Marketing Letters 10:3 (1999): 205–217
© 1999 Kluwer Academic Publishers, Manufactured in The Netherlands

Combining Sources of Preference Data for Modeling Complex Decision Processes

JORDAN J. LOUVIERE
Faculty of Economics, University of Sydney

ROBERT J. MEYER
Wharton School of Business, University of Pennsylvania

DAVID S. BUNCH
Graduate School of Management, University of California, Davis

RICHARD CARSON
Department of Economics, University of California, San Diego

BENEDICT DELLAERT
Center for Economic Research, Tilburg University

W. MICHAEL HANEMANN
Department of Agricultural and Resource Economics, University of California, Berkeley

DAVID HENSHER
Faculty of Economics, University of Sydney

JULIE IRWIN
Wharton School of Business, University of Pennsylvania

Abstract

We review current state-of-the-art practices for combining preference data from multiple sources and discuss future research possibilities. A central theme is that any *one* data source (e.g., a scanner panel source) is often insufficient to support tests of complex theories of choice and decision making. Hence, analysts may need to embrace a wider variety of data types and measurement tools than traditionally have been considered in applied decision making and choice research. We discuss the viability of preference-stationarity assumptions usually made when pooling data, as well as random-utility theory-based approaches for combining data sources. We also discuss types of models and data sources likely to be required to make inferences about and estimate models that describe choice dynamics. The latter discussion is speculative insofar as the body of literature on this topic is small.

Key words: Choice modeling, started preference data, data pooling, context effects, choice dynamics

1. Introduction

The objective of this paper is to review the state of the art and discuss future research issues associated with combining/pooling sources of preference data. Our basic premise is that any one source of data—be it a consumer panel or a cross-sectional survey—will invariably be limited in its ability to test a range of competing theories about the process that underlie observed patterns of choice and judgment. Thus, decision making and choice research could benefit significantly from an ability to combine multiple sources of data, each possessing different strengths and weaknesses in illuminating different aspects of decision making. This paper represents a modest step in what we believe should be an on-going dialogue about research design issues for understanding and modeling choice and decision processes. Toward that end the following general topics are discussed in subsequent sections:

- Theoretical requirements for combining/pooling preference data sources;
- Random utility theory-based approarches for combining/pooling sources of preference data for conditional-choice problems; and
- Issues associated with designing and/or obtaining data appropriate for gaining insights useful for modelling dynamic choice processes.

Background: Preference, choice, and identification problems

Theoretical and applied research in decision-making and choice behavior attempts to solve what might seem to be a straightforward modeling problem: find a functional mapping that associates observed choice responses (e.g., frequencies) with measures of the attributes of choice options and of decision makers. Although many theoretical bases for deriving such mappings exist, random utility theory (RUT) has proved useful and widely applicable (e.g., McFadden, 1974). That is, within a given choice set we assume that a stochastic ordering of preferences over alternatives exists, which can be scaled in terms of their attributes. The probability that a given option is chosen is the probability that it will be the highest-ranked alternative at the time of choice.

Of course, at this level of generality random utility theory (RUT) offers little prescriptive content. It gains substance only if we impose identifying restrictions on the probabilistic scaling function; i.e., impose restrictions to permit the theory to be uniquely applied to a particular source of data. Yet, it is precisely at this point that the modern literature of decision-making and choice behavior hits a stumbling block. While one might envision a wide variety of functional forms that could described and/or explain how choices or preferences are formed and evolve over time, most commonly available data sets allow tests of only simple model specifications.

