EFFECTS OF STATED CHOICE DESIGN DIMENSIONS ON ESTIMATES[1]

Phani Kumar Chintakayala
Accent & ITS, University of Leeds
v.p.k.chintakayala@its.leeds.ac.uk

Stephane Hess
ITS, University of Leeds
s.hess@its.leeds.ac.uk

John Rose
ITLS, University of Sydney
johnr@itls.usyd.edu.au

Mark Wardman
ITS, University of Leeds
m.r.wardman@its.leeds.ac.uk

ABSTRACT

There have always been concerns about task complexity and respondent burden in the context of stated choice (SC) studies, with calls to limit the number of alternatives, attributes and choice sets. At the same time, some researchers have also made the case that too simplistic a design might be counterproductive given that such designs may result in issues of omitting important decision variables. This paper aims to take another look at the effects of design complexity on model results. Specifically, we make use of an approach devised by Hensher (2004)[2] in which different respondents in the study are presented with designs of different complexity, and look specifically at effects on model scale in a UK context, adding to previous Chilean evidence by Caussade et al. (2005). Our results indicate that the impact of design complexity may be somewhat lower than anticipated, and that more complex designs may not necessarily lead to poorer results. In fact, some of the more complex designs lead to higher scale in the models. Overall, our findings suggest that respondents can cope adequately with high numbers of attributes, alternatives and choice sets. The implications for practical research are potentially significant, given the widespread use, especially in Europe, of stated choice designs with a limited number of alternatives and attributes.

INTRODUCTION

One of the main advantages of Stated Choice (SC) methods over the cruder Stated Intentions (SI) and Willingness to Pay (WTP) approaches is that the range of variables included makes the true purpose of the study less apparent, thereby potentially reducing the risk of response bias. From this perspective, it could be argued that,

[1] This paper builds on earlier work by the authors presented at the 2008 European Transport Conference, but makes use of dataset from which a number of coding errors have been removed, significantly changing the results.
[2] See also Hensher (2006a, 2006b).

even if the analyst is only interested in, say, the value of time, and therefore only needs respondents to trade time against cost, there might be reasons for offering a broader range of attributes to conceal the purpose of the study. Indeed, this argument was commonly used in the early days of SC applications by advocates of the method to counter the scepticism of some transport planners towards hypothetical questioning methods in general and the SI approach in particular with its reputation for yielding exaggerated behavioural responses.

Moreover, since real world decision making rarely involves choices between alternatives with only a limited number of attributes, it was argued both on the grounds of realism and for suitability for real-world forecasting that the set of key variables in an SC choice exercise should not be artificially restricted.

SC methods as first employed in Great Britain were *imported* from marketing research applications in the United States (Sheldon and Steer, 1982). The view in that field at that time was that a large number of attributes and/or alternatives did not have an adverse impact on model validity (Green and Wind, 1973; Malhotra, 1982; McCullough and Best, 1979; Scott and Wright, 1976). More recently, the view that simplicity is best is regarded as 'urban myth' by some and, according to Louviere (2001), there is no evidence to suggest that simpler SC exercises produce different valuations due to reduced task complexity.

Despite these various considerations, especially in an applied context, and even more so in the UK and Europe by extension, there is often a prevailing opinion that SC exercises must be kept relatively straightforward if reliable responses are to be obtained. This view has resulted in a trend towards simpler SC exercises with fewer variables. Hence SC exercises with two alternatives and three or four attributes are common, with many examples of the most simplistic two-alternative two attribute-designs. To some extent, this is based on evidence in the transport context that indicates task complexity can influence valuations through simplifying choice rules, ignoring attributes or inability to process the information correctly (Timmermans, 1993; Widlert, 1998; Arentze et al., 2003). There are a number of other explanations of the trend to and preference for simpler designs.

Work closely associated with Fowkes (e.g., Fowkes et al., 1993) has involved the use of 'boundary rays' which graphically represent the trade-off between attributes in terms of their implied valuations. This is largely a development of the 'Beesleygraph' concept of the central importance of sensible trade-offs between attributes. This has intuitive appeal, although not necessarily statistical support. This method was fairly widely used in consultancy studies in Great Britain in the 1990s. The need to represent the problem in two-dimensional space tended to limit the SC design to just three attributes.

Another stimulus to simple designs, recognised by Kroes and Sheldon (1988), is that self completion questionnaires require more straightforward SC exercise than interviewer led or computer based exercises. The cost effectiveness of this means of data collection means that in suitable contexts it is very attractive, thereby increasing the reliance on simple exercises.

Recently, analysts have moved away from SC exercises based around real-world choice contexts when the purpose of the study is valuation. Instead, choices are offered amongst unlabelled alternatives (A and B), whereupon there can be no confounding effects from a real world choice context. This loosens the restriction to offer the principal variables influencing choice with instead a focus on just those main attributes of interest.

There has been a tendency in Great Britain, which has been reflected in mainland Europe, towards simpler exercises. Not only is this our impression from familiarity with a great many studies, but it can be at least partially evidenced. For example, in a review of a large amount of British empirical evidence, Wardman (1998) quantifies trends towards simpler forms of SC exercise, with choice experiments replacing ranking exercises and the number of choice scenarios evaluated reducing over time.

As far as what are often termed national value of time studies are concerned, two have been conducted in Britain. The first of these consisted of five SC exercises (MVA et al., 1987). For urban bus and inter-urban car users (both unlabelled SC exercises), only three variables were offered. This was increased to four in the case of urban car (route choice) and inter-urban public transport (unlabelled). The largest number was offered in a mode choice exercise involving rail and coach for commuting journeys into London (main in-vehicle time, other in-vehicle time, walk time, wait time and fare).

