ITS

WORKING PAPER
ITS-WP-99-12

Are More Profiles Better than Less? Searching for Parsimony and Relevance in Stated Choice Experiments

By

Peter Stopher
& David A. Hensher

June,1999

ISSN 1440-3501

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

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University

Introduction

Stated Choice experiments are recognized as a rich framework within which to study the preferences of individuals *[1], [2]*. Whether used as an enrichment devise for revealed preference (RP) models *[3], [4], [5], [6]* or as a stand alone empirical approach, there remains a degree of skepticism (often unwritten) about the ability of respondents to comprehend and respond to choice designs that involve many alternatives, many attributes and many treatments. Typically, a design with more than two alternatives, three attributes per alternative and four treatments is often perceived as being "too complex" for a respondent. Analysts frequently ponder on the implications of simplified SC experiments in contrast to statistically more rigorous designs in respect of the goodness-of-fit and behavioral responsiveness (e.g., elasticities and values of travel time savings). However the opportunity to investigate the potential loss of benefit has not been taken up.

The question that we focus on is: Are there any statistical and behaviorally substantive differences between the results from a stated choice model as we vary the number of treatments that each sampled individual evaluates? Holding the set of attributes and choice set constant, we investigate the implications of 4, 8, 12, 24 and 32 treatment blocks under two blocking strategies – accumulation and non-accumulation.

The paper is organized as follows. We begin with some general principles on experimental design that provide guidelines on the design of choice experiments. This is followed by the design of a choice experiment focused on the choice between airlines and class of travel for international services between Australia and New Zealand (including the null alternative – no travel). The survey instrument, data collection method and respondent burden considerations are presented prior to an overview of the data. A series of discrete choice models and their associated fare elasticities are then presented and interpreted as the basis of establishing the extent to which fewer treatments capture the behavioral essence of an increasing number of treatments.

Designing Stated Choice Experiments

Stated choice data are a derivative of an experimental design in which all possible combinations of attribute levels (between and within alternatives) are systematically reduced to a fraction that preserves much of the rich statistical properties of the full design. As the fraction becomes smaller for a given attribute set, the amount of information captured by the variability offered within the design decreases, with the consequence that the ability to investigate the statistical importance of each attribute and possible interactions with other design attributes and covariates decreases. As we approach the smallest feasible fraction, we may be limiting the model estimation to a main effects specification that is linear additive in the design attributes. Adding additional complexity to the estimation function (e.g., quadratic main effect and two-way interactions) may induce statistically significant pair-wise correlation between the effects.

Transportation choice modelers tend to select the smallest fraction for reasons that are in part linked to the limitations imposed on them by the laptop software used to implement an

SC experiment and/or a belief that respondents can only handle a limited number of attributes and alternatives. Although there is no specific "formula" used in selecting the dimensions of the design, it appears to be popular practice in transportation to limit the number of treatments to less than nine and typically between two and four. This is in contrast to the increasing number of applications in other disciplines such as marketing where applications routinely require each respondent to evaluate 16 to 32 SC treatments *[2]*.

To investigate the extent to which there are any substantive gains from more complex SC designs which impose increasing demands on the respondent, we developed a design that might reasonably be described as significantly more complex than what we observe in transportation studies. We initially selected a context in which we have 11 alternatives (two airlines – Qantas and Ansett – and a null alternative), each defined by varying numbers of attributes. Eight alternatives have four attributes (a total of 32) and two alternatives have 9 attributes (a total of 18). The null alternative has no design attributes.

The experiment can be conceptualized as a $(4 \times 2^6) \times (4 \times 2^8) \times (4 \times 2^6) = 4^3 \times 2^{20}$, giving 8 + 20 = 29 degrees of freedom to be estimated for each competitor (Qantas and Ansett), a total of 58 degrees of freedom. This can be reduced by collapsing one or more four-level attributes to two levels (there are other possibilities, as well). For example, if we reduce full economy fare to two levels ($500 & $950), cancellation penalty to two (0% & 100%) and the premium for business class to two (20% & 40%), the design is a 2^{23} for each competitor (46 degrees of freedom). Further reductions can come from only varying extra leg room and/or a hot meal in full economy. If we impose a 100% cancellation penalty, and provide no other amenities, this reduces the design to a 2^{13} for each competitor, or a 2^{26} total, for 26 degrees of freedom. We have selected the latter design in the current application. The task is to complete all 32 scenarios in the survey, and provide a small amount of background data. The latter is used to accommodate individual differences in choice models. This is a complex design of airline choice on a city-pair routing, allowing for a wide variety of nonlinear effects.

Two blocking strategies are specified as ways of investigating the behavioral implications of the number of treatments shown to a respondent. We limited the exposure to a maximum set of 32 treatments and partitioned in into sets of 4, 8, 12, 24 and 32 treatments. In all cases, the same treatments were preserved in sequence 1 to 32; thus treatments 1 to 4 were common to all five blocks, treatments 5 to 8 were common to the last four blocks etc. This structure was selected on purpose in order to establish some exact commonality of the task up to a specific number of treatments. Two empirical blocking strategies are considered - an accumulation blocking strategy (ABS) and a non-accumulation blocking strategy (NABS). The former adds each block of treatments sequentially to previous blocks; for example we estimate a model of the 4-block treatment set, then we add in the 8-block treatment. The non-accumulating strategy (of potentially greater interest) involves independent assessment of each of the five blocks.

The Empirical Context

The empirical setting for the implementation of the SC experiment is international air travel between Australia and New Zealand. A convenience sample of 500 individuals on the Institute of Transport Studies mailing list resident in Australia and New Zealand were sent the survey form. The five versions were divided equally amongst the sample. Some of the attributes do not vary but are included in the SC card as a contextual effect.