To illustrate, consider a one-time forced choice from a small set of options. RUT posits that the probability that a given option i will be chosen from this set is the probability that its latent (unobserved) utility U_i is the highest at the time of choice. Latent utility, in turn, is then decomposed into a systematic component (V) and a random component (ε), represented by the mapping $U = V + \varepsilon$, where $\mathrm{E}(\varepsilon)$ is usually assumed to be zero as an identifying restriction. The strict component of utility V is then typically modeled as a function of an observed attribute vector X_i, expressed as the mapping $V(X_i)$. What assumptions should one make about the form of the function $V(X_i)$ and the distribution of ε, $f(\varepsilon)$? Lacking strong priors, we would clearly prefer to specify $V(X_i)$. and $f(\varepsilon)$ as generally as possible, and let a (restricted) form emerge from data analysis and model testing. Unfortunately, while computational procedures for estimating complex RUT models exist, efficient estimates of model parameters are frequently unobtainable given the limitations of most normally-available data sets (including longitudinal panels).

As an example of the problem this limited efficiency poses for analysts, consider the hypothesis that the strict function $V(X_i)$ will frequently be nonadditive in form, or display interactions among attributes. There are a number of theoretical reasons for suspecting that such a hypothesis would hold true for many consumer markets. Interactions would arise, for example, if consumers have limited information about product qualities and infer values from brand names (a process that would lead to marketing-mix effects varying by brand), or if consumers form overall impressions of brand value using something other than an additive composition rule (see. e.g., Cooper and Nakanishi, 1988; McClelland and Judd, 1993). In light of this, it is perhaps surprising to note that few published applications of choice models report statistical support for estimated interactions, or when found, report their absolute sizes to be small (e.g., Johnson and Meyer, 1984).

Why is this the case? This question was recently pursued by McClelland and Judd (1993), who argued that failures to reject additivity in choice and judgement research most likely were due to the lack of power of reported interaction tests rather than any an inherent validity of additivity assumptions. Specifically, because marketing actions by firms will tend to be both correlated and vary over a restricted range, large data sets commonly have far less statistical information than one might assume. McClelland and Judd (1993), in particular, argue that that even in cases where psychological scales are measured and modeled directly, the power of attribute interaction tests in field databases may be as little as 20% or less of that associated with replicated factorial designs. This limitation, in turn, virtually guarantees that a researcher will be unable to reject additivity assumptions even when they are false.

This problem is exacerbated when the researcher seeks to estimate a latent preference structure from the information contained in data on *discrete* choices—the usual problem in random utility theory. A good illustration is the difficulty that arises when one attempt to relax the widely-used assumption that the random component of utility, ε, is independently and identically distributed among options. It should be recalled that the distribution $f(\varepsilon)$ essentially characterizes the modeler's ignorance about the drivers of utility in a given setting. Hence, the less a modeler understands (and/or can measure the drivers of) a choice process, the more choice predictions will be driven by the assumptions that are made about the form of $f(\varepsilon)$. In unfamiliar applied settings, therefore, the analyst would prefer to impose as few restrictions as possible on the structure of the random component—such as by allowing variances to vary across options (reflecting differential precision in our ability to predict utility), and errors to be correlated among options.

How easy is it to efficiently estimate models such as this? Ongoing work by Bunch (1999) explores this issue by examining a "small" problem involving a designed stated choice experiment with two brands (plus a "no choice" option), three generic attributes (price, plus two binary features), and an additive function for the systematic component. Monte Carlo experiments were performed to examine estimation behavior for RUT models with alternate covariance structures (e.g., simple IID errors, general covariance structure for brands, covariance for brands plus price heterogeneity). An experiment consisted of generating many data sets from a given "true model" using a specific sample size (number of independent choices), and then recovering the estimates. Biases, mean square errors, and empirical distributions of standard statistical tests (e.g., t-statistics, likelihood ratios) were obtained for a range of sample sizes. Simple IID probit models (analogous to multinomial logit models) exhibit stable asymptotic behavior for 400 independent observations (a rule of thumb familiar to many researchers); however, comparable behavior for similar models requiring estimation of two covariance parameters associated with brands requires a four-fold increase in the number of observations. Other models containing what would seem to be simple variance components required even more observations (3200 or more), indicating that model identification is likely to be a major issue, even when data are produced from experimental designs.

