JOHN BATES SERVICES

1 Pickler's Hill. ABINGDON OX14 2BA Tel: Abingdon (01235)-526702

An Essay on Cost Damping and Transport User Benefits

Final Report for Department for Transport
July 2025

Contents

Executive Summary

In Transport Appraisal the User Benefits are calculated according to the Rule of a Half [RoH] approximation, but concerns have been raised as to whether this is still valid when cost damping is present. This Essay was commissioned to examine this.

Beginning with a brief account of how the current situation regarding modelling and appraisal developed over time, the evidence for cost damping – or more generally, reduced sensitivity to time and cost for longer distance travel – is examined. The three main sources are airport access mode choice studies, detailed model estimation exercises carried out by RAND Europe and its associates, and national studies into the Value of Travel Time Savings [VTTS].

The conclusions are not as clear as might be hoped. The most consistent result to emerge is that there are reduced sensitivity effects with **both** cost and time increases. Generally – though not always – the cost effects are larger, and this leads to VTTS increasing with "distance".

The theoretical consequences for user benefit are examined in connection with the current TAG recommendations for cost damping, leading to some proposals for tests to see whether the use of the RoH involves significant inaccuracy or whether it can be tolerated as a satisfactory approximation.

A number of modelling issues are discussed, with special attention to the model of mode and destination choice which is the critical determinant of transport demand. It would appear that the reduced sensitivity effect is more relevant to mode choice than destination choice but there is no current evidence as to whether this is the case. It is suggested that the cross-nested logit [CNL] model might be able to reconcile this. In addition, assignment and unimodal models are discussed.

Various proposals are made for testing a number of relevant empirical issues using available datasets: the possible inaccuracy from the RoH, the correlation between time, cost and distance, and the sensitivity of highway assignment to VTTS.

0 Introduction

This Essay was commissioned as a thinkpiece in the light of concerns about possible incompatibility between cost damping (as recommended in TAG M2.1 §3.3) and the calculation of user benefits using the Rule of a Half [RoH] approximation. The detailed Terms of Reference are attached as Appendix A.

A large number of issues are raised in the Terms of Reference, and within the resources available not all of them have been addressed, in particular those relating to Wider Impacts (TAG A2). It was decided to concentrate on the two related issues – Cost damping, and Consistency between modelling and appraisal. Can cost damping be captured by appropriate non-linear utility functions of time and cost, or – implicitly – VTTS, in a way that can be (more or less) consistently transferred to appraisal? There are already inconsistencies between the modelling of mode and destination choice and that of route choice within the assignment procedure – partly because of software constraints: will these be exacerbated by a more non-linear approach to utility or can they be resolved, at least in part?

Chapter 1 sets out the origins of the underlying approach to transport modelling and appraisal, and how it became reconciled with developing economic theory. As long as we can proceed with a linear combination of time and cost, few problems arise. However, (as the Terms of Reference note) "There is strong empirical evidence that the sensitivity of demand responses to changes in generalised cost reduces with increasing trip length", which implies some kind of non-linearity.

Accordingly, Chapter 2 reviews the evidence for this reduced sensitivity. There are three main sources, and these are discussed in some detail: airport access mode choice studies, the detailed model estimation exercises carried out by RAND Europe and its associates, and national studies into the Value of Travel Time Savings [VTTS]. While the review concentrates on studies with which the author is familiar (mainly in the UK and The Netherlands), it is believed that these are among the most important sources.

The conclusions are not as clear as might be hoped. The most consistent result to emerge is that there are reduced sensitivity effects with **both** cost and time increases. Generally – though not always – the cost effects are larger, and this leads to VTTS increasing with "distance".

Chapter 3 then reviews the current TAG recommendations for cost damping, noting the implications for the specification of utility/generalised cost, and what the theoretical consequences for user benefit might be.

Given that it is the model of mode and destination choice which is the critical determinant of transport demand, Chapter 4 considers in some depth what the implications are for utility theory in respect of the two choices, and how they might be reconciled.

Chapter 5 then addresses two other modelling issues, relating to assignment, where there are no current recommendations for cost damping, and the rather different topic of unimodal models, which are particularly relevant to rail appraisal.

While the theory does suggest that the standard RoH may not always be appropriate in the presence of certain types of cost damping, it remains an open question as to whether this is a serious issue in practice. In Chapter 6 various proposals are made for testing this and other empirical issues using available datasets. It is hoped that these can be progressed, but they are outside the scope of this Essay.

Finally, Chapter 7 draws some general conclusions and in Chapter 8 a full Bibliography is provided.

Author's Note

I was involved in the first UK National VTTS study as Technical Adviser and was a member of the ITS Leeds team which re-analysed the second study carried out by Accent and Hague Consulting Group [AHCG]. More recently, I was part of the Significance team carrying out the 2008 Dutch VTTS study and the third UK VTTS study (Arup, with ITS Leeds and Accent) in 2015.

I have acquired extensive knowledge of the RAND Europe modelling approach, having been part of the Transport Research Laboratory team that carried out a technical audit of the original version of the Dutch National Model [LMS], and since then I have advised the Dutch Ministry of Transport [Rijkswaterstaat] about further development of LMS over the period 2014-22. In addition, I advised Transport for London [TfL] in connection with the MoTiON model developed by Systra and RAND Europe.

Separately I have been advising Transport for the North [TfN] on the development of their NoRTMS model (carried out by Systra).

The Essay draws from this experience.

1 Preliminaries

1.1 Background

It took some time for the separate disciplines of transport models, micro-economic theory of the consumer and discrete choice modelling to reach a satisfactory concordance. In the 1960s and early 1970s, transport models were largely based on simple mathematical functions and physical analogies, with little statistical rigour. Advances in micro-economic theory during the 1970s led to the concept of *indirect* utility and its relation to demand, as set out in the pioneering work by Deaton & Muellbauer (1980) [D&M], while random utility theory was being developed by McFadden and others to account for discrete choices.

While economic theory naturally concentrated on (money) prices, it was recognised early (eg by McIntosh & Quarmby (1970)) [M&Q] that for transport models, both money and time needed to be dealt with, and the concept of generalised cost [GC] was promoted as a linear combination of the two. This led naturally to the need to provide a means of scaling the two quantities, and the "value of time" (more accurately, the value of travel time savings/changes [VTTS]) was identified as a critical parameter. It was gradually recognised that "generalised cost" was a simple (negative) version of indirect utility, and that most of the transport models in use could be re-cast as discrete choice models.

As was noted in the first UK National VTTS Study (MVA *et al*, 1987)[1] [3.5.1]: "Given that the basic [microeconomic] theory and the [discrete choice] model formulations both relate to the concept of utility, it is perhaps surprising that little effort is apparent in the literature to relate functional form to the *a priori* requirements of microeconomic theory." Some aspects of this were investigated in the first VTTS study: we will consider them further in subsequent sections of this paper.

A particular issue of interest related to the measure of benefit. D&M had demonstrated that the preferred measures – Compensating Variation [CV] and Equivalent Variation [EV] – could be obtained straightforwardly from the indirect utility function V, which is a function of income Y and the price vector **p**. For example, in the case of CV. we have

$$V\left(Y+CV,\mathbf{p}'\right)=V\left(Y,\mathbf{p}\right) \qquad (1.1)$$

where the prime ($'$) indicates the "after situation". Note that these are measures of negative benefit: for a price increase the CV represents the (positive) amount of income necessary to compensate the consumer to maintain the same utility, while EV is the (positive) amount of income that a consumer would be willing to forgo to avoid a price increase. Clearly the benefit to the consumer from a price increase is negative.

While from a strict theoretical point of view, "consumer surplus" [CS] is an inferior measure to either CV or EV, it can be shown that provided the marginal utility of income is constant, which will be the case if the indirect utility function is separable between income and prices, both EV and CV resolve to the CS measure.

In relation to discrete choice theory, with a choice between discrete elements $\{i\}$, the key result is due to Small & Rosen (1981) [S&R], who show that the benefit accruing to a "representative individual" can be written as

[1] Note that the study was carried out over the period 1980-85, with some interruptions

$$\Delta S = \frac{1}{\lambda}\sum_i \int_{\mathbf{V}^0}^{\mathbf{V}'} p_i(\mathbf{V}).dV_i \qquad \text{(based}^2 \text{ on their Eq (5.5a))} \qquad (1.2)$$

where p_i is the probability of choosing discrete alternative i, and λ is the marginal utility of income (more generally, a scaling factor to convert from units of utility to monetary values). Note that there is an integrability condition for this to be valid, but this is generally met in most discrete choice applications. As is well known, in the case of the logit model, there is a closed form solution to the integral in terms of the so-called "logsum" or "composite cost" formula, which we write as V^*, with the formula:

$$\text{when } p_i(\mathbf{V}) = \frac{\exp(V_i)}{\sum_{i'}\exp(V_{i'})},\ V^* = \ln\sum_{i'}\exp(V_{i'}) \qquad (1.3)$$

A detailed account of all this is provided in Bates (2003, 2006), where it is also shown that the composite cost formula for benefit is closely approximated by the "Rule of a Half" [RoH] provided the changes in utility are not too large (and when they are, the approximation can be rescued by means of a piece-wise linearisation as recommended by Nellthorp and Hyman (2001)). In other words, with the logit model

$$\Delta S = \frac{1}{\lambda}\sum_i \int_{\mathbf{V}^0}^{\mathbf{V}'} p_i(\mathbf{V}).dV_i = \frac{1}{\lambda}\left[V'^* - V^{0*}\right] \approx ½\frac{1}{\lambda}\sum_i \left(p_i'(\mathbf{V}') + p_i^0(\mathbf{V}^0)\right).\left(V_i' - V_i^0\right) \qquad (1.4)$$

1.2 Application to Transport Models

For the most part we will concentrate on the main component of transport demand – that of mode and destination ("distribution") choice [DMS], together with issues relative to assignment (which in itself is partly a component of demand relating to route choice). The use of unimodal models, which has particular relevance to rail, raises some further problems which will be addressed later.

The notion of generalised cost put forward in MAU Note 179 [M&Q] was explicitly distinguished according to three measures – behavioural for prediction (B), behavioural for benefits (U), and resource cost (R). Apart from questions of units (where B was "usually" in time units while U was in monetary units), the essential differences between these measures was that U allowed for "alternative values of non-work time as reflections of possible social values", while R was principally to allow for the true monetary costs as opposed to those "perceived" by the user, and related particularly to "considerable evidence to suggest that people significantly underestimate the costs of running cars[3]", as well as issues of taxation.

A linear function was proposed, as can be illustrated by the suggested behavioural cost function:

---

[2] S&R Eq 5.5a is $(\Delta E)/N = -\frac{1}{\lambda}\int_{\mathbf{W}^0}^{\mathbf{W}^f}\sum_j \pi_j(\mathbf{W}).dW_j$ where E is the aggregate Expenditure function (referred to by D&M as the "cost function") and N is the number of consumers. This gives the compensating variation [CV] which, as explained above, is the negative of the benefit. Hence in Eq (1.2) above the minus sign has been removed. I am grateful to Kenneth Small for clarifying this.

[3] The following references are cited: Harrison, A. J. (1969), LGORU (1968), Quarmby, D. A. (1967)

.

B = b_1.in-vehicle time + b_2.walking time + b_3.waiting and transfer time + b_4.travel cost (1.5)

Using standard notation with i = origin/production, j = destination/attraction, and m = mode, we re-write this as

$$B_{ijm} = \beta_t.IVT_{ijm} + \beta_w.Walk_{ijm} + \beta_h Wait_{ijm} + \beta_c.Cost_{ijm} \qquad (1.6)$$

In like spirit, the indirect utility functions for discrete choice analysis were typically linear in time and cost. On this basis, applying the RoH approximation to the S&R formula, as in Eq (1.4) above, the benefit (change in consumer surplus) can be written:

$$\Delta S \approx -½ \frac{1}{\beta_c} \sum_{ijm} \left(T'_{ijm}(\mathbf{B}') + T_i^0(\mathbf{B}^0) \right) . \left(B'_{ijm} - B^0_{ijm} \right) \qquad (1.7)$$

where the minus sign deals with the fact that the generalised cost is negative utility.

In line with the S&R approach, this uses the "behavioural" values (in other words, the negative indirect utility function) as it relates to the components which can be affected by the scheme/policy under consideration.

The linearity of the generalised cost formula allows the benefit to be allocated to the separate components (here IVT, Walk, Wait and Cost). It must be noted that, as TAG A1-3 (paragraph 3.1.5) points out, "This approach relates the breakdown of benefits to the mode of transport where the change in cost has occurred, and not to particular groups of travellers[4]." Note that by contrast the logsum is not decomposable into the constituent elements of "utility", and hence conveys less information than may be considered desirable.

A further advantage of the linearity is that for appraisal purposes agencies may wish to allocate different weights to the elements of generalised cost[5] from those that are being used in the demand model, even though this may give rise to inconsistencies, as pointed out, e.g., by Pearce & Nash, 1981:

"This inconsistency could lead to misallocation of resources; for example a scheme which gives the poor time savings at an increased money cost of travel could be selected in circumstances in which they would rather forgo the time savings for the sake of cheaper travel." (182)

A similar example, but from the opposite end of the income spectrum, is given by Sugden (1999), who argues strongly against such practice. Essentially, this is a political decision. The primary issue relates to the value of time [VTTS], where the value proposed for use in the benefit calculation may be different from the ratio $\frac{\beta_t}{\beta_c}$ in the behavioural cost formula. To avoid the impact of such inconsistencies, most recent models have tended to use the appraisal weights in the demand utility formulation. While this resolves the issue of incompatibility, it could lead to a less appropriate demand model, with potential consequences for forecasts.

[4] A fuller discussion is provided in Appendix E of the Common Appraisal Framework (MVA et al (1994))
[5] essentially, reflecting the differences between B and U in the McIntosh & Quarmby work.

On both theoretical and intuitive grounds, we can expect VTTS to vary with the income of the respondent, and this was recognised right from the start, in principle in relation to the cost coefficient. However, as noted in the 1987 VTTS study [§7.3], the early empirical evidence was surprisingly weak: it was only in the later phases of that study that significant results were first obtained. All subsequent studies, NB based on Stated Preference [SP] data, have confirmed a strong income effect.

Since income is a characteristic of the traveller rather than the journey, it can be dealt with by *segmentation*, and this is indeed allowed for in current modelling guidance [TAG M2.1], especially when the schemes/policies being tested involve charging. This does, nonetheless, cause some issues for appraisal which to date have not really been resolved.

Greater problems attend possible non-linearities in terms of the time and cost variables, and we now turn to this.

1.3 Non-linearities

As we will discuss later in more detail, we are talking here as to whether the derivatives of V with respect to time and/or cost are constant: clearly, inasfar as this may not be the case, there will be implications for VTTS. In the 1987 VTTS study, one of the possible Hypotheses put forward [H9] was that "The value of journey time savings may be related to the duration of the journey", and it was noted [4.2.12] that "It is plausible that a small saving on a short journey is more appreciated (or more easily perceived) than on a long one." The general question of non-linearity was further addressed in §4.9, and it is instructive to reflect on what was said there. The issues were illustrated in Figure 4.3 from that study, and the subsequent text, reproduced here:

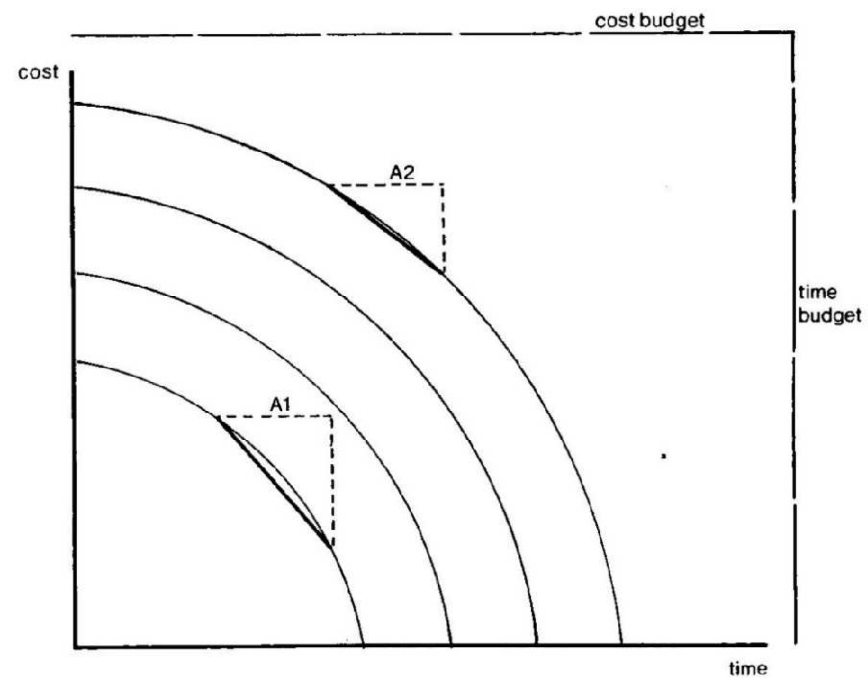

Figure 4.3 Lines of equal (Indirect) utility

66

Figure 1.1 – taken from MVA *et al* (1987)

"4.9.5 Now in our application of random utility theory, we deal with indirect utility functions, and the analysis of indifference curves is not strictly relevant. However, it is possible to construct iso-(indirect)-utility functions, in terms of travel expenditure in money and time, which have a formal resemblance to indifference curves. Since increases in time or money expenditure detract from the maximum utility that can be obtained with a given budget, the corresponding properties for these 'indifference curves' is that they should be concave to the origin, and that (indirect) utility decreases, the further we move from the origin. We can illustrate this by means of Figure 4.3….

