8th WCTR

ANYTHING YOU CAN DO, WE CAN DO BETTER: A PROVOCATIVE INTRODUCTION TO A NEW APPROACH TO STATED PREFERENCE DESIGN

JEREMY P TONER
Institute for Transport Studies
University of Leeds
Leeds, LS2 9JT, UK

STEPHEN D CLARK
Institute for Transport Studies
University of Leeds
Leeds, LS2 9JT, UK

SUSAN M GRANT-MULLER
Institute for Transport Studies
University of Leeds
Leeds, LS2 9JT, UK

ANTHONY S FOWKES
Institute for Transport Studies
University of Leeds
Leeds, LS2 9JT, UK

Abstract

This paper presents a new method for designing stated preference (SP) experiments. Currently, many SP designs used in practice maintain orthogonality between the attribute levels, because it is thought this will produce lower standard errors associated with the parameter estimates. We however contend that the standard errors of coefficient estimates derived from a logit model are not necessarily minimised when using an orthogonal design. We have obtained expressions for minimising the variance of the estimated parameters of a logit model which indicate that choices should be offered to respondents which have probabilities of being chosen of 0.917 and 0.083. We have also derived a limit for the t statistic associated with each parameter, so that any design can be assessed against the theoretical optimum in terms of the t statistics it produces. Clearly, an optimal design depends on some foreknowledge of the parameters, but when we subjected the technique to empirical application, we found that our advanced design outperformed a standard orthogonal design notwithstanding that the orthogonal design actually produced estimated parameters much closer to the design points than the advanced design.

INTRODUCTION

Stated Preference (SP) experiments present individuals with hypothetical scenarios characterised by relevant variables or attributes. The individuals are then asked to choose between the different scenarios. By offering trade-offs amongst the levels of the attributes, the responses supplied allow the estimation of the importance attached to each of the attributes in decision-making. This approach to transport demand modelling has been used increasingly in recent years. After initial use of rating scales and ranking experiments, it has been found more satisfactory to use binary response experiments which are now by far the most commonly used in transport applications. We therefore concentrate on binary response experiments. In such experiments, each respondent is presented with a small (typically between 8 and 16) number of choices or replications. Each choice consists of a pair of scenarios and the respondent is invited to express a preference for one scenario or the other.

In this paper, we examine some recent advances in procedures for designing SP experiments. Section 2 reviews design principles which are in current use. In section 3, we set out the theory for designing statistically efficient experiments. The outworking of this theory is presented in section 4, where we show the potential improvements in efficiency which can be obtained and demonstrate some actual efficiency gains in a recent application to motorists' parking choices. Our concluding remarks are in section 5.

CURRENT SP DESIGN PRINCIPLES

SP experiments in transport are typically used for one of two purposes:

- (i) estimating relative values, such as the money value of time;
- (ii) forecasting.

In case (i), it is the ratios of the parameter estimates which are of interest, while in case (ii) it is the set of parameter estimates, which drive the forecasts, which are of primary interest. This distinction between parameter estimates and ratios of parameter estimates is an important one which needs to be maintained when considering the efficiency of different statistical designs. In particular, the most efficient design to capture parameter estimates may not be the most efficient design to capture parameter ratios. The efficiency of the statistical design is important because it has a direct impact on either the cost of the data collection or the accuracy of the information gained or both.

At the outset, we must clearly define what we mean by statistical design and distinguish it from the more general term of experimental design. Experimental design can be taken to represent the entire process of creating an SP choice exercise. Hensher (1994) lists several key stages in experimental design. These include: the identification of the set of attributes; the selection of the measurement units for each attribute; the specification of the number and magnitude of the attribute levels; and statistical design itself. The latter was defined as follows: *"Statistical design is where the attribute levels are combined into an experiment"*.

A principal advantage of SP methods is that they allow the levels of the independent variables and how they are combined to be determined by the analyst. It is therefore important that this feature is exploited in order to achieve the greatest benefit in terms of the precision of the parameter estimates. Louviere and Woodworth (1983) state: *"Thus, we suggest that for both practical and academic applications, main* *effects fractional factorial design plans are an efficient choice for the design of discrete choice or resource allocation studies".*

Fractional factorial plans are by far the most commonly used statistical designs. The principal feature of such designs is orthogonality, that is, the absence of correlations between independent variables. The use of such plans for a given number of attributes and given numbers of levels of the attributes will determine the number of replications required. The main attractions claimed for this approach are:

i) No confounding effects between variables;
ii) The standard errors of parameter estimates are lower than they would otherwise be;
iii) The design plans are straightforward to implement.

