elrod, T. and Keane, M. (1995) A Factor Analytic Probit Model for Representing the Market Structure in Panel Data, Journal of Marketing Research, 32:1-16.

Englin, J. and Cameron, T.A. (1996) Augmenting travel cost models with contingent behavior data, Environmental and Resource Economics, 7, 133-147.

Gopinath, D. and Ben-Akiva, M. (1995) Estimation of Randomly Distributed Value of Time, Working Paper, Dept. of Civil Engineering, Massachusetts Institute of Technology, October 1995, 36 pp.

Green, P. and Srinivasan, V. (1978) Conjoint Analysis in Consumer Research: Issues and Outlook, Journal of Consumer Research, 1:61-68.

Green, P. and Srinivasan, V. (1990) Conjoint Analysis in Marketing Research: New Developments and Directions, Journal of Marketing, 54(4):3-19.

Green, P. and Wind, Y. (1971) Multiattribute Decisions in Marketing: A Measurement Approach, Hinsdale, Dryden Press, Illinois.

Greene, W. (1996) Heteroskedastic extreme value model for discrete choice, New York University.

Hanemann, M. and Kanninen, B. (1997) The statistical analysis of discrete-response CV data, in I. J. Bateman and K.G. Willis (eds.) Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries. Oxford University Press, Oxford.

Hausman, J. and D. Wise (1978), AFDC Participation: Measured Variables or Unobserved Characteristics, Permanent or Transitory, Working Paper, Dept. of Economics, MIT, Cambridge, MA.

Heckman, J. (1981) Statistical Models for Discrete Panel Data, in Manski, C.F. and McFadden, D. (eds.), Structural Analysis of Discrete Data with Econometric Applications, Cambridge, MA: MIT Press, 114-178..

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

Hensher, D.A. (1996a) Extending Valuation to Controlled Value Functions and Non-Uniform Scaling with Generalised Unobserved Variances, invited paper presented at the International Conference on the Theoretical Foundations of Travel Choice Models, Stockholm, August 7-11.

Hensher, D.A. (1996b) A practical approach to identifying the market for high speed rail: a case study in the Sydney-Canberra corridor, Institute of Transport Studies, The University of Sydney, July.

Hensher, D.A. and Barnard, P.O. (1990) The Orthogonality Issue in Stated Choice Designs, in Fischer, M. Nijkamp, P. and Papageorgiou, Y. (Eds.) in Spatial Choices and Processes, North-Holland, Amsterdam; 265-278.

Hensher, D.A. and Bradley, M. (1993) Using stated response data to enrich revealed preference discrete choice models, Marketing Letters, 4(2):39-152.

Hensher, D. and Johnson, L. (1981) Applied Discrete Choice Modelling, London: Croom Helm, London.

Hensher, D.A. and Louviere, J.J. (1983) Identifying Individual Preferences for Alternative International Air Fares: An Application of Functional Measurement Theory, Journal of Transport Economics and Policy, Vol. XVII (2), 225-245.

Hensher, D.A., Smith, N.C., Milthorpe, F.W. and Barnard, P.O. (1992) Dimensions of Automobile Demand: A Longitudinal Analysis, North Holland, Amsterdam.

Huber, J. and Zwerina, K. (1996) The Importance of Utility Balance in Efficient Choice Set Designs, Journal of Marketing Research, 33:307-317.

Kahneman, D. and Tversky, A. (1979) Prospect Theory: an Analysis of Decisions Under Risk, Econometrica, 47, 263-291.

Kahneman, D. and Tversky, A. (1984) Choices, Values and Frames, American Psychologist, 39, 341-350.

Keane, M. (1995) Modelling Heterogeneity and State Dependence in Consumer Choice Behavior, Working Paper, July 1995, Dept. of Economics, University of Minnesota.

Keller, K. L. (1993) Conceptualizing, Measuring and Managing Customer-Based Brand Equity, Journal of Marketing, 57 (1), 1-22.

Krantz, D.H. and Tversky, A. (1971) Conjoint-Measurement analysis of Composition Rules in Psychology, Psychological Review, 78, 151-169.

Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A. (1971) Foundations of Measurement. New York: Academic Press.

Lancaster, K. (1966) A New Approach to Consumer Theory, Journal of Political Economy, 74:132-157.

Louviere, J.J. (1974) Predicting the Evaluation of Real Stimulus Objects From an Abstract Evaluation of Their Attributes: The Case of Trout Streams, Journal of Applied Psychology, 59(5):572-577.

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

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

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

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

Louviere, J.J., Fox, M. and Moore, W. (1993) Cross-Task Validity Comparisons of Stated Preference Choice Models, Marketing Letters, 4(3):205-213.

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

Louviere, J.J. and Swait, J. (1996a) Best/Worst Conjoint, Working Paper, Dept. of Marketing, Faculty of Economics, University of Sydney, August 1996.

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 1996.

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

Luce, R. D. (1959) Individual Choice Behavior: A Theoretical Analysis, New York:Wiley.

Lynch, J. (1985) Uniqueness Issues in the Decompositional Modeling of Multiattribute Overall Evaluations, Journal of Marketing Research, 22:1-19.

Machina, M.J. (1987) Choice Under Uncertainty: Problems Solved and Unsolved, Economic Perspectives, 1 (1), 121-154.

Meyer, R., Levin, I. and Louviere, J. (1984) Functional Analysis of Mode Choice, Transportation Research Record, 673:1-7.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour, in P. Zarembka (ed.) Frontiers in Econometrics, New York:Academic Press, 105-142.

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., Tye, W. and Train, K. (1976) An Application of Diagnostic Tests for Independence from Irrelevant Alternatives Property of the Multinomial Logit Model, Transportation Research Record, 534.

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 RP/SP Combined Estimation Method, Transportation, 21(2):153-166.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) Numerical Recipes, Cambridge:Cambridge University Press.

Restle, F. (1961) Psychology of Judgment and Choice. New York: Wiley.

Rushton, G. (1969) Analysis of Spatial Behaviour by Revealed Space Preference, Annals of the Association of American Geographers, 59, 391-400.

Small, K. (1987) A Discrete Choice Model for Ordered Alternatives, Econometrica, 55(2):409-424.

Small, K. (1994) Approximate Generalized Extreme Value Models of Discrete Choice, Journal of Econometrics, 62:351-382.

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., Erdem, T., Louviere, J.J. and Dubelaar, C. (1993) The Equalization Price: A Measure of Consumer-Perceived Brand Equity, International Journal of Research in Marketing, 10:23-45.

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

Swait, J., Louviere, J.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.

Swait, J. and Naik, P. (1996) Consumer Preference Evolution and Sequential Choice Behavior, Working Paper, Dept. of Marketing, College of Business Administration, University of Florida, June 1996, 44 pp.

Swait, J. and Stacey, E.C. (1996) Consumer Brand Assessment and Assessment Confidence in Models of Longitudinal Choice Behavior, presented at the 1996 INFORMS Marketing Science Conference, March 7-10, 1996, Gainesville, FL.

Swait, J. and Sweeney, J. (1996) Perceived Value and Its Impact on Choice Behavior in a Retail Setting, Working Paper, Dept. of Marketing, College of Business Administration, University of Florida, August 1996.

Thurstone, L. (1927) A Law of Comparative Judgment, Psychological Review, 34:273-286.

ITS

**INSTITUTE OF
TRANSPORT STUDIES**

The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University