The response rate was impressive for a mail-out, mail-back survey, with an overall response rate of 33 percent. There was also relatively little difference in response rate, except for the shortest survey form (4 treatments) which had the highest response rate. Table 1 shows the numbers of usable responses for each of the subsets of the sample, which are also percentage response rates, because 100 surveys were sent out for each of the subsets of the sample

A few statistics are useful to examine to see the nature of the sample. The primary distinguishing factor between each alternative in a treatment is the fare. Figure 1 shows the distribution of fares across the sample. There is a spike for not applicable, which is the fare result for the choice of not traveling at all. There is also a spike at each of $500 and $950, because these were used as the full fares in all samples. Otherwise, the sample shows no strong biases in the fares associated with the different treatments. Figure 2 shows the same graph, but with separate lines for each of the subsets of the sample. This shows that there are only minor differences between the subsets, and that each subset follows an almost identical pattern of distribution of fares across the alternatives.

Table 1: Distribution of Responses by Numbers of Treatments

Number of Treatments	4	6	16	24	32
Useable Responses	44	30	31	29	32

The distribution of the choices across the treatments is shown in Table 2, which shows both the number of times the alternative was chosen and the percentage for the treatment subset. Two important conclusions can be drawn from this table. First, while the percentages vary somewhat across the treatment subsets for each alternative, there is remarkable similarity in the percentages across the treatment subsets. The Qantas 29-day advance purchase fare seems to be the most popular within all treatment subsets, followed by the Ansett/Air New Zealand 29-day advance purchase. The least popular fares appear to be the 8-day advance purchase fares, which provide only a very small discount against the full fare. Because full fare receives some other conveniences that are not available to discount economy passengers, the full fare is more popular than the 8-day advance purchase. Second, there are very small numbers of respondents selecting the 8-day advance purchase fares. This causes some difficulties in fitting some of the models with restricted subsets of the data.

Table 2: Distribution of Chosen Alternatives by Number of Treatments

Alternative Chosen	Number of Treatments									
	4		8		16		24		32	
Ansett/Air NZ 29-Day	33	18.8	52	21.7	105	21.4	173	24.9	330	32.5
Ansett/Air NZ 15-Day	8	4.5	9	3.8	41	8.4	35	5.0	20	2.0
Ansett/Air NZ 8-Day	1	0.6	10	4.2	3	0.6	2	0.3	13	1.3
Qantas 29-Day	82	46.6	118	49.4	167	34.1	197	28.4	356	35.0
Qantas 15-Day	6	3.4	2	0.8	27	5.5	34	4.9	29	2.9
Qantas 8-Day	1	0.6	14	5.8	1	0.2	21	3.0	3	0.3
Ansett/Air NZ Full Economy	10	5.7	5	2.1	36	7.3	42	6.1	34	3.3
Qantas Full Economy	10	5.7	13	5.4	28	5.7	51	7.3	46	4.5
Ansett/Air NZ Business Class	0	0.0	1	0.4	12	2.4	19	2.7	40	3.9
Qantas Business Class	15	8.5	12	5.0	28	5.7	63	9.1	74	7.3
Would not Travel	10	5.7	4	1.7	42	8.6	57	8.2	71	7.0
Total	176		240		490		694		1016	