The second British study (Hague Consulting Group et al., 1999) used even simpler designs. The main design had only two attributes, time and cost, and is similar in nature to designs used in major national studies in the Netherlands (Gunn et al., 1999), Denmark (Fosgerau et al., 2007) and Switzerland (Algers et al., 1995) but not, as far as we are aware, in studies outside Europe. However, it is probably fair to state that the reasoning here was not simplicity but to offer a specific form of SC design with desired and clear trade-offs between time and cost. Four other SC exercises were covered in the study, one of which also had two attributes, two of which had three attributes and the remaining one had four attributes.

From the data set used in meta-analysis of British empirical evidence relating to valuations of time and service quality covering the period 1963 to 2000 (Wardman, 2004), 155 studies used SC methods. Of these, 6% contained only two variables, 19% contained three variables, 59% contained four variables, 15% contained five variables and 1% contained six variables. Whilst designs with four attributes are most common, and in some cases are sufficient to address the key variables in the choice context of interest, there remains a widespread concern to limit the number of attributes contained in an SC exercise to avoid over-burdening respondents. In some cases, however, this may artificially limit the variables that are covered, ignore what for some are important influences on choice and generally reduce the value of the results obtained.

Current viewpoints are set against a background of little direct empirical evidence that demonstrates the effect of the dimensions of the SC designs on the results obtained in the British context. We here report on a repeat of the survey that has been previously applied in Australia, Chile and Taiwan (Hensher, 2004; Caussade et al., 2005; Rose et al., 2008). The purpose here is to contribute to the evidence base as it relates to Britain rather than to undertake a comparative analysis.

DATA

Description of Survey

For readers unfamiliar with the original work (Hensher, 2004), we here outline the approach to survey design.

The survey is aimed at understanding the influence of design dimensionality on behavioural response. The SC experiment has 16 designs embedded in one master design. The different designs have different numbers of alternatives, choice sets and attributes, as well as varying numbers of attribute levels and attribute level ranges (i.e., narrow range, medium range and wide range). The design characteristics are shown in Table 1. Prior to undertaking the SC experiment, respondents are first asked to provide information on a recent trip that they have undertaken. The SC alternatives are then constructed with the values of the attributes constructed as percentage variations around the values for the respondent's reported reference trip. Different percentages were used depending upon whether the respondent was assigned to a narrow, middle or wide range design. Narrow range designs varied the percentages between +/- 5% of the respondent's reported attribute values, medium range designs varied the percentages by up to +/- 20% whilst wide range designs varied the percentages by up to +/- 60%.

Each of the designs is computer-generated, with the objective in constructing them being to minimise the variance-covariance matrix of the parameters obtained from models estimated on data collected based on the survey instrument (Rose et al., 2008). Six attributes have been selected based on earlier studies (Hensher 2000, 2001). These are: a- free flow time (FFT), b- slowed down time (SDT), c- stop/start time (SST), d- Uncertainty Time (UT) i.e. the buffer time/extra time people need to cover any uncertainties like accidents, road works etc. so as to reach their destinations in time, e- toll cost (TLC), and f- running cost (RC). Given that the 'number of attributes' dimension has four levels, ranging from three to six, the following combinations of the six attributes were selected.

3: (a+b+c)[3], d, (e+f)[4]

4: a,(b+c)[5], d, (e+f)

5: a,b,c,d, (e+f)

6: a,b,c,d,e,f

Take in Table 1

[3] Referred to as Total Time
[4] Referred to as Total costs
[5] Referred to as combined Slowed down and Stop Start (SSST)

Attempts were made to prepare level balanced designs. For attributes with odd numbers of levels, the centre pivot level is 0% with levels evenly spaced +/- either side with a maximum range of +/- 20% change from the reference alternative. For attributes with even numbers of levels, similar ranges are used but the 0% level is not used. Balancing the levels in this way avoids potential biases that can arise when unbalanced designs entice respondents to focus too much on those attribute levels that are presented more frequently.

The elements of the design plan are manipulated according to a master plan. As is apparent from Table 1, the master plan has 16 runs, that is, 16 different designs are constructed to test the impact of these five design elements. In addition to the linear and quadratic effects for the five dimensions, some interactions between those elements can also be estimated. The master plan design allows for the interaction between the number of choice sets and number of alternatives as well as between the number of alternatives and number of attributes.

The DoD SC experiment is 16 Designs embedded in one design. That is, there are 16 designs in the background, each with two versions. Since these designs do not have the same number of alternatives, choice sets and attributes, and since they do not refer to the same number of attribute levels, neither do they refer to the same levels (narrow range, medium range and wide range), all this is made interactive.

Given the nature of the overall design process, not all of the 32 designs used display attribute level balance, however where balance is not possible, attempts were made to make the designs as balanced as possible. Where attribute level balance was not possible, typically the midpoint levels are those displayed the least.

Respondents are randomly allocated to one of the 16 sub designs. As such, the levels applied to the choice tasks given to individual respondents differ depending on the range of attribute levels as well as on the number of levels for each attribute based on the design that they have been allocated to. Once a respondent is allocated to a design, the number of attributes, alternatives and attribute level range, are kept constant for that respondent. Variations in the design dimensions thus occur not within respondent but rather between respondents. An example SC screen is illustrated in Figure 1.

Respondents are initially asked to choose among alternatives that include their current trip as one of the alternatives. If one chooses the current trip, then he/she is forced to make a secondary choice among the non-base alternatives available in that choice task. In the present paper, we focus on an analysis of the primary choices.

Take in Figure 1

Description of Sample

The survey was conducted in the UK by Accent Marketing Research using Computer Aided Personal Interview (CAPI) technology. Face to face interviews with 300 commuters were undertaken in Bristol, Edinburgh and London in early May 2008.