"4.9.6 Although Figure 4.3 can be interpreted in many ways, it is convenient to consider it as referring to a particular journey by a particular mode (eg , travelling to work by train), and assume that there is some notional budget (not necessarily binding) for this journey in terms of "acceptable" cost and time. Then, if we have strict concavity, as illustrated, a number of useful results are immediately obtained. The value of time is given by the (negative) slope of the curve at any point in the (cost, time) plane. As cost increases towards the boundary of acceptability, the value of time falls, because travellers are increasingly unwilling to incur further money expenditure. Conversely, as time increases towards the boundary of acceptability, the value of time increases.

"4.9.7 Next consider two individuals, at the points Al and A2, where we may assume that because of home location, individual 2 has a longer and more expensive journey to work than individual 1. Suppose now that both individuals are faced with a time increase of x minutes; what reduction in cost would leave them with the same utility? Clearly, the answer is obtained by moving along the indifference curve until the time co-ordinate is equal to the existing journey time plus x minutes; the change in cost indicates the amount by which they would require compensation. The slope of the line joining the two points is an indication of the (non-marginal) value of time. Provided only that the indifference curves are reasonably 'parallel', the slope will be lower for individual 2, since the change in his travel time is proportionately less. These kinds of effects cannot be dealt with linear indirect utility functions, which imply constant values of time."

Unfortunately, it was not considered possible to pursue this line of thought beyond "fairly simple deviations from the constant value of time formulation, including piecewise linear functions:" [4.9.10], and no effects of significance were found in the 1987 VTTS study. We return to this in Chapter 3.

2 Empirical Evidence for Non-Linearity

2.1 Airport surface mode studies

Independently of the first VTTS study, the British Railways Operation Research unit were using the 1984 CAA [Civil Aviation Authority] Passenger survey to estimate a logit model for airport surface access mode choice for Stansted. They found that the sensitivity to changes varied with distance D, and proposed a utility function in which generalised cost (NB excluding interchange penalty) was scaled by $D^{-0.25}$, provided that $D \geq 16$ miles. The final model was "calibrated on a subset of the 1984 data by British Rail, assisted and monitored by MVA", 1986-87.

The model was re-estimated by MVA in 1994, on 1991 CAA data, where the (negative) distance exponent was increased to 0.5. A subsequent re-estimation by SKM in 2006, on 2003 CAA data, kept the distance exponent at 0.5 for Leisure, but reduced it to 0.4 for Business. Initial work on 2009 CAA data confirmed these results, but further work was terminated in April 2010. It is not known whether subsequent re-estimation has been carried out.

Essentially the model can be written as

$$V_{ijm} = \alpha_m + f(\text{interchange}) + \frac{\beta_t.IVT_{ijm} + \beta_w.WalkWait_{ijm} + \beta_c.Cost_{ijm}}{D_{ij}^{\theta}} \qquad (2.1)$$

where D_{ij} is defined as the highway distance between origin i and airport j, m is the access mode, and θ is the exponent on distance. Note that this formulation does not impact on the (implied) VTTS. Tests to investigate whether the interchange function should be included in the scope of the distance effect have consistently produced a negative result.

2.2 "RAND" models

Dutch National Model[6]

Another independent investigation was being carried out in the Netherlands for the development of the Dutch National Model [LMS – Landelijk Model Systeem]. Following extensive exploratory work[7], Hague Consulting Group[8], led by Andrew Daly and Hugh Gunn, were commissioned in 1983 to develop LMS and the initial version was released in 1986.

We will concentrate on the mode/destination choice model, which was estimated on OVD data (though subsequent models have used the Dutch equivalent of the NTS). Models were defined for five purposes, and for four modes (Car driver, Car passenger, Transit and Slow). Unusually, this was an MNL model (thus no hierarchy between mode and destination

[6] Most of the information in this section is drawn from unpublished papers/reports which have been made available to the author in the course of his work on the LMS: as a result, no formal references are given, though the documents are identified.

[7] Zuidvleugel Study (1980), Overdraagbaarheidstudie [OVD] (Transferability) (1985).

[8] Hague Consulting Group later became RAND Europe based in UK, with a "sister" organisation "Significance" in the Netherlands.

choice), though later versions have generally made use of a nested logit [NL] specification[9]. Ignoring the slow mode, which used a measure of distance, the equivalent of generalised cost for the motorised modes was as follows:

$$V_{ijm} = \alpha_m + \beta_t^m.IVT_{ijm} + \beta_w.Walk_{ijm} + \beta_h.Wait_{ijm} + \beta_c.\ln\left(Cost_{ijm} + 1\right) \qquad (2.2)$$

The IVT coefficient varied by mode, and for car driver (for purposes other than education) there was an additional term when the round-trip time was greater than a given "cut-off", defined as 40 minutes for most purposes, but 80 minutes in the case of "other work" (EB). For all purposes other than "other work", this was negative, thus increasing the disutility for longer (in duration) journeys. Time was measured in minutes and cost in Dutch guilders (*fl*, 1977 prices). To discourage longer distance car passengers, an additional distance squared term was included for that mode. For car journeys up to 10 Km, an additional positive constant was included in the utility of both car modes.

With regard to the cost term, the addition of *fl*1 is, of course, to avoid the logarithm of zero, though, as we will see, there are problems associated with this. Regarding the log transformation, the documentation[10] states:

> *The models use a transformation of cost, taking the natural logarithm as the appropriate variable. This specification was originally suggested after examination of the elasticity of car-driver demand to petrol prices; it was found that with the conventional linear specification, the impacts of price rises/falls were much greater on long-distance trips than on short distance ones. The same was true of public transport travel. In fact, the actual experience of historic price rises was not felt to bear out such an effect: rather, the effect was reasonably uniform over all distances. The log specification (which would tend to produce such a uniform result over all distance classes) was accordingly tested against the linear alternative, and was found to provide a statistically superior model for <u>all five travel purposes</u>.*

As is well understood, this form of the cost function means that VTTS is not constant, but has the form $VTTS = \frac{\beta_t^m}{\beta_c}\left(Cost_{ijm} + \delta\right)$ (where δ was set equal to 1) so that it increases linearly with cost. There is no evidence in the LMS documentation that a log transform was tried for time.

It is important to note that the log transform is very sensitive to the assumption made about the addition of the constant δ, so that it can be expected that the estimated coefficient β_c will reflect this. The figure below shows how different the log Cost function looks for different assumptions about δ: note that in terms of the scale, the X-axis is in 1977 guilders and *fl*15 is compatible with the LMS assumption of the cost for a round-trip journey by car of 100 Km.

---

[9] By 2001, the structure had mode choice above destination choice for HB Shopping and HB Other as well as NHB Business, but destination choice above mode choice for Commuting and HB Business. Other purposes (HB Education and other "child" purposes retained MNL). This endured till LMS7: in later versions, there were no instances of destination choice above mode choice.

[10] Resource Papers for the Landelijk Model vol 2 August 1989, Paper 1

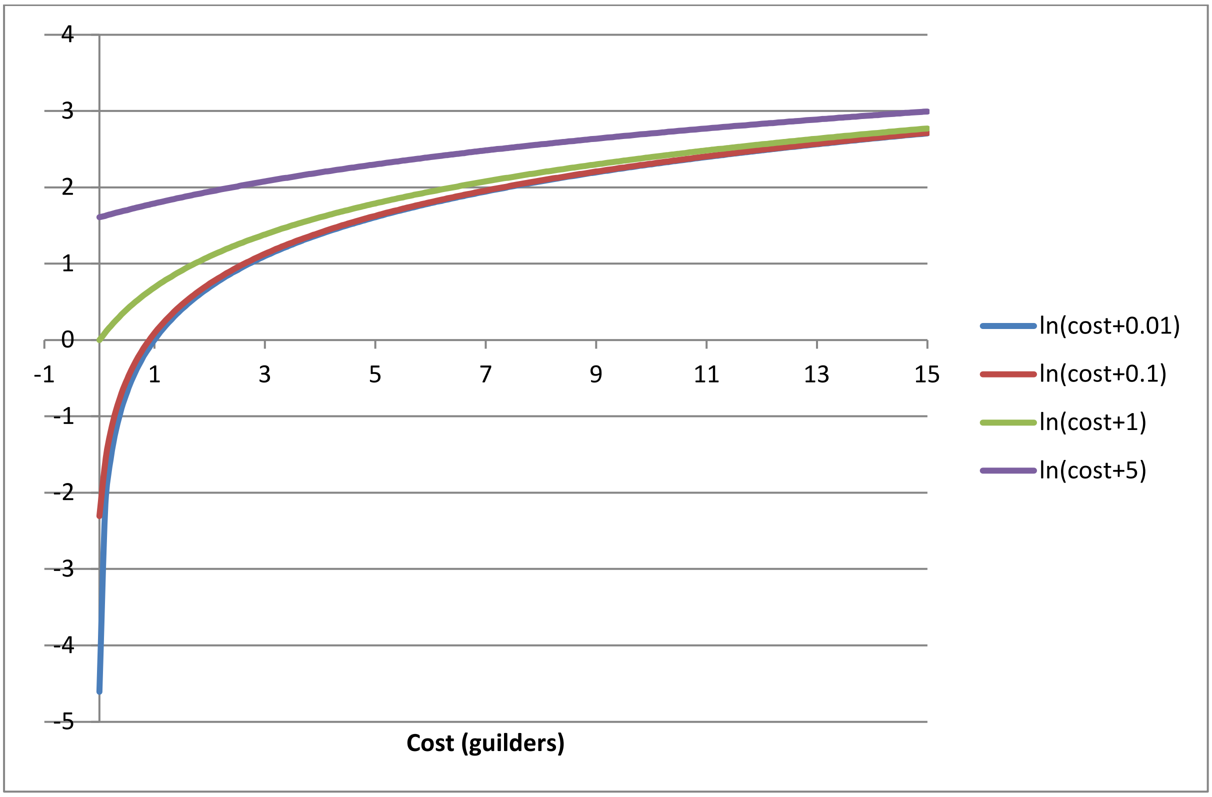

Figure 2.1: Variation in cost function with different assumptions about δ

The LMS has been frequently updated since the original version: the range of purposes and modes has been expanded. Nevertheless, with minor variations (addition of income segmentation to the cost coefficient for some purposes), the essential specification as it relates to generalised cost was maintained up to and including LMS7. In moving to the 2010 version [LMS2010], the cost function was investigated and attention was paid to the constant δ (termed, for some unexplained reason, the "Tukey-constant"). The following text is translated from the Dutch[11].

> "The reason [for the investigation] is mainly because of the use of a so-called Tukey-constant in the logarithmic cost function. ….The logarithmic form allows for cost damping in the model: the effect that the cost sensitivity reduces (and so the value-of-time increases) with the length of a tour. The Tukey-constant is intended to prevent the cost sensitivity from being too high for very short tours. Thus the Tukey-constant acts as a kind of intermediary between a fully logarithmic and a fully linear cost function.
>
> "The logarithmic cost function suffers from the problem that the value of the Tukey-constant cannot be satisfactorily supported, while it also has a large influence on the cost elasticity. Therefore alternative specifications were tested…. The most important alternatives were:
> 1. Both linear and logarithmic cost function (without Tukey-constant) in het model;
> 2. Linear and logarithmic cost function with a gamma-mixfactor;
> 3. logarithmic with cut-off for low costs;
> 4. Linear for low costs, transferring to logarithmic for higher costs.
> The last named specification was chosen….. with the following specification:
> $u = \ldots + \beta_1 \text{ cost}/e + \ldots$ if cost < e = 2.718 (Euro)
> $u = \ldots + \beta_1 \ln(\text{cost}) + \ldots$ if cost ≥ e "

[11] "Schattingen van keuzemodellen voor het LMS 2011"§ 8.4.4, Significance 17 June 2011

While it is clearly necessary to avoid a discontinuity at the cut-off point, and this is achieved, specifying this point at €2.718 seems unnecessarily restrictive[12]. In subsequent versions (GM3, GM4) this cost specification was abandoned in favour of the "gamma mix-factor" model proposed by Fox *et al* (2009), as follows:

$$V_i^{\text{cost}} = \beta_{\text{cost}} . \left[\gamma.\text{cost} + (1 - \gamma).\ln(\text{cost}). \frac{\text{E[cost]}}{\text{E[ln(cost)]}} \right] \qquad (2.3)$$

To avoid taking the log of zero, a minimum value of cost is imposed (currently €0.01), though for those with free public transport, the entire cost function is set to zero[13].

The "mix-factor" γ must lie in the range [0,1]: a value of γ = 1 implies a linear cost function, while a value of γ = 0 implies a logarithmic cost function. The resulting formula for VTTS is:

$$VTTS = \frac{\beta_t}{\beta_c \left[\gamma + \frac{(1\text{-}\gamma)}{\text{cost}} . \frac{\text{E[cost]}}{\text{E[ln(cost)]}} \right]} \qquad (2.4)$$

When γ ≠ 1, VTTS is a function of cost.

According to Fox *et al* (2009), "The factor giving the ratio of the mean costs is necessary to normalise the contribution of the log cost term to be on the same scale as the linear cost term.)". However, if we consider a general function of the form $\beta_c.\text{cost} + \beta_L.\ln(\text{cost})$, **provided** both coefficients have the same sign (negative)[14], the two formulations are equivalent, since $\beta_c = \gamma.\beta_{\text{cost}}$ and $\beta_L = \alpha \frac{(1\text{-}\gamma)}{\gamma}.\beta_c$ (where α is the ratio of mean cost to mean log cost), and we can solve for γ.

In practice, with the Fox formulation, γ is obtained by means of a grid search[15]. But the Fox formulation does have the advantage of **ensuring** that both coefficients have the same sign, given the restriction on the range of γ.

For the general function with cost and logcost, $$VTTS = \frac{\beta_t}{\beta_c + \frac{\beta_L}{\text{cost}}} \qquad (2.5)$$

Manchester Motorway Box[16]

This was a study carried out by RAND Europe in conjunction with Mott MacDonald and The Denvil Coombe Consultancy for DfT over the period 2005-10. A combination of linear and

[12] If K is the cutoff, the linear part of the function can be defined as +β1 (ln(K)/K).cost
[13] I am grateful to Gerard de Jong of Significance for providing this information
[14] which, as we discuss later, is necessary to ensure $\frac{\partial V}{\partial c} < 0$
[15] In one of the LMS documents it is claimed (translation from Dutch, [10055-R1- Herijking LMS_v3.pdf], Significance 2010): "Maximising the log likelihood by allowing the logarithmic and linear cost coefficients to be freely estimated can deliver quite unreasonable values for elasticities and/or values of time."
[16] Fox and Daly (2013)

log cost terms were tested. For the models based on household interview surveys[17], the authors report the following [§6.1]:

> *For most model purposes, the best model fit was obtained with cost entering the utilities in separately linear and log-cost terms. The log-cost term has the most effect at the short-distance trip range. For employer's business, where trip lengths are longer and the volume of data is lower, it was not possible to identify both linear and log-cost terms; the final model contains a log-cost term only.*

However, when the household data was pooled with data from roadside and public transport intercept data, this resulted in some changes [§7.2]:

> *For business, the log-cost only formulation resulted in a positive car time parameter; when a linear-cost only formulation was tested instead, the car time parameter improved but PT in-vehicle time became insignificant. The final model specification used linear-cost only, and a separate PT in-vehicle time parameter for the PT intercept data, which was identified as the cause of the difficulties with the cost and time parameters. In the PT intercept model for business, reported in 5.2 a positive PT in-vehicle time parameter was also obtained.*
> *For other travel, the linear cost parameter was positive in the pooled model and was therefore dropped. The results with log-cost alone were plausible.*

*West Midlands PRISM (**P**olicy **R**esponsive **I**ntegrated **S**trategy **M**odel)*[18]

This model was estimated on choice data from Household interviews collected between 2009 and 2012. An earlier version of the model had been constructed between 2002 and 2004 based on 2001 data.