Research into SP experimental design can be regarded as having followed one of three strands: the pursuit of orthogonality; the use of non-orthogonal designs, especially for obtaining relative valuations; and statistical methods. We address each of these in turn.

Orthogonality

In the early SP applications it was felt sufficient to assign values to the levels of each variable in the orthogonal plan according to reasonable percentage changes to some base values of the variables. It was soon realised that consideration must be given to the nature of the trade-offs. Here the concept of boundary values proves useful. Fowkes and Wardman (1988) provided an early discussion of this: *"The choice of the particular attribute values must take into account the relative valuations at which individuals would be indifferent between options. In order to achieve a satisfactory design, the set of these boundary or equi-utility relative valuations should cover a reasonable range....... In order to obtain accurate estimates of the respondent's relative valuation, we must present sufficient boundary values to make the inter-boundary value distance acceptably small".*

A boundary value of time (BVoT) is thus the value of time (or the `price' of time) at which, given the attribute levels presented in any given replication, an individual would be indifferent between the scenarios. Following Fowkes and Wardman (1988), designers have sought to incorporate a range of boundary values clustered around some *a priori* expected value of time or to include a range of plausible values with the greatest concentration of boundary values at or about the expected value of time. An example of a two-variable design, taken from Fowkes and Nash (1991), is given in table 1. In this design, scenario B is always the faster but more expensive option. This need not always be the case, a mixture of either scenario being the faster, more expensive is possible. It is also possible to have one of the scenarios being both the faster and cheaper option, in other words a dominated choice *ceteris paribus*. While the incorporation of a limited number of dominated choices may not be too problematical, in general practitioners have agreed with Bates (1994), who states that: *"in SP the comparison of two alternatives where one 'dominates' the other is generally uninformative: this is not the case in more general scientific experimental work. This is now recognised, and the result has been a considerably more careful and thorough approach to SP design".*

The aim has been to obtain responses where, for any given replication, each scenario has a non-negligible chance of being preferred. For the two-attribute design in table 1, the boundary values of time (BVoT) are simply given by (minus) the COST difference divided by TIME difference. Thus for replication 1, if an individual's value of time is 1.5 pence per minute then they will indifferent between scenarios A and B in this replication. If their value of time is less than 1.5, then they would be expected to choose the slower but cheaper option (scenario A). If their value of time is greater than 1.5, then alternative B should be the preferred scenario. It is possible to plot a graphical representation of the spread of BVoTs. This plot for the design in table 1 is given in figure 1.

Table 1 : A two-variable binary choice SP design

Replication	Scenario A		Scenario B		Difference		
	COST (pence)	TIME (min)	COST (pence)	TIME (min)	COST (pence)	TIME (min)	BVoT (pence/min)
1	100	30	115	20	15	-10	1.50
2	100	30	125	20	25	-10	2.50
3	100	30	140	20	40	-10	4.00
4	150	45	165	30	15	-15	1.00
5	150	45	175	30	25	-15	1.67
6	150	45	190	30	40	-15	2.67
7	200	60	215	40	15	-20	0.75
8	200	60	225	40	25	-20	1.25
9	200	60	240	40	40	-20	2.00

Figure 1 begins to show how effective a design should be at recovering a range of values of time. In the discussion which follows, a perfect knowledge on the part of the respondent is assumed (ie deterministic choice). In practice, we use a random utility model to model choices, so that even if someone's true value of time is greater than the BVoT in a given replication, there is a non-zero probability of choosing (apparently "wrongly") the slower, cheaper scenario.