Fourteen of the 300 respondents were removed due to errors made during initial data entry. Stratified random sampling was employed, with screening questions used to establish eligibility in respect to commuting by car. Three trip length quotas were imposed: less than 30 minutes, 30-60 minutes and more than 60 minutes. Sample characteristics are given in Table 2. Given the emphasis on commuting, it is not surprising that the vast majority are in full time employment whilst this could contribute to a greater proportion of males in the sample along with higher car ownership levels amongst male commuters. The distributions of travel times to work and age also seem reasonable.

Take in Table 2

Summary information on the distribution of the sample across choice sets, alternatives, attributes, levels and range is provided in Table 3. A good spread across the design dimensions has been achieved.

Take in Table 3

BASE MODEL

The results of our base model are shown in Table 4. This model makes use of a linear-in-attributes specification of a Multinomial Logit (MNL) model, with coefficients estimated for all attributes, along with alternative specific constants (ASC) for the first four alternatives.

Take in Table 4

A jackknife procedure was used to allow for the repeat observations nature of the data. As is typically the case, this procedure had very little impact on the coefficient estimates themselves but there are some noticeable reductions in the precision of these estimates to the extent that five coefficients are not significant at the 5% level. Two of these are ASCs and we would not necessarily expect these to be significant, representing as they do an ordering effect. Nonetheless, a positive and highly significant ASC was detected for the first alternative, reflecting an inertia towards the current alternative.

Comparison across coefficients at this stage is hampered by possible differences in scale across the different designs. For example, total cost and total time appear together but are not necessarily comparable with the coefficient estimates for their constituent parts that enter different designs. Nonetheless, it is encouraging that there is a monotonic relationship of the expected form between free flow time, slowed down time and stop start time, although it is normally found that toll cost is assigned a higher weight than running cost and this is not the case here. Table 5 shows the *t*-ratios for the time parameter differences based on the model presented in Table 4. It can be seen that free flow time is significantly different from stop start and combined slow and stop start time and not far from significantly different from slowed down time. However, none of the other differences are significant except for those relating to uncertain time.

Take in Table 5

The WTP values estimated using the running cost and total cost, and the ratios of the time parameters from the model in Table 4 are shown in Table 6. None of the valuations of time in terms of running cost are significant, and a key factor here is that the running cost coefficient is not significant. The value of time ranges from £4.28 per hour (7.13 pence per minute) for free flow time through to £10.15 per hour (16.91 pence per minute) for stop-start time. These numbers surround official Department for Transport values of, essentially an unspecified type of time, of around 10.5 pence per minute, for commuting.

Take in Table 6

ACCOUNTING FOR DESIGN DIMENSIONS

In this section, we investigate the impacts of design complexity by testing for differences in the relative weights of the observed and unobserved utility components as a function of the design. It is done by estimating scales for each dimension. When estimating models on data with different sources, it is important to allow for differences in scale across data sources. Indeed, the relative weight of the observed and unobserved parts of utility may vary significantly across datasets, for example as a result of more or less randomness in the choice processes. We can hypothesise that the amount of randomness in choices will be greater as task complexity increases.

The models estimated in this section all need to be compared to the base model from the previous section.

Allowing for differences in scale

One can use the Bradley-Daly (1992) nested logit model technique for estimating scale parameters. In this paper, however, we employ a more direct method of estimating scale parameters, using BIOGEME as explained below.

Let $U_{i,d}$ give the utility of alternative i with sample d, where d = 1, ... ,D. Then, we have:

$U_{i,1} = V_{i,1} + \varepsilon_{i,1}$

...

$U_{i,d} = V_{d,1} + \varepsilon_{i,d}$

...

$U_{i,D} = V_{i,D} + \varepsilon_{i,D}$ (1)

where var $(\varepsilon_{i,d}) = \pi^2/6\mu_d^2$

Using sample 1 as the arbitrary base, we then multiply the utility functions in group d by α_d, where we set $\alpha_1 = 1$. In detail, we then have:

$$
\begin{aligned}
&U_{i,1} = V_{i,1} + \varepsilon_{i,1} \\
&\ldots \\
&\alpha_d U_{i,d} = \alpha_d V_{i,d} + \alpha_d \varepsilon_{i,d} \\
&\ldots \\
&\alpha_D U_{i,D} = \alpha_D V_{i,D} + \alpha_D \varepsilon_{i,D} \qquad (2)
\end{aligned}
$$

For estimation as a homoscedastic model, we thus obtain that:

$$\text{var}(\varepsilon_{i,1}) = \alpha_2 \text{var}(\varepsilon_{i,d}) \qquad (3)$$

This gives us that:

$$\alpha^2_d = \text{var}(\varepsilon_{i,1}) / \text{var}(\varepsilon_{i,d}) \qquad (4)$$

which in turn reduces to:

$$\alpha_d = \mu_d / \mu_1 \qquad (5)$$

This means that if the estimate for α_d is larger than 1, then the variance of the unobserved utility components in sample d is smaller than in the base sample, with the converse applying if α_d is smaller than 1.