Toward a solution: a framework for data pooling

Although such problems of data inadequacy are unsolvable within any one data set, they *might* be solvable if the analyst could *pool* the information contained in multiple databases and measurement instruments. However, to be viable any exercise in pooling data sources requires the validity of at least one strong (but testable) assumption, namely that underlying all data is a common latent-preference structure. How valid will such an assumption be in general? Some skepticism would be understandable. There is, after all, a large literature in behavioral decision theory that purports to show that preferences are often context-dependent, constructed in response to the observed features of a choice problem (e.g., Payne, Bettman, and Johnson, 1992; 1993; Simonson and Tversky, 1992). If latent preferences were indeed responsive to task context, data-pooling would be futile, because explanations of preferences could not be generalised beyond the domains in which they are observed.

Nonetheless, we note that although this literature clearly shows that preference measures can vary across response instruments, it does not necessarily follow that the fundamental preferences of interest to choice modelers (the latent construct of utility) will vary. Quite to the contrary, an important finding in work on context effects is that differences in response patterns across instruments often follow simple re-scaling rules, implying that we often may be observing task effects on *response language*, not necessarily effects on the latent, underlying fundamental preferences themselves.

For example, a well-known context effect is the so-called "preference reversal" phenomenon, where it has been observed that survey formats that elicit values by asking decision makers to state a maximum willingness to pay for a good yield higher apparent sensitivities to price (or, equivalently, lower marginal valuation of product quality) than formats where decision makers evaluate the overall attractiveness of an option where price is one of the attributes (e.g., Bazerman, Loewenstein, and White, 1992; Irwin, 1994; Tversky, Slovic, and Kahneman, 1990). In particular, when decision makers assert a level of willingness-to-pay, they frequently betray these values when faced with choices.

Yet, there is growing evidence that this result accrues more to a simple bias in the two response languages than any fundamental difference in the latent values that are being elicited. In ongoing work, Irwin and Meyer (1999) examine the stability of preferences elicited by different response-modes over time in a setting in which subjects are trained to predict the apartment preferences of a hypothetical real-estate client using one of two response languages: either holistic judgments of acceptability or statements of willingness-to-pay. After learning the client's true trade-off function via one of these two modes, they then are asked to predict preferences using the opposite mode. Reinforcing the generality of prior work on response-mode effects (e.g., Tversky, Slovic, and Kahneman, 1990), they find that each mode is indeed associated with a systematic bias that does not vanish with experience. That is, predicted willingness to pay judgments tend to be consistently too low, whereas predicted overall evaluations of acceptability tend to be consistently too high. The important feature of the data, however, is that *after this constant language bias is controlled for*, both methods yield identical inferences about the marginal value of different non-price attributes. In short, they reach the conclusion that the two methods are not revealing fundamentally different preferences: they are revealing the same preferences, but just expressed on different scales.

A wider survey of the literature on cross-task validity of preference measures also suggests a more optimistic conclusion about the potential to pool preference data. For example, there is extensive evidence that responses to stated preference experiments can predict actual parallel marketplace choices (Carson, et al., 1996, Swait, et al., 1995, Blamey, et al., 1999), despite the potential for significant context effects. Likewise, there is also evidence that responses to hypothetical choice options in experiments are quite similar across stimulus-presentation formats (e.g., Burke, Harlam, Kahn and Lodish, 1992), and model parameters estimated from choice experiments can be highly correlated with parallel process-tracing measures of decision making, such as the length of time individuals spend looking at the attributes of different options (e.g., Johnson, Meyer, Hardie and Walsh, 1998).

A case example: Pooling stated preference (SP) and revealed preference (RP) data in cross-sectional random utility theory based models

Even if one knows that data from two different sources resulted from a common underlying preference process, one still must know the re-scaling rule that relates both data sets. Although re-scaling rules could be complex, evidence suggests that in many cases they may be remarkably simple, indeed, as simple as a single parameter. Morikawa (1989) provided the first evidence for this by proposing and testing the hypothesis that there should be a simple relationship between sources of preference data across data contexts if the underlying indirect utility functions were the same (see also Ben-Akiva and

Morikawa, 1990; Ben-Akiva et al., 1994). Morikawa (1989). He reminded us that RUT choice models have embedded scale constants in the vectors of utility parameters that are inversely proportional to the variance of the random component, and cannot be separately identified in any one source of preference data. That is, if the preference parameters of the indirect utility functions that characterise each data source are the same, the utility parameter vectors should be proportional to one another, and the constants of proportionality should equal the scale constant ratios.