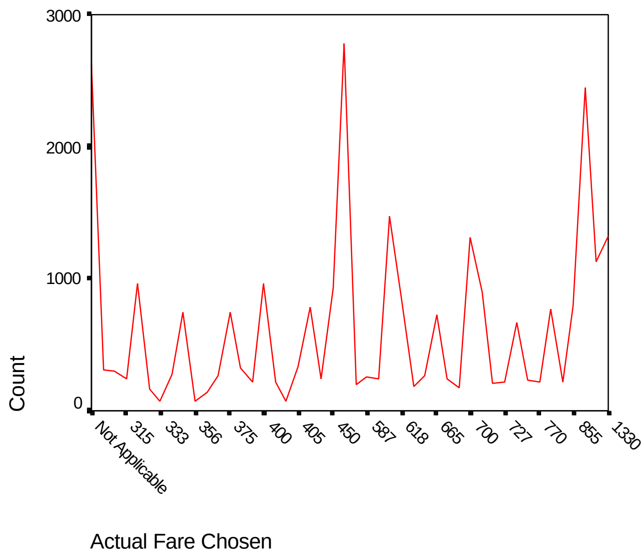

Figure 1: Distribution of Fares Chosen by Respondents

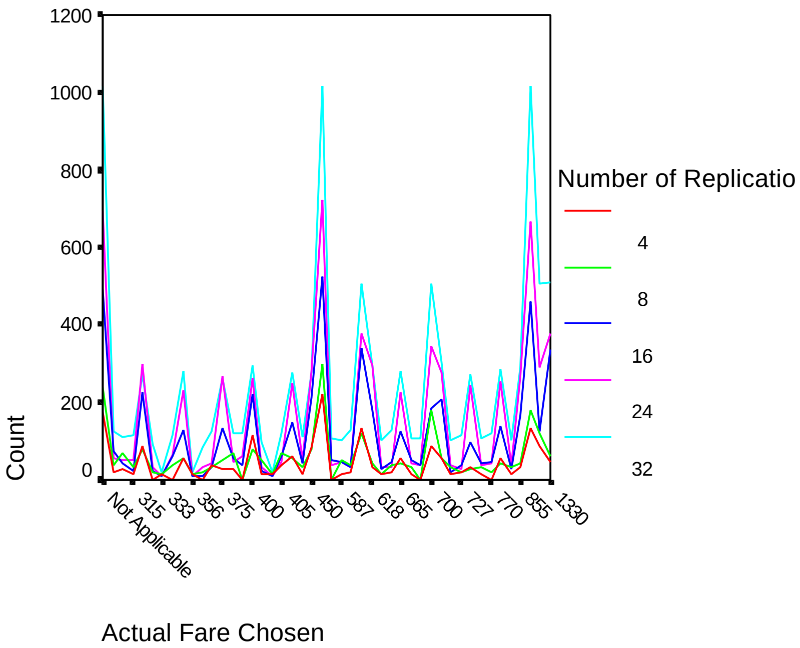

Figure 2: Distribution of Fares by Treatment Subsets

One of the issues that is frequently brought up in SP experiments is respondent burden. This is generally an issue in survey design *[7]*, but is raised more particularly in SP surveys because of the potentially large number of choice tasks that are presented to respondents. In this regard, an important concern might be that as the number of treatments increases, there might be a growing tendency for respondents to mark the same alternative for each treatment, indicating that the task becomes burdensome and less attention is paid to the details of the individual treatments. Table 3 shows the number of respondents in each treatment set who chose the same alternative from each treatment. The evidence suggests that respondent fatigue is not present. Because the response rates, shown in Table 1, are very even across the treatment subsets, it is apparent that the size of the task also had very little effect on response rates. What is apparent from Table 3 is that there is actually a decline in the number of cases in which a respondent chose the same alternative across all treatments. This suggests that the fewer treatments used, the less the respondent is able to discriminate among the profiles. In other words, having more treatments presented allows the respondent to find more occasions for choosing a different alternative.

Table 3: Respondents Choosing the Same Alternative by Number of Treatments

Number of Treatments	4	8	16	24	32
Number of Respondents Choosing the Same Alternative	13	4	0	2	1
Percent of Respondents Choosing the Same Alternative	29.5%	13.3%	0.0%	6.9%	3.1%

One further area that might indicate respondent burden would be failure to select an alternative for a treatment. There were a few instances of a treatment being missed out. There are, in fact, very few cases of this occurring, with none at all in the 4 and 8 treatment subsets, 6 in the 16-treatment subset, 2 in the 24 treatment subset, and 8 in the 32 treatment subset. These represent less than 1 percent in all cases, therefore indicating probable lack of any systematic respondent-burden effect.

Finally, it is interesting to look at whether there might be any systematic differences between the subsets of the sample in their choices, after controlling for membership in frequent flyer programs. Table 4 shows that, while there are interesting variations in the choices, there is no appearance of any systematic bias. There is some indication that those subjected to the larger numbers of treatments discriminated more resulting in less dominance of a particular alternative. However, this is not clearly established throughout the table. The table also shows the expected biases towards the airline for which the individual held frequent flyer membership, with less indication of biases to an airline in those with no membership and those with both memberships.

Table 4: Percent Choosing Alternatives by Treatment Subset and Frequent Flyer Membership

Frequent Flyer	Alternative	Treatment Subset					
		4	8	16	24	32	Total
None	AN 29	26.7%	24.2%	23.4%	28.4%	31.9%	28.4%
	AN 15+AN 8	8.3%	2.5%	7.3%	9.8%	3.7%	6.0%
	Q 29	40.0%	56.7%	36.5%	24.2%	36.2%	35.8%
	Q 15+Q 8	3.3%	0.8%	3.1%	7.2%	6.5%	5.3%
	AN FY	10.0%	2.5%	6.8%	7.2%	3.6%	5.3%
	Q FY	3.3%	5.8%	6.8%	3.8%	5.1%	5.1%
	AN BC	0.0%	0.8%	4.7%	5.3%	1.1%	2.7%
	Q BC	3.3%	5.0%	5.7%	8.3%	4.5%	5.6%
	Would not Travel	5.0%	1.7%	4.2%	5.7%	6.0%	5.1%
Qantas	AN 29	7.7%	11.1%	18.1%	11.4%	21.9%	16.0%
	AN 15+AN 8	1.9%	2.8%	4.4%	3.8%	0.0%	2.4%
	Q 29	63.5%	55.6%	35.6%	33.0%	49.3%	42.9%
	Q 15+Q 8	1.9%	12.5%	7.5%	11.4%	0.0%	6.2%
	AN FY	1.9%	2.8%	4.4%	1.1%	3.1%	2.6%
	Q FY	5.8%	8.3%	7.5%	8.3%	5.6%	7.1%
	AN BC	0.0%	0.0%	1.9%	1.9%	3.1%	2.0%
	Q BC	11.5%	5.6%	8.1%	15.5%	8.3%	10.5%
	Would not Travel	5.8%	1.4%	10.6%	13.6%	8.0%	9.6%
Ansett/Air NZ	AN 29	27.3%	20.8%	6.3%	42.7%	61.3%	46.2%
	AN 15+AN 8	6.8%	45.9%	50.0%	0.0%	3.1%	8.0%
	Q 29	40.9%	29.2%	12.5%	27.1%	24.4%	27.1%
	Q 15+Q 8	2.3%	4.2%	31.3%	0.0%	0.0%	2.1%
	AN FY	4.5%	0.0%	0.0%	16.7%	0.0%	5.3%
	Q FY	9.1%	0.0%	0.0%	6.3%	0.0%	2.9%
	AN BC	0.0%	0.0%	0.0%	0.0%	3.1%	1.5%
	Q BC	2.3%	0.0%	0.0%	0.0%	3.1%	1.8%
	Would not Travel	6.8%	0.0%	0.0%	5.2%	5.0%	4.7%
Both	AN 29	5.0%	41.7%	23.4%	37.5%	20.3%	25.3%
	AN 15+AN 8	0.0%	12.5%	11.7%	1.4%	8.6%	8.1%
	Q 29	35.0%	12.5%	29.7%	27.8%	10.2%	21.8%
	Q 15+Q 8	15.0%	20.8%	3.9%	8.3%	2.3%	6.0%
	AN FY	5.0%	0.0%	12.5%	5.6%	7.0%	8.1%
	Q FY	5.0%	0.0%	2.3%	18.1%	5.5%	6.5%
	AN BC	0.0%	0.0%	0.0%	0.0%	16.4%	5.6%
	Q BC	30.0%	8.3%	3.1%	0.0%	19.5%	9.9%
	Would not Travel	5.0%	4.2%	13.3%	1.4%	9.4%	8.6%

