IMPROVING UNDERSTANDING OF CHOICE EXPERIMENTS TO ESTIMATE VALUES OF TRAVEL TIME

Andrew Daly, RAND Europe and ITS Leeds
Flavia Tsang, RAND Europe

1. INTRODUCTION

The value of travel time (VOT) is perhaps the single most important variable in the analysis and planning of transport systems. Knowledge of the VOT of the travelling population is essential for decision makers to be able to justify investment in transport improvements that save travellers time. Further, VOT is required as an input to several forecasting processes.

VOT is an important example of the more general concept of willingness-to-pay (WTP), which is central to evidence-based decision-making on public investment. Decision makers justify expenditure on the basis that the public would be willing to pay for infrastructure or improved services: the challenge is to quantify WTP.

The most common approach currently adopted for the estimation of population VOT (as for a number of other WTP parameters) is to conduct stated choice (SC) experiments. These experiments involve offering a sample of respondents a series of sets of alternatives, differing in time and cost. Most often, these sets contain just two alternatives. In several studies, including some recent ones, part of the experiment has been based on an experimental design that balances the gains and losses in time and cost but does not include any other variables. A design of this nature seems to have first been used by MVA *et al*., (1987).

This design, though well tried and tested, e.g. in subsequent national VOT studies in the UK, The Netherlands and Denmark, has been criticised as being too simple, as not giving sufficient variety of choice and of inviting inconsistency in responses (Hess *et al*., 2009). Sometimes it has proved quite difficult to obtain satisfactory results from specific data sets. However, sophisticated analyses of data of this form (e.g. Fosgerau *et al*., 2006a) are able to obtain quite robust estimates of VOT. It is therefore interesting to consider whether the problems that have been encountered might be due to the excessively simple form of analysis that has conventionally been used. Further, if a different form of analysis is required, can this be performed in a reasonably simple way, or is the full sophistication of the Fosgerau *et al*. approach required?

The conventional assumption is that all the choices are equally accurately assessed, an assumption that is also widely applied in other choice modelling contexts. However, as we shall see, this assumption is just one among a number of equally simple assumptions that might have been made and the question then arises as to which of the alternative assumptions is most appropriate.

The paper therefore addresses these issues by presenting a range of possible simple assumptions that might be made, discussing what their theoretical foundations might be and assessing their success on two data sets that have been used in important VOT studies. Conclusions are then drawn for future practice, taking particular account of the constraints under which practical studies have to be conducted in a commercial environment, where budget, time and even expertise are often lacking.

The following section of the paper discusses the ways in which simple assumptions can be made that lead to different analyses of stated choice VOT data. Section 3 presents briefly the data sets that are used for analysis and Section 4 gives the modelling results. Section 5 summarises the findings, both theoretical and empirical.

2. FORMULATION OF MODELS FOR ANALYSIS

2.1 Random utility models

For the discussion in this paper, we shall assume that travellers make choices, whether in their actual behaviour or in response to SC interviews, to maximise utility. This approach can be seen as the simplest available, in the sense that the attractiveness of each alternative is summarised as a single measure and the assumption is made that choice is based on that measure alone. While this characterisation of behaviour is known to have limitations, other approaches usually require a more complicated description of choice behaviour. It is important, however, to recognise that however good the data collected, the behaviour of different individuals (and even perhaps of the same individual) will vary because of variables affecting choice that are unknown to the analyst (and perhaps to the individual).

To allow for unmeasured variation between individuals or between successive responses by the same individual we formulate a random utility model, which gives the probability p_j of choosing a single alternative j from the J available to him or her:

$$p_j = \Pr\left\{U_j \geq U_k, k = 1..J\right\} \tag{1}$$

where U gives the utility of the alternatives. This is the standard formulation of random utility model that has been used in so many studies. Providing the distribution of U is continuous, the probability of equality of utilities is zero and can be neglected.