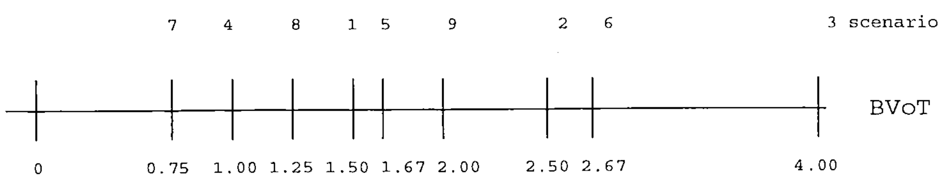

Figure 1 : Boundary value map of table 1 design

With reference to the example design, if the value of time is greater than 4.00 pence per minute, then, *ceteris paribus*, all the respondents will always chose the faster, more expensive scenario. Thus the only clear result will be that the lower bound of value of time is 4.00. If the value is between 2.67 and 4.00 then all the respondents will chose the faster, more expensive scenario, except for replication 3, in which case they would select the slower, cheaper scenario. In this case there is both a lower and upper bound on the value of time. The interval is however wide, at 1.33 pence per minute. If the value of time is between 1.50 and 1.67 then the choice will be the faster, more expensive mode for replications 7, 4, 8 and 1 and the slower, cheaper mode for replications 5, 9, 6 and 3. The interval is also narrow at 0.16. Conventional wisdom suggests that this design would perform well at recovering true values of time in the range 1.00 to 2.00 pence per minute.

It is possible to extend these design principles to cover cases with three attributes, where the equivalent concept is boundary rays. Considerable effort has been expended on determining where to put the boundary rays in order to improve the accuracy of estimation, in particular on how best to cover the area (in parameter ratio space) where it is anticipated that the true monetary valuations of attributes lie. For example, Fowkes (1991) and Holden (1993) demonstrated the possibility of using such techniques to permit improvements in the accuracy of the estimates compared with those obtained from fractional factorial SP designs.

Boundary rays do not, though, tell the whole story. For any boundary ray and for a given combination of true parameters, there is a probability of the choice made being one side or the other of the boundary ray. We therefore examined how this probability changed for different parameter combinations for each boundary ray of an orthogonal design in turn. In our design, there were nine scenarios, but it happened that two of the nine boundary rays coincided. However, a given location in parameter space away from these two rays did not lead to the same probability of choosing the better scenario for both rays. Thus the boundary ray cannot be considered in isolation from the magnitude of the parameters.

Non Orthogonal design to improve value of time estimation

Various pieces of research have suggested moving away from orthogonal designs. In some cases, this might be for pragmatic reasons, such as realism, but in others the aim has been to improve the efficiency of value of time estimates.

When the purpose of the exercise is to determine monetary valuations, Fowkes (1991) advocated orthogonality amongst non-cost attributes but correlations between the cost and non-cost attributes. This was because the formula for the variance of the value of time, where the value of time is obtained as a ratio, contains a covariance term which, if suitable correlation was introduced, would reduce the variance of the value of time. The approximate variance of the value of time derived as the ratio of time and cost coefficients (β_t and β_c) is (Fowkes *et al.*, 1993):

$$Var\left(\frac{\beta_t}{\beta_c}\right) = \frac{\beta_t^2}{\beta_c^2}\left(\frac{Var(\beta_t)}{\beta_t^2} + \frac{Var(\beta_c)}{\beta_c^2} - \frac{2\,Cov(\beta_t,\beta_c)}{\beta_t\beta_c}\right) \tag{1}$$

Fowkes *et al.* (1993) showed in the two variable case that, although correlations between variables increase the variances of their coefficient estimates, it can lead to appreciable reductions in the variance of the value of time. These results were obtained from regression theory. For the linear regression equation with two variables, it is possible to express the variances of the parameters and their covariances in terms of the correlation between COST and TIME. By differentiating the variance of the parameter ratio (β_t/β_c) with respect to the correlation between COST and TIME and setting the derivative to zero, an optimal degree of correlation between the attributes, that is one which minimises the variance of the value of time, can be obtained.

To summarise Fowkes *et al.*'s results, reductions in the variance of the value of time of up to 50% were obtained using non-orthogonal designs rather than the traditional orthogonal design, although it was accepted that there may be reasons of a practical, contextual or plausibility nature which might restrain the degree of non-orthogonality acceptable in a design. Having demonstrated the desirability of non-orthogonality in a logit regression context, simulations were carried out based on a discrete choice logit model to test the transferability of the results. This simulation, where the degree of correlation between the attributes was close to the optimal correlation when using regression, showed the predicted improvement to the accuracy of the estimation of the value of time compared with the orthogonal design, but (Fowkes *et al.*, 1993): *"The coefficient estimates were, as expected, less precise in the non-orthogonal case".*

However, Watson *et al.* (1996) found that this is not a hard and fast rule; the introduction of correlation between independent variables can be consistent with reductions in the variances of the coefficient estimates themselves in disaggregate logit models. Thus, the accepted wisdom derived from logit regression that departures from orthogonality are acceptable (and desirable) when estimating relative valuations but not when using the model for forecasting is correct; but the result does not transfer to disaggregate logit modelling, when it is not necessarily the case that orthogonal designs are superior for either valuation or forecasting.