The Fox formulation for the cost function was used for this model, noting the following (p 41):

> *When the 2006 base version of PRISM was validated against the guidance elasticity values in WebTAG, the fuel cost elasticities were observed to be lower than values recommended in WebTAG. As a result, in 2009 the PRISM mode-destination models were re-estimated with both log and linear cost terms in the utility specifications. For most model purposes, a combined log and linear cost specification resulted in an improvement in model fit, plausible values of time, and higher and more plausible fuel cost elasticities, and therefore the models with both linear and log cost terms were incorporated into the 2006 base version of PRISM. However, for other travel, the linear cost term was not significant with this formulation, and the log-cost only formulation remained too inelastic to changes in fuel cost. Therefore a procedure was used to impose a mixture of linear and log cost into the model which allowed a log-linear mixture to be introduced into the model as a single term. This procedure was termed the 'gamma formulation'…*

In fact, the model estimation started with separate linear and log coefficients, but rejected these in favour of the Fox formulation in order to achieve satisfactory VTTS and demand

[17] Income data was not available in these surveys
[18] Fox *et al* (2014)

elasticities with respect to cost, as well as cases of unacceptable positive linear coefficients. The following results were obtained:

Table 51: Summary of cost specifications

Purpose	Cost specification
Home-based tours	
commuting	gamma specification, gamma = 0.45
home–business	VoTs imported from WebTAG with a distance damping
home–primary education	log-cost only
home–secondary education	log-cost only
home–tertiary education	gamma specification, gamma = 0.25
home–shopping	gamma specification, gamma = 0.1
home–escort	gamma specification, gamma = 0.4
home–other travel	gamma specification, gamma = 0.15
Primary Destination [PD]-based tours	
work-related PD to work-related SD	log-cost only
work-related PD to other SD	log-cost only
other PD to other SD	VoTs imported from WebTAG, linear generalised time formulation
"detours"	
during work-related tours to work-related SDs	gamma specification, gamma = 0.55
during work-related tours to other SDs	gamma specification, gamma = 0.5
during other tours to other SDs	gamma specification, gamma = 0.3

In some cases, these final results had been modified from earlier findings because of changes in overall utility specification: it is clear that the choice of gamma is sensitive to this. Overall, as the discussion in the Report makes clear, considerable judgment is required.

Note that for some purposes it was possible to include income segmentation for the cost coefficient.

Transport for London

The Fox formulation was also retained for the TfL model MoTiON, though the more general form was used as long as both coefficients had the correct negative sign. The work built on the earlier PRISM experience. Two versions of the model were produced, the first (Phase 1/2) based on household data for 2010–2012 while the second (Phase 3) was based on household data for 2015/16–2017/18.

In the first model, for commuting and HB tertiary education, separate linear and log cost coefficients could be estimated, but for HB Business, it was necessary to move to the Fox formulation with a low value of γ (0.01). For the home–shopping, home–escort and home–other travel models it was not possible to estimate linear and logarithmic terms nor models using a gamma cost specification that gave rise to plausible values of time. In particular, the cost parameters were weakly estimated leading to implausibly high VTTS. Given these differences, cost sensitivity information was imported from the commute model, with adjustments for VTTS based on TAG.

For all of the NHB models except the work-work detour model, VTTS could not be estimated and was therefore imported from WebTAG using the distance function given in the WebTAG databook. VTTS was used to convert costs into time units, and sensitivities to generalised time were then estimated separately by mode. For work-work tours the WebTAG distance function for car travel in work time was used, taken from Table A1.3.1 of the WebTAG databook):

$$VTTS = \frac{A}{\left(1 + e^{\frac{x_{mid} - D}{k}}\right)} \tag{2.6}$$

For the non-work purposes, using distance varying VTTS was found to give a better fit to the data than using fixed VTTS. The distance varying VTTS formulation was taken from WebTAG Unit M2.1:

$$VTTS_d = VTTS\left(\frac{d}{d_0}\right)^{\eta_c}. \tag{2.7}$$

For the work–work detour model, a pure log-cost formulation was used instead of importing VTTS from WebTAG.

In the Phase 3 model, the commuting model used a Fox formulation with a relatively high γ value (0.6), while HB Business had a value of 0.05. As before, it was not possible to freely estimate models with acceptable cost sensitivities for home–tertiary education, home–shopping, home–escort, home–other, work–other tours and other–other tours. But this time it was found that the distance variations from TAG Unit M2 produced cost elasticities that were too high. As in the earlier models, cost sensitivity information was imported from the commute model, with adjustments for VTTS based on TAG. Note that, given the high value of γ for the commuting model, the extent of cost damping is relatively low.

For the work–work tour purpose models with using business VoTs incorporating distance variation from TAG were tested but yielded unacceptably high fuel cost elasticities, even when additional generalised time damping was introduced. Hence cost sensitivity information was imported from the HB business model. For the three NHB detour models it was possible to estimate models with acceptable cost sensitivity parameters using either freely estimated linear and log cost parameters or Fox formulations.

Conclusion on the RAND models

Overall, while all these models have included some form of damping for the cost term, we may note that a) it has not always been possible to estimate the (finally) preferred form, b) hence, different forms of damping have been implemented and c) the final forms have not always been consistent between successive estimations on different data sets. A considerable amount of judgment is required in order to achieve acceptable tradeoffs between model fit, credible VTTS and the TAG realism tests. Finally, a residual question is whether the damping applies more to mode choice or to destination choice (or equally to both): the RAND models do not provide any evidence on this as they have all assumed that any possible non-linearity relates to the cost function which is common to both choices.

2.3 VTTS studies

UK 1980s

As noted earlier, the MVA *et al* (1987) study did not produce any evidence of variation in VTTS with distance (etc.). However, subsequent studies **have** found consistent evidence. These studies, based almost entirely on binary SP experiments which may be termed "route choice", have used a variety of estimation methods, but can all broadly be classified as having a generalised cost formulation of $\Delta V_m = \beta_t^m.\Delta IVT_m + \beta_c.\Delta Cost_m$. The variations on Cost and Time were in all cases presented relative to "reference values" C_{ref} and T_{ref} based on the respondent's actual journey. While income effects have always been investigated, they have been applied to the cost coefficient, as theory would suggest. Since these are uncontroversial, we ignore them in the following discussion (though we note that there is some correlation between income and distance travelled).

AHCG 1995 study

While this study reported in 1999[19], the DfT had some difficulties in implementing its recommendations, and decided to review the analysis. The subsequent re-estimation of the AHCG data (which applied to the car mode only) by ITS *et al*[20], carried out over the period 2001-03, examined possible effects on VTTS due to "distance", though it must be pointed out that the actual distance of the reference journey was not available in the data. An early conclusion was that while there were strong non-linearities in the utility formulation, these were associated with the reference values rather than the implied absolute values presented. In particular, the form $\Delta V = \beta_t.\Delta IVT + \beta_L.\Delta[\ln\left(Cost_{ref} + \Delta Cost\right)]$ performed quite badly.

No significant damping effects were found for the time variable, but those for the cost variable were significant (for all purposes investigated), and the preferred utility form was

$$\Delta V = \beta_t.\Delta IVT + \beta_c.\left(\frac{C_{ref}}{C_0}\right)^{\lambda_C} \Delta Cost, \qquad (2.5)^{21}$$

where C_0 is an arbitrary base value to stabilise the estimation. The elasticity λ_c is negative, so that the absolute value of the sensitivity to cost declines with the reference cost. Hence VTTS increases with cost. The elasticities were 0.42 for Commuting and 0.32 for Other.

Dutch VTTS Study begun 2007

Over the period 2007-13, a corresponding VTTS study was carried out in the Netherlands[22]. For passengers, the main form of the utility function can be written as

$$\Delta V = \beta_t.\left(\frac{T_{ref}}{T_0}\right)^{\lambda_T} \Delta IVT + \beta_c.\left(\frac{C_{ref}}{C_0}\right)^{\lambda_C} \Delta Cost, \qquad (2.6)^{23}$$

[19] AHCG (1999)
[20] Mackie *et al* (2003)
[21] An income elasticity was included for the cost coefficient
[22] Significance *et al* (2013)
[23] An income elasticity was included for the cost coefficient

and this time significant elasticities were found for both variables. Once again, the "ref" subscript denotes the reference value (ie the cost and time of the journey around which the SP choices are designed, while the "0" subscript is an arbitrary value to stabilise the estimation.

The value for λ_C ranges from -0.16 to -0.55 with most estimates approximately equal to -0.35, while the value for λ_T ranges from -0.19 to -0.63 with most estimates approximately equal to -0.4. The fact that both elasticities are negative implies that as the reference time or cost increases people will be less sensitive to changes in these variables.

The estimated values for λ_C and λ_T are as follows:

Mode	Car			Train			Other PT			Plane	
Purpose	Com	EB	Oth	Com	EB	Oth	Com	EB	Oth	EB	Oth
λ_C	-0.350	-0.356	-0.345	-0.303	-0.358	-0.161	-0.246	-0.479	-0.232	-0.569	-0.615
λ_T	-0.381	-0.391	-0.411	-0.411	-0.382	-0.222	-0.585	-0.194*	-0.448	-0.713	-0.654

* not significant (t-ratio = 1.1)

It is clear that the effect on VTTS will depend to a considerable extent[24] on the **difference** between the two elasticities ($\lambda_T - \lambda_C$), since

$$VTTS = \frac{\beta_t}{\beta_c} \cdot \left(\frac{T_{ref}}{T_0}\right)^{\lambda_T} \cdot \left(\frac{C_{ref}}{C_0}\right)^{-\lambda_C}. \qquad (2.7)$$

In most cases, however, it can be seen that in absolute value terms $|\lambda_T| > |\lambda_C|$, so inasfar as both time and cost are linearly dependent on distance (which may of course **not** be the case), this would imply that VTTS **falls** with distance. However, the difference may not be significant (the report does not provide information about the correlation between the two elasticities).

UK VTTS Study 2014

The method of estimation for the 2014 Arup *et al* study made very different assumptions about the error structure etc., which resulted in significant improvements in overall model fit, as well as incorporating important theoretical developments (including multiplicative error terms and the use of 'random valuation' as opposed to 'random utility' models).

Ignoring the issue of sign and size effects, which complicates the analysis somewhat, the essential formulation can be written as follows:

Starting with $\Delta V = \beta_t \Delta T + \beta_c \Delta C$, where the difference Δ is defined as the less expensive (and slower) option *minus* the more expensive (and faster), this implies that the least expensive option will be chosen whenever $\Delta V = \beta_t \Delta T + \beta_c \Delta C > 0$, and since ΔT is positive

[24] The extent depends on the correlation between T_{ref} and C_{ref}: if they are completely correlated, the effect will depend entirely on the difference. Generally, we expect considerable correlation of both variables with distance.

by definition, we have $VTTS = \frac{\beta_t}{\beta_c} > -\frac{\Delta C}{\Delta T}$. In place of the standard additive error term applied to V, the analysis used a logarithmic transform so that the model for estimation was $\Delta V = \mu.\ln\left(\frac{-\Delta C}{VTTS.\Delta T}\right) + \varepsilon$ while VTTS was expanded to

$$VTTS = VTTS_{ref}.\left(\frac{Y_{ref}}{Y_0}\right)^{\lambda_Y}.\left(\frac{C_{ref}}{C_0}\right)^{\lambda_C}.\left(\frac{T_{ref}}{T_0}\right)^{\lambda_T}.\left(\frac{D_{ref}}{D_0}\right)^{\lambda_D} \quad (2.8)$$

to take account of variations with the reference values. Here Y is income and D is distance.

In spite of the different approach, it is useful from a cost damping point of view to assume the same utility function as in Eq (2.6), together with an income elasticity and a distance elasticity. While Eq (2.8) makes clear that all the elasticities were in fact applied to VTTS, we will act as if the time and cost elasticities are applied to the time and cost terms respectively, which means that the sign on λ_C is changed[25]. We will interpret the distance elasticity as applying to the cost coefficient, again changing the sign[26]. As with the previous studies, we are not discussing the income effect.

The estimated values are as follows:

Mode	Car			Train			Other PT			Bus	
Purpose	Com	EB	Oth	Com	EB	Oth	Com	EB	Oth	Com	Oth
λ_C	-0.679	-0.451	-1.049	-0.664	-0.743	-0.598	-0.409	-	-0.210	-0.523	-0.565
λ_T	-0.624	-0.454	-0.927	-0.275	-0.348	-0.541	-0.267	-0.488	-0.217	-0.576	-0.347
λ_D	-	-0.239	-	-	-0.06*	-	-	-	-	-0.15*	-0.07*

* not significant (t-ratio = 1.15, 1.88, 1.36)

In most cases, in absolute value terms $|\lambda_C| > |\lambda_T|$, so that there is an implication that VTTS rises with distance. However, there are exceptions: for Car EB $|\lambda_T| > |\lambda_C|$ by a small amount, though the negative distance elasticity will rescue this; Other PT EB did not recover any significant value for λ_C, so it implies VTTS will fall with distance; Other PT Other has $|\lambda_T| > |\lambda_C|$ by a small amount, this time with no compensating distance elasticity, and for Bus Commute $|\lambda_T| > |\lambda_C|$, though this will be slightly offset by the relatively weak negative distance elasticity.

Citing from the VTTS study:

(7.6.4) *In all cases, the elasticities to time and cost are more or less equal and opposite: they are substantially higher than was found in the 2003 work (where the variation related to cost only), suggesting a large proportionate response. However, typically these two coefficients are very highly (negatively) correlated: e.g. for car the correlations were -0.928 for commute*

[25] In the discussion of this study we have throughout changed the sign of λ_C as given in the published documents

[26] Given the signs of the cost and time elasticities, it seems reasonable to interpret the distance elasticity as reducing the sensitivity to cost, rather than increasing the sensitivity to time.

and -0.936 for other. This suggests that the absolute levels are less reliable than the difference between the absolute values.

In fact, though, the correlations are not so high for other modes. The table below gives the correlations[27] and calculates the accuracy of the difference in $(\lambda_T - \lambda_C)$[28]:

	correlation	$\lambda_T - \lambda_C$	var =	t=
Car commute	0.92793	0.0549	0.009284	0.57
Car other	0.93625	0.12187	0.005336	1.67
Rail commute	0.65015	0.38865	0.007252	4.6
Rail other	0.63295	0.05767	0.003482	0.98
OtherPT commute	0.29544	0.14146	0.018254	1.05
OtherPT other	0.29106	−0.00719	0.012285	−0.06
Bus commute	0.16773	−0.05311	0.037622	−0.3
Bus other	0.4008	0.21826	0.024217	1.4

In two cases the difference is of the wrong sign, and only in the case of rail commute is it significant.

In the study, it was felt that a more appropriate way of understanding the practical impact of distance and income was to use the NTS Implementation Tool to calculate the VTTS for each record in the NTS data, unweighted by distance, income, or other weighting configuration. Then a regression analysis of the form

$$\ln(\text{VTTS}) = \text{constant} + \omega_D.\ln(\text{trip\_distance}) + \omega_Y.\ln(\text{income}) \qquad (2.9)$$

was conducted, thus directly estimating the distance and income elasticities. The calculations in the tool for VTTS took account of most of the relevant variables estimated in the models, and thus also of the observed correlation between income and distance (as well as other covariates). These produced a rather stronger relationship with distance than the above table would imply, and it was concluded that these estimates from the tool were more reliable (especially as they do not require the assumption that time and cost are linear with distance and they account for how other covariates vary with time and cost).