Morikawa (1989) and Ben-Akiva and Morikawa (1990) tested this hypothesis by comparing choice model parameters estimated from an SP experiment involving choice of bus and train for inter-city trips in the Netherlands with actual mode choices. They could not reject the hypothesis of parameter proportionality between data sources. There have been numerous tests of this basic hypothesis for two or more sources of preference data since 1990; most tests failed to reject the hypothesis or rejected it for only one or a small subset of the attributes in the indirect utility function (e.g., Swait and Louviere, 1993; Hensher and Bradley, 1993; Louviere, Fox and Moore, 1993; Louviere, 1995; Swait, Louviere and Williams, 1995; Adamowicz, Louviere, and Williams, 1994).

The Morikawa (1989) hypothesis is often seen narrowly (especially in transport) as a way to combine and test the equality of RP and SP data sources. In fact, this result can be applied to any two or more preference data sources consistent with the Luce and Suppes (1965) ranking theorem. Hence, it can be used to combine and compare many different preference data sources. For example, Louviere et al. (1993) combined and compared six different sources of RP and SP data; Swait et al. (1995) combined and compared RP and SP data for three cities; and more recently Cameron et al. (1998) combined and compared six preference elicitation procedures.

Side bar: Whither context effects and market segments?

Besides establishing the viability of estimating models by pooling data sets, the above research stream also suggests that past research in decision-making and choice behaviour may overstate the degree to which choice processes vary across decision makers and contexts. For example, many methods for modelling heterogeneity in choice-model parameters in populations (e.g., latent-class methods) assume that random component variances are homogeneous over the population (e.g., Kamakura and Russell, 1989). If this is false, results will overstate degrees of preference heterogeneity in samples because variance differences are confounded with utility-function mean parameter differences.

This same risk of confounding may be even more acute in attempts to form inferences about the degree to which changes in choice context (e.g., choice set size) induce changes in latent preferences. As earlier suggested, the existence of stationary latent preferences does not necessarily imply that an agent will always be observed making the same choices regardless of context. This easily can be seen by taking any initial systematic and random components in some one choice context, and then, in a second context, applying this same systematic component but letting the variance of the random component grow very large relative to that of the systematic component. In this instance choices will always be more equally distributed among the alternatives than initially—an instance where context affects choice but does not affect latent preference.

This idea was more fully developed by Albernini (1992), who notes that in many published demonstrations of preference invariance across contexts (e.g., Tversky, Slovic, and Kahneman, 1990) involved tests that implicitly that different elicitation procedures and tasks were characterized by symmetric error distributions with equal variances, an assumption that is quite possibly false. To wit, Olsen, Swait, and Louviere (1995) have found that context manipulations (e.g., set size) impact not only the means of distributions, but also random component variances. Once random component differences were taken into account mean differences were eliminated.

Of course, we do not mean to suggest that all segment differences and/or context effects results (e.g., preference reversals) have random component variance explanations. To the contrary, there are good theoretical and empirical reasons to suspect that in some cases changes in task structure will induce genuine changes in latent preference orders (e.g., changes in choice objectives). Instead, our argument is simply that to be valid and useful research must differentiate context effects whose locus is primarily on response languages and variances from those that induce real changes in preference structures. We note that this differentiation is rarely undertaken; thus, it remains unclear how to interpret many published findings without ruling out such competing explanations.