The Evidence

The focus of the modeling is a series of discrete choice models in which the accumulating and nonaccumulating treatment samples were evaluated. However, in the 4 treatments data set, no person chose Ansett/Air New Zealand Business Class, and very few chose either of the 8-day advance purchase tickets. As a result, it was not possible to fit models solely for the 4-treatment data set, or for the first four treatments in all subsets of the sample. In the following results, dummy variables are used when treatment subsets are combined, in order to account for variability in these.

A number of attempts were made to find a reasonably well-specified model. However, since the major focus of the research is on the effects of the number of treatments, in which the desire is to fit models with identical specifications for each subset of the data, the pursuit of a "best" model was deemed unnecessary. Two versions of model specifications were used. In the first version, the fare was entered as a simple variable with no interactions or adjustments. This was done to permit easy calculation of elasticities as one of the measures to be examined. The second version used a fare that was adjusted by centering, i.e., subtracting the mean fare from each value of fare, and then also adding in a quadratic term for the centered fare. Apart from this, the two model specifications were identical. The variables used were:

- ACTFARE – the fare for the alternative, including any applicable discounts, present in all alternatives except NONE
- FAREC – the centered fare, present in all alternatives except NONE
- FARESQC – the squared centered fare, present in all alternatives except NONE
- ANSETT – a dummy variable for membership in the Ansett/Air New Zealand frequent flyer program, present for all Ansett/Air New Zealand alternatives
- QANTAS – a dummy variable for membership in the Qantas frequent flyer program, present for all Qantas alternatives
- MWKDISC – a dummy variable indicating if a midweek discount fare was chosen, present for all discount fare alternatives
- SATNIT – a dummy variable indicating if a reduced fare for a Saturday night stay was chosen, present for all discount fare alternatives
- YADDLEG – a dummy variable indicating increased leg room in full economy seating (not available to discount economy passengers), present for full economy fare alternatives only
- YFOODFL – a dummy variable indicating availability of better food menus and free drinks in full economy seating (not available to discount economy passengers), present for full economy fare alternatives only
- FREELIMO – availability of free limousine service to and from the airport for business class passengers, present for business class alternatives only

- SATADULT – interaction term of Saturday night discount fare with number of adults in the household, present for all discount fare alternatives
- HHINC – household income, present for the business class alternatives only
- ADULTSHH – number of adults in the household, present for the alternative NONE only
- FREQFLY – a variable indicating whether the person was a member of no frequent flyer program, Ansett, Qantas, or both, present for the alternative NONE only

Because the main focus of this paper is on the number of treatments needed for well-specified models, we do not provide a discussion of the rationale for the selection of variables in the models. Comments are made later concerning the implications of the sign and magnitude of coefficients, but further discussion is not germane to this paper. The dependent variable defined eleven possible choices: three each of discount economy fares for each of Ansett/Air New Zealand and Qantas, full fare on each of the two airlines, business class fare on each of the two airlines, and the decision not to make the trip. A full set of alternative-specific constants were used, with not making the trip being the null case.

Table 5 shows the results of the two pooled models, using all of the data, irrespective of the number of treatments. There are two model versions, one with the simple fare variable to permit computation of elasticities easily, and the second having the more complex fare variables, with the quadratic term, yielding a better statistical model. In the "Simple Fare Model," fare has, as expected, a negative sign and is very significant. The alternative-specific constants are all positive and all significant. Membership in the frequent flyer programs has a pronounced effect on choice of the respective airlines. The negative sign on adults per household indicates that the more adults there are in the household, the less the respondent is likely to not make the trip. Household income has a positive, though small, effect on the probability of a respondent choosing business class. In general, these results are much as one would expect. The models exhibit reasonable fit to the data, the adjusted R^2 is acceptable, and the chi-square values are large and significant.