Statistical approaches

This strand of research in the transport area has its roots in the work of Gunn (1981) who set out a theoretical analysis of the statistical aspects of value of time estimation and who was concerned with improving the precision of value of time estimates by calibrating models on data points which had the greatest informational content. Gunn concluded that data points which gave choice probabilities close to 50/50 were not desirable. Unfortunately, this approach has been somewhat neglected. However, our research is in this vein and the concept is fairly straightforward. If we wish to obtain estimates with the minimum variance possible then we should choose the independent variables so that, taken together, the variances of the parameters of interest are minimised. We consider the further development of this approach in section 3.

A summary of the conventional approach to SP design

The design principles followed by most practitioners can be conveniently summarised as follows:

- When calibrating models for forecasting purposes, where the parameters are of primary importance, orthogonal designs should be used;
- Departures from orthogonality can be beneficial when the purpose of the exercise is to obtain monetary valuations. In particular, correlations between cost and the other attributes is helpful, but orthogonality should be maintained between non-cost attributes;
- Boundary values are helpful in determining appropriate departures from orthogonality. Designers should seek to incorporate "difficult" or marginal choices, since that is where the information content of the responses is greatest.

A NEW APPROACH TO SP DESIGN

Since the efficiency of parameter estimates is the principal theme of this paper, we here present some variance formulae for the logit model as that is the model most often used to analyse the responses obtained from SP discrete choice experiments. The logit model expresses the probability of an individual I choosing alternative j from amongst the m available alternatives as a function of their relative utilities (U) as:

$$P_{ij} = \frac{e^{\Omega U_j}}{\sum_{k=1}^{m} e^{\Omega U_k}} \qquad (2)$$

where Ω is a scale factor whose purpose is to account for the effect of unobserved factors on choices and which is related to the standard deviation of the error term (σ_k) as:

$$\Omega = \frac{\pi}{\sqrt{6}\,\sigma_k} \qquad (3)$$

In the case of just two alternatives (1 and 2), the model simplifies to:

$$P_{i1} = \frac{1}{1 + e^{\Omega(U_2 - U_1)}} \tag{4}$$

The observable or deterministic utility for any alternative k is related to observable variables (X_k) which influence choice:

$$U_k = f(\beta_k, X_k) \tag{5}$$

The parameters (β_k) of the discrete choice logit model are estimated by maximum likelihood and the statistical design will have a direct bearing on the precision with which they are estimated. The most common form of utility function is linear additive; for two variables, and henceforth suppressing the scale factor, this is expressed as:

$$U_k = \beta_{1k}X_{1k} + \beta_{2k}X_{2k} \tag{6}$$

Constraining the coefficients to be generic allows the utility function to be rewritten in difference form and, using lower case x's to denote differences, the variance and covariance expressions across n choice observations are:

$$Var(\beta_1) = \frac{\sum_n \hat{P}_n(1-\hat{P}_n)x_{n2}^2}{D} \tag{7}$$

$$Var(\beta_2) = \frac{\sum_n \hat{P}_n(1-\hat{P}_n)x_{n1}^2}{D} \tag{8}$$

$$Cov(\beta_1,\beta_2) = \frac{-\sum_n \hat{P}_n(1-\hat{P}_n)x_{n1}x_{n2}}{D} \tag{9}$$

where $\hat{P}_n$ is the probability of choosing one option over the other in the *n*th replication and

$$D = \sum_n \hat{P}_n(1-\hat{P}_n)x_{n1}^2 \sum_n \hat{P}_n(1-\hat{P}_n)x_{n2}^2 - \left(\sum_n \hat{P}_n(1-\hat{P}_n)x_{n1}x_{n2}\right)^2 \tag{10}$$

These formulae appear similar to those obtained from ordinary least squares regression, with the exception of the additional $\hat{P}_n(1-\hat{P}_n)$ term within each summation. However, in the comparable formulae from regression, the x's are deviations from the mean and not, as here, the actual values. Inspection of the covariance matrix output by logit estimation packages shows that the correlations between estimated coefficients are not zero even when the input variables are independent of each other. It is theoretically possible (though unlikely) to obtain zero correlations between estimated coefficients in logit models; but there is no relationship between this event and correlation between the independent variables. In contrast, the correlation between two regression coefficient estimates is simply the negative of the correlation between the independent variables in the two independent variable case. Care therefore needs to be taken in drawing conclusions for disaggregate logit models by analogy with regression theory.

