Never mind the Data – Feel the Model*

A Keynote Paper for an International Conference on Transport Survey Quality

August 5-10, 2001, Kruger National Park, South Africa

A.J. Richardson,
The Urban Transport Institute
PO Box 363, Alexandra, Victoria, 3714, Australia
Ph: (613) 5774 7617   email: Tony.Richardson@tuti.com.au

Abstract

The rapid development of choice modelling procedures, especially using Stated Preference data, over the past twenty years, has created an opportunity for an imbalance between the models and the data used for construction of the models. This paper examines this imbalance and stresses the need for more attention to be paid to the needs of the respondent when collecting data for choice modelling. Drawing on the classic ideas of Alonso on the sources of error in modelling studies, the paper highlights the fact that the sophistication of the models and the data need to be commensurate. The paper gives some contrasting examples of studies in which sophisticated analysis has not been supported by equally sophisticated data collection, compared to studies where the emphasis has been placed on respondent-friendly designs without compromising the analytical outputs of the study. A major conclusion to be drawn from this paper is the absolute need to pilot test SP questions that could possibly be seen as being difficult for respondents to comprehend and answer.

* with due apologies to the British comedy series "Never Mind the Quality - Feel the Width", screened on ITV between 1967 and 1971, in which the differences in perspective between two London tailors are chronicled and analysed.

Introduction

The 2000 Presidential Election in the United States had at least one beneficial, if unintended, outcome. It clearly demonstrated that the detailed design of questionnaires can make a significant difference to the outcome of surveys (or elections). Ignoring all the debate about “dimpled chads” and “dangling chads” and so on, a single design error clearly resulted in the election of the wrong person to the office of President. This design error related to the use of the so-called “butterfly ballot” in the County of Palm Beach in Florida.

Given that each county in the USA is allowed to design their own voting system for Presidential elections, the County of Palm Beach used a system designed by election official Theresa LePore. Ms LePore was concerned about the eyesight of voters in her County and hence designed a form where the names of the candidates were printed in larger type and spread over two pages, with a single row of punch-holes down the middle of the two-page spread. This “butterfly ballot”, as it came to be known, is shown in Figure 1.

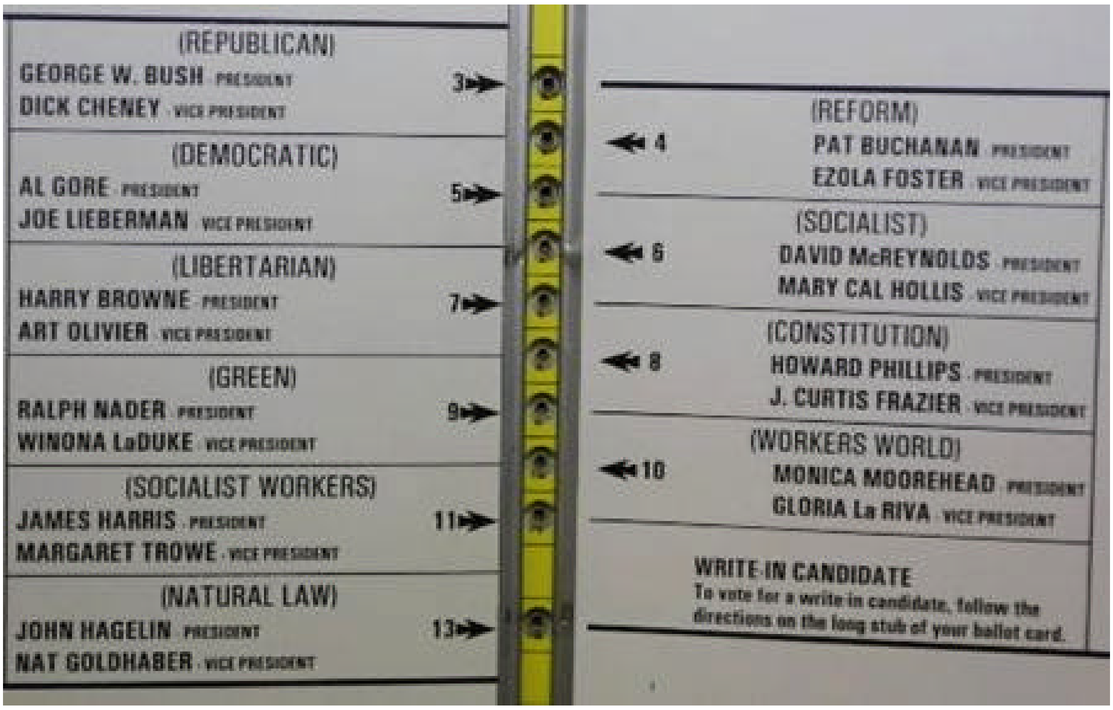

Figure 1 The Butterfly Ballot

In Ms LePore’s mind, and in hindsight, it is fairly obvious how the ballot was supposed to be used. Voters would find the candidate they wanted to vote for, then follow the little numbered arrow in that box to find the hole they were supposed to punch. Hence, people intending to vote for Al Gore would follow the arrow numbered 5 to punch the hole that was third from the top in Figure 1. Unfortunately, many voters did not perceive the ballot in the way that Ms LePore intended. Instead they saw that Al Gore’s name was second from the top of the list on the left-hand page and simply punched the second hole. Clearly, they had misunderstood what they were supposed to do, but the question remains as to whether it made any real difference. Surely, voters (survey respondents) make mistakes all the time, but if they are random errors then they will simply cancel each other out. In the case of the Palm Beach vote, however, we can demonstrate clearly that such cancellation of errors did not occur.

Looking at Figure 1, it can be seen that anyone punching the second hole to vote for Al Gore would instead have been voting for Pat Buchanan. Given that Pat Buchanan and George W. Bush are on the same side of the political spectrum, it might be expected that in counties where Bush polled well so too would Buchanan, and vice versa. In general, this is the case, as shown in Figure 2 (using results from the Florida Electoral Office). There are, however, two notable exceptions. The County of Miami-Dade had a very low vote for Buchanan, while the County of Palm Beach had a very high vote for Buchanan. The Miami-Dade result can be explained by the fact that Miami-Dade was strongly pro-Gore (and hence anti-Buchanan, who is at the extreme opposite end of the political spectrum from Gore). The Palm Beach result can only be explained by the use of the Butterfly Ballot. While Buchanan polled 0.3% overall in Florida, and only 0.1% in other pro-Gore counties such as Miami-Dade and Broward, he polled 0.8% in Palm Beach. This gave an extra 3000 votes to Buchanan, and took 3000 votes away from Gore. Given that the final margin in Florida, on which the Presidency rested, was less than 1000, it can clearly be seen that the use of the Butterfly Ballot resulted in the election of the wrong person to the office of President.