The resulting Distance elasticity values are set out in the Table below (taken from Table 7.15 of Arup *et al* (2014)):

[27] The evidence for these is in the author's archive – he has never seen the corresponding results for Employees' Business, and it has not been possible to locate these

[28] using the formula: var(x+y) = var(x) + 2cov(x,y) + var(y)

Mode	Purpose	Estimated by summing distance, time and cost elasticities	Estimated from regression analysis of VTTS output by the Tool
Car	Commuting	0.055	0.179
	Employees' business	0.236	0.340
	Other non-work	0.122	0.298
Rail	Commuting	0.389	0.306
	Employees' business	0.453	0.370
	Other non-work	0.058	0.088
Bus	Commuting	0.099	–0.037
	Other non-work	0.290	0.063
'Other PT'	Commuting	0.141	0.043
	Employees' business	–0.488	0.0154**
	Other non-work	–0.007	–0.072

** Insignificantly different from zero at the 5% level

At least for Car, the elasticities from the tool are rather stronger for all purposes, though for other modes it is less clear. It would be useful to know how much of the difference is due to the possible errors in the assumption of time and cost increasing linearly with distance, and how much is due to the influence of other co-variates, in particular income. Certainly there is some evidence of distance increasing with income, as the graph below[29], based on data from NTS 2002-10, shows.

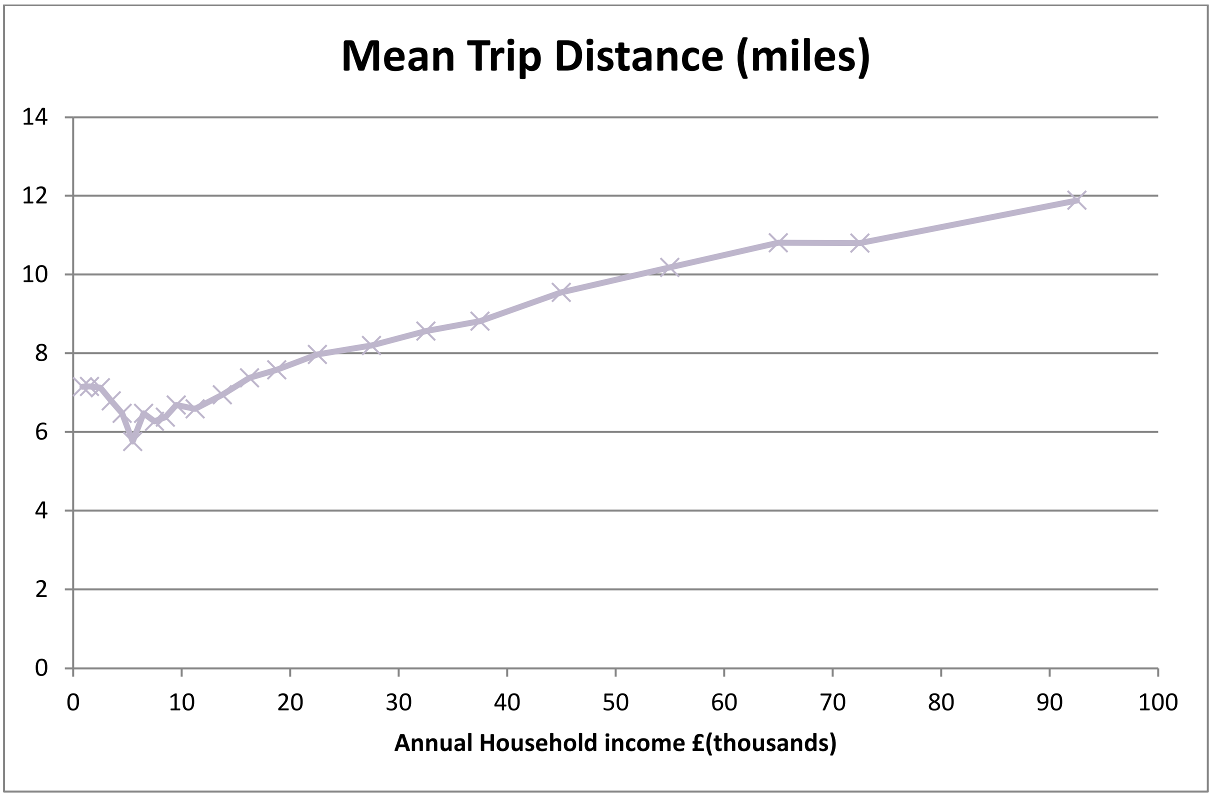

Figure 2.2: relationship between average trip length and household income

[29] author's own analysis of NTS data for 2002–2010, kindly provided by the UK Data Archive at the University of Essex

The VTTS report notes that the values for the distance elasticity taken from the 2003 study (NB for car only) were 0.421 (commute) and 0.315 (other non-work), so the new car values are lower, though the 2003 study values in WebTAG were really giving a cost elasticity (not distance) and were based on models that exclude several other covariates (but include income). It suggests that the "distance" elasticities from the 2003 study "were high, as the evidence base (e.g. Abrantes and Wardman, 2011) is reasonably well aligned to the implied elasticities in [the] Table".

It should also be noted that in addition to the effects noted the study allowed for a random distribution of VTTS, using a log-uniform distribution, revealing significant residual heterogeneity for all purposes, with all coefficients of variation approximately equal to 0.75.

Dutch VTTS Study 2022

The most recent VTTS study in the Netherlands was carried out in 2022[30].

Concentrating again on the GC specification and ignoring the income effect[31], the essential utility (in WTP space) can be written as:

$$V=\left[\mu.\left(\frac{T_{ref}}{T_0}\right)^{\lambda_{T,\mu}}.\left(\frac{C_{ref}}{C_0}\right)^{-\lambda_{C,\mu}}\right].\left(\Delta C+VTTS.\Delta T.\left(\frac{T_{ref}}{T_0}\right)^{\lambda_T}.\left(\frac{C_{ref}}{C_0}\right)^{-\lambda_C}\right) \qquad (2.10)$$

where for comparability we have again changed the sign of λ_C as given in the published document.

If we set $\mu=\beta_C$ and interpret VTTS as $\dfrac{\beta_T}{\beta_C}$, we can recast this (in preference space) as

$$V=\left[\left(\frac{T_{ref}}{T_0}\right)^{\lambda_{T,\mu}}.\left(\frac{C_{ref}}{C_0}\right)^{-(\lambda_{C,\mu}+\lambda_C)}\right].\left(\beta_C.\left(\frac{C_{ref}}{C_0}\right)^{\lambda_C}.\Delta C+\beta_T.\left(\frac{T_{ref}}{T_0}\right)^{\lambda_T}.\Delta T\right) \qquad (2.11)$$

Note that in this case the elasticities relating to the reference values affect not only the estimated VTTS but also the scaling parameter for utility: essentially, therefore, this detects cost and time sensitivity effects on generalised cost as well.

Unlike the UK study, the elasticities were not allowed to vary with purpose (though the VTTS itself was). Separate models were estimated by the three main modes (Car, Train and "Local PT"). As in the UK study, the study allowed for a random distribution of VTTS, in this case using a log-normal distribution.

Taking the results presented in Table 15 [Estimated coefficients for joint SP1A/2A models for car, train, local public transport], we obtain the following:

---

[30] Significance (2023)

[31] An income elasticity for VTTS was included in the specification.

Mode	Car	Train	Local PT
λ_C	-0.3339	-0.3142	-0.4374
λ_T	0.04922**	-0.0191**	-0.0882**
$\lambda_{C,\mu}$	0.2365	0.2721	0.2596
$\lambda_{T,\mu}$	-0.8851	-0.7302	-0.7215
Hence			
$\lambda_T - \lambda_C$	0.3831	0.2951	0.3492
$-\left(\lambda_{C,\mu} + \lambda_C\right)$	0.0974	0.0421	0.1778

** Insignificantly different from zero at the 5% level

The sign of $\lambda_T - \lambda_C$ suggests a significantly increasing distance effect on VTTS. In relation to these values, the authors comment: "The VTT and VTTR depend strongly on the BaseCost (elasticity between 0.3 and 0.4), but they hardly depend on BaseTime (elasticity between -0.1 and 0.1, with very low t-ratios). However, the BaseCost and BaseTime values themselves are strongly correlated, so the VTT has a strong correlation with BaseTime, as was found in other studies."

The report contains no commentary on the elasticity effects on the **scale**. However, we may note that here it is the **time** elasticity ($\lambda_{T,\mu}$, strongly negative) which suggests a decreasing scale with distance, leading to greater randomness in the model: by contrast, the effect from cost [$-\left(\lambda_{C,\mu} + \lambda_C\right)$] is slightly positive, though probably not significant.

Conclusions from the VTTS studies

In all the studies investigated (apart from the earliest UK study), the response to changes in time and cost (NB relative to a reference value) has been found to some extent to vary with some indicator of the reference journey length. The relative strength of the effect seems variable, and given the likely (though unknown[32]) high correlation between journey time, journey cost and journey distance, there are potential difficulties in being sure about the allocation of the size of the effect between them. Only the latest Dutch study has attempted to distinguish between separate effects on the (changes in) cost and time and the overall scaling factor for the utility function.

2.4 Overall Conclusion from empirical evidence

While the review has concentrated on British and Dutch work (especially that in which the author has had some involvement), these are probably the most valuable sources of evidence available. The most consistent result to emerge is that there are what we may call reduced sensitivity effects with both cost and time increases[33]. Generally – though not always – the cost effects are larger, and this leads to VTTS increasing with "distance". This effect has been picked up in the RAND models by means of the "log cost" formulation but we should note that it has been entirely attributed to the cost variable **by assumption**. The 2022 Dutch

[32] It would be valuable to investigate this correlation, for example using the 2014 UK VTTS data.

[33] As we discuss in §4.2, this is in contradiction to the "budget" effect from micro-economic theory

VTTS study again raises the possibility that there could be a more general effect on the scale parameter, which – interestingly – reflects the airport access studies.

Daly & Carrasco (2009), which we will discuss later, cite other evidence: VTTS studies in Sweden and Norway as well as other "RAND models" in Paris, Sydney and West Midlands. The 2004 Swiss VTTS study[34] also found VTTS increasing with distance, though again this was from an elasticity (for distance) applied to the cost coefficient only.

In respect of the revealed preference studies, it is worth noting that there is often considerable uncertainty relating to the cost variable in the case of the car mode: usually some assumption is made about the likely fuel cost, shared among the occupants, but in practice the costs relating to an individual journey may be perceived very differently. Likewise, for public transport, there may be issues relating to special discount fares, and the allocation of period tickets (e.g. season tickets) to individual trips.

As we will discuss later, the reduced sensitivity makes more sense in respect of mode choice than it does for destination choice. However, making such a distinction does not fit well with the standard nested logit model. We will consider whether a cross-nested logit model [CNL] could deal with this.

[34] König *et al.* (2004)

3 Cost Damping – The Current Recommendations for Modelling

3.1 Background

Following on from the experience of the Multi-Modal models and the limited modelling guidance then available in 2000, DfT recognised the need to provide more detailed guidance, leading to WebTAG, later TAG [Transport Analysis Guidance], in 2004. An important aspect of the modelling guidance, designed to ensure some consistency between models, was a set of "realism tests" for forecasting, so that model sensitivity was in line with best evidence on certain elasticities.

Based largely on the evidence from the airport access models and AHCG, DfT was aware of possible non-linearities in the standard generalised cost model formulation. In September 2007, they commissioned a study "Guidance on Cost Damping and Realism Testing in Demand Modelling", noting that "The need to address these topics together arises because cost damping mechanisms, which are already widely adopted, are increasingly seen as a way to address WebTag requirements for realism testing." Following further discussion led by Andrew Daly, a revised Brief was issued in January 2008 with the following additional text:

> "The main objective of cost damping is to improve the representation of traveller behaviour by recognising the diminishing sensitivity to marginal changes in cost or generalised cost as trip lengths increase….
> "Up to nine cost damping mechanisms have been identified, which differ principally in two ways:
> - whether the damping mechanism applies to the whole generalised cost function or just to part (e.g. through the value of time);
> - whether the damping impact is fixed for given journeys (e.g. it applies a factor on a distance or an OD basis) or works by applying a transformation to the costs (e.g. by applying a power function)."

Draft Guidance was produced in mid-2009, and apart from some minor adjustments, this remains largely unchanged.

3.2 Guidance on Cost Damping

The current guidance is presented in TAG M2.1, section 3.3. It may be noted that while the term "cost damping" implies that damping should be applied to the (money) cost, in practice the guidance interprets this more widely as "generalised cost" in some cases. Note that the guidance assumes that models will usually make use of generalised cost in time units ("generalised time"), so G = T + C/VTTS.

The general guidance on cost damping is presented in the following paragraphs:

> 3.3.1 There is strong empirical evidence that the sensitivity of demand responses to changes in generalised cost reduces with increasing trip length (see, for example, Daly (2008, 2010)). In order to ensure that a model meets the requirements of the realism tests specified in section 6, it may be necessary to include this variation. The mechanisms by which this may be achieved are generally referred to as 'cost damping'.

3.3.4 If cost damping is employed, it should apply to all person demand responses. The same cost damping function should be applied to both car (private) and public transport costs. While the starting position should be that the same cost damping parameter values are used for both modes, it may be necessary to vary the cost damping parameters between the modes in order to achieve satisfactory realism test results. It may also be necessary to vary cost damping parameters by trip purpose. However, these variations by mode and purpose should be avoided unless it is essential to achieve acceptable model performance (and always reported).

There are four recommended measurements, which we will refer to as A-D.

A (3.3.6-3.3.10) Use $VOT(dist) = VOT.\left(\dfrac{\max(dist, distcutoff)}{dist_0}\right)^{\eta_c}$, $0 \leq \eta_d \leq 1$

dist should "be calculated by skimming distances along minimum distance paths built between all origin-destination pairs using a base year network. In forecasting, there would only be a need to recalculate these distances if the structure of the network changed significantly between base and forecast years."

B. (3.3.11-3.3.15) Apply scale on G of $\left(\dfrac{dist}{dist_0}\right)^{-\alpha}$, $0 \leq \alpha \leq 1$, or 1 if *dist* ≤ *distcutoff*

C (3.3.16-3.3.18) Raise G to power of β, $0 \leq \beta \leq 1$ and adjust scale factor

Combinations of these are possible.

D (3.3.21-3.3.22) replace cost term by $\left[\beta^{I,C}_{cost} cost + \beta^{I,C}_{Logcost} * \ln(cost + \delta)\right]$ where coefficients must have the same sign, and δ is "a small constant (e.g. 1 pence)".

3.3.23 If cost damping is employed, the generalised costs used at the bottom of the choice hierarchy should be those obtained by the application of cost damping. At each higher level in the choice hierarchy, the composite costs should be calculated in the standard manner, so that the cost damping effects are reflected automatically throughout the variable demand modelling process.

Method D is typically associated with cases where demand models have been estimated[35] from survey data, while the other three would be more appropriate to cases where models have been constructed using TAG-suggested parameters.

The general evidence suggests that Method B is still by far the most widely used, though Method D is used in all the RAND Models, and in some cases Method A. Both A and B have been implemented in DIADEM.

We now discuss the implications of each of these formulations.

In principle, from the point of view of the (nested logit) demand model, it is possible to work throughout with "utility", however formulated, and apply the structural coefficients ["θ"] for

[35] in practice, these will be the "RAND models" discussed in §2.2: note that these have also used variants of Method A when Method D has not been found appropriate.

the hierarchy. Hence, maintaining the standard assumption of "mode above destination", we have[36]:

$$p_{j|im} = \frac{\exp(V_{imj})}{\sum_{j'} \exp(V_{imj'})}; V_{im*} = \ln \sum_{j'} \exp(V_{imj'}); \; p_{m|i} = \frac{\exp(\theta.V_{im*})}{\sum_{m'} \exp(\theta.V_{im'*})} \tag{3.1}$$

In line with S&R, the appraisal question is then how to convert/rescale from utility to monetary units. The TAG recommendation is to work with generalised cost in units of time, as noted, but essentially the same issue arises. In what follows, we ignore for simplicity the different time components, and just work with time (t) and cost (c).

For Method A, we have $V_{imj} = -\lambda.\left(t_{imj} + c_{imj} / VOT(d_{ij})\right)$. In many ways, this is the most straightforward approach, but it seems that it has hardly been used in practice, other than as a "back-up" method in the estimation of the RAND models when the Fox formulation fails to produce acceptable results.

In terms of re-scaling utility, this method raises no issues for time, but it does for cost, as the scale depends on distance. Thus we can convert the overall benefit to time units, but, presumably, we then have to assume some average value of time to convert to money, and this could lead to inconsistency.