Problems in Combining Data in Models of Choice Over Time

Issues in pooling data sources to estimate cross-sectional choice models seem to be reasonably well-understood, but this is not true for more general, and potentially more important problems involving choice dynamics. In this case, one faces difficult problems of how to combine data or information from several different levels of aggregation, such as individual-level choices of brands over time and space; aggregate weekly choices of brands (stock-keeping units—*SKU*s) from scanner panel sources; advertising flights and schedules, including *GRP*s, *TARP*s; etc. Unresolved econometric issues remain in the analysis of such sources of information in combined data sources, such as how to weight each source of data in model estimation (classically or Bayesian), or deal with different sources of errors in different data sets, such as independent versus serially-correlated observations.

For example, suppose that one has brand choice data from a choice experiment in which one varies prices of brands systematically, and wants to combine these data with parallel weekly unit sales data from one or more supermarkets. Even if one can assume that the underlying preferences are identical in both sets of data, the re-scaling rules required to allow one to pool them to estimate a common indirect utility function and the potential and actual effects of confounding serial and panel unobserved brand and price effects remain unclear (e.g., Blamey, Morrison, Bennett and Louviere, 1998).

A more basic problem is that knowledge of choice process dynamics is much less well-developed than our understanding of static or cross-sectional choices. For example, a time series of choice observations is unlikely to result from independent realisations of a single, time-invariant, utility-maximization process. Thus, a more fundamental concern in pooling data sources to better understand and model choice dynamics is that we lack a widely-accepted theoretical structure within which to combine data, or even a structure to prescribe the types of data that should be collected in certain situations.

To illustrate the problem and associated issues, let us consider how one might go about forecasting demand over time for a truly new technology like a compact satellite phone. One might approach the problem by first building a model of conditional choice among potential phone options that vary on attributes (e.g., price, features, voice clarity) using a choice (or conjoint-analysis) experiment conducted at one point in time. One might further assume that this model could be used to predict market shares at future points in time for a wide array of market options and total demand levels. Unfortunately, such an approach probably is inadequate in at least three respects:

1. Decisions about purchase timing (and volume) may be a function of the same factors that drive brand choices (attributes and options) as well as expectations about the future states of these factors. We do not yet know how best to design experiments or surveys to model purchase timing decisions. Also, most past empirical choice models in marketing and transport focused on brand/mode choices conditional on category choices (and fixed levels of demand), hence provide fewer insights for modelling category choices (do consumers want wireless phones at all?) and purchase timing. One must take such conditional choices into account, or risk sample selection bias. In turn, this suggests that we need to ask if share elasticities estimated from lower-level choices can be interpreted as ordinary or compensated demand elasticities.
2. One must take into account possible feedback loops between consumer choice processes and sellers' strategies to exploit these decisions (i.e., structural endogeneity). That is, many argue that parameter estimates from cross-sectional or panel data (e.g., price elasticities) reflect not only consumers' fundamental sensitivity to price (an unobserved structural parameter), but also strategic expectations about sellers' likely future prices. Thus, one can use non-structural parameters revealed by choice models to predict responses to product attribute changes only if they do not alter consumer expectations about future choice sets. The latter assumption may be untenable in rapidly evolving markets like new consumer packaged goods, PC's and the like.
3. One must recognise and represent learning effects in choice because aggregate choices at any one point in time are a mixture of trial and repeat conditional on trial choices. Both are random utility processes with systematic and random components, and both components are likely to change over time in response to changes in individual consumer circumstances, market evolution and seller actions. To date, only changes in systematic components have been examined, but random components also are likely to change over time in response to market and consumer evolution, and we need to understand this better.

Overcoming these concerns is challenging because researchers must specify the parameters of a complete choice and market system, a problem with which we have had little prior experience (although see Geweke and Keane, 1998). That is, such a system would need to explicitly model at least four classes of functional elements:

1. How consumers make purchase timing decisions at a given point in time t;
2. How consumers make brand or model-type choices conditional on a decision to purchase at t;
3. How the set of options available to consumers and their attributes (e.g., prices, communication levels) evolve over time; and
4. How the choice processes represented in (1) and (2) evolve over time in response to expected changes in (3).

Presently, these issues cannot be modelled easily, but we believe that progress can be made if we can find theoretically and statistically appropriate ways to pool data that provides insights into issues that cannot be obtained from typical single data sources currently available.