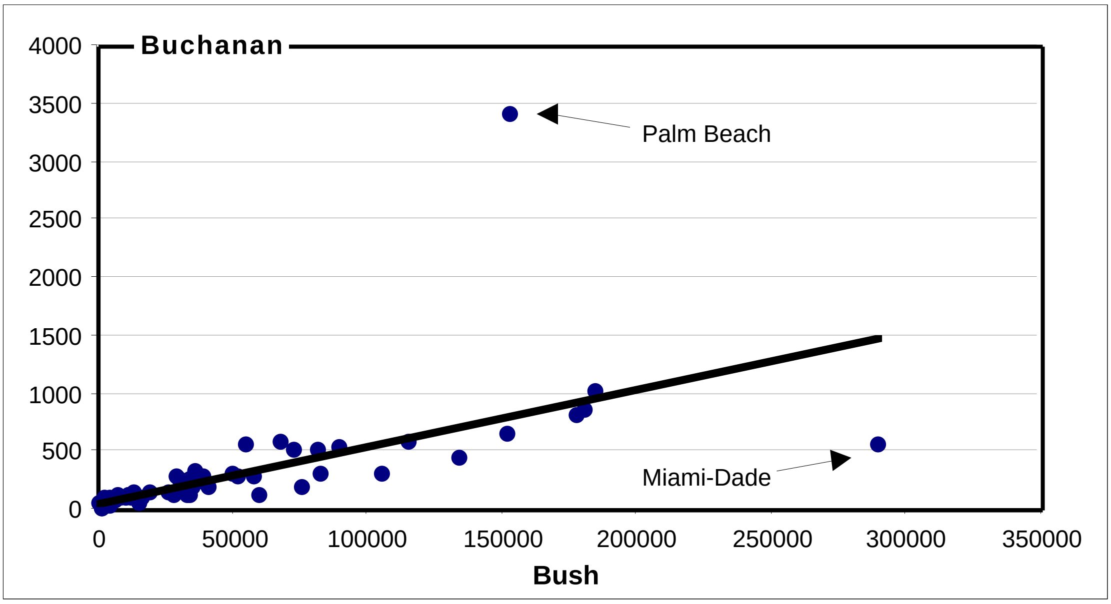

Figure 2 Votes for Pat Buchanan by County in Florida

The above description of the 2000 U.S. Presidential Election is not meant to carry any political overtones. It is meant only to show, in a very public and important "survey", that bad questionnaire design can greatly affect the results of a survey. The Butterfly Ballot was never pilot tested to see whether it would work as intended. The only "pilot testing" was on the day of the election, when Ms LePore learnt that the butterfly ballot was not working properly and frantically tried to phone voting supervisors to have them warn voters of how to use the ballot. The design of the butterfly ballot was not politically inspired, since Ms LePore herself was a registered Democrat. However, she was not a survey designer; she was an experienced electoral official who thought that designing survey forms was an easy matter. She had clearly never read any books on survey design such as Payne (1951), Sudman and Bradburn (1986), Schwarz and Sudman (1996), Sudman, Bradburn and Schwarz (1996) or Tourangeau et al. (2000).

While the above example is in the very public sphere of politics, it is the contention of this paper that similar problems are occurring in the more private sphere of transport planning. Furthermore, it is the contention of this paper that one of the reasons for such problems in survey design in transport is that too much emphasis is placed on modelling by some professionals and not enough emphasis on survey design from the viewpoint of the respondent. Just as Ms LePore was an experienced electoral official who didn't pay enough attention to survey designs which considered the needs and abilities of her voters, so we believe there are transport modellers who don't pay enough attention to survey designs which consider the needs and abilities of their respondents.

In the introduction to a recent book on Stated Choice Methods (Louviere et al., 2000), it is stated that "it is the premise…of this book that measurement in the absence of theory is at best uninterpretable, and at worst meaningless". While this is completely true, one could take the opposite view and assert that "theory in the absence of valid measurement is at best uninterpretable, and at worst meaningless". The premise of this paper is that theory and measurement are both important components of transport planning, and should be treated in a balanced manner.

The Interaction between Data and Model Error

The concept of balance is one proposed many years ago by Alonso (1968). In the context of modelling and data, there are three major sources of error: specification error, sampling error and measurement error. Specification error in modelling occurs because of the inclusion of unnecessary variables or the exclusion of necessary variables from the model. It can also arise from the use or combination of variables in an inappropriate manner (e.g. assuming linear relationships, when non-linear relationships are more correct). While exhaustive testing of the structure of the model may eliminate unnecessary variables from the model, it is never possible to determine the extent to which necessary variables have been omitted (because if we knew about the extent of variables wrongly omitted, we would make efforts to include them).

Sampling error occurs because we are generally dealing with a sample survey (not a census) and this inevitably introduces a degree of sampling error. Taking another sample in the same way from a population will result in a different sample and hence different results. The extent to which results vary across different samples is a measure of sampling error.

Measurement error occurs because of inaccuracies in the measurements made to obtain the data on which the model is built. While sampling error can be computed analytically or empirically, and reduced by taking a larger sample, measurement error often cannot be determined precisely, particularly when humans are the subjects of the data collection and model building. Importantly, measurement error cannot be reduced by taking a larger sample.

Measurement error is a property of the data and cannot be significantly reduced in the modelling process, with the exception of error propagation (i.e. the means by which a model magnifies or diminishes errors in different variables). Specification error is minimised by careful selection and treatment of variables in the model through prior exploratory statistical analysis and visual examination of the data, together with the development of a carefully reasoned hypothesis of structure and variable content.

The trade-off between specification and measurement error provides an interesting perspective on the relative effort that should be spent on data collection and modelling in any transport study. Alonso (1968) has postulated that, although specification error cannot be measured exactly, it is likely that specification error will drop rapidly as the complexity of a model increases as shown in Figure 3 (complexity is defined as being measured by the number and structure of relevant explanatory variables included in the model).

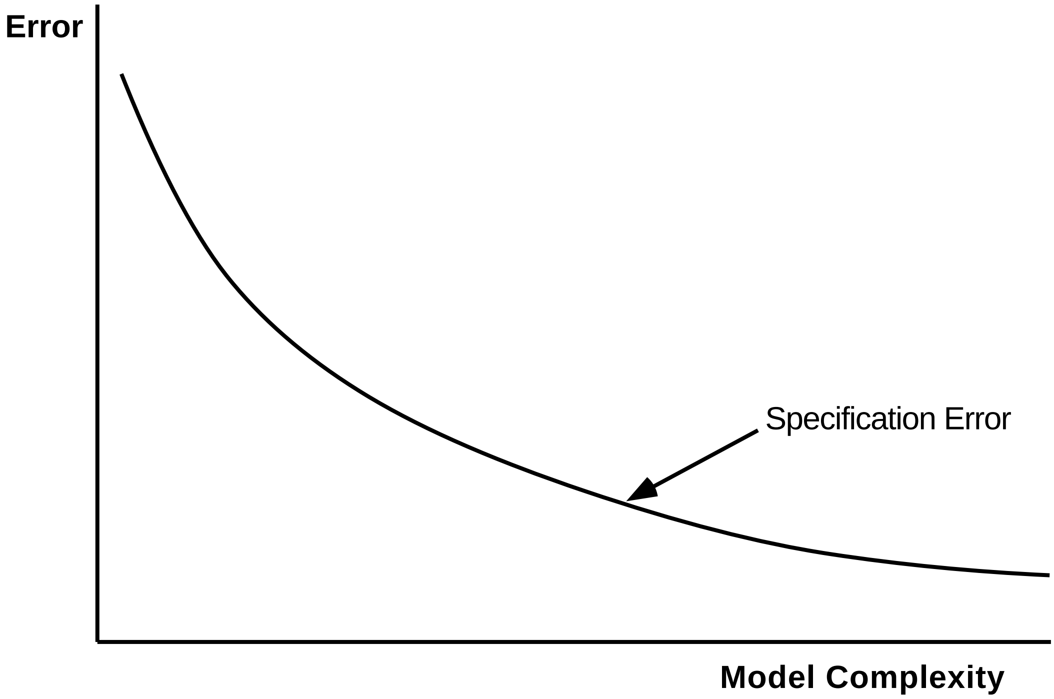

Figure 3 Specification Error and Model Complexity

However, each variable included in the model will have a degree of measurement error associated with it, so that the inclusion of more variables will mean that there is a greater total amount of measurement error in the model, as shown in Figure 4.

Since the total error in the model includes both specification and measurement error (ignoring calibration error for the moment), then the total error can be obtained by:

$$E_{total} = \sqrt{e^2_{spec} + e^2_{meas}} \qquad (1)$$

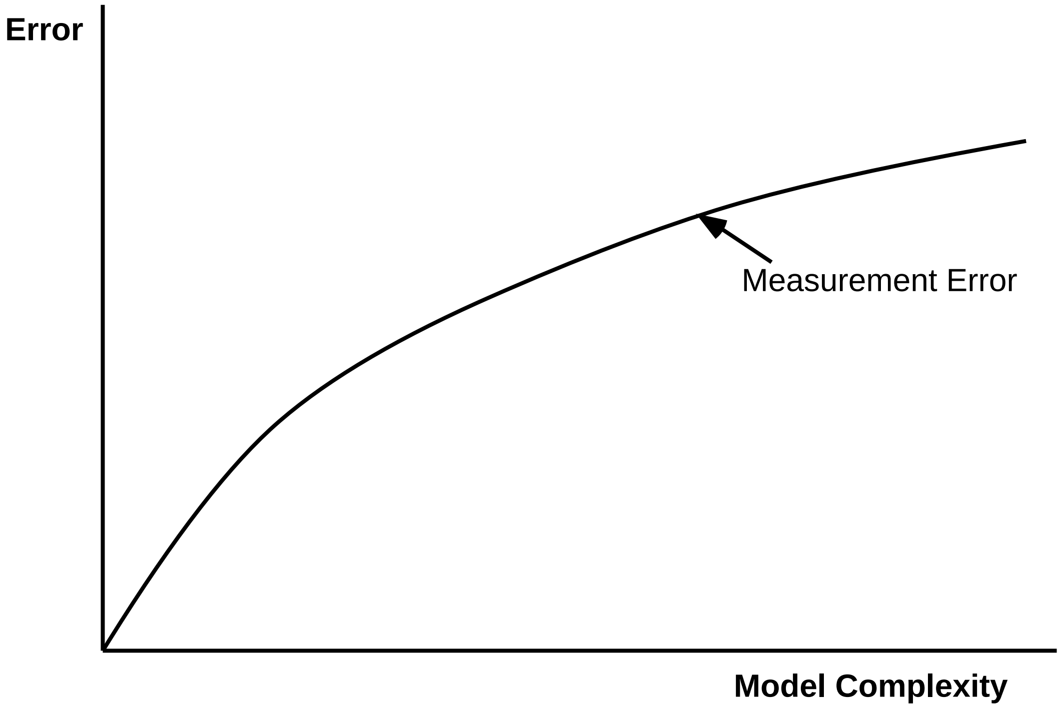

Figure 4 Measurement Error and Model Complexity

This total error is shown in Figure 5, wherein it can be seen that, because of the counter-balancing effects of each source of error, there is in fact an optimum degree of model complexity in order to obtain the minimum total error in the model.