Table 5: Results for Pooled Models

Variable	Simple Fare Model		Complex Fare Model	
	Coefficient	t-score	Coefficient	t-score
ACTFARE	-0.00471	19.18	N/A	N/A
FAREC	N/A	N/A	-0.008801	16.89
FARESQC	N/A	N/A	0.007669	9.79
SATADULT	-0.196	2.72	-0.1995	2.75
MWKDIS	-0.00229	7.79	0.00346	4.98
SATNIT	0.316	1.87	0.3313	1.94
ANSETT	1.1635	7.98	1.1922	8.09
QANTAS	0.9722	9.09	0.9926	9.16
YADDLEG	0.53	3.85	0.6074	4.03
YFOODFL	0.5551	4.14	0.5985	4.14
FREELIMO	0.559	4.25	0.545	4.17
FREQFLY	0.148	2.13	0.1545	2.22
HHINC	0.00941	3.15	0.008965	3
ADULTSHH	-0.4559	3.65	-0.4607	3.67
GROUP8	1.3219	2.56	1.2939	2.5
AN29	3.6839	11.49	-2.5499	5.14
AN15	2.2692	6.71	-3.9372	7.8
AN8	1.3183	3.47	-4.9141	9.15
Q29	3.7755	11.78	-2.4679	4.97
Q15	1.9748	5.79	-4.2356	8.36
Q8	1.483	4.03	-4.7482	8.96
ANFY	1.361	3.7	-1.8176	5.19
QFY	1.4769	4.11	-1.6744	5.03
ANBC	1.3027	2.98	-1.4877	3.75
QBC	2.0509	4.78	-0.7347	1.89
Adjusted R^2	0.3119		0.3197	
Chi-Square	1204.9		1301.9	
Number of Observations	2616		2616	

Table 6 shows the results from the simple model with the two blocking strategies, the accumulated blocking strategy (ABS) and the non-accumulating blocking strategy (NABS). In each case, the table shows the values of the direct fare elasticities, together with summary model statistics of the adjusted R^2, the chi-square, and the number of observations. It should be recalled that there are insufficient observations on some choices to permit the 4-treatment subset to be analyzed on its own. In the accumulated blocking strategy, the column headed 8 uses the first eight treatments for all of those with eight or more treatments and all four of those from those who were given only four treatments. Similarly, the column headed 16 uses the first 16 treatments for all those with 16 or more treatments, and all the treatments from the 4- and 8-treatment groups. The column headed 32 reports the results of using all the data.

In the non-accumulating block strategy, the column headed ? 16 uses those respondents who were given 4, 8, or 16 treatments, while the column headed 16 uses only those who received 16 treatments. The column headed 32 uses only those who received the 32 treatments. Because of lack of choosers of the Ansett/Air New Zealand business class fare, models could not be estimated for 4 treatments alone, or 4 and 8 treatments together. It was necessary to include the 16 treatments as well. Dummy variables were included for each of group 8 and group 16 in this model.

Considering the differences in the data represented by the different columns in Table 6, the results show surprising consistency across the blocks. From the ABS strategy, there appears to be very little difference once 16 treatments are included. The addition of the extra 8 treatments for the 24-treatment block and of a further 8 for the 32-treatment block achieves very little in change in model parameters or in elasticity estimates. Variability is more marked in the NABS strategy, where most elasticities vary quite notably from one block to the next, although without the appearance of convergence to a specific set of estimates. For the cases of 16 and 24 treatments, there are also sample-size issues that may be playing a substantial role in modifying the results. The adjusted R^2 is much lower for these two cases, the chi-square is lower, and the sample sizes are small.

Table 6: Summaries of Elasticities and Overall Model Statistics for the Two Sets of Blocking Strategies

Measure/ Alternative	Accumulating Block Strategy				Non-Accumulating Block Strategy			
	£8	£16	£24	£32	£16	16	24	32
AN29 elasticity	-1.505	-1.244	-1.282	-1.259	-1.754	-1.861	-1.021	-1.217
AN 15 elasticity	-2.035	-1.9	-1.998	-2.015	-2.427	-2.491	-1.568	-2.318
AN 8 elasticity	-2.293	-2.191	-2.27	-2.288	-2.837	-3.054	-1.819	-2.529
Q 29 elasticity	-0.884	-1.016	-1.098	-1.12	-1.202	-1.48	-1.01	-1.244
Q 15 elasticity	-1.816	-1.902	-1.985	-2.018	-2.396	-2.577	-1.575	-2.326
Q8 elasticity	-2.067	-2.117	-2.232	-2.26	-2.865	-3.008	-1.768	-2.601
ANFY elasticity	-2.511	-2.543	-2.642	-2.687	-2.956	-2.959	-2.182	-3.107
QFY elasticity	-2.373	-2.518	-2.672	-2.724	-2.947	-3.129	-2.217	-3.235
ANBC elasticity	-3.828	-3.271	-3.424	-3.374	-4.164	-3.968	-2.803	-3.543
QBC elasticity	-2.77	-2.786	-3.014	-3.047	-3.362	-3.466	-2.563	-3.383
Adjusted R^2	0.306	0.303	0.306	0.312	0.315	0.298	0.254	0.391
Chi-square	391.6	850.7	1084.1	1204.9	398.4	256.8	312.8	579.9
Number of Observations	1149	1880	2363	2616	906	490	694	1016

Table 7 shows a comparison of the coefficients across the treatment subgroups for each of the two blocking strategies. In the interests of readability, the t-scores are not included, but indications are given of the level of significance of the coefficients. Unless otherwise noted, all coefficients are statistically significant beyond 99 percent. Only the models using the more complex fare variables are reported in Table 7.