A First Step Toward Pooling: An Expanded Domain of Data Sources

We believe that progress can be made on such complex modelling problems if researchers embrace a wider variety of data types than previously considered and develop model specifications and estimation methods suitable to such data. For example, if we continue our phone example, one might try to collect and pool data sources such as:

1. Cross-sectional surveys and choice experiments to measure immediate conditional phone attribute trade-offs from restricted choice set sizes;
2. Corresponding panel data to track actual phone purchases and switching patterns, or field tests;
3. Forecasts of likely future changes in aggregate demand based on demographics and/or published forecasts of attributes, features, and technological changes,
4. Descriptive information about the attributes of both current and future choice alternatives;
5. Forecasts of likely changes in choice sets, either derived normatively (i.e., from game-theoretic models) or behaviorally (i.e., from modelling past seller actions or choices in designed experiments); and
6. Parallel data on sales evolution from similar products in the same and/or other markets.

None of these sources of data in and of themselves will be sufficient to model the entire behavioural system of interacting consumers and sellers (not to mention middle-men), or even any one of its components.

To illustrate, consider modeling consumer choice strategies. Economic theory suggests a theoretical solution for such problems, which is not viable in many field settings. Specifically, one could attempt to build a dynamic-structural model of choice from a time series of consumer choices and managerial actions (e.g., pricing decisions). Such a model might assume that consumer choices are the consequence of a multi-period utility maximization process, in which choices are based on rational expectations of both sellers' future actions and buyer's expected response to those actions (e.g., Erdem and Keane, 1996; Gonul and Srinivasan, 1996; Wolpin, 1996; Geweke and Keane, 1998). Unfortunately, although conceptually attractive, there are two major barriers to developing such a model:

1. Dynamic structural model estimation currently faces formidable computational barriers, hence applications will lag the reduction of these barriers. Experience suggests that dynamic-structural models are difficult to specify even when they can be estimated, and as complexity increases even fundamental identifiability is not obvious (e.g., Gonul and Srinivassan, 1996).
2. Even though one can specify and estimate such a model, it may not capture the real behavioural process. That is, many argue that market behavior at any point in time reflects strategic interactions between buyers and sellers, which are unlikely to be closely (or even roughly) approximated by optimal dynamic-programs. However, Geweke and Keane (1998) recently formulated and tested a structural model that does not assume that humans can solve dynamic optimum programs, a welcome development.

The foregoing discussion suggests that potentially fruitful approaches should try to find solutions to the following two problems:

1. Identify general forms of boundedly-rational models of dynamic interactions between buyers and sellers; and
2. Use some set of convergent methods to estimate their parameters.

For example, one way to approach the first problem might be to develop a reduced-form choice model that simulates buyers' adaptations to sellers' strategies. Such a model would recognise the effects of expectations of future states of the world as explanatory variables, but not necessarily endogenously generate optimal responses to them. On the other hand, it may be a challenge to obtain sources of data from which to estimate such models. A naive starting point might be to try to pool cross-sectional survey data measuring consumer expectations with panel data choices over time. These data should provide insights into associations between cross-sectional sales-response patterns relate and consumer expectations of sellers' actions, but would not yield insights into how changes in consumer choices lead to changes in sellers' product-attribute strategies.

Addressing the latter questions probably requires data from yet other sources, such as dynamic laboratory experiments. Dynamic laboratory experiments record how choices in simulated competitive markets change over time in response to systematic changes in seller's strategic product actions, which would provide insights into changes in consumers' choice strategies in response to such strategic actions. Such experiments often are used to test theory in experimental economics (e.g., Kagel and Roth, 1995), but have been used in field applications only recently. For example, Brewer and Hensher (1998) developed an experimental market to study interactive telecommuting decisions. Justifiably, many are skeptical that consumer behaviour in experimental, time-compressed markets can simulate the behavioural changes that occur in real markets over time. This concern almost certainly will be true at least to some extent, although there is evidence that the two can be similar

(Burke et al., 1992). Like many problems previously discussed in pooling preference data, one hopes that the differences are a matter of scale instead of fundamental process, and if the latter is true, pooling of cross-sectional and/or panel data bases would be feasible.