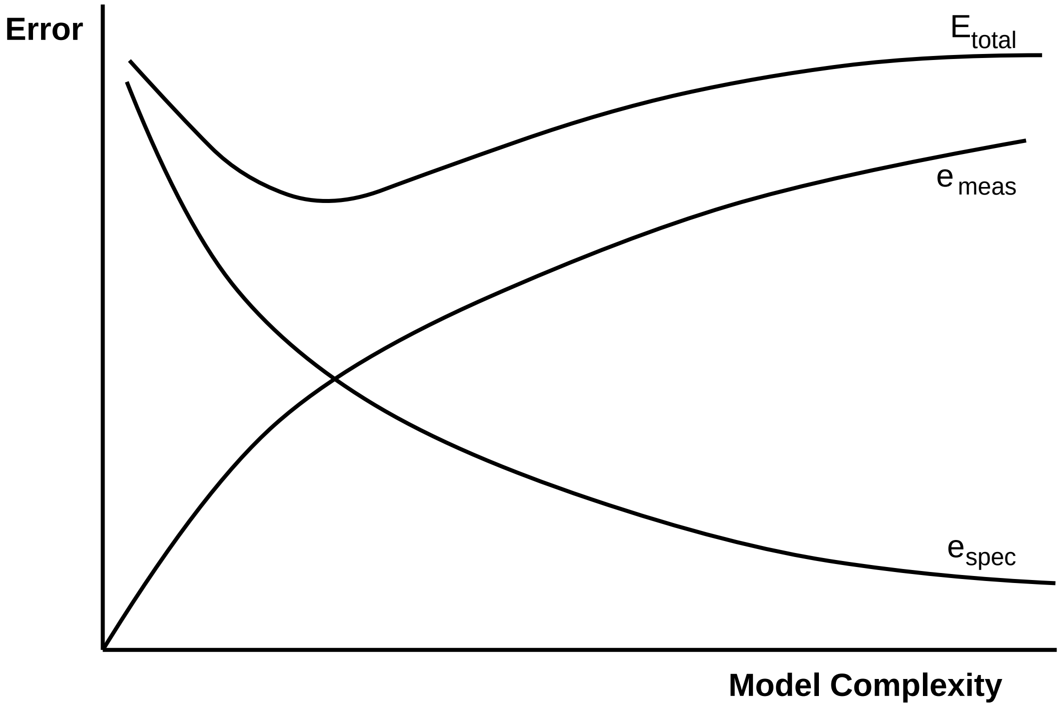

Figure 5 Relationship between Total Error and Model Complexity

The implications of Figure 5 are important enough to bear repeating. The best model is not necessarily the most complex model. This is due to the fact that error arises from both specification and measurement error. While a more complex model will reduce the specification error, it will also increase the measurement error. At some point, the inclusion of more variables into the model will increase the measurement error more than it will reduce the specification error.

This trade-off between specification error and measurement error can be further demonstrated by considering the use of a dataset which has a higher degree of measurement error (perhaps because we spent less time and effort on quality control in the data collection process, or because we posed questions that could not realistically be answered by our respondents). Under these conditions, the measurement error will be higher at all levels of model complexity, as will be the total error, as shown in Figure 6.

The really important feature of Figure 6, however, is that apart from simply being higher, the total error curve is minimised at a lower level of model complexity (i.e. the valley in this total error curve is shifted to the left). The implication of this is that if you have worse data, then you should also use simpler models. This finding runs counter to actual modelling practice in many cases, where some modellers believe that they can overcome the effects of bad data by using more complex models. Alternatively, they may believe that their models are so good that they don’t need to pay much attention to the collection of the data or the needs of the respondents. While it may be difficult to quantify the curves in Figure 6, the overall implication is clear: using more complex models with bad data simply increases the total error in the model.

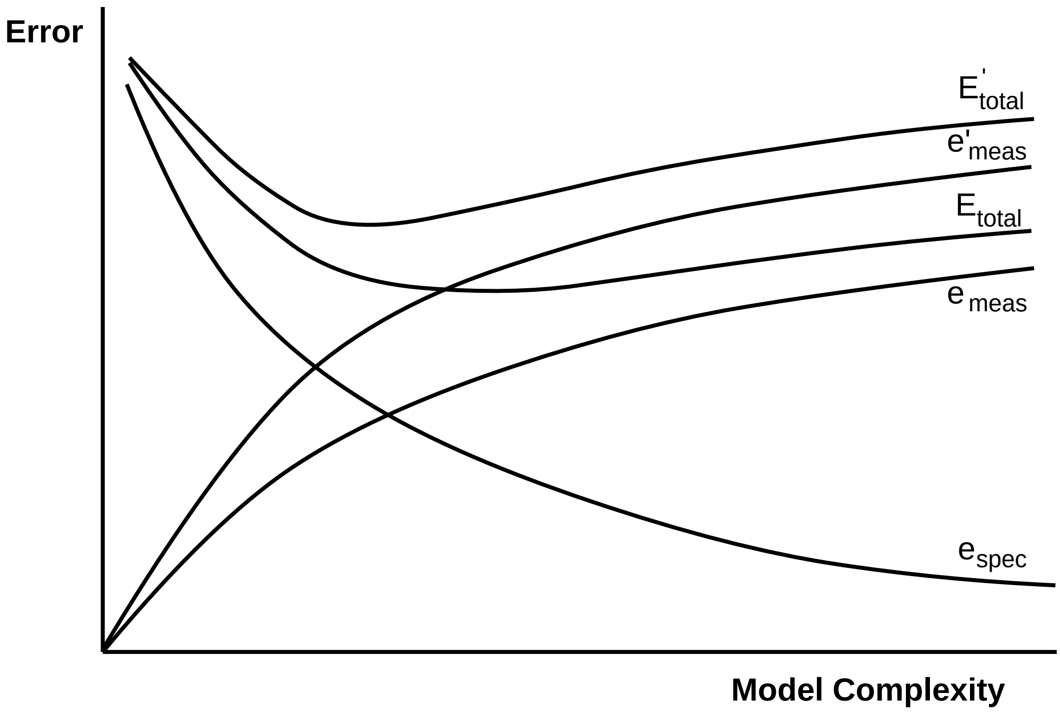

Figure 6 The Effect of Bad Data on Total Model Error

The policy lesson that comes from consideration of the above arguments is that there must be a balance between the time and effort spent on data collection and model building. You cannot compensate for poor quality data by doing better modelling. On the other hand, having collected high quality data, you should use more complex models to obtain the most information out of this dataset if you wish to minimise the total error.

While the ideas proposed by Alonso have been around for a long time, it seems that they have not been fully absorbed by the modelling fraternity. The rapid development of revealed choice modelling in the 1980s, and stated choice modelling in the 1990s, has given rise to many complex forms of choice model which have relaxed some of the restrictive theoretical assumptions of earlier models. However, these new models need high quality data to be fully useful. While parallel advances have been made in survey design, it seems that the two streams of intellectual endeavour have not coalesced.

Some Contrasts in Stated Preference Travel Survey Designs

The purpose of this section of the paper is to describe two studies whose modelling aims were similar, though not identical, but whose approach to survey design is very different. Both studies use Stated Preference modelling techniques, but adopt different approaches to collecting the required data from respondents.