Variations in Scale with Design Dimension

We now turn away from accounting for differences in scales across all 16 designs, focussing instead on differences as a function of the number of choice sets, alternatives, attributes, levels of attributes and the range of levels (narrow or wide) on the choices. Here, it should be noted that this comparison is not perfect as other factors might come into it, although the master design should have largely dealt with this. Table 7 presents the estimates of scale parameters and Table 8 reports the estimates of the study parameters.

i. Model with scales for different designs: There is a significant improvement in the model fit when the designs are treated separately. Chi-squared test suggests that the model with scales for different designs is superior to the base model. All but three scale parameters are significantly different from and less than one indicating that the choices are less sensitive to the observed utilities when compared to the base design, which was one of the more complex designs, with 15 choice sets, 4 alternatives, 4 attributes, and 3 levels. However, no clear pattern in the scales as a function of design can be observed.

ii. Model with scales for number of choice sets: The scale parameter for the group that has data related to 6 choice sets is normalised to one. The model fit marginally improved from the base model. However, none of the scale parameters is significantly different from one, indicating that the number of choice sets seems to have little influence on the relative weight of the observed and unobserved utility components.

iii. Model with scales for number of alternatives: The scale parameter for the group that has data related to 3 alternatives is set to one. The model fit has marginally improved from the base model and a Chi-squared test suggests that this model is significantly better than the base model. Both remaining scale parameters are not significantly different from 1 indicating that the number of alternatives in this case has little or no impact on the observed utility components.

iv. Model with scales for number of attributes: The scale parameter for the group that has data related to 3 attributes is set to one. The model fit has marginally improved from the base model. Again, a Chi-squared test suggests that this model is significantly better than the base model. Only the scale parameter for the data with 6 attributes is significantly different from and less than 1, indicating more randomness for 6 attributes. This potentially suggests an upper limit to the acceptable number of attributes.

Take in Table 7

v. Model with scales for number of levels: The scale parameter for the group that has data related to 2 levels is set to one. The model fit has improved considerably from the base model. Of the remaining scale parameters, the scale parameter representing 3 levels is significantly different from and greater than 1, indicating more weight to the utilities, and the 4 levels scale being insignificant does not convey any message. There is no clear pattern to suggest.

vi. Model with scales for range: The scale parameter for the group that has data from designs using the medium range is set to one. The model fit has marginally improved from the base model. The remaining scale parameters are significantly different from and less than 1, suggesting that narrow and wide ranges are less desirable. Nonetheless, the magnitude of the effect is minor.

vii. Model with scales for choice set size (i.e. number of items in a set): The scale parameter for the group that has data from design using the minimum number of items (i.e. 9 items resulting from 3 alternatives times 3 attributes) is set to one. The model fit has improved significantly from the base model. Two scale parameters are significantly different from and less than 1, suggesting that the choices are less sensitive to the observed utilities when compared to the base. The remaining scale parameters are insignificant. No clear pattern could be observed.

The above analysis suggests that accounting for scale differences for various designs improves the model performance but that few design dimension effects are statistically significant and where they are the scale is not greatly different from one. That increases in design complexity do not necessarily lead to a reduction in scale dispels some prevailing concerns.

Take in Table 8

Variations in WTP values

We now focus on the effect of accounting for scale difference for different design dimensions on the WTP values. Table 9 presents the WTP values estimated from different models that account for different design dimensions.

Take in Table 9

There seem to be not much variation in the WTP values for the parameters across models with exceptions of models that account for different designs, attributes and set size. Free flow time is valued lowest as expected while the parameter combined slow and stop start time is valued highest in models that account for scale differences for variation in the number of choice sets, alternatives, levels and width of levels, it is valued second highest to stop start time in models that account for scale differences for variation in the number of designs, attributes and size of set.

SEPARATE MODELS BY GROUP

Having established that design dimension has little bearing on the scale of a model, we turn to its impact on model fit and WTP valuations through the estimation of separate models for different subgroups of the data, i.e., data collected from surveys with 3 alternatives, 4 alternatives, etc. As a first result, the adjusted ρ^2 measure for the various models is given in Figure 2

Take in Figure 2

The results show that the number of alternatives has little or no effect on model fit. The findings on the number of attributes and levels are consistent with those from the scale analysis, with 6 attributes and uneven numbers of levels leading to lower model fit. However, unlike the results showing lower scale with narrower or wider ranges, we here observe higher model fit. This could be an indication that with the narrower or wider ranges, the choice becomes more deterministic without however increasing the weight for the attributes, for example as a result of explaining more behavior through the constants.

Table 10 shows a summary of WTP measures from the different models estimated in this section. Here, we can see that while there is consistency across surveys for some of the indicators, there are also some significant differences.

Indeed, the results show that the WTP for free flow time is between £5/hr and £9.3/hr, slowed down time is between £6.5/hr and £8.2/hr, stop start time is between £6.2/hr and £13.6/hr, combined slow stop start time is between £3.8/hr and £12.9/hr with £46.2/hr for 5 alternatives data set, uncertainty time is between £2.6/hr and £5.1/hr and Travel time WTP is between £2.8/hr and £16.2/hr with £22.9/hr value for 9 choice sets data set.

To some extent, these differences are also a result of small sample sizes in some groups and lower parameter significance, where with hindsight this clearly potentially also impacts our earlier findings in terms of scale. In any case, while the results show some impact by the design on the WTP indicators, there is no clear pattern suggesting that more complex designs yield less reliable results.

Take in Table 10

CONCLUSIONS

With a view to demonstrate the effect of the dimensions of the SC designs on the results obtained in the British context, we have repeated the survey that has been previously applied in Australia, Chile and Taiwan. We have attempted to capture the effect of the design dimensions by allowing for differences in the scales that represent variations in the dimensions.

The model with scale differences for designs showed no clear indication to suggest that more complex designs are difficult to cope with. The model with scale differences for number of choice sets indicates that the number of choice sets seems to have little influence on the relative weight of the observed and unobserved utility components. The model with scale differences for number of alternatives indicates that the number of alternatives seems to have little or no impact on the relative weight of the observed and unobserved utility components. The model with scale differences for number of attributes provides evidence that the randomness is more when the number of attributes is 6 than when the number of attributes is less than 6. However, there is no clear pattern. The model with scale differences for number of levels indicates that when the number of levels is 3 choices are more sensitive to observed utilities. Again, no clear pattern could be observed. The model with scale differences for level range shows significant effects from wide and narrow range, but the size of the effects is minor.