For quantitative analysis, the only reasonable procedure seems to be to make an approximation V to the utility, based on measured variables, so that, taking account of the fact that utility is an ordinal rather than a cardinal concept, i.e.

$$\{U_j \geq U_k\} \text{ can be approximated by } \{V_j \geq V_k\} \tag{2}$$

Formulated in this way, it is clear that the scale of V is not defined in any absolute sense. Of course the analyst will want to maintain reasonable simplicity in the functional form, so that it is very common to specify V as a linear function of measured variables, but these functions have no special justification. What is important is that the direction of impact is correct for the components of V. For some components of V it is clear what the direction of impact will be, e.g. the impact of increased cost will always be negative. But as far as functional form is concerned, V need not be linear, as in principle any function V that preserves the direction of impact of these components will be satisfactory.

2.2 Measured and random components

The statement of the approximation (2) of the model of choice is incomplete, as it fails to deal with the exact nature of the approximation. To complete the model it is necessary to add an error term which will allow for failure of the analyst's approximation. It turns out there are several ways in which this can be done.

2.2.1 Difference formulation

The most common approach to completing the model (2) is to represent the *difference* between the approximation and the true utility by a random number ε, giving the utilities by

$$U_j = V_j + \varepsilon_j \tag{3}$$

and thus the probabilities are given by

$$p_j = \Pr\{\varepsilon_j - \varepsilon_k \geq V_k - V_j, k = 1..J\} \tag{4}$$

which, provided the distributions are known, can be calculated from the joint distributions of ε; this calculation exercise may be onerous, depending on the form of the distribution.

This additive form is not, of course, a restriction on the model, as we are simply defining the error to be the difference between the true utility and the approximation. Other specifications could be used, though these would imply different and perhaps less convenient functional forms.

2.2.2 Multiplicative formulation

An alternative formulation of the model, instead of (3), would be to use the multiplicative form

$$U_j = V_j \varepsilon_j \tag{5}$$

This specification was probably first proposed by Harris and Tanner (1974) and has recently been studied by Fosgerau and Bierlaire (2009). Here ε is defined as the *ratio* of the true and approximate utilities. But as pointed out by Fosgerau and Bierlaire (providing the terms of (5) are positive) simply taking the log of equation (5) – and any positive monotonic transformation can always be applied to utility – we obtain

$$U_j^* = \log V_j + \eta_j \qquad (6)$$

$$p_j = \Pr\{\eta_k - \eta_j \geq \log V_k - \log V_j, k = 1..J\} \qquad (7)$$

where U^* is the transformed utility and $\eta = \log \varepsilon$, thus obtaining the additive form again, at the expense of requiring the analyst's approximation to be formulated as a log.

What is shown by the possibility of formulating the model either additively or multiplicatively is therefore not so much that the model can be set up in different ways, as that the scales of V and ε are interdependent. While the analyst will wish to avoid excessive complication in the scale of V, because of the need to specify a distribution it is usually more important to keep a simple specification for ε. Thus Fosgerau and Bierlaire suggest using a formulation like (6) and (7), with η following the Gumbel distribution to obtain a logit model with the modification that V has the log form.

Again, this formulation is not a restriction of the model, as it simply defines the error to be the ratio of the true and approximated utilities.

2.2.3 Generalised formulation

Forms generated by starting with multiplicative and additive models are just two of many specifications. A range of differing specifications can be obtained by suitable transformations of V, allowing the simple additive form and simple assumptions for the distribution of ε to be maintained. For example, we may apply a Box-Cox transformation to V:

$$U_j = V_j^{(\alpha)} + \varepsilon_j \qquad (8)$$

where

$$V^{(\alpha)} = \frac{V^{\alpha} - 1}{\alpha}, \quad when \quad \alpha \neq 0$$
$$V^{(\alpha)} = \log(V), \quad when \quad \alpha = 0$$

as explained in the Appendix, allowing a range of transformations to be applied to V (it is shown in the Appendix that all of these transformations preserve the sign of the slope of V). Among these transformations are the conventional additive (linear) model (1), which is obtained when $\alpha = 1$, and the multiplicative (log) model (4), which is obtained when $\alpha = 0$. Clearly a range of models may be obtained by selecting differing values of α, all of which can be paired with the simple Gumbel assumption for the distribution of ε. The issue is which of these are most suited to explaining the actual and stated travel choices we observe.