Table 7: Coefficients for the Various Models Tested within the Two Blocking Strategies

Variable	Accumulating Block Strategy				Non-Accumulating Block Strategy			
	£8	£16	£24	£32	£16	16	24	32
FAREC	-0.0098	-0.009	-0.01	-0.009	-0.01	-0.008	-0.01	-0.0126
FARESQC	0.0109	0.0077	0.007	0.0077	0.007	0.00507*	0.005	0.0123
SATADULT	-0.275	-0.301	-0.258	-0.2	-0.389	-0.689	-0.474	0.319*
MWKDIS	0.0056	0.0037	0.003	0.0035	0.004	0.00341†	0.00008†	0.0077
SATNIT	-0.106†	0.433*	0.468	0.331†	0.489†	1.169	1.07	-0.797*
ANSETT	0.916	0.911	1.072	1.192	0.719	0.923†	0.939	1.635
QANTAS	1.098	1.075	1.051	0.993	0.904	0.679	1.328	0.888
YADDLEG	0.632*	0.63	0.587	0.607	0.397†	0.479†	0.359†	1.128
YFOODFL	0.300†	0.515	0.562	0.599	0.531*	0.570†	0.368†	0.992
FREELIMO	0.96	0.661	0.618	0.545	0.966	1.062	0.361†	0.325†
FREQFLY	0.144†	0.183*	0.148*	0.155*	0.258*	0.320*	0.0725†	0.255*
HHINC	0.00460†	0.00555†	0.00632*	0.009	0.00040†	0.00222†	0.00838†	0.0356
ADULTSHH	-0.479*	-0.452	-0.532	-0.461	-0.547*	-0.722*	-0.998	0.290†
GROUP8	1.021†	1.240*	1.286*	1.294*	0.840†	–	–	–
GROUP16	–	–	–	–	-0.634†	–	–	–
AN29	-3.158	-2.476	-2.465	-2.55	-2.392*	-2.849*	-1.921	-3.199
AN15	-4.187	-3.634	-3.745	-3.937	-3.18	-3.354	-3.096	-5.528
AN8	-4.912	-4.626	-4.698	-4.914	-4.235	-5.58	-5.57	-5.602
Q29	-2.947	-2.424	-2.409	-2.468	-2.172*	-2.633*	-2.14	-3.052
Q15	-4.702	-4.267	-4.155	-4.236	-4.034	-4.001	-3.46	-5.089
Q8	-4.963	-4.542	-4.61	-4.748	-4.46	-6.937	-3.528	-7.019
ANFY	-1.72	-1.657	-1.833	-1.818	-1.072†	-1.637*	-2.389	-1.220†
QFY	-1.598	-1.593	-1.757	-1.674	-1.298*	-2.012	-2.428	-0.600†
ANBC	-1.634*	-1.127*	-1.484	-1.488	-1.066†	-1.567†	-1.518*	-1.245†
QBC	-0.408†	-0.478†	-0.721†	-0.735†	-0.119†	-1.141†	-0.697†	-0.548†
Adjusted R^2	0.315	0.311	0.313	0.32	0.319	0.3	0.258	0.407
Chi-Square	440.8	920.9	1164.9	1301.9	416.6	261.8	326.3	657.8
Number of Observations	1149	1880	2363	2616	906	490	694	1016

* Indicates significance at 95% but not at 99%
† Indicates lack of significance at 95%
All other coefficients are significant beyond 99%

Looking first at the ABS results, which show the effect of adding more treatments to the initial four or eight, it appears that most coefficient values converge fairly quickly on the values obtained with the pooled data. By the time the 16 treatments are included, there is little subsequent change in the values of most coefficients, and few results differ in significance. Because of variation in sample size, it is not entirely clear how much of the variation in values to be seen here is a function of changing sample size and how much to omission of treatments. In the NABS results, however, it is less clear, but the results are also more substantially affected by changes in sample size. In these results, where the individual treatment subsets of the sample are analyzed separately, there is no indication of convergence of values. It should be noted that the column labeled ≤16 in this strategy is a different sample from the one labeled ≤16 in the ABS results. The latter uses all of those who received 4, 8, or 16 treatments, and the first 16 treatments of those who received the 24 and 32 treatment surveys. The former, however, uses just the combination of the 4-treatment, 8-treatment, and 16-treatment subsets. No information is used from the 24- and 32-treatment subsets in this case.

In an effort to isolate the effects of sample size on the modeling results, a second set of model estimates was prepared in which the number of treatments used in each model was approximately the same. This was done by subsampling individuals from the data set until the number of treatments was approximately equal to the smallest number of each of the ABS and NABS strategies. The controlling number for ABS was the 1149 observations for the ≤8 group, while for NABS, it was the 490 treatments in the 16 treatments subset. The model estimates are summarized in Table 9. Comparing Tables 7 and 9, it is apparent that many of the coefficients are now quite different with the equal samples for the ABS, than for the full available data (Table 7). The mode-specific constants are markedly different, and many of the coefficients are different, including substantial differences in which variables are significant at 95 percent or better. There is relatively little evidence of trends in the values from one treatment subset to another. Also, the model goodness-of-fit statistics are clearly on a trend of significant improvement from the ≤8 to the ≤32 treatment groups. In the case of NABS, much the same thing can be concluded. The constant sample size leads to much more variability in the results and there is little evidence of trends in coefficient values. The goodness-of-fit statistics are again much better for the data from the 32 treatments than from any of the other subgroups.

The comparison of coefficients in a logit model is less interesting than a comparison of behavioral outputs such as elasticities. Indeed, as shown in Louviere et al. *[2]*, the coefficients of a logit model are essentially uninformative and what really has meaning is the marginal effect of a variable or its elasticity. The elasticity results for the full sample and the equal sample data are shown in Tables 6 and 8 respectively, and graphed in Figure 3.