For Method B, we have $V = -\lambda.\left(\frac{dist_{ij}}{dist_0}\right)^{-\alpha} \left(t_{imj} + c_{imj} / VOT\right)$: this corresponds with the approach of the airport surface access mode choice models reviewed in §2.1 At the time of drafting the Guidance, this was certainly the most widely applied method, and remains so. In this case, there is no fixed scale to convert from utility to either time or cost. The composite generalised cost is presumably calculated as

$$G_{im*} = -\frac{1}{\lambda}.\ln \sum_{j'} \exp\left[\left(\frac{dist_{ij'}}{dist_0}\right)^{-\alpha} \left(t_{imj'} + c_{imj'} / VOT\right)\right], \tag{3.2}$$

but this is not strictly in time units, as the damping factor varies with each ij pair. A "literal" interpretation of the S&R result would imply:

$$\Delta S \approx -½.VOT \sum_{ijm} \left(p'_{ijm} + p^0_{ijm}\right).\left(\frac{dist_{ij'}}{dist_0}\right)^{-\alpha} \left(\Delta t_{ijm} + \Delta c_{ijm} / VOT\right) \tag{3.3}$$

For Method C, we have $V = -\lambda.\left(t_{imj} + c_{imj} / VOT\right)^{-\alpha}$: this method was proposed in some of the models prepared by WSP (in particular, PTOLEMY for Leicester/Nottingham/Derby), but was only used for destination choice, not for mode choice, and – in contrast to TAG – had destination choice above mode choice. It is not clear whether models compatible with TAG have ever used this structure. A literal interpretation of the S&R result would imply:

$$\Delta S \approx -½.\frac{1}{\lambda} \sum_{ijm} \left(p'_{ijm} + p^0_{ijm}\right).\left[\left(t'_{imj} + c'_{imj} / VOT\right)^{-\alpha} - \left(t^0_{ijm} + c^0_{ijm} / VOT\right)^{-\alpha}\right] \tag{3.4}$$

[36] NB This formulation assumes that utility V is defined at the **lower** level (destination choice). As we will see in the discussion about the CNL model, there are good reasons to define it at the upper level, in which case the utility for destination choice becomes V/θ.

though the units are unclear. It is worth considering whether this method should actually be retained.

Method D is the approach generally used in the "RAND models", and in contrast to the other methods, which start with a pre-defined generalised cost formulation, this method is always applied at the time of model estimation, and thus directly reflects the data underlying the model construction, rather than a subsequent attempt to satisfy the realism tests. In this case, $V_{imj} = -\lambda.\left(t_{imj} + \left[\beta_{cost} c_{imj} + \beta_{Logcost}.\ln(c_{imj} + \delta)\right] / (-\lambda)\right)$. A literal interpretation of the S&R result would imply:

$$\Delta S \approx -\tfrac{1}{2} \sum_{ijm} \left(p'_{ijm} + p^0_{ijm} \right).(-\lambda) \left(\Delta t_{ijm} + \frac{\beta_{cost}}{-\lambda} \Delta c_{ijm} + \frac{\beta_{Logcost}}{-\lambda} \left[\ln\left(\frac{c'_{ijm} + \delta}{c_{ijm} + \delta} \right) \right] \right) \qquad (3.5)$$

thus with two terms relating to cost. Note that while the guidance implies that a small value of δ "(e.g. 1 pence)" should always be added to cost, the more recent RAND practice, as noted earlier, is to substitute [max (cost, δ)] for the cost, and to drop the cost function entirely in the case of free public transport.

For all methods, the difficulty lies in the fact that one or both of $\frac{\partial V}{\partial T}, \frac{\partial V}{\partial C}$ is not constant. Working with the logsum, de Jong *et al* (2007) propose two methods to address this. While they are working specifically in the context of a RAND model (hence Method D), the proposal has wider relevance. In the case where utility can be scaled to time units (i.e. $\frac{\partial V}{\partial T}$ is constant, which applies to Methods A and D), calculate the result according to the logsum and multiply by an estimate of VTTS – this is their Method 1. They note that there are potential issues of inconsistency here if the chosen value of VTTS is not compatible with that implied by the model. In Method 2 they calculate the **average** value of $\frac{\partial V}{\partial C}$ and use this to scale the logsum[37]. The particular form of *V* that they discuss is the earliest form used in the LMS, with $V = \beta_T T + \beta_L.\ln(C)$ so $\frac{\partial V}{\partial C} = \beta_L.\frac{1}{C}$, but it should be generally applicable. The implication of their approach is that for any particular segment the average $E\left[\frac{\partial V}{\partial C}\right]$ is given as $E\left[\frac{\partial V}{\partial C}\right] = \frac{\sum_{ijm} T_{ijm} \frac{\partial V_{ijm}}{\partial C_{ijm}}}{\sum_{ijm} T_{ijm}}$. It would seem that this calculation needs to be made separately for both before and after logsums, prior to taking the difference, though de Jong *et al* (2007) do not explicitly note this.

It is clear that whichever form of cost damping is chosen raises potential questions of how far we are willing to depart from the S&R theory. There is thus an empirical question as to how important this is. If the current approach to the calculation of benefits, based on a single value of time (for a given purpose, etc.) and the change in costs and times, can be viewed as an acceptable approximation to the S&R benefits, there is no strong need to take explicit account

[37] They note that "the use of the expectation…is only approximately correct"

of the theory[38]. Such a conclusion might, in addition, be conditional on the choice of cost damping Method.

In a later section, we set out a proposal for some empirical testing, making use of Transport for the North [TfN]'s NoRTMS model.

[38] In a recent email exchange, Ken Small suggested that the RoH should remain valid if the change is small enough so that the demand curve is approximately linear over the relevant range. The same point was made, in my interpretation, by Andrew Daly in his exchange with Iven Stead in December 2023. However, while this is certainly true for a constant generalised cost formulation, it remains to be seen if it holds up under cost damping and other non-linearities.

4 Utility in the Context of Mode and Destination Choice

4.1 Further on the Value of Time[39]

The economic theory of time allocation derives primarily from the goods leisure tradeoff within the theory of the labour market. The standard analysis is indicated in Figure 4.1, whereby individuals are assumed to have an indifference between different quantities of money and leisure time, with the shape shown. As leisure time is reduced, individuals become more reluctant to give up additional leisure time to work unless they are compensated by a higher wage rate (implying a rationale for "overtime" rates). Conversely, at low incomes and large amounts of "leisure" (not necessarily voluntary!), individuals will be willing to accept work (ie reduce leisure time) in return for relatively low wage rates.

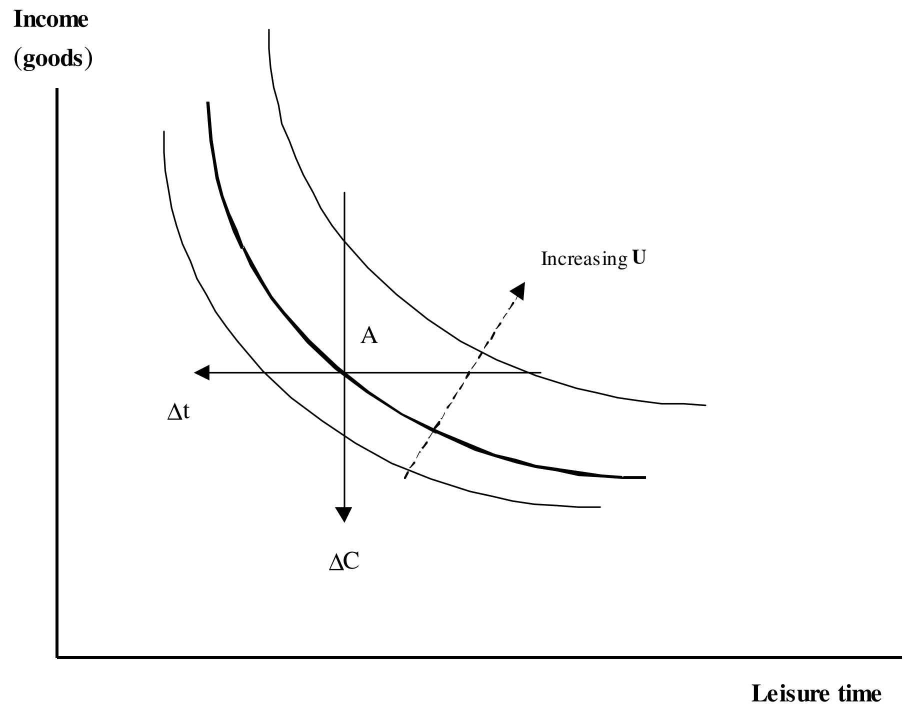

Figure 3: Tradeoff between income and leisure

Figure 4.1 Tradeoff between income and leisure

From the transport point of view, we are interested in how the balance is affected by changes in the cost and time of **travel**. For a given current position A in Figure 4.1, we can re-interpret the figure in transport terms to produce the shape illustrated in Figure 4.2, in respect of an individual trip. An increase in travel cost acts in the opposite direction to an increase in income, and an increase in travel time acts in the opposite direction to an increase in leisure time. Hence, the figure is (more or less) inverted. Clearly the indifference curve through the origin must be entirely in the NW and SE quadrants, in which the changes in time and cost are of opposite sign.

[39] The text in the next two sections derives primarily from an earlier document prepared for the AHCG study.

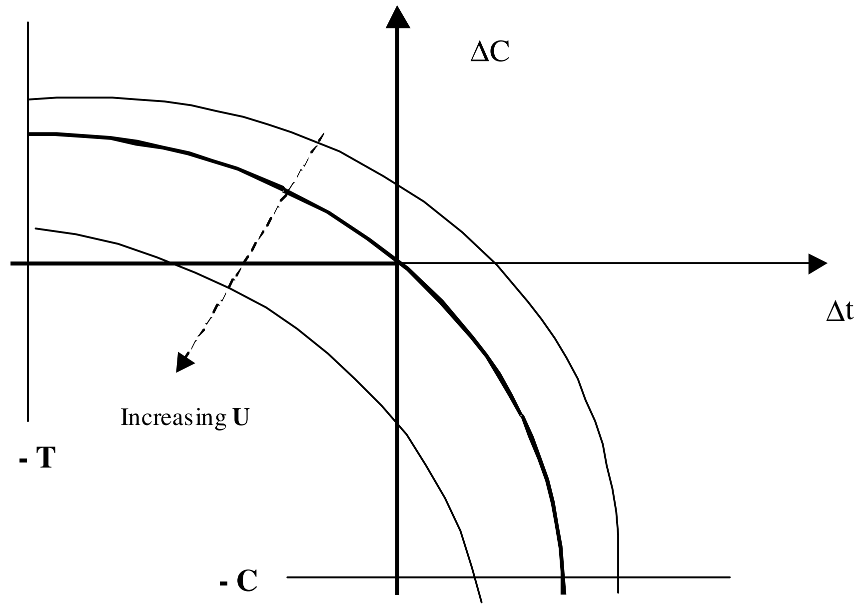

Figure 4: Tradeoff between cost and time for given trip

Figure 4.2: Tradeoff between cost and time for a given trip

For an individual making a specified journey with cost C and time T, the basic requirement of the analysis is to deliver a family of indifference curves U(c,t) = k. The value of time at any point c,t can be found by:

a determining on which indifference curve (c,t) lies (ie the value of k)

b along this curve (ie holding k constant), taking the ratio of marginal utilities:

$$v = \left(\frac{\partial U_k}{\partial t} / \frac{\partial U_k}{\partial c}\right)$$

This is related to the fundamental differential equation along the indifference curve:

$$dU_k = \frac{\partial U_k}{\partial t}.dt + \frac{\partial U_k}{\partial c}.dc = 0 \qquad (4.1)$$

Hence the slope dc/dt along the indifference curve is given by $-\frac{\partial U_k}{\partial t} / \frac{\partial U_k}{\partial c}$. The (negative) slope of the indifference curve thus gives the "value of time" (tradeoff, or marginal rate of substitution, between time and cost). To remain at the same utility level, an increase in cost must be matched by a decrease in time, and *vice versa*. This allows a time loss or gain to be valued along the lines set out above.

4.2 Conditions on non-linear forms

The direction of the curvature should be clear. From a given base, the greater the cost increase ($\Delta c > 0$), the greater the relative compensating reduction in time must be, leading to a fall in the value of time (flatter curve), as the (money) budget constraint begins to bind. Likewise, the greater the time increase ($\Delta t > 0$), the greater the relative compensating reduction in money must be, leading to an increase in the value of time (steeper curve), as the “time budget” constraint begins to bind.

If non-linearity exists, then the value of time, which is the tangent to the indifference curve, will fall as cost increases and time decreases, and conversely. In other words, it should have the shape shown in Figure 4.2. For an individual, the question thus turns on the functional form of marginal utility with respect to money and time. On general theoretical grounds we expect

a both $\frac{\partial U_k}{\partial t}$ and $\frac{\partial U_k}{\partial c} < 0$

b both $\frac{\partial^2 U_k}{\partial t^2}$ and $\frac{\partial^2 U_k}{\partial c^2} \leq 0$ (4.2)

The condition on the second derivatives reflects both the "satiety" axiom[40] (in reverse!) and the constraints on the overall time and money budgets.

If both second derivatives are zero, then the value of time is constant, and the indifference curve **at that level** is a straight line. In practice it is unreasonable to expect either derivative to be zero: however, what is at issue is the change in the marginal utilities over the range of (c,t), both in the SP experiment, **and** in an appraisal.

The simplest form of allowing for this theoretical variation is to use a form

$$\frac{\partial U_k}{\partial t} = -\varphi - \zeta/(X-t) \qquad (4.3)$$

where X is some "travel (time) budget"; $\varphi, \zeta > 0$

Integrating, $U_k = -\varphi t + \zeta \ln(X-t)$ + terms in other variables. We can deal with costs in a corresponding way, to obtain $U_k = -\lambda c + \xi \ln(Y-c) + ...$, where Y can either be total income, or some "travel (money) budget".

Unfortunately, unless we know the budgets X and Y this is not much help! One possibility is to try to estimate them, or make an assumption about their distribution, and this could be given further consideration. Alternatively, we can expand terms of the form ln (Y–c) as follows:

$$\ln(Y-c) = \ln Y + \ln(1-c/Y) \approx \ln Y - c/Y - ½ (c/Y)^2 ...... \qquad (4.4)$$

Since (for a single individual) Y is a constant (as is X), and we are only interested in relative utilities, we can re-write the effective utility function as:

$$U_k = -\varphi t - \zeta [t/X + ½ (t/X)^2 ......] - \lambda c - \xi [c/Y + ½ (c/Y)^2 ......] + ... \qquad (4.5)$$

Collecting terms and re-defining the coefficients,

$$U_k = [-\varphi - \zeta/X] t - [½ \zeta / X^2] t^2 ...... [-\lambda - \xi/Y] c - [½ \xi / Y^2] c^2 ...... + ...$$

$$= -\varphi' t - \zeta' t^2 - \lambda' c - \xi' c^2 + ... \qquad (4.6)$$

[40] The satiety axiom (or better, axiom of non-satiation) assumes that the utility function is non-decreasing in all its arguments

The quadratic coefficients ζ' and ξ' will need to be significantly different from 0 to justify a departure from linearity. The original conditions on the positive signs of (φ, ζ, λ, ξ) to satisfy the 2^{nd} order conditions imply that the transformed coefficients (φ', ζ', λ', ξ') must also all be positive.

Since both the linear and quadratic coefficients, in the transformed version for estimation, are functions of the travel time and cost budgets, we can expect that variations in these coefficients across the sample will be found corresponding to different budgets.

Note that although the curvature as shown in Figure 4.2 has intuitive validity, the **scale** is not clear. That is, while on theoretical grounds we expect the slope to increase (in negative terms) as Δt increases, we have no immediate expectations as to what size of increase in Δt is required to lead to a **significant** change in slope. We might expect little departure from linearity for the majority of changes in cost or time that can be realistically associated with a journey. This suggests a possible line of enquiry. It should be noted, however, that there is no empirical evidence that the budget effect has affected VTTS.

4.3 Distance effects on VTTS

Although the empirical evidence is not completely consistent, there is a general impression that VTTS increases with distance. In this section we consider why this might be so. We will work with a general specification of generalised cost/utility, implying a linear combination of time and cost functions, noting that both time (T) and cost (C) are functions of distance (D), with the expectation that both $\frac{\partial T}{\partial D}$ and $\frac{\partial C}{\partial D} > 0$:

$$V = f(T) + g(C) \tag{4.7}$$

As usual, $VTTS = \frac{\partial V}{\partial T} / \frac{\partial V}{\partial C}$ where both $\frac{\partial V}{\partial T}$ and $\frac{\partial V}{\partial C} < 0$.