Appreciation of the importance of the $\hat{P}_n(1-\hat{P}_n)$ term led us to investigate further the theoretical expressions for the variances of the parameters estimated from a logit model. Differentiating the variance of $\hat{\beta}_1$ with respect to x_{i1}, the value of x_1 in the ith replication, and with respect to x_{i2} and equating both to zero yields, in the case of a two variable utility function, the first order (necessary) condition for the variance of $\hat{\beta}_1$ to be minimised as (suppressing subscripts I):

$$(x_1 Var(\beta_1) + x_2 Cov(\beta_1,\beta_2))[(1 - 2\hat{P})(\beta_1 x_1 + \beta_2 x_2) + 2] = 0 \quad (11)$$

for each pairwise choice, where again the x's denote differences in the variables between alternatives. A similar expression can be derived to minimise the variance of $\hat{\beta}_2$.

For equation 11 to hold, either the first term in the expression equals zero, which requires that x_1 and x_2 be in fixed proportion throughout the statistical design (a constant boundary value of time), in which event the model cannot be calibrated; or the second term equals zero. In this latter case, design becomes a matter of choosing x_1 and x_2 to fit the condition for an assumed set of β_1 and β_2. When following this latter rule, it is therefore important to know the actual magnitude of the parameters it is expected to find (as indicated by Louviere and Woodworth, 1983), rather than just the relative value of the coefficients as with the boundary value ray approach. These assumed parameters have been termed the design points. In order for the necessary condition to be satisfied, we need to choose x_1 and x_2 so that:

$$(1 - 2\hat{P})(\beta_1 x_1 + \beta_2 x_2) + 2 = 0 \quad (12)$$

where again the x's are expressed as differences between alternatives. The solution to this equation is found where:

$$\beta_1 x_1 + \beta_2 x_2 = \pm 2.399 \quad (13)$$

Equation 13 simply states that the (scaled) utility difference between the two alternatives equals ±2.399. This then gives the choice probabilities for the two alternatives as 0.917 and 0.083. Once we have these "magic P's", it is possible to demonstrate (though we have not yet derived a formal proof) that there is a limit to the t statistic ("magic t") for any particular parameter given by:

$$|t^*| = \frac{\sqrt{n}}{2}[\{\log(\frac{1 - P^*}{P^*})\}^2 - 4]^{0.5} \quad (14)$$

where n is the number of replications and P^* is the magic P. It is not possible for all t statistics to approach this limit simultaneously; only one t at a time can be "magic". There are consequently a number of approaches to achieving an overall best design. We have adopted an iterative approach optimising each of the t statistics in turn, stopping the process when non optimised t statistics were deemed satisfactory. This involved judgment as to whether the non-optimised t's were sufficiently close to magic t's and whether the x's were reasonable. When it is the ratio of parameters which is of interest, the minimum value of the variance is zero. It is therefore not possible to perform an unconstrained optimisation on the t ratio of the money value of time.

These theoretical developments yielded three important findings which contradict established beliefs:

- Marginal (close to 50/50) choices are not desirable for efficiency.
- Fractional factorial orthogonal plans do not necessarily provide coefficient estimates in disaggregate logit models with least variance; indeed they can be regarded as a very special case.
- Boundary ray maps can be misleading.

Furthermore, the design principles outlined here readily generalise to any number of independent variables, which is a considerable advantage over the boundary ray approach.

Testing of the new approach

We have tested our new approach in a number of ways (Clark and Toner, 1996). The first was the sensitivity of a new-style design to the optimality criterion used. While the design obtained does depend on the optimality criterion, it seems that the first-order conditions hold (ie magic P's) for (among others): successive optimisation of each parameter's t value in turn; maximising the sum of the absolute t values of the parameters; minimising the sum of the squared deviations of the t values from magic t; and a version of C-optimality.

One of the crucial elements in design is choosing a set of parameters on which to base the design. We tested three situations:

(i) where there is a reasonable degree of confidence about the location of the true parameters;
(ii) where we have a small number of different combinations of parameters on which to base the design;
(iii) where we do not have much idea what the true parameters are.

