This file contains additional information, which could not be formatted as part of a text-only readme file, about the data used in the paper

> David Hensher "How do Respondents Process Stated Choice Experiments? - Attribute Consideration under Varying Information Load", *Journal of Applied Econometrics.*

Background Source of Data

The data are drawn from a larger study, reported in Hensher (JAE), in which 16 stated choice sub-designs have been embedded in one overall design, with each sub-design being used in surveying a sample of car commuter trips in Sydney in 2002. Each commuter evaluated one randomly assigned sub-design; however, across the full set of stated choice experiments, the designs differed in terms of the number, range and levels of attributes, the number of alternatives and the number of choice sets. The combination of these dimensions of each design is often seen as the source of design 'complexity', and it is within this setting that we have varied the number of attributes that each respondent is asked to evaluate.

Previous studies were used to identify candidate design dimensions. The five design dimensions are shown in Table 1.

Table 1: Dimensionality of the design plan

Choice set size	Number of alternatives	Number of attributes	Number of attribute levels	Range of attribute levels
6	2	3	2	Narrower than base
9	3	4	3	Base
12	4	5	4	Wider than base
15	----	6	----	----

Six attributes were selected for each alternative, based on previous evidence (see Hensher, in press), to characterise the options: free-flow time, slowed down time, stop/start time, variability of trip time, toll cost and running costs. To explore how varying the number of attributes affects information processing, the attributes were grouped according to the following patterns, noting that aggregated attributes are combinations of existing attributes[1]:

- *designs with three attributes:* total time (free flow + slowed down + stop/start time), trip time variability, total costs (toll + running cost);
- *designs with four attributes:* free flow time, congestion time (slowed down + stop/start), trip time variability, total costs;
- *designs with five attributes:* free flow time, slowed down time, stop/start time, trip time variability, total costs;
- *designs with six attributes:* free flow time, slowed down time, stop/start time, trip time variability, toll cost, running cost.

We have selected a generic design (i.e., unlabeled alternatives) to avoid confounding the effect of the number of alternatives with the labeling (e.g., car, train). The 16 sub-design dimensions are shown in Table 2.

Table 2: The sub-designs of the overall design

[1] This is an important point because we did not want the analysis to be confounded by extra attribute dimensions.

Choice set of size	Number of alternatives	Number of attributes	Number of levels of attributes	Range of attribute levels
15	3	4	3	Base
12	3	4	4	Wider than base
15	2	5	2	Wider than base
9	2	5	4	Base
6	2	3	3	Wider than base
15	2	3	4	Narrower than base
6	3	6	2	Narrower than base
9	4	3	4	Wider than base
15	4	6	4	Base
6	4	6	3	Wider than base
6	3	5	4	Narrower than base
9	4	4	2	Narrower than base
12	3	6	2	Base
12	2	3	3	Narrower than base
9	2	4	2	Base
12	4	5	3	Narrower than base

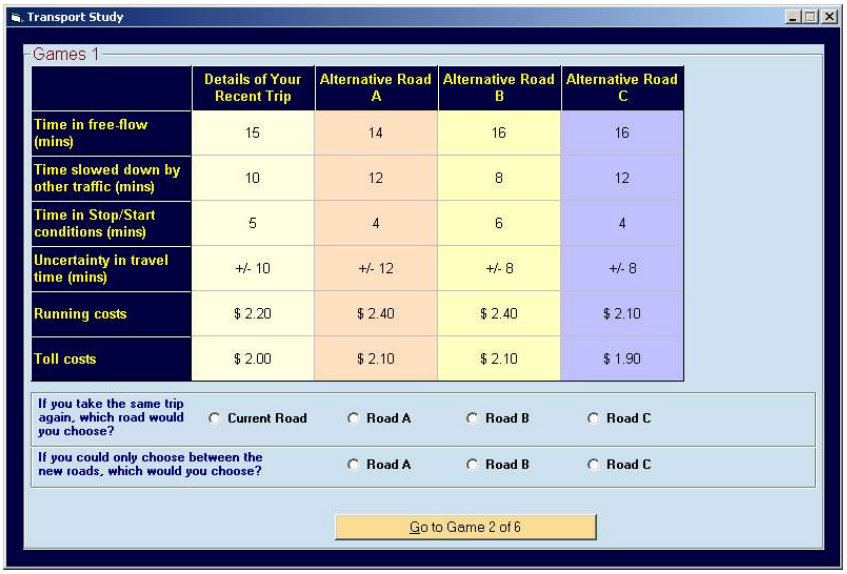
Note: Column 1 refers to the number of choice sets. The 16 rows represent the set of designs (referred to as Des0, Des1,…..,Des15 in model estimation).

As a generic design, the added alternatives are exactly the same. That is, for two design alternatives, we should not expect to find the parameter for an attribute (e.g., 'free flow travel time') to be different for the set of non-reference alternatives. Therefore we do not need the attribute 'free flow time one' to be orthogonal to the attribute 'free flow time two' etc up to 'free flow time *J*-1'. We need to ensure that the attribute 'free flow time' representing all non-reference alternatives is perfectly[2] orthogonal to the other attributes (such as slow down time, etc.).

The designs are computer-generated. A preferred choice experiment design is one that maximizes the determinant of the covariance matrix, which is itself a function of the estimated parameters. Knowledge of the parameters or at least some priors (such as signs) for each attribute, from past studies, provides a useful input. We found that in so doing, the search eliminates dominant alternatives. The method used finds the D-optimality plan very quickly.

The *actual* levels of the attributes shown to respondents are calculated relative to those of the experienced reference alternative – a recent car commuter trip. The levels applied to the choice task differ depending on the range of attribute levels and the number of levels for each attribute. The design dimensions are translated into SC screens, illustrated in Figure 1. The number and range of attribute levels only vary *across* designs. Each sampled commuter is given a varying number of choice sets (or scenarios), but the number of attributes and alternatives remain fixed. All analysis reported herein uses the elicitation response in Figure 1 associated with a choice set that excludes the recent trip.

[2] *Approximately* orthogonal is also acceptable given that some designs cannot guarantee complete orthogonality without loss of structure in terms of cognitive efficiency (in contrast to statistical efficiency).



Transport Study

Games 1

	Details of Your Recent Trip	Alternative Road A	Alternative Road B	Alternative Road C
Time in free-flow (mins)	15	14	16	16
Time slowed down by other traffic (mins)	10	12	8	12
Time in Stop/Start conditions (mins)	5	4	6	4
Uncertainty in travel time (mins)	+/- 10	+/- 12	+/- 8	+/- 8
Running costs	$ 2.20	$ 2.40	$ 2.40	$ 2.10
Toll costs	$ 2.00	$ 2.10	$ 2.10	$ 1.90

If you take the same trip again, which road would you choose? ○ Current Road ○ Road A ○ Road B ○ Road C

If you could only choose between the new roads, which would you choose? ○ Road A ○ Road B ○ Road C

Go to Game 2 of 6

Figure1. An example of a stated choice screen