Hence, differentiating with respect to distance and using the chain rule,

$$\frac{\partial [VTTS]}{\partial D} = \frac{\frac{\partial V}{\partial C}.\frac{\partial^2 V}{\partial T^2.}\frac{\partial T}{\partial D} - \frac{\partial V}{\partial T}.\frac{\partial^2 V}{\partial C^2.}\frac{\partial C}{\partial D}}{\left(\frac{\partial V}{\partial C}\right)^2} \tag{4.8}$$

For this to be positive, this requires that $\frac{\partial V}{\partial C}.\frac{\partial^2 V}{\partial T^2.}\frac{\partial T}{\partial D} > \frac{\partial V}{\partial T}.\frac{\partial^2 V}{\partial C^2.}\frac{\partial C}{\partial D}$

Dividing through by $\frac{\partial V}{\partial C}$ (which is negative) and changing the sign of the inequality, we get

$$\frac{\partial^2 V}{\partial T^2} < VTTS.\frac{\partial^2 V}{\partial C^2.}\left(\frac{\partial C}{\partial D} / \frac{\partial T}{\partial D}\right) \tag{4.9}$$

Note that for the "RAND formulation" which we can write as $V = \beta_{\mathrm{T}}.T + \beta_{\mathrm{C}}.[\gamma.\mathrm{C} + \alpha.(1\text{-}\gamma).\ln(\mathrm{C})]$, $\dfrac{\partial^2 V}{\partial T^2}$ =0, so provided $\dfrac{\partial^2 V}{\partial C^2}$ is always positive, this condition will be met, regardless of the shape of the cost and time functions with respect to distance. $\dfrac{\partial V}{\partial C} = \beta_{\mathrm{c}}.\left[\gamma + \dfrac{\alpha.(1\text{-}\gamma)}{C}\right]$, so $\dfrac{\partial^2 V}{\partial C^2} = -\beta_{\mathrm{c}}.\dfrac{\alpha.(1\text{-}\gamma)}{C^2}$ > 0 since β_{c} <0. We note, of course, that the positive sign of $\dfrac{\partial^2 V}{\partial C^2}$ is in contradiction to that implied by the satiety axiom discussed in §4.2.

According to Daly & Carrasco (2009): "The issue of increasing VOT was considered in the context of large-scale modelling by Ben-Akiva *et al*. (1987)"… "An important difficulty with a formulation that allows the marginal impedance of cost to decrease with trip length – whether by a log function or by any other non-linear transformation – is that it is difficult to explain in the context of an economic theory of utility maximisation subject to a budget constraint. An increasing marginal disutility of money expenditure could be explained, in terms of an approach to the ultimate budget constraint, but the empirical results point strongly to decreasing marginal disutility."

They consider 5 hypotheses advanced by Ben-Akiva *et al*., plus a further 8 of their own, with various levels of plausibility, making the critical observation that "While each of the postulated causes could have an impact in the 'right' direction, it is necessary to explain why each would apply to one of the time and cost variables and not to the other."

They test a linear version of Generalised Cost but with random coefficients β_T and β_C, using data from 3 datasets – DMS models from Sydney and Paris, and Dutch VTTS SP data[41] – arguing that this will lead to increased randomness for the utility function as C and T increase. The Sydney data provides little evidence of improved model fit, but for Paris there is strong evidence for a random cost effect (much less for time), and even stronger for the Dutch data, where both effects were significant though the cost effect was stronger. In all cases, the random coefficient was specified as having a normal distribution (so that the mean and standard deviation were estimated). Unfortunately, the estimated coefficients are not presented, and in cases where **both** random components are estimated (this was not done for the Dutch data), no indication is given of correlation. For the two DMS datasets, cost was separately included in both linear and logarithmic form: for the Paris data, but not for Sydney, the random cost specification with linear cost was an improvement on the fixed cost coefficient with log cost.

Their main conclusions can be cited as follows:

> "Significant heterogeneity of preference exists in all the data sets analysed and can be represented as heteroskedasticity in time or in cost: this was found to be significant in most of the models tested. Heteroskedasticity in both time and cost was tested only for the Paris models and gave little further improvement. Non-linear time variables (tested for the Paris data only) were not found to be significant.

[41] Note that these were much earlier VTTS studies conducted in 1988 and 1997

> "For all the models with linear cost functions and three of the four models with log cost functions, heteroskedasticity with respect to cost gave a greater improvement than heteroskedasticity with respect to time. In terms of the hypotheses, the most likely mechanisms causing the increase are, in order of support by the results:
> self-selection by choice of mode, poor perception of costs, and proportionate varying valuation of cost (but not of time).
>
> "Tests of the approach on other data sets would be useful in clarifying the extent to which these findings are transferable to a wide range of contexts."

They also propose the rather strong conclusion that "the underlying VOT is therefore not increasing with distance at an individual level".

Random coefficients ("mixed logit") have generally not been used in DMS models, essentially because of model complexity and run-time implications. They are more appropriate to VTTS studies where the aim is to make general recommendations about values, though it is only relatively recently that serious attempts have been made to estimate these: both the UK 2014 study and the Dutch 2022 included them in the final recommended models, in both cases on the VTTS directly, rather than on one or both of the cost and time coefficients (this is partly because of the chosen estimation method). It is noteworthy that in spite of the random VTTS, a distance effect could still be estimated in the UK study: the Dutch study appears silent on this matter.

If we revert to the Dutch VTTS specification [Eq 2.6], we can again investigate the conditions for VTTS to increase with distance. We now have:

$$VTTS = \beta_t.\left(\frac{T_{ref}(D)}{T_0}\right)^{\lambda_T} / \beta_c.\left(\frac{C_{ref}(D)}{C_0}\right)^{\lambda_C} \text{ so that}$$

$$\frac{\partial VTTS}{\partial D} = \frac{\beta_c.\left(\frac{C_{ref}(D)}{C_0}\right)^{\lambda_C}.\beta_t.\lambda_T.\left(\frac{T_{ref}(D)}{T_0}\right)^{\lambda_T-1}\frac{1}{T_0}\frac{\partial T_{ref}}{\partial D} - \beta_t.\left(\frac{T_{ref}(D)}{T_0}\right)^{\lambda_T}.\beta_c.\lambda_C.\left(\frac{C_{ref}(D)}{C_0}\right)^{\lambda_C-1}\frac{1}{C_0}\frac{\partial C_{ref}}{\partial D}}{\beta_c^{\,2}.\left(\frac{C_{ref}(D)}{C_0}\right)^{2.\lambda_C}}$$

$$= \frac{\beta_c.\beta_t\left(\frac{C_{ref}(D)}{C_0}\right)^{\lambda_C}\left(\frac{T_{ref}(D)}{T_0}\right)^{\lambda_T}\left[\lambda_T.\left(\frac{T_{ref}(D)}{T_0}\right)^{-1}\frac{1}{T_0}\frac{\partial T_{ref}}{\partial D} - \lambda_C.\left(\frac{C_{ref}(D)}{C_0}\right)^{-1}\frac{1}{C_0}\frac{\partial C_{ref}}{\partial D}\right]}{\beta_c^{\,2}.\left(\frac{C_{ref}(D)}{C_0}\right)^{2.\lambda_C}}$$

All the terms outside the square bracket are positive, so for VTTS to increase with D we must have $\lambda_T.\left(T_{ref}(D)\right)^{-1}\frac{\partial T_{ref}}{\partial D} > \lambda_C.\left(C_{ref}(D)\right)^{-1}\frac{\partial C_{ref}}{\partial D}$ (4.10)

Hence the result depends not only on the elasticities (which have usually been shown to be negative) but also on the way in which T_{ref} and C_{ref} increase with distance. However, in the special case where both are linear in D, so that $T_{ref}(D) = \phi.D$ and $C_{ref}(D) = \omega.D$, then we can see that condition (4.10) resolves to $(\lambda_T - \lambda_C) > 0$.

As noted previously, the negative values of the elasticities are not compatible with the budget effect: rather, they are compatible with a "framing effect" whereby the response becomes less sensitive given an increase in prices (time and money) with distance: implying both $\partial^2 V / \partial c^2 > 0$ and $\partial^2 V / \partial t^2 > 0$ (the opposite inequalities would be expected due to the budget effect).

4.4 Implications for destination and mode choice

We now consider the general question of modelling mode and destination choice. While clearly we will wish to construct a single model which deals with both choices, it will help initially to think of them separately.

In fact, the discussion of how they should be linked was a recurring issue in the early history of transport modelling. From a behavioural point of view, it makes more intuitive sense to decide on a destination and then, given the destination, to choose the appropriate mode and/or route to reach the destination. It is important to realise, however, that this is not equivalent to "mode choice conditional on destination" in a nested logit model.

In the context of nested logit models, Ortúzar (2001) notes that "the application of the NL model in disaggregate form was first undertaken by Ben-Akiva (1974) and its later theoretical derivation should be attributed to Williams (1977) and Daly and Zachary (1978), with the generalized extreme value generalization being provided by McFadden (1978)."

According to the theory, it is the composite utility ("logsum") which provides the link between the two models, and the ordering of the two models is dictated by the structural parameter which measures the relative size of the error variances, which must increase as we descend the structure. Bates (1998) suggested: "The arguments advanced so far imply that the variability of destination utilities is the key factor determining trip length. Although we have kept the discussion "unimodal" for simplicity, we could expect that the variability of destination utility would "dominate" that of modal utility. In terms of the nested or hierarchical logit model, this implies that mode is below destination choice in the hierarchy. Inasfar as this is so, we can simply substitute the unimodal GC terms in the theoretical development with terms that are composite over all available modes."

However, the majority of estimated models have concluded that the error variance is in fact greater for mode choice than for destination choice (though sometimes they are not found to be significantly different), and this underlies the TAG recommendations of destination choice conditional on mode choice.

For destination choice, the essential question is whether the additional utility from a further destination outweighs the additional cost. While there are individual-specific factors which will influence the destination utility, as well as the modelling issue of how to represent the average utility (by means of "size variables", Daly (1981)), the general correlation of both cost and time with distance would seem to imply that there is less error associated with generalised cost[42]. As for the possibility of variation in VTTS with distance, it is very unclear in the context of destination choice why this should be so, apart from a possible budget effect (which, as we have seen, implies that both $\partial^2 V / \partial c^2 < 0$ and $\partial^2 V / \partial t^2 < 0$): for an increase in VTTS with distance, we would require the condition in Eq (4.9).

In the case of mode choice, however, there are many factors which will influence the choice other than the time and cost, and in the case of public transport there is further possibility for heterogeneity in the response to the various journey components (frequency, access/egress etc.). Perhaps it is this which outweighs the level of error associated with the destination utilities?

As Bates (2017) points out, the standard nested logit approach has the property that a change to the generalised cost of one mode to one destination will impact on the choice of mode for all destinations to the same extent proportionally, rather than being concentrated on the particular destination affected. He proposes the use of the Cross-nested Logit [CNL] to avoid this. In such a model, in addition to the standard TAG "destination below mode" structure it is possible to include the alternative "mode below destination" structure. For this model, it would be possible to include the "framing effect" whereby the response becomes less sensitive given an increase in prices (time and money) with distance: implying both $\partial^2 V / \partial c^2 > 0$ and $\partial^2 V / \partial t^2 > 0$. The figures below give an illustration of the implied shapes for $\partial V / \partial c$ and V: the formulae used are $\frac{\partial V}{\partial C} = \alpha . \left(\frac{C}{25}\right)^{\lambda}$ whence $V = \frac{25.\alpha}{\lambda + 1} . \left(\frac{C}{25}\right)^{\lambda + 1}$, with α = –0.1 and λ = [0.3, 0, –0.3] corresponding to $\partial^2 V / \partial C^2$ having values [<0, 0 (linear), >0] respectively.

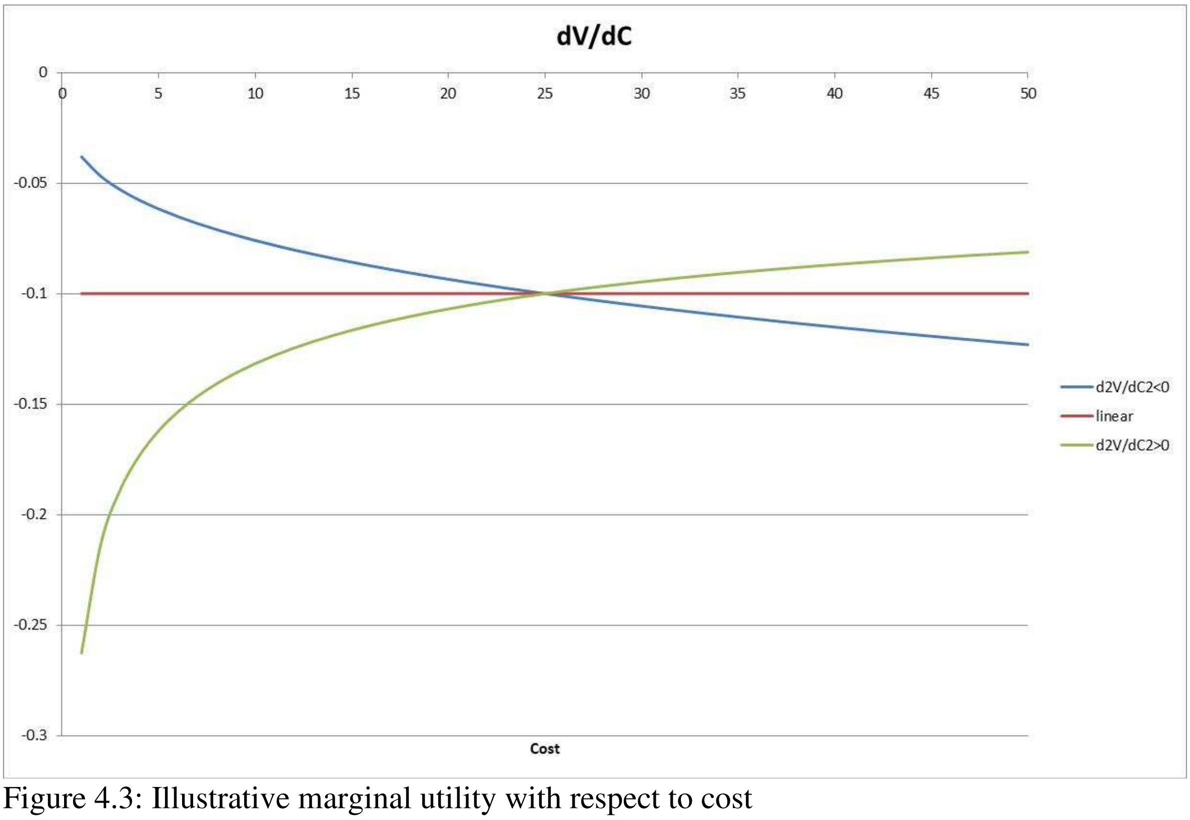

Figure 4.3: Illustrative marginal utility with respect to cost

[42] more importantly, with the **differences** in generalised cost between destinations