A full-profile, complex question approach

The first example is a project that focuses on the development of Service Quality Indicators (SQI) for urban bus services (Prioni and Hensher, 1999).[1] The purpose of this project was to develop "a revised cost model in which output heterogeneity explicitly takes into account the level of service as perceived by users which impacts on costs. A service quality indicator (SQI) is derived from a revealed preference discrete choice model, enriched by a stated preference experiment, as a representation of the quality level currently on offer. Changes in service levels have a direct influence on costs as quality represents an input which has costs and an indirect influence via SQI and its parameterisation, which influences the demand for bus travel, which itself influences costs in a model in which physical output is defined by final demand. An important feature of our approach is the distinction between those attributes in an SQI which are directly linked to the cost of service provision and enter as additional arguments into the cost function, and those that are indirectly linked to the cost through their influence on final demand. The complete model system includes a cost function, a demand function for passenger travel, and a discrete choice model to identify the choice of bus service quality under alternative potential offerings." (Prioni and Hensher, 1999).

"The data for the empirical model system is sourced from a sample of bus operators and their passengers in New South Wales", and we would argue that the data collection process falls short of the sophistication of the analysis. One specific aspect of the data collection process will be highlighted here; the design of the Stated Choice questions.

It seems clear that the Stated Preference questions used by Prioni and Hensher (1999) are at the difficult end of the spectrum of complexity, from the viewpoint of the respondent. While each respondent is only asked to answer three SP questions, each question contains 3 Options, each described by 13 Attributes. Thus, respondents are asked to deal with 39 pieces of information in order to come to a decision. In addition, as shown in the example below, two of the Options are described in full and the third (that of the current operator) is left to the memory of the respondent.

[1] The choice of this specific project is not meant to convey any particular criticism on the researchers involved. It was chosen because it involved a prominent modeller whose modelling credentials are excellent. All the points on survey design discussed in the current paper have previously been discussed with David Hensher.

BUS SCENARIOS

Bus Operators try to provide a package of services, which appeals to passengers, but recognize that there will be trade-offs in order to be able to justify the cost of the service. Please compare the two bus service packages offered by the Bus Companies A and B with the trip service level of the current Bus Company and choose which of the three service packages you most prefer. Indicate also which of the two new packages is the most preferred. [***SET1***]

SCENARIO 1

SERVICE FEATURE	BUS PACKAGE OF THE BUS COMPANY A	BUS PACKAGE OF THE BUS COMPANY B
Reliability	**10 minutes late**	**on time**
One-way fare	**same as now**	**same as now**
Walking distance to the bus stop	**5 minutes more than now**	**5 minutes more than now**
Personal Safety at the bus stop	**reasonably unsafe**	**reasonably safe**
Travel Time	**25% longer than the current travel time**	**25% quicker than the current travel time**
Bus stop facilities	**No shelter or seats at all**	**Seats only**
Air conditioning	**Available with a surcharge of 20% on existing one-way fare**	**Available with a surcharge of 20% on existing one-way fare**
Information at the bus stop	**Timetable but no map**	**Timetable but no map**
Frequency	**Every 15 minutes**	**Every 30 minutes**
Safety on board	**The ride is jerky; sudden braking occurs often**	**The ride is jerky; sudden braking occurs often**
Cleanliness of seats	**Clean enough**	**Clean enough**
Ease of access to the bus	**Wide entry with no steps inside the bus**	**Wide entry with 2 steps inside the bus**
Driver behaviour	**Friendly enough**	**Very friendly**

32. Which bus service would you use, if the BUS PACKAGES of the Bus Companies A and B were both available today for your journey?

☐ BUS PACKAGE of the Bus Company A ☐ BUS PACKAGE of the Bus Company B ☐ BUS PACKAGE OF YOUR CURRENT BUS COMPANY

33. If <u>ONLY</u> the BUS PACKAGES of the Bus Companies A and B were available today for your journey, which bus service would you use?

☐ BUS PACKAGE of the Bus Company A ☐ BUS PACKAGE of the Bus Company B

It is true that travellers may have to perform complex mental tasks when actually choosing a service in practice. However, when trying to identify key variables that need improvement from the perspective of the operator, it may not be particularly helpful to try to replicate the full complexity of the decision process in a single survey question. This leaves the survey designer in somewhat of a quandary –should they use a less complicated, easier, and to some extent less realistic choice task on paper, or a more complicated, more difficult and more realistic choice task on paper.

The issue of question complexity in SP studies has been the subject of debate for some time. Hensher (2001) notes that "there is controversy over the ability of respondents to comprehend more complex SC [stated choice] experiments". However, he also states that "Louviere and Hensher (2000) have looked into this issue and found no evidence to flatly reject specific design strategies". The discussion of survey complexity in Louviere and Hensher (2000) focuses on three aspects of SP survey design: the number of scenarios (or choice sets or games or profiles) evaluated by respondents; the number of attributes considered; and the number of alternatives from which respondents can choose.

However, most of the studies cited by Louviere and Hensher (2000) refer only to the number of scenarios evaluated by respondents (e.g. Brazell and Louviere, 1995; Johnson and Orme, 1996; Stopher and Hensher, 2000). Indeed, from these studies the general conclusion is that respondents can play many more games (or scenarios, or profiles or choices) than previously considered. For example, many SP studies use a maximum of nine games for each respondent, while the studies cited conclude that respondents can handle 20 to 30 games without problem (providing of course that they can handle each of the games individually without problem). Only one of the cited references (Louviere et al. 1993) appeared to deal with the issue of the number of attributes and alternatives faced by respondents in a single game. In that paper, the authors list many conjoint studies which have used more than 10 attributes. The paper judged that these designs were acceptable on the basis of the statistical significance of the models estimated with the data obtained from these designs. Thus the acceptability of the question design was judged in hindsight from the viewpoint of the modeller. The ability of respondents to understand and answer complex SP questions does not seem to have been systematically assessed using cognitive pretesting techniques, as has been done with many other types of survey question (e.g. Forsyth & Lessler, 1991; Timmermans, 1993; Bickart & Felcher, 1996; DeMaio & Rothgeb, 1996).

Until such cognitive pretesting is comprehensively applied to SP questions, one must judge the "answerability" of complex SP questions on the basis of research performed in other areas of survey design and cognitive processing. On this basis, it is considered that the task required of respondents in answering questions such as that described by Prioni and Hensher (1999) is simply too much to reasonably expect respondents to deal with properly. To quote Orme (1997), "Psychologists have long recognized that people have difficulty processing more than about six pieces of information at the same time (Miller 1956). Full-profile conjoint studies often press the limits of how much information can be successfully evaluated before respondents either quit, glaze over, or start to employ sub-optimal shortcut methods for making choices. The number of attributes is an important factor in task complexity - but not the only one. Complex or wordy attribute levels and respondents' familiarity with the attributes or product category also impact how effectively they can manage full-profile tasks. Indeed, it may be possible to construct a full-profile conjoint study with only six attributes that is too complex for some survey populations".

Having only a limited number of SP questions is, of course, no advantage if respondents cannot even understand the first of the questions. It could be argued that it is acknowledged that respondents cannot handle all this information and that they will take short-cuts in coming to a decision. However, the problem is that it is not known which short-cuts are being taken by which of the respondents. Krosnick (1999) has described in some detail the process of satisficing (Simon, 1990) which respondents might employ when faced with complex questions. Krosnick (1991) identifies three factors that contribute to satisficing, namely the degree of task difficulty, the cognitive ability of the respondent, and the level of motivation for the respondent to answer fully. Payne et al. (1992) also note that when faced with multiple alternatives from which to choose, people are more likely to adopt non-compensatory choice strategies. Foerster (1979) and Golob and Richardson (1981) have identified a wide range of non-compensatory choice rules that might be employed in complex decisions, including maximin, maximax, lexicographic, conjunctive and disjunctive processes.

The problem with posing SP questions that are highly complex is that different people may employ very different decision rules to arrive at an answer. The choice of these rules is generally out of the hands of, and unknowable to, the investigator. A alternative option is to reduce the complexity of the SP questions to a more manageable level for respondents. Additionally, if the questions are of a complex nature, then some steps can be taken in the survey design to account for possible response strategies by respondents. For example, Krosnick and Alwin (1987) have shown that when presented with a long written list of options or responses, respondents tend to give more weight to those options presented earlier in the list (the primacy effect). One way of countering this primacy effect, therefore, is to systematically rotate the responses presented so that each response appears with equal probability at various points in the list. In this way, the primacy effect is cancelled across all the respondents. Unfortunately, rotation of responses was not used in the SQI survey.