Overall, our study shows that there is little or no impact on scales when design dimension are accounted for. Our results also suggest that people are not having problems with more complex designs.

The estimates for the study parameters indicate that all the parameters are having signs as expected and the alternative specific constant for the base alternative (ASC1a) suggests that there is a strong inertia towards the base alternative. As for the WTP values, the pattern shows that the Free flow time is valued lowest as expected while the parameter combined slow and stop start time is valued highest in models that account for scale differences for variation in the number of choice sets, alternatives, levels and width of levels, it is valued second highest to stop start time in models that account for scale differences for variation in the number of designs, attributes and size of set.

The estimation of separate models for different subgroups of the data suggests the number of alternatives has little or no effect on model fit. The findings on the number of attributes and levels are consistent with those from the scale analysis, with 6 attributes and uneven numbers of levels leading to lower model fit. However, unlike the results showing lower scale with narrower or wider ranges, we here observe higher model fit.

While our results show some significant differences across designs in terms of model scale and substantive results such as WTP indicators, there is little or no suggestion that the results from the more complex designs are less reliable than those from the more simplistic designs. In fact, especially when looking at the results for scale, the opposite is regularly the case.

ACKNOWLEDGMENTS

The authors wish to thank Accent Marketing & Research Ltd. for funding the data collection in the UK. The second author thanks the Leverhulme Trust for the financial support in the form of a Leverhulme Early Career Fellowship.

REFERENCES

Algers, S., Lindqvist-Dillén, J. and Widlert, S. (1995), "The National Swedish Value of Time Study", Proceedings of Seminar F, PTRC, 1995.

Arentze, T., Borgers, A., Timmermans, H. and DelMistro, R. (2003) "Transport Stated Choice Responses: Effects of Task Complexity", Presentation Format and Literacy. Transportation Research 39E (3), pp. 229-244.

Bradley, M.A. and Daly A.J. (1992). "Uses of the logit scaling approach in stated preference analysis", paper presented at the 7th World Conference on Transport Research, Lyon, July.

Caussade, S., Ortúzar, J. de D., Rizzi, L.I. and Hensher, D.A. (2005) "Assessing the influence of design dimensions on stated choice experiment estimates". Transportation Research B, 39, pp. 621-640.

Fowkes, A.S, Wardman, M. and Holden, D. (1993) "Non-Orthogonal Stated Preference Design", PTRC Summer Annual Meeting, Manchester.

Green, P.E. and Wind, Y. (1973) "Multiattribute Decisions in Marketing: A Measurement Approach", The Dryden Press, Hinsdale, Illinois.

Gunn, H.F., Tuinenga, J.G., Cheung, Y.H.F. and Kleijn, H.J. (1999) "Value of Dutch Travel Time Savings in 1997", Proceedings of the 8th World Conference on Transport Research, Transport Modelling/Assessment, Vol. 3 pp.513-526. Edited by Meersman, H., Van de Voorde, E. and Winkelmans, W., Pergamon, Amsterdam.

Hague Consulting Group, Accent Marketing and Research, Department for Transport (1999) "The Value of Travel Time on UK Roads", The Hague.

Hensher, D.A. (2004) "Accounting for stated choice design dimensionality in willingness to pay for travel time savings", Journal of Transport Economics and Policy, 38, pp. 425-446.

Hensher, D.A. (2006a) "Revealing differences in behavioural response due to the dimensionality of stated choice designs: an initial assessment", Environmental and Resource Economics, 34, pp. 7-44.

Hensher, D.A. (2006b) "How do respondents process stated choice experiments? attribute consideration under varying information load", Journal of Applied Econometrics, 21, pp. 861-878

Kroes, E. and Sheldon, R. (1988) "Stated Preference Methods: An Introduction", Journal of Transport Economics and Policy Vol. 22 No. 1, pp. 11-26

Louviere, J.J. (2001) "Choice Experiments: An Overview of Concepts and Issues", In Bennett, J. and Blamey, R. (Eds) The Choice Modelling Approach to Environmental Valuation. Edward Elgar, Cheltenham.

Malhotra, N.K. (1982) "Structural Reliability and Stability of Nonmetric Conjoint Analysis", Journal of Marketing Research, 19, pp. 199-207.

McCullough, J. and Best, R. (1979) "Conjoint Measurement: Temporal Stability and Structural Reliability", Journal of Marketing Research, 18, pp. 80-86.

MVA Consultancy, ITS University of Leeds, TSU Oxford University (1987) "Value of Travel Time Savings", Policy Journals. Newbury, Berkshire.

Rose, J.M, Bliemer, M.C.J, Hensher, D.A, and Collins, T. A. (2008) "Designing efficient stated choice experiments in the presence of reference alternatives". Transportation Research Part B: Vol. 42 No. 4, pp. 395-406

Rose, J.M., Hensher, D.A., Caussade, S., Ortuzar, J. de D., and Jou, R.-C. (2008), "Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis", Journal of Transport Geography, Vol. 17 Issue 1, pp. 21-29.

Scott, J.E. and Wright, P. (1976) "Modelling an Organisational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations", Journal of Marketing Research 13, pp. 211-24.

Sheldon, R. and Steer, J. (1982) "The Use of Conjoint Analysis in Transport Research", Paper presented at the PTRC Summer Annual Meeting, PTRC, London

Wardman, M. (1998) "The Value of Travel Time: A Review of British Evidence", Journal of Transport Economics and Policy, 32(3), pp. 285-315.