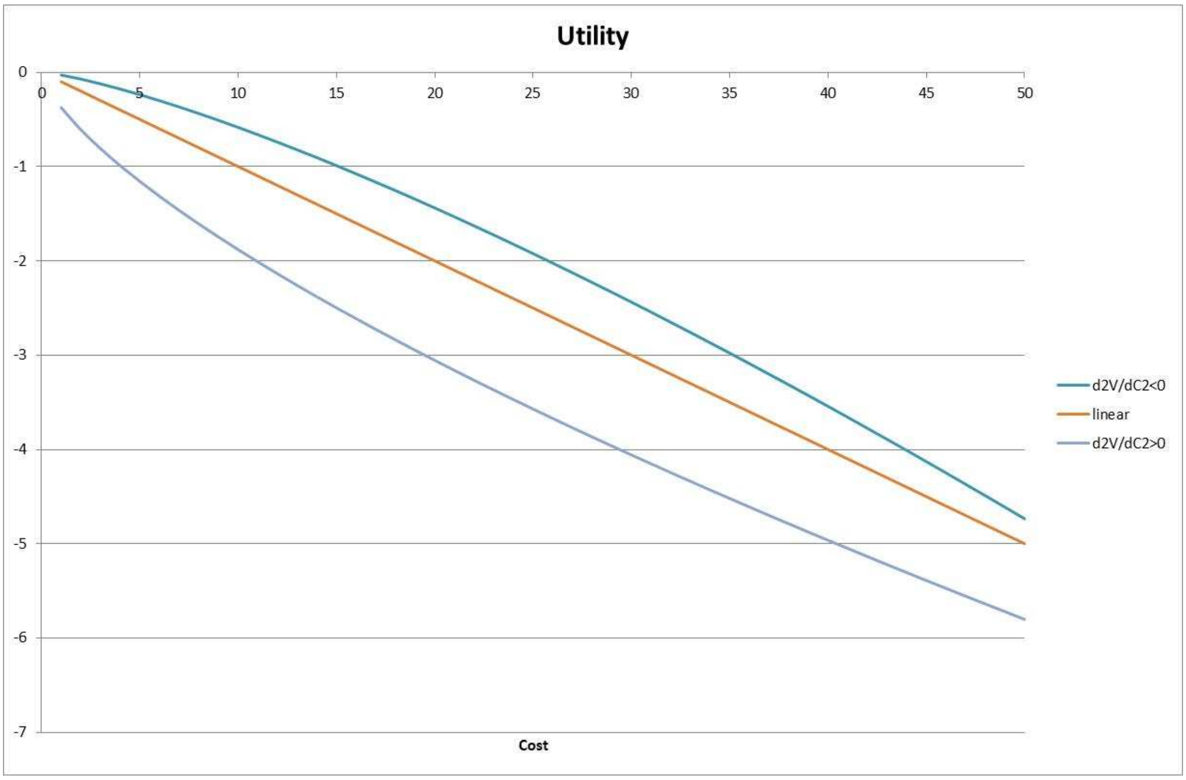

Figure 4.4: Illustrative utility with respect to cost

Bates (2024) proposes the following CNL structure, with a set of nests for each destination (here 1, 2 3), in addition to the standard modal nests (here h for highway and p for public transport), and each mode/destination "elemental" alternative only in the one mode nest and the one destination nest to which it relates:

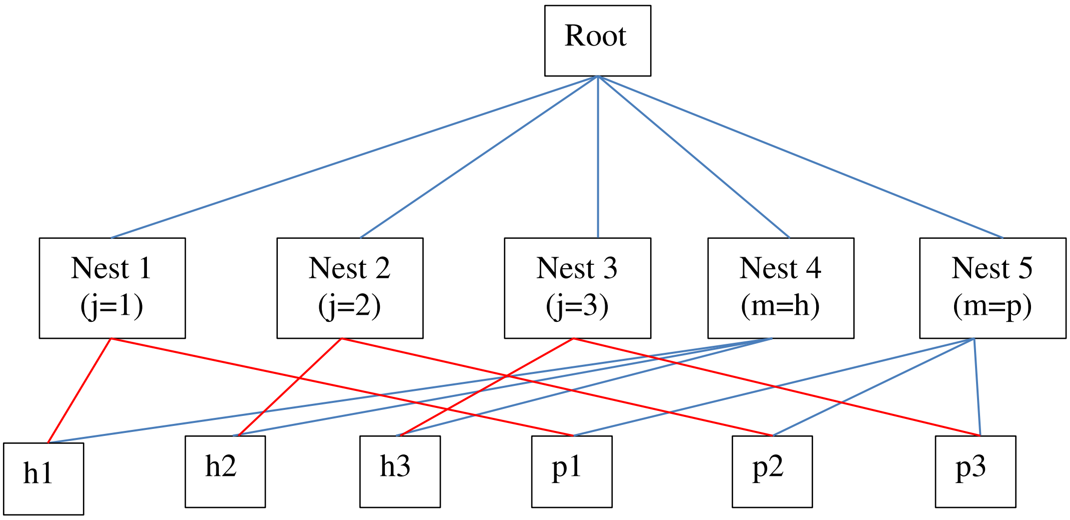

Figure 4.5: Proposed CNL structure for DMS (Bates (2024))

The red lines indicate the mode choice structure for each destination. Note that if we ignore the red connections, the first three (destination) nests become redundant, and we revert to the standard nested logit structure, with mode above destination. If the structure is viewed as identifying decreasing variance as we descend, then the **furthest** destinations would need to be higher in the structure. Concentrating on the destination nests, and assuming that the destinations are in increasing order of distance from the origin, this implies a structure like:

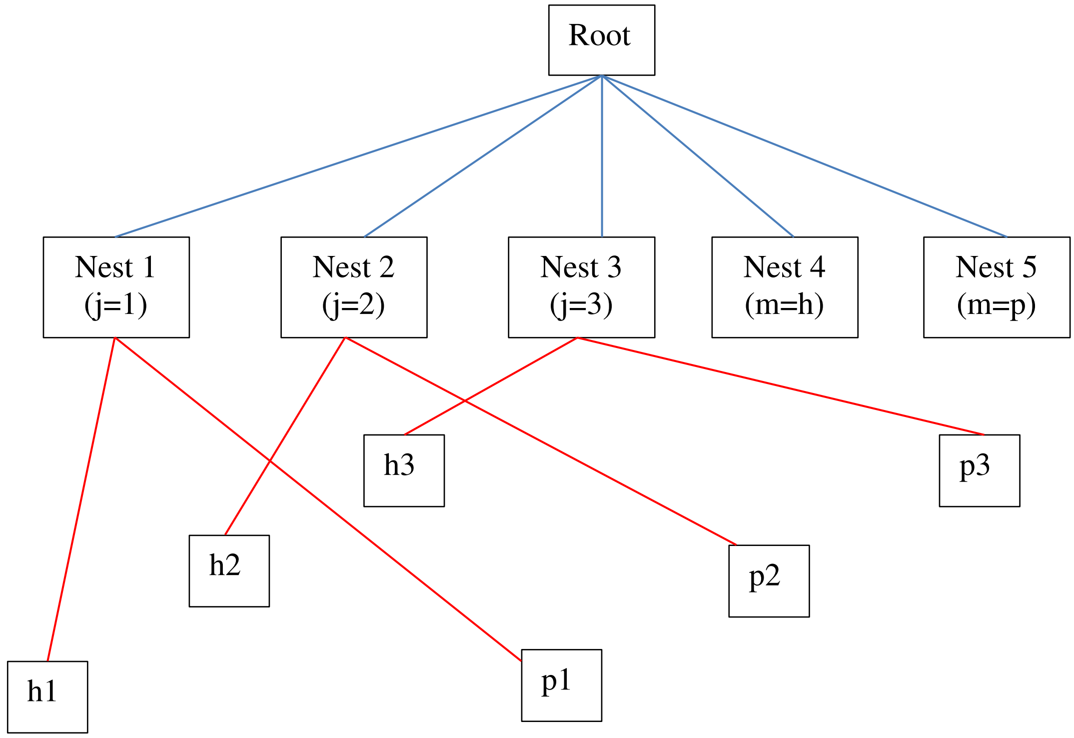

Figure 4.6: Destination nests in CNL model

In the notation of Bates (2024), this would be achieved by ensuring that structural coefficients $\theta_{n(j)}$ (relating to destination nest j) increase (in the required range [0,1]) with the distance of destination j from the origin[43]. Note that this implies that the structural coefficients would be different for each origin.

It can be seen that this is in line with the above discussion on mode and destination choice, and is in some way a version of damping mechanism B. It does not imply that the VTTS increases with distance, however.

In the general DMS context, it is of interest to examine the actual distribution of trips by distance from NTS[44]. For all modes and purposes combined, the data has been grouped into distance bands of various sizes, and the proportion of data in each band has been divided by the width of the band, to convert it approximately to a probability distribution. Note that given the lack of detailed location in the NTS data, it is not possible to take account of the number of attractions/opportunities at each distance, as would be required when constructing the "size variables" for a destination choice model: the figures should be viewed with this qualification in mind.

According to this data, 81% of all trips are under 10 miles, 94% are under 25 miles, and 98% are under 50 miles. As can be seen, the actual data has a peak at about 1.5 miles, reflecting

[43] In the CNL formulation, the utility is defined at the root level, and is divided by the structural coefficients
[44] Again, this is based on the author's own analysis of NTS data for 2002–2010, from the UK Data Archive

the fact that journeys are made to achieve activities which cannot be met in the home. Attempting to fit either the negative exponential or the gamma functions (which, in terms of the logit function, correspond with linear or log-linear utilities) reveals some aspects of interest: clearly the negative exponential cannot pick up the peak, though the fit over the range 2 to 25 miles is good, while the gamma function picks up the peak well, but falls off too rapidly for distances over 5 miles. Although it cannot easily be seen from the graph, neither function performs well at long distances, falling off far too fast. Note also that in order to pick up the peak, the gamma function requires a positive coefficient on the log term, which would lead to positive marginal utility over some range if applied to the cost or time variables.

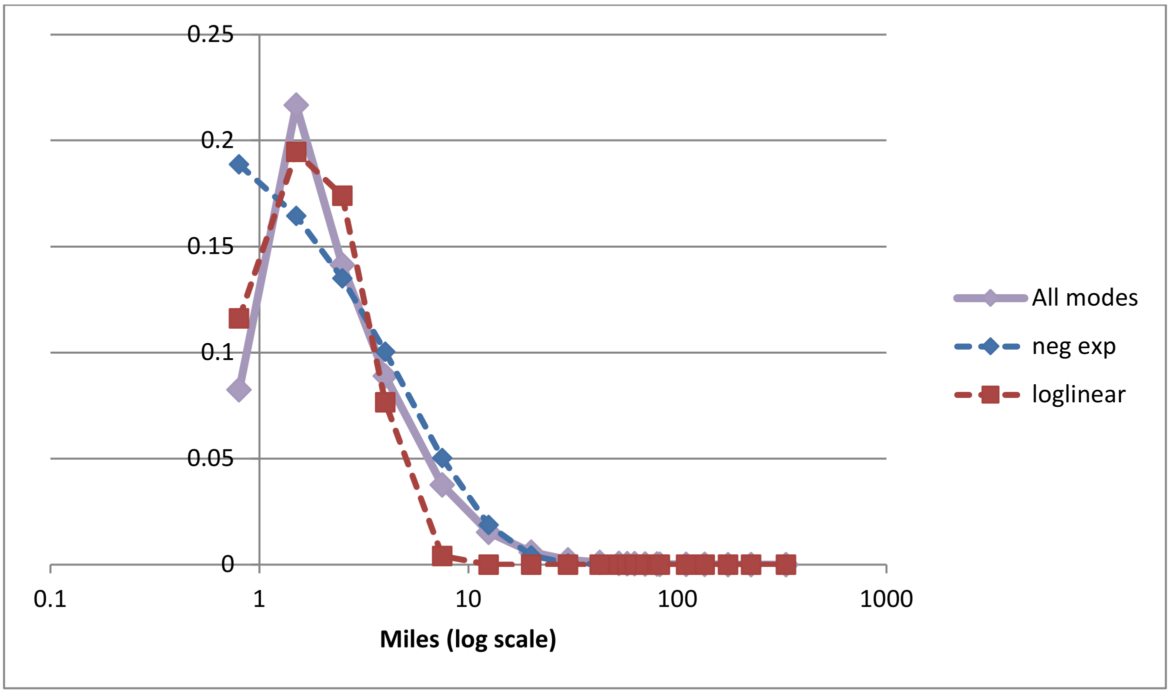

Figure 4.7: NTS distance distribution, with fitted curves

Turning to the mode proportions, using the same data, we see a highly characteristic pattern. For trips under 1 mile, the shares of car driver and “slow” (walk + cycle) are almost equal, thereafter showing a more or less mirror image up to 25 miles, when the slow mode becomes negligible. The car driver mode share then starts to fall, with increasing share for rail and, at the longest distances, air. The car passenger mode is more or less constant at about 28% over the range of 2 to 50 miles, then rising slightly until the very longest journeys (250 miles or more). The bus mode reaches a peak in the 3-5 miles band, falling to about 2% in the 35-120 miles range, though thereafter there is a slight increase owing to long distance coach.

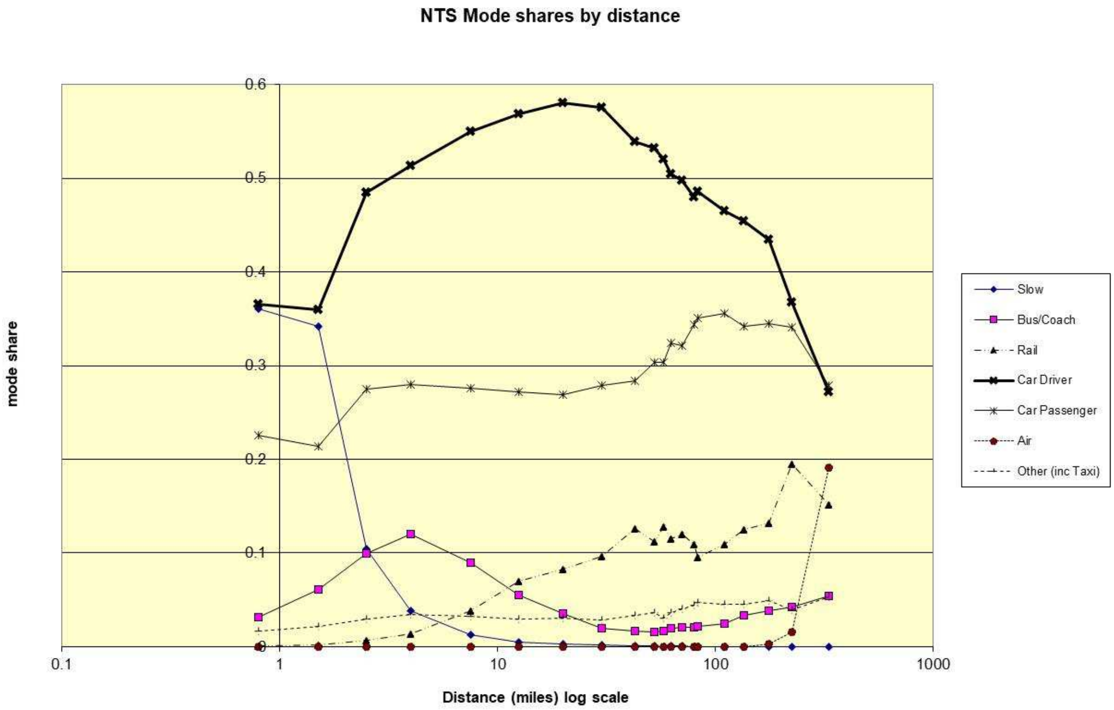

Figure 4.8: NTS mode shares by distance (all trips)

The ability to pick up these patterns is a crucial aspect of the estimation of the DMS model. It is hoped that some of these variants could be tested in the forthcoming estimation for the Northern Behavioural Survey for TfN.

In responding to the 2014 VTTS study, the DfT asked Paul Hanson and Bryan Whittaker for a thinkpiece on the impact that new values of time savings would have on existing and new transport modelling practice: this was included as Appendix C in the response document[45]. They particularly noted "the statistically significant variation in VTTS for business travel with distance", while also recognising the "framing effect" as in their Paragraph 5.2:

> A typical thought experiment is to consider how a traveller would respond to a one minute or £1 increase in their journey time or cost. It is reasonable to expect that there would be a much larger response for a short 20 minute journey that costs £3 than for a much longer 6 hour journey that costs £100.

Accordingly they concluded that model utility functions should "imply that sensitivity cost and time declines with travel distance [and] that the value of travel time savings increases with travel distance." Going further, they advocate "**that the default recommendations should be for a non-linear utility function** and that the current requirement to justify the use of cost dampening should be withdrawn." While these proposals are not unreasonable, they should be subject to further careful examination: while current evidence (as reviewed in Chapter 2) generally suggests VTTS increasing with distance, the reasons for this remain debatable, while the best specification of the framing effect should again be subject to empirical investigation.

[45] DfT (2016)

5 Other Issues

5.1 Assignment

The suggestion of a "framing effect" for mode choice should, in principle, apply to assignment (route choice) as well. Current practice in highway assignment[46] is to assume a deterministic choice process based on minimising a combination of time and money, with

$$GC = T + C/(VTTS.Occ) \qquad (5.1)$$

The (vehicle) cost can be calculated using the TAG formula as a function of distance *D* and speed *v*:

$$C = D.\left(\mathrm{P}.\left(a/v + b + c.v + d.v^2\right) + a_1 + b_1/v\right) \qquad (5.2)$$

where P is the fuel price per litre and the last two terms in the bracket relate to non-fuel costs, which only apply to vehicles in course of work[47].