The modelling processes employed in the construction of the SQI values, and in the remainder of the SQI study, are of high complexity and sophistication (see the report by Prioni and Hensher, 1999). However, the point of the current paper is that sophisticated analysis will not compensate for data of questionable quality. The tailor must have good materials to make a fine suit.

A less complex survey approach

The other study reported here is the Singapore Value of Time Study, conducted for the Singapore Land Transport Authority by Scott Wilson Pty Ltd and The Urban Transport Institute[2]. In this study, relatively simple methods of analysis are used, but with more attention paid to the needs of the respondents. As will be seen, this combination can produce results that are credible, and not lacking in sophistication.

Like the SQI study, the Singapore VOT study had as its main objective the measurement of relative valuations of importance of various attributes to travellers. Like the SQI study, it did not seek to build a predictive model of choice or usage, but merely to measure the relative value of parameters to users (in this case, the value of travel time differences relative to cost differences). For users of public modes of transport, it was desired to obtain values of time for in-vehicle time, waiting time, walking time and transfers between vehicles. For users of private modes, it was desired to obtain a value of time for door-to-door travel time. It was also desired to obtain perceived values of various costs involved for private modes, such as fuel, parking, Electronic Road Pricing (ERP) and the Certificate of Entitlement (CoE) which must be paid in order to initially obtain a vehicle in Singapore.

[2] The Singapore Value of Time Study is chosen only as an example of how SP surveys might be designed from the viewpoint of respondent-friendliness, without sacrificing the ability to conduct detailed analyses. Many other well-designed SP surveys could have been used as examples. For example, Louviere et al. (1993) describe Hierarchical Information Integration as a means of dealing with many attributes in a more respondent-friendly manner , and this method has been used by many others (e.g. Perrey, 1998). However, the author is more familiar with the details of the Singapore Value of Time Study.

Given the nature of the study, a stated preference survey was the logical candidate. However, a prime concern was to design a Stated Preference survey which would be respondent-friendly, and which would not overtax the cognitive capabilities of the respondents. Given the manner in which the results were to be subsequently analysed, involving the estimation of Values of Time for many different market segments, it was also desirable if the results could be obtain at a highly disaggregated level, i.e. at the level of the individual. It was felt that conventional SP surveys did not adequately fulfil all of these requirements (although clearly there is no question that many conventional SP surveys do fulfil some of these requirements). As a result, the chosen method was a variation on a technique used widely in market research called Adaptive Conjoint Analysis (ACA). To distinguish the specific method used in the Singapore Value of Time Study, we refer to it as Adaptive Stated Preference (ASP).

ASP surveys differ from conventional SP surveys in four major ways:

- The options presented to the respondent in an ASP game depend on the responses given by the respondent to previous games;
- The individual ASP games generally have fewer options and fewer attributes presented to the respondent in one question than in conventional SP games;
- The respondent is often presented with more games in an ASP survey than in a conventional SP survey;
- At the end of an ASP survey, it is possible to obtain estimates of the parameters of interest (such as Value of Time) for that individual. Conventional SP surveys require that data from many respondents be aggregated and the parameters estimated as the result of a model-fitting process.

ASP games start with an initial estimate of the parameters of interest and then present the respondent with ASP games that progressively refine the estimates of the parameters of interest. The ASP survey is designed to deliberately present the respondent with games that force them to trade-off increases in one attribute against reductions in another attribute, and to do so with closer decisions as the survey progresses (i..e the respondent gets closer and closer to their point of indifference between the two options).

In the ASP survey used in the Singapore study, each respondent answered 18 ASP questions (as well as questions on the nature of the household, the person and the trip). However, each of the ASP questions was fairly simple and easily understood, being only a trade-off between two attributes across two options. A typical screen from the computerised interview is shown in Figure 7, for a trade-off of total travel time against fare for a public transport trip.

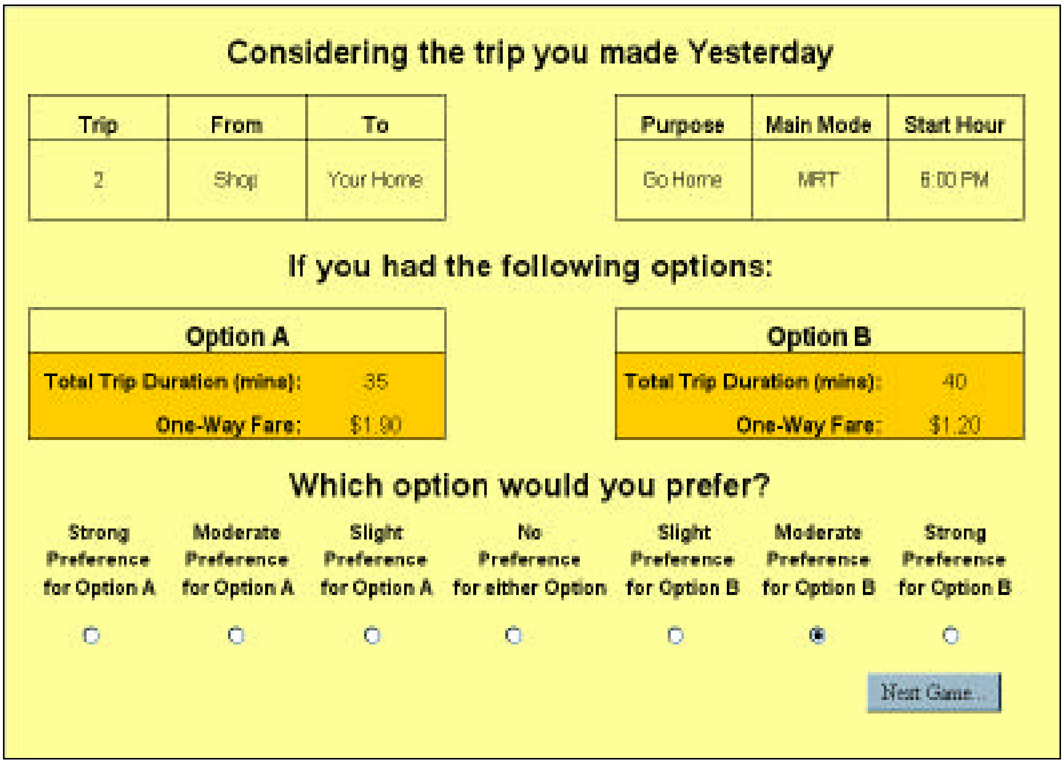

Figure 7 A Typical Question from the Adaptive Stated Preference Survey

The ideas of ASP and simplified questions are not entirely new to transport. Ampt et al. (2000)[3] note that "In the early days of conjoint analysis, models were produced for each individual, and the practice is still common in market research studies. In transport, however, this idea was largely swept aside by the rise of random utility theory and the use of logit models in the 1980s. Variation in preference weights is handled by segmenting respondents into groups who are expected to behave similarly. This still loses distributional information about choice behaviour, which can sometimes be important (such as in toll studies)". Interestingly, a rapidly expanding new area in choice modelling is the development of Mixed Logit Models in an attempt to recapture the information on the heterogeneity in tastes across the population of travellers (e.g. Bhat, 1998; Revelt and Train, 1998; Brownstone, Bunch and Train, 2000; Brownstone, 2000; Hensher, 2001a, 2001b).

Polak and Jones (1997, pg 190) describe adaptive SP surveys as "very appealing intuitively" and state that they "have become popular among practitioners, not least because they tend to reduce the amount of questioning that needs to be carried out and because respondents typically report adaptive exercises to be easier and more interesting to complete". Wang, Oppewal and Timmermans (2000) also report on the use of a simplified SP design which, while not being adaptive as in the Singapore survey, varies only two attributes at a time in an

[3] There are several references to the book *Stated Preference Modelling Techniques* published by PTRC in 2000. This book is a compilation of papers on SP presented at several PTRC Conferences during the 1990s. The year of publication of all these papers will be given as 2000, even though they were presented in several different years.

attempt to avoid the "information overload problem typical of the conventional fractional factorial design strategy".