The transformations that are made in practice are to set

$$U_j = V_j^{\alpha} + \varepsilon_j \text{ with } \alpha \neq 0, \quad \text{or} \quad U_j = \log V_j + \varepsilon_j \qquad (9)$$

The switch from the strict Box-Cox form to the power function was made for simplicity in implementation. For a given α the difference between Box-Cox and power function is simply a linear transformation and therefore does not affect the modelling.

The Box-Cox formulation (8) or power/log function (9) is again a fully general model. That is, like the additive or multiplicative form the error is simply defined to be the difference between the true utility (for which we can only observe ordering) and a formulation of the measured utility. The issue of which of these formulations is most suitable can be determined by which of them gives the best explanation of the data when we use a simple form for the distribution of the error.

2.3 Scaling of the utility function

The use of power functions such as the Box-Cox transformation change the *shape* of the utility function. We may also think about changes to the *scale* of the model, which may be related to measured variables or random. The possibility of random variation has been studied by Train and Weeks (2005), who specified a random distribution of scales for this variance, citing an earlier unpublished note by Louviere. In the present paper, however, we focus on relating the scale of the model to measured aspects of the choice situation, rather than varying randomly.

The approach that is taken is to retain the Gumbel distribution for the random component of error but to seek relatively simple transformations to the scale of the non-random component that may give a better explanation of the behaviour we observe. Specifically, we investigate whether the standard deviation of the error term might be related to specific components of the utility, e.g. in this case cost or time. If this is the case, then the standard additive model (3) is incorrect, because of heteroskedasticity induced by different levels of cost. That is, we suppose that (3) is generalised to give

$$\text{var}(\varepsilon_j) = kc_j^{2\alpha} + \frac{\pi^2}{6} \qquad (10)$$

where k is a constant, with part of the error arising from uncertainty concerning the cost and the remainder arising because of general error in the choice process, represented by the Gumbel term with variance $\pi^2/6$. The heteroskedasticity can be corrected by scaling the measured part of the utility function by $c_j^{-\alpha}$, arriving at a homoskedastic utility with variance $k + \frac{\pi^2}{6}$.

$$U_j = \frac{V_j}{c_j^{\alpha}} + \varepsilon_j\,, \qquad \text{var}(\varepsilon_j) = k + \frac{\pi^2}{6} \qquad (11)$$

In a choice model, alternatives are compared with each other and variation in the value attached to (say) cost will apply to the cost of each of the alternatives. The scaling described above should then be calculated using the *difference* in cost between the alternatives. In a general choice context, this is not possible, but for binary choices it is quite straightforward.

The transformation (11) is made using cost but it is clearly possible to make analogous transformations using time or, indeed, another attribute.

The formulation of the error variance (10) resembles that of a mixed logit formulation with a random cost parameter. However, the formulation we propose differs from the mixed logit because

- we investigate values of α both equal and unequal to 1, which would be the value required for mixed logit;
- the approach we propose does not require sampling, just a simple scaling of the utility function.

Of course, if sampling is used, then the mixed logit has the further advantage that various other effects can be introduced into the model. In particular, correlation between the responses for a single individual can be introduced, when we have panel data. In the tests described below comparisons are made of scaled models both with and without this correlation.

3. TESTING OF THE HYPOTHESES

In the previous section we outlined two broad hypotheses that generalise restrictive assumptions made in previous work in analysing VOT data:

a. that the combination of measured and random utility components is made additively or, less often, multiplicatively; this is generalised by using the power/log transformation (9) which includes the additive and multiplicative specifications as special cases;
b. that the scale of the error term is constant or, less often, purely random; this is generalised by scaling the measured part of the utility function by powers of time and cost.

These hypotheses were investigated by making tests on two data sets which have been the subject of intensive previous study.

In this section we describe first the detailed specification of the models that will be used and specify the tests to be made, then go on to discuss the data sets used.

3.1 Specification of the models

When SC methods are used to make estimates of WTP, respondents are generally presented with two alternative ‘scenarios’ and asked to indicate which they prefer. These scenarios contain a cost variable and other variables for which we wish to know the value. For the simplest VOT studies, there is just a single time variable in addition to cost.