Conclusion: Current Accomplishments and Future Problems

What can be learned by using single data sources to study consumer decision making and choice processes? If one's objectives are modest, such as describing and predicting cross-sectional or stationary panel behaviour, using simple reduced from statistical models to approximate such behaviour are likely to suffice. Indeed, most published models fit data and/or predict short-term behaviour reasonably well (i.e., significantly better than chance). In contrast, if one wants to understand and predict more complex behavioural systems, such as demand for new product innovations like the satellite phone discussed earlier, the outlook seems much bleaker. Many ideas have been advanced for theoretically appealing processes that underlie complex choice systems (e.g., learning), but these ideas have outpaced our ability to test them. We believe that alleviating this imbalance requires a fundamental shift in how we select our modelling problems. That is, instead of asking. "what model forms can a certain class of data best support (e.g., scanner-panel data)?", we should ask, "what sources of data do we need to best support a certain class of models?"

The objective of this paper was to initiate the latter discussion; and we think that issues involving pooling sources of preference data to enhance our understanding of process and improve model estimation should be a key discussion focus. A key issue then is the extent to there is preference invariance across data sources and collection methods. Our principal conclusion is that, contrary to much popular wisdom in behavioral-decision making and related research areas, consumers seem to have fundamental preferences and values that can be revealed by a variety of forms of preference measures and tasks. As well, many preference measures often exhibit significant cross-task robustness, especially after taking random component variance differences in scale into account.

If, as argued above, there is a basis for pooling sources of preference data, recent results from pooling SP and RP data are a very simple instance. That is, a common theoretical model is assumed to underlie both data types, the data have similar structural forms and there is a theoretical basis by which to pool the data sources (variance-scale ratio parameters). However, for more complex problems, such as pooling very different data sources to model dynamics in new markets, difficulties seem much more formidable.

A good illustration is the recent widespread availability of click-screen data for Internet sessions (West et al., 1999). Although many think that such data may eventually provide useful insights into diverse issues like how choice sets are formed, information is gathered, and even how preferences themselves are formed, such potential has yet to be realised. Currently, such data mainly are used to create simple summaries of Internet use patterns; hence their under-utilization is easy to explain. That is, there are no formal theoretical models of choice in interactive environments that can be used to understand and explain the choices underlying the data.

In contrast, the opposite is true for choice dynamics. There are a plethora of formal theories about dynamics and strategic adjustments, few of which have been tested against available data sources. However, little thought seems to have been given to possibilities of estimating and testing models that reflect aspects of theories by combining market simulations involving interactive agents with information-accelerators or other similarly innovative data pooling efforts.

These concerns, therefore, lead us to two important final observations. First, to address many unsolved problems in decision making and choice behaviour successfully will require much closer linkage between and co-development of theory and data than has been evident in previous research. Scanner-panel data represent an historical example of such co-development. That is, probabilistic discrete-choice models were regarded as a sound theoretical way to explain how consumers make conditional choices from sets of offerings, and single-source data bases were constructed to provide the data needed to estimate applied models (e.g., Guadagni and Little, 1983). Yet, these history lessons seem to have been forgotten. That is, as we develop more complex models that address ever more complex issues (e.g., learning), we do not want to ask what types of data we need to pool to address these issues. Instead, we should ask how we can obtain useful information from data sources not designed to provide it directly (e.g., scanner data). Thus, our overarching objective is to remind researchers of the importance of co-development of theory and data.

Finally, we also hope that researchers interested in decision-making and choice behavior will recognize that statistics and econometrics, while inherently useful to our endeavors, are not the end, but rather the means to support our continuing quest to reach the end. Far too much effort has been expended to develop complex models for their own sake, and far too little to develop models that parsimoniously approximate real processes. Prediction success and good model fits do not equal understanding, and understanding is unlikely to come from pedantically overly complex statistical models that demonstrate mathematical and statistical ability but little understanding of theory and substance. Like much of contemporary society, research in decision making and choice behaviour could benefit from less substance abuse.