While recognising the benefits of ASP to the respondents, Polak and Jones (1997) also note that there are some potential disadvantages. First, they note that ASP will result in non-orthogonal data, which reduces a traditional advantage of SP data (although Watson et al., (2000) note that non-orthogonality can be a positive advantage for model estimation). Second, they note that poorly designed ASP surveys can "run out of control", especially in the case of respondents with extreme preferences. In the Singapore ASP survey, this was accounted for by bounding the Values of Time estimated in each game so that the next game would present reasonable attribute values to the respondent (this restriction was removed in the later analysis of the data obtained from the ASP survey). Third, they note the reservations expressed by Bradley and Daly (2000) concerning correlations being introduced between the levels of the design variables and the unmeasured components of utility that are not accounted for in model estimation. This presumably is also the basis for Wardman's (2001, pg 120) description of ASP designs as being "the now generally discredited adaptive approach", where they refer to Bradley and Daly's paper.

This apparently fatal flaw of adaptive designs, however, is very much dependent on the method of analysis used with the data. Bradley and Daly (2000) analysed the data using relatively conventional choice modelling methods, estimated with the ALOGIT package, which assume that the attributes presented in the various games are orthogonal and that the decisions made in successive games are independent of each other. The non-orthogonality and serial correlations introduced by adaptive designs do indeed introduce bias and variability when performing such model estimations. This problem was avoided in the Singapore study by not estimating models for groups, under assumptions of orthogonality and independence, but rather estimating individual specific parameters. Indeed, Bradley and Daly (2000) had recognised this as an alternative approach, but appear to have discarded the method on the basis of some earlier work by Morikawa (1989) in which he suggested that the bias and variability in individual specific estimates was much higher than that obtained from a conventional modelling process with group data. However, this conclusion by Morikawa appears to have been based on an incorrect procedure whereby the two methods were not compared on a like-for-like basis. Such an observation is supported by Kroes and Cirillo (2000), who conclude that grouped models are not necessarily superior to individual specific models.

The Singapore ASP survey therefore estimated individual specific values of time (and other parameters). In doing so, it took heed of the warning given by Fowkes (2000) and Fowkes and Wardman (1998) that there is a problem with estimating Values of Time in the presence of taste variation, since conventional model estimation estimates a ratio of averages (the average time coefficient divided by the average cost coefficient) when what is really needed is the average of the ratios of time coefficient and cost coefficient for each individual. The bias and variability characteristics of individual specific parameters using ASP were tested extensively before the conduct of the survey using simulation techniques. The results of these tests are reported elsewhere (Richardson, 2001).

The Singapore survey was in the form of home interviews with approximately 3000 households selected at random throughout Singapore. The survey provides data on the socio-economic characteristics of each household and its constituent persons. One person was selected at random (but according to quotas of respondent characteristics) from each household to provide further details about trips made on a previous day. One of these trips was then selected at random (but according to quotas of trip characteristics) to form the basis of a number of ASP ‘games’, during which the respondent was asked to state his/her preference between options that involve combinations of travel time and cost. The games were adaptive, in that the options presented for each game depended upon the preferences expressed in the previous games. The survey was conducted with the aid of laptop computers, by a team of trained interviewers. The results of the ASP games were used to calculate the behavioural values of components of travel time and cost for the person concerned.

The survey field work was undertaken by MediaCorp Research Pte Ltd (MRPL). A team of approximately 40 interviewers was recruited largely from MRPL’s own register of casual staff, many of whom had been regularly employed on survey work over a number of years. The interview staff were briefed in small groups at MRPL’s offices, where they were provided with a copy of the Interviewers’ Briefing Notes. The verbal briefing was followed by practice interviews. A Pilot survey of approximately 100 interviews was carried out over a period of approximately ten days prior to the commencement of the main survey. During the pilot survey, a number of problems were encountered with the ASP program written for the laptop computers. In addition, there were the usual problems with sampling and obtaining responses that would be encountered in any such large-scale survey. Importantly, however, no problems were reported with the respondents’ understanding of the ASP questions.

Each respondent answered 18 ASP questions. After the respondent answered each question, the survey program updated the estimate of this respondent’s value of time (using a simple binary logit formulation) and generated the attribute levels for the next question in an attempt to refine the estimate. It was found that this process converged fairly quickly, with three questions needed for each parameter to be estimated. In the public transport surveys, there were six parameters to be estimated (value of in-vehicle travel time, value of waiting time, value of walking time, value of inter-vehicle transfers, and value of door-to-door travel time with travel time changes of 5 and 10 minutes). The 18 ASP questions required took about 7-10 minutes to complete.

Even though the ASP questions are easy for the respondent to complete, they generate highly detailed and disaggregate data on Values of Time (at the level of the individual respondent). These Values of Time are the same as any variable in a travel survey data set, and can be expanded to population totals using Census data and other external data sets. They can then be analysed using readily available computer programs, such as Excel or any statistical software.

Every one of the 3103 respondents in the final data set answered several questions involving a trade-off between trip cost and overall door-to-door travel time. In each of these trade-offs, the door-to-door travel time varied by five minutes either side of the current door-to-door travel time for each respondent. From this data, an overall Value of Time has been calculated for each respondent, from which an average VOT can be calculated.

An advantage of the ASP survey approach, however, is that an individual VOT is obtained for every respondent. Therefore, in addition to the average VOT noted above (which would be the normal output of a typical SP study), the ASP survey can produce a distribution of VOTs, as shown in Figure 8 (this figure is a facsimile of the actual results, pending release of the results by the client). For the total sample of respondents, a proportion have zero (or negative) VOTs, while a proportion have VOTs much greater than the average. The existence of a zero (or negative) VOT is at first somewhat surprising, mainly because an average VOT of zero (or negative) has not been seen in most previous SP studies (which do not show the distribution of VOT). However, on reflection, it simply means that a proportion of the population would not be prepared to spend money for a shorter trip time. Either they have plenty of time on their hands, and hence don't need to save time, or they have not enough money to pay for time savings (or both). Alternatively, they may actually prefer a longer trip for the various reasons described by Redmond and Mokhtarian (2001). The distributional results obtained from the ASP survey are in contrast to Mixed Logit Models that estimate the parameters of an assumed distribution of individual parameters, but do not actually estimate the distribution empirically.

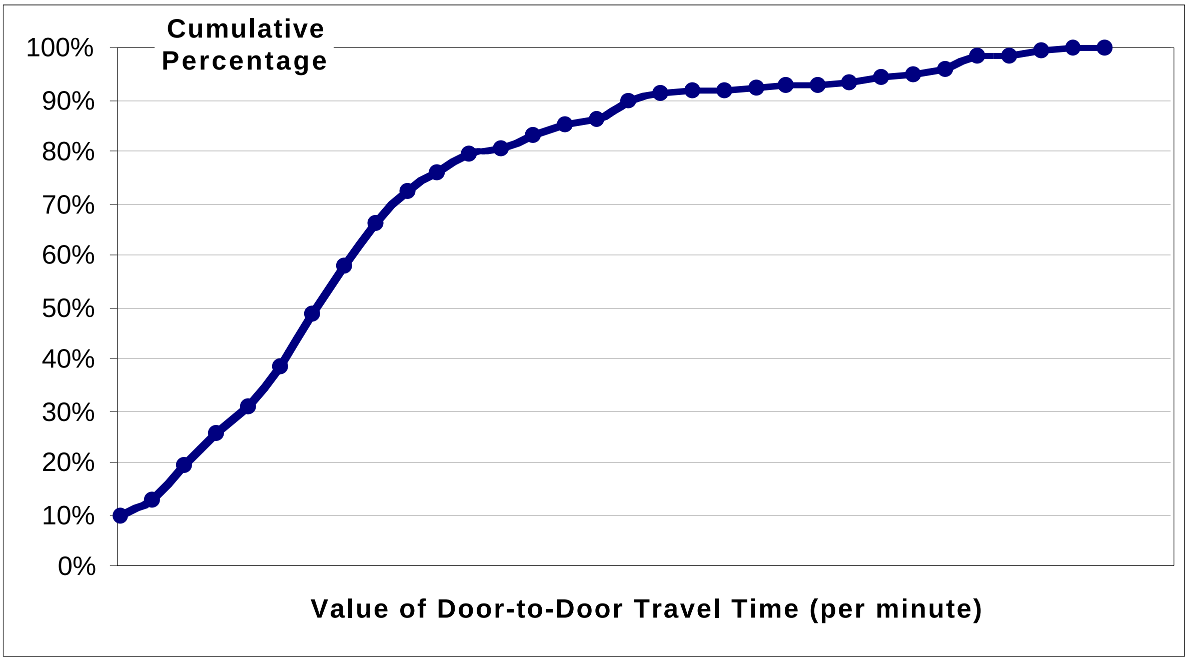

Figure 8 Distribution of the Value of Door-to-Door Travel Time

Because of the disaggregate nature of the VOT results, average VOTs can be calculated for any segment of the data set, defined by any of the variables in the data set. For example, the relationship between Personal Income and VOT is highly significant, as shown in Figure 9. Those with the highest income have the highest VOT, and vice versa. However, the VOT is not directly proportional to personal income. If one divides VOT (per hour) by hourly income, one obtains the relationship between VOT (as a proportion of personal income) and Personal Income as shown in Figure 10. It can be seen that the ratio of VOT to personal income is not constant for all income levels. It falls from a high ratio of VOT to personal income for the lowest income group down to the lowest ratio for the highest income group. The high ratio for the lowest income group does not just mean that those in the lowest income group would be willing to spend more of their income in order to save travel time. It also means that they would be prepared to accept longer travel times if they were paid an amount that was a substantial proportion of their hourly income rate. While those on lower incomes would be willing to pay or accept less in absolute terms for travel time changes, these amounts are a higher proportion of their personal income.