For analysis, the utility function for these simple studies would take a form like

$$U_j = V_j + \varepsilon_j = \beta_c c_j + \beta_t t_j + \beta_a a_j + \varepsilon_j \qquad (12)$$

where β are coefficients to be estimated;

c, t are respectively the time and cost presented in the scenario; and

a is an “as now” dummy variable, which takes the value 1 if the time and cost presented are both the same as those encountered by the respondent in his or her current journey.

The specification of the variable a is of course arbitrary, but such effects have consistently been found to be important in practice and in this example it also serves to show how further variables could be added to the model as required.

In this model, VOT may be calculated as the ratio of the marginal utilities of time and cost (both expected to be negative), i.e. β_t / β_c , while the value of the as-now attribute can be calculated as β_a / β_c .

3.1.1 Transformations of the utility function

For the transformations of the utility function described in 2.2.3 above, the tests to be made are of the form

$$U_j = \gamma\left(c_j + \nu_t t_j + \nu_a a_j\right)^{\alpha} + \varepsilon_j \text{ or } U_j = \gamma \log\left(c_j + \nu_t t_j + \nu_a a_j\right) + \varepsilon_j \qquad (13)$$

where γ gives an overall scale to the model ($\gamma = \beta_c^{\alpha}$ in the exponentiated form); and

$\nu_t = \beta_t / \beta_c , \nu_a = \beta_a / \beta_c$ give direct estimates of the values of the attributes.

This model is tested for values of α equal to 0 (the log form), 0.25, 0.5, 1 (the basic linear form) and 1.5.

An iterative estimation procedure was used to deal with the novel formulation of the utility function. Because of this complication, it was not possible to investigate inter-personal variation in the coefficients.

3.1.2 Scaling of the utility function

When respondents are presented with binary choices, it can simplify the analysis to work with utility differences.

$$\Delta U = \Delta V + \eta = \beta_c \Delta c + \beta_t \Delta t + \beta_a \Delta a + \eta \qquad (14)$$

where the Δ prefix indicates a difference between the alternatives (values for alternative 1 minus values for alternative 2) and $\eta = (\varepsilon_1 - \varepsilon_2)$ is the random error in this formulation. The formulation as differences does not affect the calculation of VOT which is still expressed by the ratio of the β coefficients, as it was in the derivation from (12).

This formulation could not be used for the tests described in the previous section because it was required to transform the utility functions separately. However, in the case of scaling, as described in 2.3, it is possible to work with differences and indeed to apply the scaling based on differences in the utility components. It is also interesting to base the scaling for these tests on differences, while the scaling in the previous section is based on transformations relating to the absolute values of the measured utility.

The tests made are then based on models defining utility differences by

$$\Delta U = \frac{\beta_c \Delta c + \beta_t \Delta t + \beta_a \Delta a}{\Delta c^{\alpha}} + \eta \qquad (15)$$

and

$$\Delta U = \frac{\beta_c \Delta c + \beta_t \Delta t + \beta_a \Delta a}{\Delta t^{\alpha}} + \eta \qquad (16)$$

Tests were made of both these formulations for values of α equal to 0, 0.5, 1 and 1.5. In this case the value 0 gives the basic linear form and then does not differ between the time-power and cost-power models.

3.1.3 Scaling with random variation

In the models with scaling, (15) and (16), it is also possible to consider random variation in cost sensitivity. This leads to models of the form

$$\Delta U = \frac{\beta_c \Delta c + \beta_t \Delta t + \beta_a \Delta a}{\Delta c^{\alpha}} + \frac{\gamma \xi \Delta c}{\Delta c^{\alpha}} + \eta \qquad (17)$$

and

$$\Delta U = \frac{\beta_c \Delta c + \beta_t \Delta t + \beta_a \Delta a}{\Delta t^{\alpha}} + \frac{\gamma \xi \Delta c}{\Delta t^{\alpha}} + \eta \qquad (18)$$

where ξ is a random term distributed normally with mean 0 and variance 1. For panel data it varies between individuals but not over the choices of one respondent.