In all cases, the comparison was of a new-style design against a standard orthogonal design. In case (i), we found that our new design was often likely to produce higher t values for the parameters, but with a small chance that it would produce substantially lower t values. However, the new design was almost always able to obtain a higher t value for the ratio of parameters. In case (ii), the new design provided worthwhile improvements in the t values of both the parameters and the parameter ratio. In case (iii), it was clear that if the true parameters were much larger than the design parameters, the orthogonal design produced more accurate estimates of the t value of the ratio of parameters for many parameter combinations, but where the new design was better it was often much better. Similarly for the parameters themselves, although at parameter combinations well away from the design point, the new design was able to out-perform the orthogonal design. In general, it seems that if there is some uncertainty, it is better to design an experiment with parameters which are rather larger than those it is hoped to recover. This is because the new-style designs typically present respondents with more extreme choices than orthogonal designs and, as the parameters get smaller, so pushing the choice probabilities towards 0.5, the divergence from magic P's is less with new-style designs.

We also extended the approach to consider four variable cases. Once again, we have magic P's and a limiting magic t. This is an important result. While it is possible to represent a three variable case on a boundary ray diagram, it is not so easy to represent designs with more independent variables, and it is extremely unclear as to what the optimisation rule should be. Our approach readily generalises to any number of independent variables; indeed, since the first-order condition is always that the utility difference between the two alternatives should be +/-2.399, then the more variables the less extreme any one variable has to be to fit in with this condition. If it is desired that "unrealistic" x's be avoided, it is perfectly possible to constrain the design to fit in with reasonable x's. Magic P's still result, though there may be some redundancy with duplication of scenarios.

One area which remains unresolved is how to design experiments where the utility function includes a constant. Here, it seems that different rules apply, and while we can produce designs which are an improvement on orthogonal designs, they do not provide magic P's.

AN APPLICATION OF THE NEW APPROACH

In the Summer of 1996, the opportunity arose to test the theoretical advances in SP design set out above in an application to car parking choices amongst members of staff at the University of Leeds. The SP exercises contained 9 pairwise comparisons of two parking options characterised by the annual cost of a parking permit (C), search time for a parking space (S) and daily walking time (W). Two designs were created, based on orthogonality (the "standard" design) and the new procedure. These designs are given in Table 2, where the variables are presented as the differences in the attributes between the two scenarios. Also reported are the boundary values of walk time on the assumption that search time is valued twice as highly as walking time.

In order to produce designs, we looked at previous evidence and decided to work with a value of walk time of about £5 a year to save 1 minute a day and a value of search time of twice that. Given this, the boundary values of the orthogonal design, with four values below the target and five above, seem broadly acceptable, although they do not fence in the target very closely. This is the sort of spread one would use if it was thought that the true value was probably about £5 but that it could range from £0 to £10. We make no claim that this is the best orthogonal design which could be obtained, it is just one which looks as though it can do the job. The new design in contrast has only two boundary value in excess of 5. Thus the conventional wisdom would be that the orthogonal design should be better at producing values if 0 to 10 is about right as the range of values of walk time, since the standard design has only three values outside this range whereas the new design has five values outside this range.

In this three variable case, the necessary condition to optimise any particular parameter is:

$$\sum_i \beta_i x_i = \pm 2.399 \qquad (15)$$

The new design used here was obtained after optimisation on the t statistics of β_{cost}, β_{search}, and β_{walk} in turn and then using the design produced. The starting point for the procedure was the standard design actually used. In fact, a further pass through the routine was obtained; this generated improvements in the non-optimised t statistics but yielded x's which were implausible in the application context. We thus used our expert judgment to decide which new design to use.

A crucial factor in the new design is the assumed values of the coefficients, the so-called design points. Taking into account other empirical evidence, the design points for the search time and walking time (both in minutes per working day) and cost (per year in pounds) coefficients were -0.5, -0.25 and -0.05. Converting costs into pence per day implies a value of search time of 4.54 pence per minute and a value of walk time of 2.27 pence per minute, which seem plausible.

For both the standard and new designs, Table 3 presents the theoretical per person t ratios for each of the coefficients and the values of search time and walk time derived as ratios. It can be seen that the new design offers the potential to achieve quite substantial improvements in the precision of both the coefficient estimates and the value of time estimates. This is so even though the correlations between cost and search, between cost and walk and between search and walk are -0.26, -0.57 and -0.32 respectively. The improvement in the precision of the parameters is broadly in line with a four-fold increase in sample size.