Table 8: Summaries of Elasticities and Overall Model Statistics for the Two Sets of Blocking Strategies – Equal Samples

Measure/ Alternative	Accumulating Block Strategy				Non-Accumulating Block Strategy			
	£8	£16	£24	£32	£16	16	24	32
AN29 elasticity	-1.508	-1.161	-1.213	-1.252	-1.671	-1.861	-0.71	-1.279
AN 15 elasticity	-2.039	-1.803	-1.96	-2.155	-2.332	-2.491	-1.098	-2.597
AN 8 elasticity	-2.297	-2.101	-2.23	-2.443	-2.756	-3.054	-1.293	-2.892
Q 29 elasticity	-0.885	-0.904	-1.01	-1.108	-1.204	-1.48	-0.669	-1.304
Q 15 elasticity	-1.82	-1.785	-1.91	-2.138	-2.299	-2.577	-1.114	-2.607
Q8 elasticity	-2.07	-2.017	-2.175	-2.394	-2.586	-3.008	-1.263	-2.914
ANFY elasticity	-2.514	-2.393	-2.536	-2.802	-2.679	-2.959	-1.617	-3.493
QFY elasticity	-2.376	-2.424	-2.601	-2.897	-2.691	-3.129	-1.644	-3.774
ANBC elasticity	-3.834	-3.194	-3.364	-3.564	-3.871	-3.968	-2.182	-4.093
QBC elasticity	-2.774	-2.742	-2.983	-3.235	-3.155	-3.466	-1.947	-4.006
Adjusted R^2	0.305	0.336	0.341	0.367	0.303	0.298	0.286	0.456
Chi-square	386.6	527.7	506.1	574.8	221.1	256.8	249	309
Number of Observations	1149	1160	1153	1171	482	490	502	505

Table 9: Coefficients for the Various Models Tested within the Two Blocking Strategies – Equal Samples

Variable	Accumulating Block Strategy				Non-Accumulating Block Strategy			
	£8	£16	£24	£32	£16	16	24	32
FAREC	-0.01	-0.008	-0.008	-0.009	-0.008	-0.008	0	-0.0137
FARESQC	0.011	0.0068	0.0066	0.0079	0.0062	0.00507*	0.003	0.0134
SATADULT	-0.275	-0.446	-0.419	-0.381	-0.604	-0.689	-0.568	0.450*
MWKDIS	0.0056	0.00237*	0.00214*	0.0033	0.00344*	0.00341†	-0.0023*	0.0081
SATNIT	-0.106†	0.755	0.753	0.605	1.035	1.169	1.413	-1.358
ANSETT	0.917	0.802	0.844	0.824	0.411†	0.923†	0.974	1.123
QANTAS	1.095	0.814	0.488	0.554	0.841	0.679	1.699	0.535*
YADDLEG	0.631*	0.814	0.624	0.993	0.592†	0.479†	0.490†	1.986
YFOODFL	0.299†	0.705	0.674	1.046	0.849	0.570†	0.448†	1.512
FREELIMO	0.961	0.778	0.801	0.686	1.159	1.062	0.383†	0.154
FREQFLY	0.169†	-0.214†	-0.011†	-0.146†	-0.316†	0.320*	0.408†	0.203†
HHINC	0.00454†	-0.00061†	-0.0113	-0.00873*	-0.0152*	0.00222†	-0.0126*	-0.0038†
ADULTSHH	-0.479*	-0.341†	0.0617†	0.0336†	-0.211†	-0.722*	-0.985	1.055
AN29	-3	-1.519*	-0.423†	-1.172†	-1.399†	-2.849*	-0.476†	-1.477†
AN15	-4.029	-2.723	-1.922*	-2.796	-2.213†	-3.354	-1.563†	-3.590*
AN8	-4.754	-4.745	-3.729	-4.36	-3.889	-5.58	-32.418†	-5.051
Q29	-2.789	-1.329†	-0.229†	-1.085†	-1.173†	-2.633*	-0.691†	-1.511†
Q15	-4.543	-3.205	-2.016	-3.122	-2.915*	-4.001	-2.308	-3.694*
Q8	-4.804	-4.244	-3.111	-4.035	-3.367	-6.937	-2.75	-33.454†
ANFY	-1.547	-1.417	-0.344†	-1.193*	-0.486†	-1.637*	-2.246	-0.662†
QFY	-1.425	-1.451	-0.384†	-0.974†	-0.555†	-2.012	-2.499	0.0011†
ANBC	-1.457*	-0.500†	1.048†	0.979†	0.817†	-1.567†	-1.543*	3.651
QBC	-0.230†	-0.0782†	1.533	1.461	1.162†	-1.141†	-0.587†	4.116
Adjusted R^2	0.314	0.342	0.346	0.374	0.307	0.3	0.288	0.469
Chi-Square	436.1	562.6	535.3	614.6	230.2	261.8	253.6	339.9
Number of Observations	1149	1160	1153	1171	482	490	502	505

* Indicates significance at 95% but not at 99%
† Indicates lack of significance at 95%
All other coefficients are significant beyond 99%