In these models, we tested the same series of values of α as in the simple scaled models.

In this case the VOT is $v_t = \dfrac{\beta_t}{\beta_c + \gamma\xi}$, which is a random number with a distribution that depends on the distribution assumed for ξ. For models of this type, the use of a normal distribution of the random term would be considered unsuitable, both because the fact that part of the distribution of VOT will be negative, clearly inappropriate, and the fact that the proximity of some values of the denominator of the VOT formula will be arbitrarily close to zero will cause problems in defining the mean and variance of the VOT (Daly *et al*., 2009). However, for the present study the interest is not in making estimates of VOT but in investigating whether the scaling proposed can give better models and the normal distribution was therefore considered acceptable. A useful way to express the variation in the VOT in such a model is through its Coefficient of Variation, i.e. γ/β_c.

3.2 The data used

We used two similar data sets to make the tests described above. The first was designed, collected and analysed by Accent and Hague Consulting Group (1996) and subsequently reanalysed by Institute for Transport Studies (Mackie *et al*., 2003), on which latter analysis the UK national time values are based. The second was designed by RAND Europe (Burge *et al*., 2004) which was subsequently analysed by Fosgerau *et al*. (2006b) and is the data on which Danish national time values are based. These data sets have been subjected to intensive analysis both in the original studies for which they were collected and subsequently for academic purposes.

In each case several data sets were collected. However, for this work we focus on commuter travel and use the simplest data that was included in the studies. These use basically the same design, a set of eight Stated Choice responses for each individual, at each choice comparing two arbitrary combinations of time and cost. These are built up from the traveller's current journey with a careful balance of gains and losses of time and cost.

The series of eight binary choices, with just two attributes presented to the respondent, represents a minimal Stated Choice experiment. However, the fact that this design has been used, apparently successfully, in several important studies means that it has the advantage of being the standard approach, from which departures would require justification. Further, in many cases governments are interested in comparing new results with previous findings and consistency of design is obviously helpful in that aim.

4. EMPIRICAL RESULTS

As set out in the previous sections, three series of runs were made. These are reported in this section.

4.1 Transformations of the utility function

Table 1 sets out the results of models based on transformations of the utility functions, first for the UK data, then for the Danish data.

Table 1a: Powers of Utility Function, UK data

Model	Final log-likelihood	VOT p/min[1]	t-ratio
Log V	-2787.7	3.67	51.7
V^{α}, with α =0.25	-2781.0	3.84	70.6
V^{α}, with α =0.5	-2774.5	3.98	34.7
V^{α}, with α =1 (Standard)	-2813.4	4.64	18.7
V^{α}, with α =1.5	-2923.9	5.38	16.5

Table 1b: Powers of Utility Function, Danish data

Model	Final log-likelihood	VOT Øre/min[2]	t-ratio
Log V	-2390.4	42.65	26.9
V^{α}, with α =0.25	-2388.8	54.23	49.2
V^{α}, with α =0.5	-2390.2	68.44	31.6
V^{α}, with α =1 (Standard)	-2409.1	92.40	20.4
V^{α}, with α =1.5	-2453.7	113.72	11.6

Looking at the fit of the models to the data we see that for both data sets all the powers of α less than 1 give better models than the standard α=1, but the higher power tested gave worse results. While the log function is considerably better than the linear function, intermediate values of α are better still. For the UK data α=0.5 is best, for the Danish data α=0.25 (among the values tested).

It appears that the values of time estimated in these models decline with α and more strongly in the Danish than in the UK data. However, the accuracy of estimation of VOT increases substantially for lower values of α, except for the value zero.

The 'as-now' parameter was estimated for the UK data but could not be estimated on the Danish data as the necessary variable was not present in the data available to us. For the log version of the model (α=0) it was not possible to make a valid estimation with the as-now variable included in the function of which the log was taken and it was therefore placed in the utility function 'outside' the log function; the reasons for this are not known.

1 UK 1p is currently worth approximately €0.011. However, the VOT data was collected in 1994; up to 2008 price inflation has been about 29% and real incomes have increased by about 50%.