Wardman, M. (2004) "Public Transport Values of Time", Transport Policy, 11, pp. 363-377.

Widlert, S. (1998) "Stated Preference Studies: The Design Affects the Results", in Ortuzar, J. de D. Hensher and S. Jara-Diaz (editors), Travel Behavior Research: Updating the state of play, chapter 7, pp. 105-123, Pergamon, UK.

Fosgerau M, Hjorth K, Lyk-Jensen V. S. (2007) "The Danish Value of Time Study: Final Report 2007".
http://www.transport.dtu.dk/Forskning/Publikationer/Publikationer%20DTF/2007.aspx

	Details of your recent trip	Alternative Road A	Alternative Road B	Alternative Road C	Alternative Road D
Time in free flow (mins)	15	12	14	16	18
Time slowed down by other traffic (mins)	10	8	12	14	6
Time in stop/start conditions (mins)	6	4	7	5	8
Uncertainty in travel time (mins)	10	12	14	8	6
Running costs	£2.00	£2.80	£1.60	£1.20	£2.40
Toll costs	£2.00	£2.20	£1.80	£2.40	£1.60
If you make the same trip again, which road would you choose?	Current Road	Route A	Route B	Route C	Route D
If you could only choose between the new roads, which would you choose?		Route A	Route B	Route C	Route D

Figure1

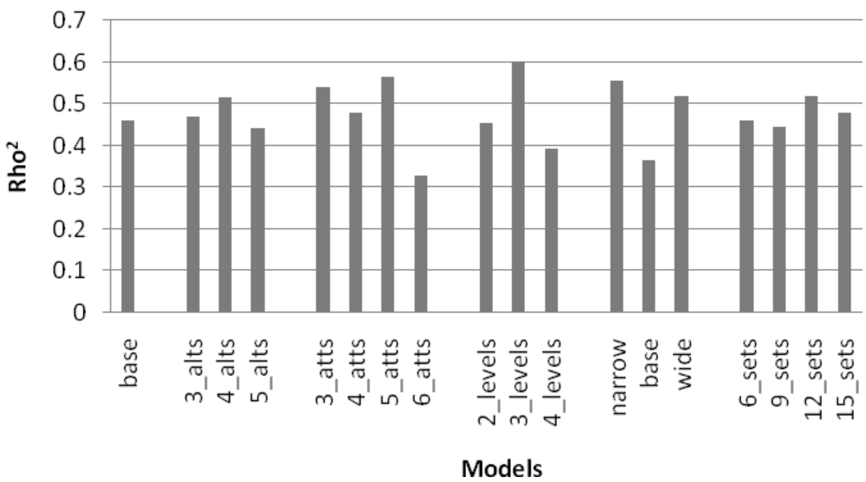

Figure 2

Figure Legend:

1. Figure 1. Sample SC screen
2. Figure 2. Adjusted ρ^2 values

Tables

Table 1. Design Characteristics

Design	No. of Choice Sets	No. of Alternatives	No. of Attributes	No. of Levels	Range of Attribute levels
1	15	4	4	3	medium
2	12	4	4	4	wide
3	15	3	5	2	wide
4	9	3	5	4	medium
5	6	3	3	3	wide
6	15	3	3	4	narrow
7	6	4	6	2	narrow
8	9	5	3	4	wide
9	15	5	6	4	medium
10	6	5	6	3	wide
11	6	4	5	4	narrow
12	9	5	4	2	narrow
13	12	4	6	2	medium
14	12	3	3	3	narrow
15	9	3	4	2	medium
16	12	5	5	3	narrow

Table 2. Sample characteristics

Characteristic	Description	Number
Gender		
	Male	187 (62%)
	Female	113 (38%)
Age		
	<35yrs	126 (42%)
	36-55yrs	130 (43%)
	>56yrs	44 (15%)
Employment		
	Full time	255 (85%)
	Part time	35 (12%)
	Casual	7 (2%)
	No (in 6 months)	3 (1%)
Trip Length (min)	<30	180 (60%)
	31-60	96 (32%)
	>60	24 (8%)

Table 3. Summary of observations information

		Number of people	Number of Obs
Choice sets	6	79	474
	9	70	630
	12	76	912
	15	75	1125
Alternatives	3	101	1122
	4	92	933
	5	107	1086
Attributes	3	63	669
	4	77	846
	5	72	732
	6	88	894
Levels	2	95	969
	3	96	963
	4	109	1200
Range	Narrow	112	1095
	Medium	99	1218
	Wide	89	828

Table 4. Results for base model

Parameter	Coefficient	*t*-stat	Coefficient	*t*-stat
	Before Jackknife		*After Jackknife*	
Free Flow Time	-0.036	2.26	-0.036	1.70
Stop Start Time	-0.084	6.42	-0.085	3.06
Combined Slow and Stop start Time	-0.096	5.24	-0.096	2.52
Slowed down time	-0.066	4.64	-0.066	2.87
Uncertain Time	-0.039	7.14	-0.039	2.93
Running Cost	-0.499	4.49	-0.500	1.90
Toll Cost	-0.416	3.36	-0.416	1.16
Total Cost	-0.961	8.31	-0.962	3.35
Total Time	-0.172	12.32	-0.172	6.90
ASC1	1.560	13.96	1.565	5.51
ASC2	0.179	1.94	0.179	2.15
ASC3	0.147	1.36	0.147	0.80
ASC4	-0.259	1.69	-0.261	1.38
Model Fits				
Observations	3,018			
Log Likelihood	-2,221.44			
Adjusted ρ^2	0.456			