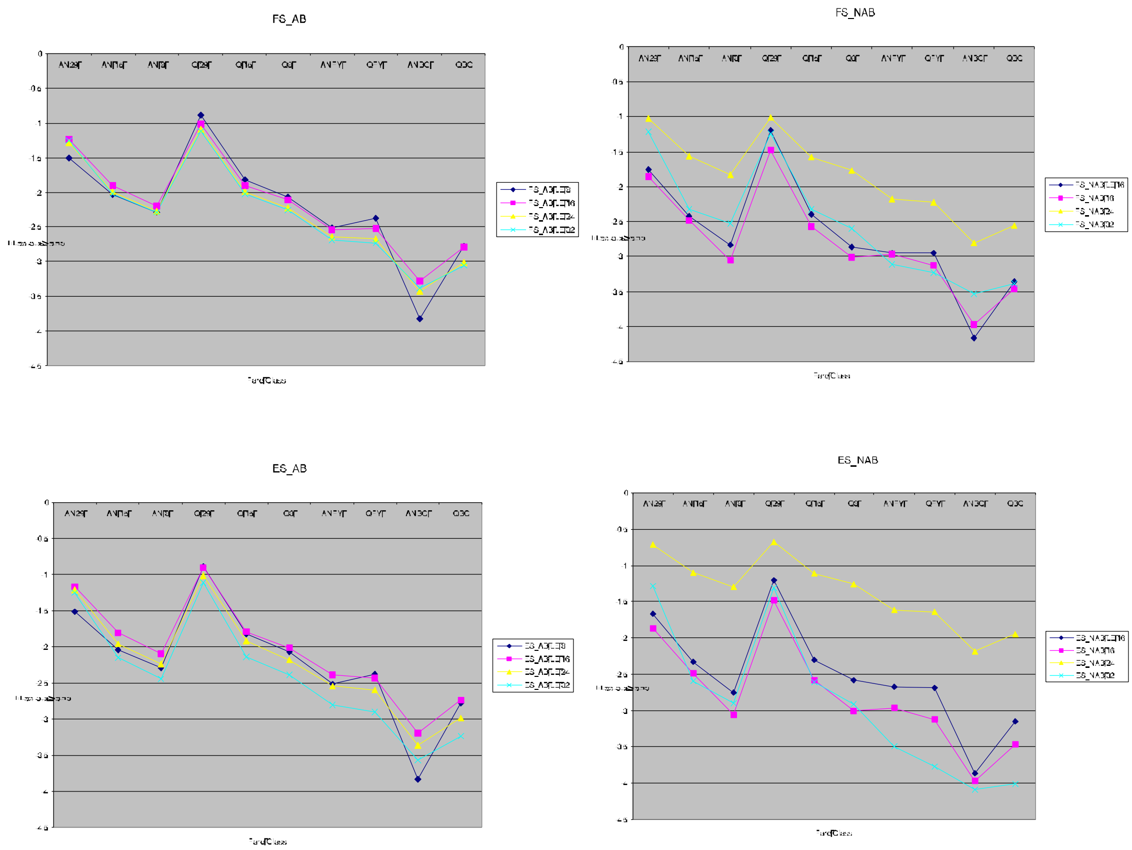

Figure 3: Comparison of Elasticities Between Equal and Full Sample for Each Blocking Strategy

Looking first at the full sample for ABS, we see very similar elasticity estimates within each fare class, with the possible exception of Ansett business class (a range of -3.37 to -3.83). However, for NABS, we see a much greater variation across the treatment sets, most noticeably between the 24 treatment and other treatments (i.e., ≤16, 16, and 32). There is no obvious reason for the 24 treatment to be different. Since the NABS format is more informative in terms of how SC studies would implement an experiment, the lesser number of treatments and the greater number of treatments (with the exception of Ansett business class) produce very similar mean elasticity estimates. This is extremely encouraging, suggesting at least for linear models (that are most commonly used in transportation) that the empirical gains from a larger number of treatments are marginal at best.

When we control for differences in sample size, we observe a greater divergence in elasticity estimates between treatment sets, but it is not substantially greater. What is of particular interest in both the full and equal sample contexts is that the greatest similarity occurs between the two extreme treatment sizes with the full-size estimates being slightly higher within the smallest treatment compared to the equal-size estimates (typically an absolute difference of 0.005 to 0.01); and the reverse for the largest treatment set.
The final comparison is between the full and equal sample elasticity estimates for a given treatment set. We find very close correspondence for the smaller treatment sets (i.e., ≤16

and 16) with noticeable differences for 24 and 32 treatment sets (except for AN29 and Q29 in the latter set), although the gap is consistently greater for the 24-treatment set, varying between 0.4 and 1 in absolute elasticity value. This is a large difference at the upper end of the number of treatments.

Conclusions

The main message from this initial empirical enquiry into the influence of varying the number of treatments in an SC experiment that a respondent is asked to evaluate is that the empirical gains within the context of a linear specification of the utility expression associated with each alternative in a discrete choice model may be quite marginal. Although this conclusion appears to hold under alternative sample size schema, a comparison of the absolute magnitudes of mean elasticities is insensitive to sample size for treatments ≤16 and 16, but very sensitive to treatments 24 and 32. This might be expected because the larger treatment sets in the current study were the ones subject to greater re-sampling. Interestingly, the 24-treatment set has mean elasticity estimates higher for all fare classes for the equal sample whereas the reverse holds for the 32-treatment set, except for two fare classes – AN29 and Q19 – which are indistinguishable.

Encouragingly, the analysis of the survey results directly showed little evidence of problems with the number of treatments placed before people, although there was some indication that 4 treatments may be too few to allow sufficient variability in responses. There was, surprisingly little evidence of fatigue effects for even the 32-treatment design.

A tentative conclusion from these results would be that around 16 treatments may be sufficient, although results can always be improved with 24 and even 32 treatments, but that once 16 treatments are used, efforts may well pay off in larger sample sizes, particularly in ensuring that there are sufficient choices of each alternative in the choice set.

References

1. Hensher, D.A. (1994) Stated preference analysis of travel choices: the state of practice, *Transportation*, 21 (2), 107-134.

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

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

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

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

6. Hensher, D.A., Louviere, J.J. and Swait, J. (1999) Combining Sources of Preference Data, *Journal of Econometrics*, 89, 197-221.

7. Stopher, P.R. (1996) Household Travel Surveys: Cutting-Edge Concepts for the Next Century, *Conference Proceedings No.10: Conference on Household Travel Surveys: New Concepts and Research Needs*,Transportation Research Board, Washington, DC, pp.17-18.

I T S

**INSTITUTE OF
TRANSPORT STUDIES**

The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University