2 DKK 0.01 (1 Øre) is currently worth about €0.0013. The data was collected in 2004.

4.2 Scaling the utility function by powers of components

Table 2 gives the results for the scaling of utility by powers of utility components, first for the UK data, then for the Danish data.

Table 2a: Scaled model, UK data

Title	Final log-likelihood	VOT p/min	t-ratio
Linear standard	-2756.1	4.67	12.5
Scale by c^{α}, with α =0.5	-2744.3	2.88	14.9
Scale by c^{α}, with α =1	-2805.5	2.15	19.0
Scale by c^{α}, with α =1.5	-2854.5	1.74	21.3
Scale by t^{α}, with α =0.5	-2733.5	4.46	11.2
Scale by t^{α}, with α =1	-2744.9	4.17	9.5
Scale by t^{α}, with α =1.5	-2778.6	3.71	7.7

Table 2b: Scaled model, Danish data

Title	Final log-likelihood	VOT Øre/min	t-ratio
Linear standard	-1840.3	101.7	10.8
Scale by c^{α}, with α =0.5	-1795.3	63.47	14.0
Scale by c^{α}, with α =1	-1817.1	34.40	13.5
Scale by c^{α}, with α =1.5	-1857.0	24.45	10.5
Scale by t^{α}, with α =0.5	-1812.8	84.78	11.6
Scale by t^{α}, with α =1	-1793.5	67.25	11.2
Scale by t^{α}, with α =1.5	-1784.4	55.20	10.8

Comparing the results for the UK data, we see that the fit is improved by a small amount of scaling by cost, but larger amounts of scaling bring a deterioration. Scaling by time is more beneficial, but again larger scale factors are less successful In both cases α=0.5 is the best parameter of those tested. For the Danish data, there is a larger benefit from scaling, so that, even though the fit deteriorates for larger cost scale factors, it increases with larger time scales and in either case is generally better than with the linear model. These results are not completely comparable with those of the previous series, because a small number of observations had to be dropped because the time or cost differences by which we were scaling were in fact zero.

In both cases the VOT declines substantially with increasing scaling, while the accuracy of estimation is improved by moderate cost-scaling in the UK model but not much changed in the Danish data.

In both data sets time scaling gives a greater improvement in fit than cost scaling.

4.3 Scaling with random component added

Table 3 gives the results for the scaling of utility by powers of utility components, with an added variation in the sensitivity to cost, first for the UK data, then for the Danish data.

The additional random component, as explained in Section 3.1.3, allows random inter-personal variation in the value of time. The models were estimated using the 'implosion' procedure (Daly *et al*., 2008). In these models, the approximate matrix of second derivatives (BHHH) matrix was used, whereas in the preceding series it was possible to use the true matrix of second derivatives of the likelihood function.

Table 3a: Scaled model with random VOT, UK data

Title	Final log-likelihood	mode VOT p/min	t-ratio of VOT mean	C. of V. of VOT	t-ratio of s.d. of VOT
Linear standard	-2468.0	3.333	20.2	0.846	17.4
Scale to c^{α}, with α =0.5	-2504.4	2.326	20.5	1.004	18.3
Scale to c^{α}, with α =1	-2544.7	2.075	24.3	1.018	20.1
Scale to c^{α}, with α =1.5	-2599.0	1.880	28.9	0.963	21.6
Scale to t^{α}, with α =0.5	-2455.4	3.543	19.3	0.836	17.8
Scale to t^{α}, with α =1	-2477.3	3.596	18.6	0.840	18.1
Scale to t^{α}, with α =1.5	-2525.2	3.424	17.4	0.859	18.1

Table 3b: Scaled model with random VOT, Danish data

Title	Final log-likelihood	mode VOT Øre/min	t-ratio of VOT mean	C. of V. of VOT	t-ratio of s.d. of VOT
Linear standard	-1699.3	58.04	22.7	0.883	14.4
Scale to c^{α}, with α =0.5	-1643.9	46.17	19.1	1.027	16.1
Scale to c^{α}, with α =1	-1667.7	35.14	18.4	1.352	16.3
Scale to c^{α}, with α =1.5	-1759.2	31.33	17.1	1.655	15.3
Scale to t^{α}, with α =0.5	-1656.9	58.16	19.6	0.866	15.0
Scale to t^{α}, with α =1	-1646.8	55.08	18.5	0.893	15.4
Scale to t^{α}, with α =1.5	-1657.2	49.76	17.0	0.946	15.4

The results for the UK models show no benefit from cost scaling, but for time scaling we see a small benefit. The Danish results largely repeat those of the previous section.