Table 5. *t*-stats for parameter differences

Parameter 1	Parameter 2	*t*-stat
Free Flow Time	Uncertain Time	0.22
Free Flow Time	Slow Down Time	1.50
Free Flow Time	Stop Start Time	2.41
Free Flow Time	Slow SST	2.40
Slow SST	Slowed down time	1.26
Slow SST	Uncertain Time	2.93
Slow Down Time	Uncertain Time	1.78
Stop Start Time	Slow SST	0.49
Stop Start Time	Slow Down Time	0.95
Stop Start Time	Uncertain Time	3.28

Table 6. WTP estimates for base model

	After JackKnifing			
Parameter	Value (£/hr)	*t*-stat	Conf.int.	
Free Flow Time	4.35	1.54	0	10.78
Stop Start Time	10.40	1.51	0	22.26
Combined Slow and Stop start time	11.80	1.47	0	26.32
Slowed down time	8.13	1.55	0	17.56
Uncertain Time	4.86	1.36	0	10.01
Total time vs Total Cost	10.84	3.16	4.1	17.37

Table 7. Scale Differences Models – Scales

DESIGNS			CHOICE SETS			ALTERNATIVES			ATTRIBUTES			LEVELS			RANGE			SET SIZE		
Parameter	Coeff.	Jack knife t-stat*	Parameter	Coeff.	Jack knife t-stat	Parameter	Coeff.	Jack knife t-stat	Parameter	Coeff.	Jack knife t-stat	Parameter	Coeff.	Jack knife t-stat	Parameter	Coeff.	Jack knife t-stat	Parameter	Coeff.	Jack knife t-stat
Design1	Base		6 choice sets	Base		3 alternatives	Base		3 attributes	Base		2 levels	Base		base	Base		Set size 9	Base	
Design 2	0.437	3.17	9 choice sets	0.842	0.92	4 alternatives	1.280	1.18	4 attributes	0.920	0.70	3 levels	1.250	1.95	narrow	0.79	2.87	Set size 12	0.577	2.22
Design 3	0.515	2.53	12 choice sets	1.060	0.40	5 alternatives	1.220	1.62	5 attributes	0.930	0.68	4 levels	0.859	1.38	wide	0.8	5.15	Set size 15	0.858	0.64
Design 4	0.400	3.28	15 choice sets	1.060	0.35				6 attributes	0.500	5.15							Set size 16	1.090	0.36
Design 5	0.419	5.34																Set size 20	0.876	0.70
Design 6	0.872	0.40																Set size 24	0.739	0.91
Design 7	0.522	3.41																Set size 25	1.610	1.55
Design 8	0.688	0.98																Set size 30	0.472	2.69
Design 9	0.242	8.90																		
Design 10	0.348	4.62																		
Design 11	0.472	2.89																		
Design 12	0.597	3.41																		
Design 13	0.404	3.36																		
Design 14	0.529	2.35																		
Design 15	0.346	5.07																		
Design 16	1.010	0.03																		
Log Likelihood	-2,107.56		Log Likelihood	-2,213.22		Log Likelihood	-2,211.96		Log Likelihood	-2,211.64		Log Likelihood	-2,194.30		Log Likelihood	-2,209.50		Log Likelihood	-2,162.44	
adjusted ρ^2	0.4800		adjusted ρ^2	0.4570		adjusted ρ^2	0.4580		adjusted ρ^2	0.4570		adjusted ρ^2	0.4620		adjusted ρ^2	0.4580		adjusted ρ^2	0.4680	

* t-stats are with respect to one

Table 8. Scale Differences Models – Study Parameters

	BASE MODEL		DESIGNS		CHOICE SETS		ALTERNATIVES		ATTRIBUTES		LEVELS		RANGE		SET SIZE	
Parameter	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat	Coeff.	After JackKnife t-stat
Free Flow Time	-0.036	1.70	-0.081	0.82	-0.034	1.63	-0.031	1.78	-0.038	1.36	-0.035	1.68	-0.045	1.66	-0.040	-1.24
Running Cost	-0.084	3.06	-1.710	1.58	-0.475	2.01	-0.401	1.63	-1.020	1.71	-0.521	1.83	-0.625	1.92	-0.972	-1.87
Stop Start Time	-0.096	2.52	-0.239	1.88	-0.081	2.84	-0.069	2.74	-0.139	2.82	-0.087	3.27	-0.101	3.00	-0.137	-2.20
Combined Slow and Stop start Time	-0.066	2.87	-0.194	2.30	-0.092	2.85	-0.081	2.02	-0.101	2.24	-0.100	3.13	-0.117	2.68	-0.094	-1.50
Slowed down time	-0.039	2.93	-0.192	2.35	-0.062	3.10	-0.056	3.25	-0.108	3.09	-0.070	3.35	-0.082	2.93	-0.106	-2.31
Toll Cost	-0.499	1.90	-3.560	2.22	-0.356	0.95	-0.281	0.90	-2.090	3.18	-0.517	1.47	-0.735	1.71	-1.960	-1.74
Total Cost	-0.416	1.16	-1.880	2.03	-1.040	2.66	-0.852	3.32	-1.000	3.53	-0.987	3.74	-1.130	3.77	-1.120	-3.37
Travel/Total Time	-0.961	3.35	-0.307	2.56	-0.172	5.36	-0.155	6.87	-0.185	7.03	-0.165	6.09	-0.189	7.38	-0.178	-6.73
Uncertain Time	-0.172	6.90	-0.115	2.47	-0.037	2.92	-0.032	2.92	-0.062	2.86	-0.042	3.63	-0.049	3.00	-0.060	-2.90
ASC1	1.560	5.51	2.430	2.70	1.530	6.01	1.370	4.28	1.700	3.90	1.470	5.93	1.720	5.11	1.510	3.33
ASC2	0.179	2.15	0.332	1.71	0.165	2.05	0.171	2.34	0.230	2.00	0.151	1.67	0.214	2.21	0.205	2.21
ASC3	0.147	0.80	0.560	1.95	0.147	0.80	0.226	1.52	0.089	0.45	0.121	0.64	0.208	0.90	0.245	1.33
ASC4	-0.259	1.38	-0.568	1.36	-0.276	1.66	-0.145	0.93	-0.568	2.22	-0.308	1.63	-0.271	1.32	-0.490	-1.79
	Log Likelihood	-2,221.44	Log Likelihood	-2,107.56	Log Likelihood	-2,213.22	Log Likelihood	-2,211.96	Log Likelihood	-2,211.64	Log Likelihood	-2,194.30	Log Likelihood	-2,209.50	Log Likelihood	-2,162.44
	adjusted ρ^2	0.4560	adjusted ρ^2	0.4800	adjusted ρ^2	0.4570	adjusted ρ^2	0.4580	adjusted ρ^2	0.4570	adjusted ρ^2	0.4620	adjusted ρ^2	0.4580	adjusted ρ^2	0.4680