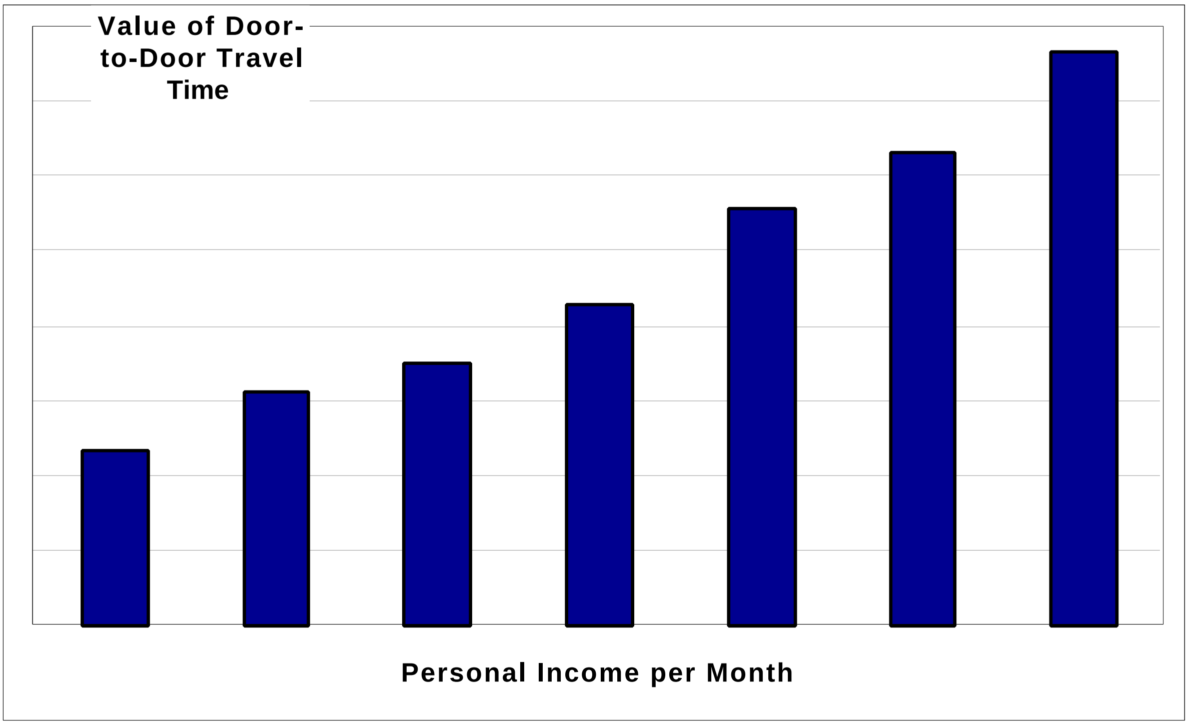

Figure 9 Average Value of Time by Personal Income

NB: the axis labels has been removed to preserve the confidentiality of the results until released by the client.

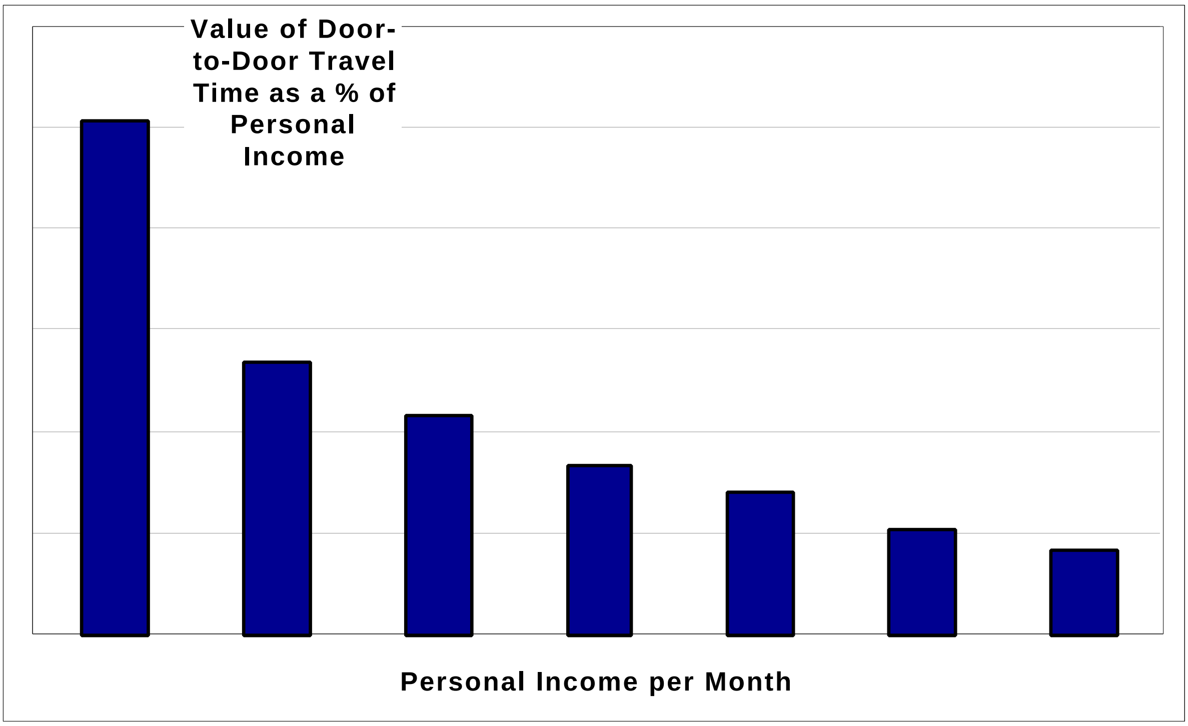

Figure 10 Average Value of Time as a Proportion of Personal Income

NB: the axis labels has been removed to preserve the confidentiality of the results until released by the client.

The richness of the results described above are typical of those obtained from the study, and demonstrate that respondent-friendly questions and relatively simple methods of analysis do not eliminate the possibility of highly detailed and relevant results. In addition, the simple nature of the questions and the analysis mean that the method and the results are more easily explainable to high-level policy makers and politicians.

Conclusions

This paper has sought to demonstrate that good survey design is just as important as sophisticated modelling, and that poor survey design can lead to models that lack credibility. As Alonso (1968) showed many years ago, there is a need for balance in the complexity of modelling and the quality of data. Highly sophisticated modelling cannot compensate for poor quality data.

The paper has highlighted that good survey design is a professional skill in its own right. People who are highly qualified in one area (such as experienced electoral officials) are not necessarily good at designing surveys, because the skills required in survey design are not the same as the skills required in other fields of endeavour.

The paper has compared two Stated Preference surveys and has shown that the production of sophisticated results does not necessarily require complex questions and complex analysis methods. Respondent-friendly surveys and relatively simple statistical analysis can produce results that are just as sophisticated and policy relevant. A major conclusion to be drawn from this paper is the absolute need to pilot test SP questions that could possibly be seen as being difficult for respondents to comprehend and answer. Various forms of pilot testing, including cognitive pretesting, would be useful in this context. As stated in Louviere et al. (2000, pg 259), *"No significant survey should ever go to field without pilot testing!"* (original emphasis and italics).