The mode of the VOT distribution decreases with increasing cost scaling. For time scaling, the changes are smaller and less clearly marked. The estimation accuracy as measured by the Coefficient of Variation does not vary widely. The VOT values in these models are not intended to indicate the distribution of values in the population, as the distributions implied by these models are quite unreasonable (Daly *et al*., 2009). However, the results suggest strongly that random variation in VOT and the scaling of the model by powers of cost and time are independent effects.

5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussion of results

The first series of runs, effectively investigating 'cost damping', shows that both data sets support the use of a 'sub-linear' power function (i.e. with $\alpha < 1$) to transform the utility function. While the logarithmic form, corresponding to the multiplicative model (5) is considerably better than the conventional additive form (3), it is clear that intermediate values of α perform better still. If α were a freely-estimated parameter, the improvement in log-likelihood relative to either linear or logarithmic form, for both data sets, is statistically significant (though not at a high level for the Danish data relative to the log form).

The improvements in likelihood obtained from the first series of runs are similar in magnitude to those obtained in the second series for the UK data, with time scaling performing better than cost scaling. It might be thought, therefore, that these could reflect the same effect, since the second series runs also implement a form of cost damping. However, this is unlikely to be true since the first series operate on the entire utility function, while the second series operate on the cost and time *differences* between the two alternatives presented. In the Danish data similar results are obtained, though in that case the improvement with time scaling continues for higher values of α. Again the improvement in log likelihood in each case is sufficient to justify a freely estimated parameter.

The third series of runs are not greatly different from the second series. though in this case no improvement at all is obtained from the Danish time-scaling runs.

It appears that the shape and scaling of the utility function can be amended with an improvement to the fit to the data. Since these models are no more or less arbitrary in principle than the ones generally in use, such changes should be considered for practical use. Because shape and scaling the transformations operate in different ways it may be possible to benefit from both of them for a single data set, though it has not been possible to test both transformations at the same time in the current study. Each of the transformations appears to give similar benefit.

The addition of a random component does not have a clear impact on the benefits of shape and scale transformations, increasing the likelihood gain for the Danish data but reducing it for the UK data. This suggests that the main benefit of the random component is the modelling of correlation of values of time, rather than allowing the error to increase with increased cost differences, since this is also modelled by the cost scaling.

5.2 Conclusions and Recommendations

If the utility theory basis is adopted for discrete choice modelling, there is no particular theoretical reason to adopt the additive, multiplicative or an alternative transformation of the utility function. The suggestion from this work is that an intermediate transformation between logarithmic (multiplicative) and linear (additive) transformation may be beneficial to the fit to the data.

Similarly, there is no theoretical reason to maintain a constant scale. The beneficial effects of the transformations are of similar magnitude to the shape transformations and do not appear to be eliminated by the inclusion of a random cost coefficient.

Neither of these changes to the model has a very large positive or negative impact on the magnitude or accuracy of the value-of-time estimates, or of the estimated variation of value of time in the population.

We conclude that the transformations – both shape and scale – offer the potential to obtain better fit to data. While they do not greatly improve the estimation accuracy, which was one of the possible benefits, they certainly do not reduce accuracy or the reasonableness of the estimates.

We recommend that these or similar transformations be tried on other data.

References

Accent Marketing and Research and Hague Consulting Group (1996). The Value of Time on UK Roads, The Hague.

Burge, P., Rohr, C., Bates, J. and Vuk, G. (2004) Review of international experience in VOT study design, presented to European Transport Conference, Strasbourg, France.

Daly, A. (2008) The relationship of cost sensitivity and trip length, presented to European Transport Conference, Noordwijkerhout, Netherlands.