References

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

Alberini, Anna (1992), *The Informational Content of Binary Responses*, Ph.D. Dissertation, University of California, San Diego.

Bazerman, Max H., Loewenstein, George F., White, Sally B. (1992), "Reversals of Preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly.* Vol 37(2), 220–240.

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

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. and Rao, V. (1994) "Combining Revealed and Stated Preferences Data" *Marketing Letters: Special Issue on the Duke Invitational Conference on Consumer Decision-Making and Choice Behavior*, 5(4).

Blamey, R. K., Bennett, J. W., Morrison, M. D. and J. J. Louviere (1998), "Divergences in revealed and stated preferences and the effect of social desirability prompts: validation of a choice experiment involving green product choice", Paper presented to the Biennial Conference of the International Society for Ecological Economics, Santiago, Chile, November 15–19.

Brewer, A. and Hensher, D. A. "Distributed Work and Travel Behaviour: The Dynamics of Interactive Agency Choices Between Employers and Employees" presented at the 8th IATBR Conference on Travel Behaviour Research, Austin, Texas, September 1997.

Bunch, David S., "Information and Sample Size Requirements for Estimating Non-IID Discrete Choice Models Using Stated-Choice Experiments," Working Paper UCD GSM 0399, Graduate School of Management, University of California, Davis, February 1999.

Burke, Raymond R. Harlam, Bari A. Kahn, Barbara E. Lodish, Leonard M. (1992), "Comparing Dynamic Consumer Choice in Real and Computer-Simulated Environments", *Journal of Consumer Research*. 19(1): 71–82.

Cameron, W. D. Schulze, R. G. Ethier, and G. L. Poe (1998), "Alternative Nonmarket Value-Elicitation Methods: Are the Underlying Preferences the Same?" working paper, Department of Economics, UCLA.

Cooper, Lee G., and Masao Nakanishi (1988), *Market-Share Analysis*. Norwell, MA: Kluwer Academic Publishers.

Erdem, Tulin. Keane, Michael P. (1996), "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets". *Marketing Science*. 15(1): 1–20.

Gonul, Fusun. Srinivasan, Kannan (1996), "Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model." *Marketing Science.* 15(3): 262–279.

Guadagni, P. M. and J. D. C. Little (1983), "A Logit Model of Brand Choice Based on Scanner Data", Marketing Science, 2(Summer), 203–238.

Geweke, J. F. and M. P. Keane (1999), "Bayesian inference for dynamic discrete choice models without the need for dynamic programming," In Mariano, Schuman and Weeks (Eds.) Simulation Based Inference and Econometrics, Cambridge, UK: Cambridge University Press, Forthcoming.

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

Hutchinson, J. W., Kamakura, W. A. and J. G. Lynch (1997) "Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research," Unpublished working paper, Dept. of Marketing, Wharton School of Business, U. of Pennsylvania, November.

Irwin, Julie R. (1994), "Buying/selling price preference reversals: Preference for environmental changes in buying versus selling modes." *Organizational Behavior & Human Decision Processes*. Vol 60(3), Dec., 431–457.

Irwin, Julie R., and Meyer, Robert J. (1999): Training the multilingual judge: the effects effects of learning on response-mode effects in multiattribute decision making", Working Paper, Department of Marketing, the Wharton School of Business.

Johnson, Eric J., Meyer, Robert J. (1984)," Compensatory choice models of noncompensatory processes: The effect of varying context". *Journal of Consumer Research*. Vol 11(1), Jun 1984, 528–541.

Johnson, Eric J., Meyer, Robert, Hardie, Bruce, and Walsh, John (1998), "Process-Assisted Choice Modeling", Working paper, Wharton School of Business, University of Pennsylvania.

Kagel, John, and Roth, Alvin (1995), *Handbook of Experimental Economics*, Princeton