Table 2: Parking Choice SP Designs

Standard				New			
Difference (scenario A – scenario B) in:			Boundary value of walk time	Difference (scenario A – scenario B) in:			Boundary value of walk time
Cost	Search	Walk		Cost	Search	Walk	
-15	5	-10	∞	-3	11	-11	0.3
-50	3	0	8.3	-101	4	21	3.5
0	3	-10	0	12	4	-20	1.0
-50	5	-5	10	-100	9	11	3.4
-15	3	-5	15	-42	7	-15	-42
0	1	-5	0	-15	5	-16	-2.5
-15	1	0	7.5	-43	-1	2	∞
-50	1	-10	-6.25	-98	14	-19	10.9
0	5	0	0	17	6	7	-3.4

Since we stopped the optimisation after maximising t_{walk} (magic t), this is the best improvement in t_{walk} which can be obtained given the design points and the starting design. Further improvements in another t statistic could have been achieved at the expense of t_{walk} and with the possibility of unacceptably extreme x's.

As we move away from the design points, that is parameters on search, walk and cost of −0.5, −0.25 and −0.05, the advantage of the new design over the standard design may be reduced and it is possible that the parameter estimates recovered would be such that the standard design performs better than the new design. Table 4 shows the sensitivity of the improvement in t ratios to the design parameters.

Table 3: Theoretical per person t statistics

Parameter	Standard	New	Improvement
Search	0.77	1.54	100%
Walk	0.96	1.99	106%
Cost	0.95	1.40	48%
Value of Search	1.24	2.54	105%
Value of Walk	0.98	1.99	103%

Table 4: Improvement in t statistics at different design points for the new design over the standard design

Parameter	ß = (−0.5, −0.25, −0.05)	ß = (−0.25, −0.125, −0.025)	ß = (−1.0, −0.5, −0.1)
Search	100%	76%	83%
Walk	106%	106%	67%
Cost	48%	48%	19%
Value of Search	105%	159%	24%
Value of Walk	103%	275%	−3%

If parameters recovered from the calibration process turned out to be half those on which the design was based, that is parameters on search, walk and cost respectively of -0.25, -0.125 and -0.025 then the new design still offers substantial improvements in the theoretical t ratios. Similarly if the calibrated parameters were twice those assumed for design purposes, although here the improvement is less extreme and, for value of walk, the standard design is theoretically better. The new design is, though, broadly robust to assumptions about the scale of the parameters.

Empirical Results

Table 5 presents the empirical findings for the 121 individuals who replied to the questionnaire which contained the standard design and the 152 who had completed the new SP design. The adjusted t statistics for the new design are the calibrated t statistics multiplied by the square root of the ratio of the sample sizes so as to correct for the advantage given to the new design by virtue of its larger sample size.

The standard design produces parameters similar to those assumed in the design process. The scale of the coefficients in the new design, however, is around a third of the scale of the standard design's results. The new design also obtains a worse fit. Both of these are consistent with more error in SP responses in the new design. This is perhaps surprising since, by definition of the design criteria, it should be easier to answer the questions in the new design since the alternatives are designed to be more different. Two possible explanations are given below.

Although the new design achieves some worthwhile improvements in t ratios after allowing for the different sample sizes in each model, the improvements are somewhat less than the theoretical maximum improvements listed in Table 3 for the design points. In part this will be because the design points were not correct, and this will always be the case to varying degrees, but it may be that a contributory factor is greater error in the responses to the new design. This might be because the new design offered values of the variables which were not 'rounded' other than to the nearest whole number and the effects of this warrants further research. Another possibility is that the new design offered values of the variables which were so extreme that people did not take the comparisons seriously.

Table 5: Results for Standard and New Designs

	Standard		New			
	Coeff	t ratio	Coeff	t ratio	Adj t	Improve
Search	-0.590	8.0	-0.177	11.1	9.91	24%
Walk	-0.278	9.9	-0.115	16.3	14.54	47%
Cost	-0.070	11.6	-0.026	13.5	12.04	4%
Value of Search	8.38	13.21	6.79	19.74	17.62	33%
Value of Walk	3.95	11.79	4.39	20.49	18.29	55%
Rho Squared	0.228		0.150			
Observations	1089		1368			

It should be pointed out that not all standard designs would give the same t-ratios. Given the chosen levels of the attributes, six possible orthogonal designs can be produced and thus, for an assumed set of parameters, there would be six possible sets of t-ratios. The standard design actually used (given the assumed parameters) is the joint best-performing out of the six at producing a t-ratio for the value of walk time and the third-best at producing a t-ratio for the value of search time. Overall, the standard design is about third best out of six for t-ratios on values. It is relatively poor at producing t-ratios on parameters, but it is worth noting that what looks like the best standard design for producing t-ratios on parameters (best on one parameter and a very close second best on two) is indisputably the worst standard design at producing t-ratios for values. Alternative orthogonal designs using different attribute levels would, of course, yield different results.