Daly, A. and Carrasco, J. (2006) The influence of trip length on marginal time and money values, presented to Conference of International Association for Travel Behaviour Research, Kyoto.

Daly, A., Choudhury, C., Burg, P. and Sivakumar, A. (2008) Dealing with repeated choices in stated preference data, presented to European Transport Conference, Noordwijkerhout, Netherlands.

Daly, A., Hess, S. and Train, K. (2009) Computing willingness to pay from random coefficients models, presented to European Transport Conference, Noordwijkerhout.

Fosgerau, M. and Bierlaire, M. (2009) Discrete choice models with multiplicative error terms, Transportation Research Part B, **43**, pp 494-505.

Fosgerau, M., Hjorth, K., and Vincent Lyk-Jensen, S. (2006a) An integrated approach to the estimation of the value of travel time, presented to European Transport Conference, Noordwijkerhout, Netherlands.

Fosgerau, M., K. Hjort, S. Vincent Lyk-Jensen (2006b) The Danish Value of Time Study, final report; DTF Report.

Gaudry, M. (2008) Non linear logit modelling developments and high speed rail profitability, Agora Jules Dupuit, Publication AJD-127.

Harris, A.J. and Tanner, J.C. (1974) Transport demand models based on personal characteristics, Transport and Road Research Laboratory Supplementary Report SR65UC, Crowthorne, UK.

Hess, S., Rose, J. and Polak, J. (2009?) Non-trading, lexicographic and inconsistent behaviour in stated choice data, accepted for publication, Transportation Research Part D.

Mackie, P., Wardman, M., Fowkes, A., Whelan, G., Nellthorp, J. and Bates, J. (2003) Values of Travel Time Savings in the UK, report to Department for Transport.

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

Train, K. and Weeks, M. (2004) Discrete Choice Models in Preference Space and Willingness-to Pay Space, Cambridge Working Papers in Economics 0443, Faculty of Economics, University of Cambridge.

Appendix Box-Cox and Box-Tukey transformations

Large-scale travel demand modelling has frequently found it necessary to make adjustments to the functional forms used the models. Very often, these adjustments take the form of non-linear transformations to all or part of the V function. Potential functions were discussed by Daly (2008) who identified Box-Cox or Box-Tukey transformations as suitable and consistent with most theoretical requirements. These functions have been studied extensively in this context by Gaudry (e.g. 2008). The functions take the form

$$x^{(\alpha,\delta)} = \frac{(x+\delta)^{\alpha}-1}{\alpha}, \quad when \quad \alpha \neq 0$$
$$x^{(\alpha,\delta)} = \log(x+\delta), \quad when \quad \alpha = 0$$

where $x \geq 0$ is the variable to be transformed;
α is the power parameter;
$\delta \geq 0$ is the Tukey shift – setting $\delta > 0$ allows x to be zero for all values of α. If $\delta = 0$ we have a Box-Cox transformation and then we need to make sure x is positive.

It may be noted that $\frac{dx^{(\alpha,\delta)}}{dx} = (x+\delta)^{\alpha-1}$ so that the transformation always has a positive slope, since $x \geq 0$.

We use the notation $x^{(\alpha)}$ for the Box-Cox transformation, i.e. $x^{(\alpha)} = x^{(\alpha,0)}$.

Box-Cox transformations differ from power functions by the inclusion of the constant 1 and factor α in the formula. The purpose of these terms is to ensure that $\lim_{\alpha \to 0} x^{(a)} = \log(x)$, i.e. the transformation is continuous with respect to α. They also to normalise the transformation, in the sense that

$x^{(a)} = 0$ and $\frac{d}{dx}\left(x^{(a)}\right) = 1$, when $x = 1$, whatever the value of α.

Thus when the objective of the analysis is to investigate the impact of differing values of α on the *shape* of the function, Box-Cox (and Box-Tukey) functions used around $x = 1$ mean that α can be used to make adjustments to the shape without changing the absolute value or the slope of the function. 'Plain' power functions inevitably change the slope when the exponent changes, though of course the value does not change with α when $x = 1$.

In order to exploit this feature of Box-Cox transformations it may be necessary to change the scale of x to be centred around the value 1.