Survey design and choice modelling have both advanced significantly in the past 25 years. The challenge for the future will be to keep a balance between modelling sophistication and data quality, so that the best quality results are produced from the combined use of quality data and appropriate modelling.

Acknowledgments

The authors are grateful to numerous colleagues for their suggestions on this paper. In particular, we wish to thank an anonymous referee for forthright comments on an earlier draft. The Singapore Land Transport Authority is thanked for the opportunity to describe the Adaptive Stated Preference survey used in the Singapore Value of Time study.

References

Alonso W (1968). "The Quality of Data and Choice and Design of Predictive Model". *Special Report 97*, Highway Research Board, pp. 178-192.

Ampt, E.S., Swanson, J. and Pearmain, D. (2000). "Stated Preference: Too Much Deference?", In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 191-203.

Bhat, C. (1998). "Accomodating variations in responsiveness to level-of-service measures in travel mode choice modeling". *Transportation Research A*, 32, pp 495-507.

Bickart, B. and Felcher, E.M. (1996). "Expanding and enhancing the use of verbal protocols in survey research". In N. Schwarz and S. Sudman (Eds), *Answering Questions*, Jossey-Bass: San Francisco, pp. 115-142.

Bradley, M. and Daly, A.J. (2000). "New Analysis Issues in Stated Preference Research", In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 121-136.

Brazell, J. and Louviere, J. (1995) "Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden". Paper presented to *INFORMS Marketing Science Conference*, Sydney, Australia 1995.

Brownstone, D. (2000). "Discrete Choice Modelling for Transportation". *Proceedings, 9th IATBR Conference*, Gold Coast, Australia, pp. 97-124.

Brownstone, D., Bunch, D.S. and Train, K. (2000). "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles". *Transportation Research B*, 34, pp 315-338.

DeMaio, T.J. and Rothgeb, J.M. (1996). “Cognitive interviewing techniques: in the lab and in the field”. In N. Schwarz and S. Sudman (Eds), *Answering Questions*, Jossey-Bass: San Francisco, pp. 177-196.

Foerster, J.F. (1979). “Mode Choice Decision Process Models: A Comparison of Compensatory and Non-Compensatory Models”, *Transportation Research A*, 13, pp. 17-28.

Forsyth, B.H.and Lessler, J.T. (1991). “Cognitive laboratory methods: a taxonomy”. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz and S. Sudman (Eds), *Measurement Error in Surveys*, Wiley: New York, pp. 393-418.

Fowkes, T. (2000). “Recent Developments in Stated Preference Techniques in Transport Research”, In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 37-52.

Fowkes, T. and Wardman, M. (1998). “The Design of Stated Preference Choice Experiments with Particular Regard to Inter-personal Taste Variations”, *Journal of Transport Economics and Policy*, 22, pp 27-44.

Golob, T.F. and Richardson, A.J. (1981). "Noncompensatory and discontinuous constructs in travel behaviour models", In *New Horizons in Travel Behavior Research*, P.R. Stopher, A.H. Meyburg and W. Brög (Eds.), Lexington Books, Chap. 19, pp. 369-384.

Hensher, D.A. (2001a). “The Valuation of Non-Commuting Travel Time Savings for Urban Car Drivers”. *24th Australasian Transport Research Forum*, Hobart.

Hensher, D.A. (2001b). “The Sensitivity of the Valuation of Travel Time Savings to the Specification of Unobserved Effects”. *Transportation Research E*, 37, pp 129-142.

Johnson, R.M. and Orme, B.K. (1996) How many questions should you ask in choice-based conjoint studies? Paper presented to the *American Marketing Association’s Advanced Research Techniques Forum*, Beaver Creek, Colorado (June).

Kroes, E. and Cirillo, C. (2000). “Individual versus Grouped SP Models – Some Experiments”. *European Transport Conference*, Cambridge, UK.

Krosnick, J.A. (1991). “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys”. *Applied Cognitive Psychology*, 5, pp 213-236.

Krosnick, J.A. (1999). “Survey Methodology”. *Annual Review of Psychology*, 50, pp 537-567

Krosnick, J.A. and Alwin, D.F. (1987). “An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement”. *Public Opinion Quarterly*, 51, pp 201-219

Redmond, L.S and Mokhtarian, P.L. (2001). “The positive utility of the commute: modeling ideal commute time and relative desired commute amount”. *Transportation*, 28(2), pp. 179-205.

Louviere J.J., Hensher D.A. and Swait J.D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge University Press: Cambridge, UK.

Louviere, J., Oppewal, H., Timmermans, H. and Thomas, T. (1993). “Handling large numbers of attributes in conjoint applications: who says existing techniques can't be applied? But if you want an alternative, how about hierarchical choice experiments?”. Paper presented to the *American Marketing Association’s Advanced Research Techniques Forum*, Beaver Creek, Colorado (June).

Louviere, J.J. and Hensher, D.A. (2000). “Combining Sources of Preference Data”. *Proceedings, 9th IATBR Conference*, Gold Coast, Australia, pp. 125-144.

Miller, G. A., “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *The Psychological Review*, (March 1956).

Morikawa, T. (1989). *Incorporating Stated Preference Data in Travel Demand Analysis (Chapter 8: Simulation Study of Models with Taste Heterogeneity)*. PhD Dissertation, Department of Civil Engineering, MIT

Orme, B. (1997). “ACA, CBC, or Both? Effective Strategies for Conjoint Research”, Sawtooth Software Working Paper (www.sawtoothsoftware.com/techpap.shtml).

Payne S.L. (1951). *The Art of Asking Questions*. Princeton University Press: Princeton, NJ.

Payne, J., Bettman, J. and Johnson, E. (1992). “Behavioural decision research: a constructuive processing perspective”. *Annual Review of Psychology*, 43, pp 87-131.

Perrey, J. (1998). *Nutzenorientierte Marketsegmentierung: ein integrativer Ansatz zum Zielgruppenmarketing im Verkehrsdienstleistungbereich*. Gabler: Wiesbaden (in German) [Utility-Oriented Market Segmentation: an integrative approach to target group marketing in the public transport service sector].

Polak, J. and Jones, P. (1997). “Using Stated-Preference Methods to Examine Traveller Preferences and Responses”, In P.Stopher and M. Lee-Gosselin (Eds.) *Understanding Travel Behaviour in an Era of Change*, Elsevier Science: UK., pp 177-207.

Prioni P. and Hensher D.A. (1999). “Measuring Service Quality and Evaluating its Influence on the Cost of Service Provision”. Working Paper ITS-WP-99-11, Institute of Transport Studies, University of Sydney.

Revelt, D. and Train, K. (1998). “Incentives for appliance efficiency in a competitive energy environment: Random-parameters logit models of households' choices”. *Review of Economics and Statistics*, 80, pp 647-657.

Richardson, A.J. (2001). “Testing the Estimation of Individual Specific Values of Time from an Adaptive Stated Preference Survey using a Simulation Approach”. *TUTI Working Paper TR9-2001*, The Urban Transport Institute, Australia

Schwarz N. and Sudman S. (1996). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. Jossey-Bass: San Francisco.

Simon, H.A. (1990). “Invariants of Human Behaviour”, *Annual Review of Psychology*, 44, pp. 1-19.

Stopher, P.R. and Hensher, D.A. (2000). “Are more profiles better than less? Searching for parsimony and relevance in stated choice experiments”. Paper presented at the *79th Annual Meeting of the Transportation Research Board*, Washington DC.

Sudman S. and Bradburn N.M. (1986). *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass: San Francisco.

Sudman S., Bradburn N.M. and Schwarz N. (1996). *Thinking about Answers: the Application of Cognitive Processes to Survey Methodology*. Jossey-Bass: San Francisco.

Timmermans, D. (1993). “The impact of task complexity on information use in multi-attribute decision making”. *Journal of Behavioural Decision Making*, 6, pp. 95-111.

Tourangeau R., Rips L.J. and Rasinski K. (2000). *The Psychology of Survey Response*. Cambridge University Press: Cambridge, UK.

Wang, D., Oppewal, H. and Timmermans, H. (2000). “Pairwise Conjoint Analysis of Activity Engagement Choice”. *Environment and Planning A*, 32, pp 805-816.

Wardman, M. (2001). “A Review of British Evidence on Time and Service Quality Valuations”, *Transportation Research E*, 37, pp 107-128.

Watson, S.M., Toner, J.P., Fowkes, T. and Wardman, M. (2000). “Efficiency Properties of Orthogonal Stated Preference Designs”, In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 91-101.