We conclude that the greater precision in estimation of both parameters and monetary valuations obtained by the new design method demonstrates the potential worth of our procedure, and provides encouragement to pursue this line of approach. We acknowledge especially the need for further empirical testing.

CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

With regard to the design of SP choice experiments, three important findings have emerged from the theoretical developments:

- Marginal (close to 50/50) choices are not particularly desirable for efficiency.
- Fractional factorial orthogonal plans do not necessarily provide coefficient estimates in disaggregate logit models with least variance.
- The design principles outlined in this paper readily generalise to any number of independent variables; this is an advantage over the boundary ray approach, which can be misleading if used in isolation.

We have here demonstrated that appreciable improvements in the precision of parameter and value of time estimates can be obtained by using a design procedure based on selecting the independent variables so as to minimise the variances of coefficient estimates. We have also shown that significant efficiency gains may be achieved in practice. This is not to deny the role of expert judgment in choosing the design points and ensuring contextually realistic values for the x's.

We consider that these are significant advances in SP design but recognise that much more research is still needed. Further research should examine:

- Whether the new design attracts more error in SP responses as seemed to be the case in our empirical example.
- How to incorporate constants into the model.
- How the approach is generalised to the case of more than two alternatives.
- How non-linear and interaction terms might best be handled.
- Issues relating to the design points and the relative efficiency of new and standard designs in the presence of uncertainty as to the parameter values.
- Appropriate optimality criteria for the new procedure.

It is also clear that we need to test, refine and extend the design principles we have developed and to demonstrate their practical applicability. These unresolved issues notwithstanding, we are satisfied that our new design procedure can offer significant theoretical efficiency gains in parameter estimation compared with a standard orthogonal design and that some of these gains can be realised in practice.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council, who funded this work as grant GR/J81242 between 1994 and 1996.

REFERENCES

Bates, J.J. (1994) "Reflections on Stated Preference Theory and Practice" **Paper presented at 7th International Conference on Travel Behaviour Research, Chile.**

Clark, S.D. and Toner, J.P. (1996), "Application of Advanced Stated Preference Design methodology", **Working Paper 485**, Institute for Transport Studies, University of Leeds

Fowkes, A.S. (1991) "Recent Developments in Stated Preference Techniques in Transport Research", **proc PTRC Summer Annual Meeting, Seminar D** pp251-263

Fowkes, A.S. and Nash, C.A. (eds) (1991) **Analysing the Demand for Rail Travel**, Avebury, Aldershot

Fowkes, AS and Wardman, M (1988) "Design of Stated Preference Travel Choice Experiments, with Special Reference to Taste Variations", **Journal of Transport Economics and Policy, 22(1)**, pp27-44

Fowkes, A.S., Wardman, M. and Holden, D.G.P. (1993) "Non-orthogonal Stated Preference Design", **proc PTRC Summer Annual Meeting, Seminar D** pp91-97

Gunn, H.F. (1981) "Value of Time Estimation", **Working Paper 157**, Institute for Transport Studies, University of Leeds.

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

Holden, D.G.P. (1993) **Design Procedures for Stated Preference Experiments** PhD Thesis, School of Business and Economic Studies, University of Leeds.

Holden, D.G.P., Fowkes, A.S. and Wardman, M. (1992) "Automatic Stated Preference design algorithms", **proc PTRC Summer Annual Meeting, Seminar D** pp153-166

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

Swanson, J., Pearmain, D. and Holden, D.G.P. (1993) "New approaches to the design of Stated Preference experiments", **proc PTRC Summer Annual Meeting, Seminar D** pp99-109

Wardman, M. and Toner, J.P. (1996), "Issues in Model Specification", invited paper **presented at PTRC International Conference on Value of Time Research, October 1996**, Wokingham

Watson, S.M., Toner, J.P., Fowkes, A.S. and Wardman, M. (1996), "Efficiency Properties of Orthogonal Stated Preference Designs" **proc PTRC Summer Annual Meeting, Seminar D**, pp1-10