Table 9. WTP values (in £/hr)

PARAMETERS	DESIGNS				CHOICE SETS				ALTERNATIVES				ATTRIBUTES				LEVELS				WIDTH				SET SIZE			
	WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)		WTP value	t value	Confidence Interval (95%)	
Free Flow Time	2.92	1.20	0.00	3.73	4.33	1.55	0.00	5.03	4.76	1.45	0.00	5.36	2.36	1.24	0.00	3.39	4.08	1.48	0.00	4.79	4.34	1.52	0.00	5.02	4.34	1.52	0.00	5.02
Stop Start Time	8.51	2.16	0.79	9.01	10.47	1.46	0.00	10.74	10.60	1.48	0.00	10.87	8.41	1.90	0.00	8.86	10.22	1.63	0.00	10.53	9.92	1.56	0.00	10.23	9.92	1.56	0.00	10.23
Combined Slow & Stop start Time	6.94	1.75	0.00	7.44	11.96	1.38	0.00	12.19	12.42	1.35	0.00	12.64	6.15	1.67	0.00	6.68	11.81	1.67	0.00	12.09	11.50	1.56	0.00	11.76	11.50	1.56	0.00	11.76
Slowed down time	6.81	2.09	0.44	7.42	7.99	1.54	0.00	8.37	8.56	1.58	0.00	8.92	6.52	1.76	0.00	7.05	8.20	1.66	0.00	8.59	8.04	1.63	0.00	8.43	8.04	1.63	0.00	8.43
Uncertain Time	4.11	1.72	0.00	4.93	4.81	1.29	0.00	5.34	4.97	1.34	0.00	5.50	3.76	1.63	0.00	4.61	4.98	1.47	0.00	5.56	4.83	1.43	0.00	5.41	4.83	1.43	0.00	5.41
Total Time vs Total Cost	9.86	3.17	3.76	10.49	9.99	3.23	3.93	10.63	11.04	3.02	3.88	11.58	11.21	2.94	3.73	11.73	10.10	3.34	4.17	10.75	10.16	3.31	4.15	10.80	10.16	3.31	4.15	10.80

Table 10. WTP values from different models

	WTP Values (£/hr) for					
Model	**FFT**	**ST**	**SST**	**SSST**	**UT**	**TT**
Base	n/s	n/s	n/s	n/s	n/s	10.84(3.16)
3_alts	n/s	N/A	N/A	N/A	N/A	8.36(3.28)
4_alts	n/s	6.699(2.96)	n/s	n/s	4.99(3.2)	N/A
5_alts	n/s	n/s	n/s	46.76(1.78*)	n/s	14.96(2.38)
3_atts	N/A	N/A	N/A	N/A	N/A	n/s
4_atts	N/A	N/A	N/A	N/A	N/A	N/A
5_atts	N/A	N/A	N/A	N/A	N/A	N/A
6_atts	8.15(1.85*)	n/s	n/s	N/A	n/s	N/A
2_levels	n/s	6.54(3.02)	9.23(2.58)	7.62(3.73)	4.44(3.08)	N/A
3_levels	5.72(1.99)	8.15(2.23)	n/s	7.72(1.74*)	4.50(2.28)	4.69(1.66*)
4_levels	n/s	n/s	n/s	n/s	n/s	16.73(1.75*)
Narrow	n/s	n/s	n/s	n/s	n/s	3.48(2.45)
Medium	9.49(1.76*)	n/s	n/s	n/s	n/s	N/A
Wide	n/s	6.61(2.00)	7.03(2.55)	n/s	n/s	14.94(2.50)
6_sets	n/s	11.46(1.93*)	n/s	N/A	3.17(1.84*)	n/s
9_sets	N/A	N/A	N/A	N/A	N/A	23.37(2.5)
12_sets	5.07(2.06)	6.54(3.03)	8.72(2.32)	n/s	4.80(2.39)	n/s
15_sets	n/s	n/s	13.66	n/s	n/s	9.05(2.19)
9_items	N/A	N/A	N/A	N/A	N/A	n/s
12_items	N/A	N/A	N/A	N/A	N/A	N/A
15_items	N/A	N/A	N/A	N/A	N/A	17.46(3.44)
20_items	N/A	N/A	N/A	N/A	N/A	N/A
24_items	5.44(4.15)	6.57(3.21)	7.74(2.16)	N/A	4.89(2.39)	N/A
25_items	N/A	N/A	N/A	N/A	N/A	N/A
30_items	9.27(1.65*)	n/s	n/s	N/A	n/s	N/A

n/s- not significant; N/A – either the parameter or running cost is not present in the data, * significant at 90% confidence level