In principle, C could be calculated separately for each link in the network and summed over the route. In practice, this is never done, and a simpler formula based merely on time and distance is typically used:

$$GC = T + D.PPK/PPM \qquad (5.3)$$

where PPK is "pence per Km" and PPM is "pence per minute". On convergence, for use in the demand model, the money cost is typically re-calculated using equation (5.2) but with a generic assumption about speed *v*: clearly there is some inconsistency here.

In line with the framing effect, it would be extremely inconvenient if separate assumptions about VTTS (or PPM) were to be made for each O-D pair. The authors of Appendix C to the DfT VTTS Response document cited earlier also note that while it would seem appropriate to "adopt model forms that represent how routeing behaviour varies with travel distance", this will require software development. In the short term, they suggest relying on segmentation using different VTTS.

The question arises, however, as to how far the addition of the distance term actually contributes to route choice – would the routes chosen be significantly different if the assignment was done on time alone? Since the specification of Eq (5.3) will vary by purpose (because of VTTS and Occupancy underlying the parameter PPM), some preliminary impression could be obtained by looking at the route choice variation by purpose. If this is only marginally significant, it would seem unreasonable to consider varying VTTS by distance in the assignment, and a time-only procedure could be considered.

[46] We omit here any discussion of congestion, or supply effects.

[47] The assumption that non-fuel costs are "unperceived" by travellers whose purpose is not business goes back (at least!) to MAU Note 179: see footnote 3. As noted in the Brief, recent research by Ricardo for DfT points to a significant upwards adjustment in these costs, which would further call into question the current TAG assumption that these costs are unperceived for non-work travel.

This will not, of course, deal adequately with tolled routes. However, this is potentially a more significant issue for assignment, and probably needs to be dealt with using a distributed VTTS[48]. Pending development of appropriate techniques (NB there is a relatively untested SATURN option "STOLL" which could be used), minor tolled links could be dealt with by *ad hoc* means making use of a penalty).

In relation to public transport, it may be noted that most examples of assignment in practical models do not in fact make use of fares – thus route choice relies on other aspects of the service such as frequency and the need to interchange as well as journey time. This is partly justified by existing fare structures (e.g. zonal fares, or rail fares which permit alternative routes). The fares are only then added when the network times are introduced to the demand model (which could include station choice).

The Brief makes reference to the handling of combined sub-mode PT assignment, noting that the TUBA manual currently recommends a flow weighted average costs but hints at the possible use of composite costs. If composite costs were used, these would embed behavioural/modelled VTTS which may differ from the desired appraisal VTTS. This only applies, however, if fares are included in the PT assignment. As the authors of Appendix C note: "Even where fares are represented in public transport assignment models (rather than the associated demand model), few urban public transport models distinguish between concession and standard fares, let alone season and the variety of fare types available."

On this basis, it is proposed that the impact of different VTTS on (highway) assignment is investigated: if it turns out to be marginal, it would be possible to carry out assignments based on time only, and then to calculate the cost using the TAG formula when introducing time to the demand model.

5.2 Unimodal models

Unimodal models, and elasticity models in particular, are widely used in rail modelling, and are recommended by TAG [Unit M4, §8]. The application of such models requires an appropriate base demand, together with elasticities relating to the variables that are changing. In keeping with the rest of this document, we restrict these to time and cost, but in the context of rail, we expand the definition of "time" to the rail construct of Generalised Journey Time [GJT].

Exogenous changes to demand are provided by the DfT's EDGE [Exogenous Demand Growth Estimator] model, while the response to changes in fares and GJT are dealt with by means of PDFH-recommended elasticities: $D_{RS} = D_{RS}^{0}\left(Fare_{RS}\right)^{\eta_F}.\left(GJT_{RS}\right)^{\eta_G}$, where RS denotes a pair of stations and D is demand. As the Brief notes, this is not consistent with the GC approach used in multimodal demand modelling. For the general case where demand is a function of GC – i.e. where we have a linear combination of time and money – it is well known that VTTS is equal to the ratio of the time and cost elasticities multiplied by the ratio of cost to time. If we have $D = f(VTTS.T + C)$, then $\eta_T = \dfrac{\partial f}{\partial T}.\dfrac{T}{D} = VTTS.f'(VTTS.T + C).\dfrac{T}{D}$

[48] this is also the recommendation of the authors of Appendix C of the DfT's VTTS Response (DfT, 2016).

and $\eta_C = \frac{\partial f}{\partial C}.\frac{C}{D} = f'(VTTS.T + C).\frac{C}{D}$. Hence $\frac{\eta_T}{\eta_C} = \frac{VTTS.f'(VTTS.T + C)}{f'(VTTS.T + C)}.\frac{T}{D}.\frac{D}{C} = VTTS.\frac{T}{C}$. However, this property is only valid if demand is a function of GC: the PDFH function does not combine Fare and GJT in a linear manner, and no deduction about VTTS is possible from this functional form.

Indeed, it is not clear whether the PDFH function can be made consistent with utility theory. Nevertheless, it should be possible to apply the RoH by linearising the demand function, **provided** that an appropriate VTTS can be assumed: this is akin to Method 1 suggested by de Jong *et al* (2007).

The use of elasticities applied to a matrix of station-to-station flows implicitly allows for changes of mode and frequency, but not for destination choice. Attempts have been made to see whether the RUDD data, which underlies most of the current recommendations on elasticities, can provide any evidence on destination choice, but these have largely been unsuccessful.

This suggests that we might be able to re-cast the forecasting equation in a discrete choice context, firstly as a mode choice and then embedding it within a frequency choice. For mode choice, we would have $p_{rail,RS} = \frac{\exp(V_{rail,RS})}{\exp(V_{rail,RS}) + \exp(V_{other,RS})}$ where $V_{rail,RS} = f\left(Fare_{RS}, GJT_{RS}\right)$.

If we then postulate a frequency model as a choice between travelling by rail and not travelling, this can be formulated as follows:

$$p_{travel,RS} = \frac{exp\left(\theta.V_{RS}^*\right)}{A + exp\left(\theta.V_{RS}^*\right)} = \frac{1}{1 + A.exp\left(-\theta.V_{RS}^*\right)} \qquad (5.4)$$

where ln(A) is the utility of not travelling, $0 \le \theta \le 1$ and $V_{RS}^* = \ln\left[\exp(V_{rail,RS}) + \exp(V_{other,RS})\right]$.

The question is then whether this can be calibrated to reproduce the required elasticities to a satisfactory approximation.

The use of the composite GJT, combining in-vehicle time, an estimate of waiting time/schedule delay based on the service interval and possible interchange penalty, is of long-standing in PDFH, though recent work (for PDFH6.1) has attempted to identify separate effects and/or revised weights for the components. Comparable elements would typically be included in the utility for multi-modal models.

The constant elasticity approach is compatible with the framing effect[49], in that for a given change in Fare or GJT, the proportional response in demand is lower the larger the base level. It has been the mainstay of PDFH recommendations, although in a few cases – such as disruption – a semi-elasticity approach has been proposed. Nonetheless, it is possible that other functional forms (such as log + linear) could be tried, as well as the discrete choice approach just suggested. There have also been attempts by Wardman to estimate GC elasticities directly, and these could be further investigated: this involves adding the GJT and Fare terms after weighting with an independently assumed VTTS.

[49] We may note that the airport access work reviewed in §2.1 did not find any framing effect in respect of the interchange penalty.

In relation to this, the Brief also touches on the issue of the growth in VTTS over time, with an implication for the units in which GC should be measured (specifically, money or time). It is worth recalling what was said about this in the first UK VTTS study:

> Under the assumption that β [the scaling factor on GC] remains constant, it then appears to matter what units are used. It is worth while quoting McIntosh & Quarmby on this point (p 15):
>
> 2.2.42 "At first sight it would seem obvious to scale the parameters so that the behavioural costs (b) are in money units. However, projecting values to a future date, for forecasting travel patterns, exposes a problem of consistency and comparability. As incomes rise, so a given cost will carry less weight; it may be better to scale the parameters so that the units of behavioural cost retain some absolute value over time. It can be argued that time has much the same value in terms of personal utility to people of different incomes, and to people living now and at some future date. There are, of course, arguments against this proposition, but at least it is probably more tenable than scaling on cost.
>
> "At the present time, therefore, it is recommended that in the forecasting procedures all the behavioural cost functions (b) are scaled on time."
>
> 2.2.43 It is clear that this is only a tentative recommendation, and a considerable debate ensued (Goodwin (1974, 1978), Grey (1978a, b), Searle (1978)), essentially on the philosophical merits of the two contenders (Generalized Cost vs Generalized Time). However, the question demands empirical resolution rather than philosophical discussion. The kind of questions to be asked are: for two alternatives which differ only in respect of travel time (costs being equal), is there evidence that the proportion choosing alternative 1 increases or decreases over time; a corresponding question could be asked of alternatives with the same times but different costs. There is no a priori reason to expect any one outcome over any other. For instance, in terms of the utility formulation given in Eq (2.2)[50], the ratio of μ to λ might remain constant, while the absolute sizes of the coefficients increased or decreased; in such a case, the units would not be important, but nonetheless the random effects would change, with an implied change in the value of β. The most likely outcome is that both coefficients, and their ratio, would change!

The dilemma was explained, though not resolved, by Gunn (1983), who stated: "*It is not sufficient to predict the relative values of time and money for the forecast year; the absolute value of the standard deviation of the distribution of the 'other factors'* [i.e. the error term] *must also be supplied, in whatever unit system is adopted.*" The lack of evidence as to how this term

[50] The equation referenced was $U_i = \alpha_i + \lambda c_i + \mu t_i$ (+ ….)

(equivalent to the assumption about β) means that practice has stayed with the original McIntosh & Quarmby recommendation, choosing units of time for GC.

The problem arises, of course, from the assumed increase in VTTS with (real) income growth over time. It seems reasonable to argue that the marginal utility of cost will diminish (in absolute terms) with increased income, but there are many possible influences on the marginal utility of time, so that the actual direction of VTTS over time is uncertain, as is clearly set out in §7.8 of the 2015 VTTS Report. For both working and non-working time, the Report concluded, with some hesitation, that the existing assumption of a unit elasticity with *per capita* GDP should be maintained, but that further investigation was needed. It may be noted that the recent 2022 Dutch study found that "The VTTs obtained for passenger transport in this study are approximately 5-20% lower than was expected based on the results from the previous Dutch VTT 2009/2011 study (corrected for inflation and for 50% of the real income change)." With respect to rail demand, it might be expected that the fares elasticity would decline over time with increasing income, but no such effect has been found. It is not clear what the implications of this are.

Overall, there is more work to be done to produce a satisfactory reconciliation between the rail approach and the multi-modal approach. This becomes particularly important when rail improvements potentially lead to a change in choice of station, which the existing elasticity approach is not well set up to handle. There has been some tendency to redefine the rail model on a zonal basis, and to introduce access and egress. This leads to models of the type suggested by Wardman *et al* (2007)[51]:

$$Q_{RS} = GC(F_{RS}.GJT_{RS})^{\varepsilon_G}\left(.\sum_i Pop_i^{\varepsilon_{PO}}.Acc_{iR}^{\varepsilon_A}\right)\left(\sum_j Pop_j^{\varepsilon_{PD}}.Egr_{Sj}^{\varepsilon_E}\right) \quad \text{(Eq W5)}$$

$$Q_{iRSj} = K.GJTAE_{iRSj}^{\varepsilon_G}.F_{RS}^{\varepsilon_F}.Pop_i^{\varepsilon_{PO}}.Pop_j^{\varepsilon_{PD}} \quad \text{(Eq W10)}$$

where R and S are stations, and i and j are zones.

In (Eq W5) the population in potential catchment area zones (raised to a power) is multiplied by Access/Egress times (also raised to a power) and then summed over all potential areas. By contrast, in (Eq W10), access and egress are combined with GJT. Clearly alternative formulations can be proposed, and there is also the issue of whether separate GJT and fare elasticities are appropriate or whether some definition of GC can be used (as implied in (Eq 5)). But dominating both types of models is the definition of zones within the catchment area which, in standard terminology, is the area from which the station derives potential demand, and the literature displays various assumptions made in this regard, typically using a radius of influence (within x minutes etc.). In many cases, however, this is difficult to define: there may be competing stations, and there may be features relative to the journey being made (eg fast services versus stopping services).

[51] The equation numbers that follow are aligned with those in Wardman *et al*, though minor changes have been made in the notation.

6 Proposals for further investigation

In this Chapter, we set out some investigations which would throw further light on the issues raised in this Thinkpiece:

- Use of NoRTMS to investigate impact of different forms of Cost Damping on User Benefits, and relation to benefits using Small &Rosen theory

- Relationship between time, money and distance, using 2014 VTTS data

- Variations in highway assignment route choice in NoHAM from using a) different values of VTTS and b) time only.

6.1 Impact of different forms of Cost Damping on User Benefits

The NoRTMS model is a relatively conventional multi-modal model structured in line with TAG, and making use of cost damping Method B, as described in the Model Development Report – Tranche 3a Iteration 2e[52].

7.3.1. Cost Damping in the internal area has been applied following the advice in TAG Unit M2, damping generalised cost by a function of distance using the following formulation:

$$G' = \left(\frac{d}{k}\right)^{-\alpha} G \qquad [33]$$

Where

G'	is the damped generalised cost;
G	is the generalised cost;
d	is the trip length;
α	is a parameter between 0 and 1;
k [30km]	is a parameter of the model, that must be positive and in the same units as d. [We used the example value in TAG];

subject to a cut off-distance

d' [=30km]	is a cut-off distance below which no damping is applied.

7.3.3. In any testing of cost damping that we have undertaken, we have used the same cost damping parameters for all modes and purposes, in line with the TAG recommendation.

Testing the impact on benefits of different cost damping methods can be done using a single demand segment (though, since cost damping in NoRTMS only applies to journeys over 30 Km, it is important to ensure adequate demand[53]), and there is no need to allow for supply-

[52] Note that the damping parameters have been subsequently updated

[53] It is also probably sensible to avoid doubly constrained purposes

side effects. Hence, we need to set up a process whereby we have a base set of costs and demand for a chosen segment, we modify the costs (ideally, using some realistic scheme), re-calculate the demand (ie the choice of mode and destination), and carry out a benefit calculation for a single year.

In carrying out the tests, it is important that the different versions of the demand model (according to the cost damping Method being tested) have the same overall impact on demand, as measured by the realism tests. The existing NoRTMS model (with cost damping Method B) has already been calibrated to satisfy these tests: however, since it is not intended to use the supply model, the first requirement is to measure the implied first-round elasticities (essentially, the fuel price Km-elasticity and the GJT elasticity) for the chosen segment. Next, the RoH benefits using the current approach should be calculated.

As an alternative which is worth considering, a "sketch model" was built by Systra with a view of testing possible CNL specifications for mode and destination choice. While further investigation would be necessary, it is worth considering whether this could provide a suitable basis for carrying out the tests suggested below.

In reverting to the S&R benefit calculation, we have the potential issue of the scale. However, since the calculations are identical for distances below 30 Km, the effect will merely be to reduce the longer distance benefits by the factor $\left(\frac{dist_{ij}}{dist_0}\right)^{-\alpha}$.

We now turn to Method A. The Guidance suggests that the VTTS elasticities should be fixed at the recommended values, and that calibration resides essentially in the cut-off value, though if there is any recorded experience with this method, this should be consulted. The most time-consuming part will be the calibration to reproduce the realism elasticities, which could potentially involve changes to the lambda and theta parameters. Once a successful calibration is achieved, the RoH benefits should be calculated using both the current approach and the S&R approach. On the assumption that a realistic value of the distance cut-off has been obtained, the scale value issue can again be avoided by ensuring that the benefits are the same for distances lower than the cut-off.

It is proposed that any investigation involving Method C is postponed, since there is no evidence that this method has actually been used since the Guidance was issued.

Turning finally to Method D, this essentially repeats the procedures required for Method A in terms of calibration and subsequent benefit calculations. It would probably be sensible carrying out some preliminary investigation to see how well the log and linear cost function can reproduce the distance elasticity of VTTS used for Method A: this will give some ideas of the appropriate ratios for $\frac{-\lambda}{\beta_{cost}}$ and $\frac{\beta_{Logcost}}{\beta_{cost}}$.

The result of the investigation will to provide some idea of the scale of the difference between the current approach (merely using the demand model to predict the changes, while acting as if a standard generalised cost is being used for appraisal purposes) and the theoretical S&R approach.

6.2 Relationship between time, money and distance

While it is expected that there is a close correlation between all three of these variables, the actual functional form is uncertain, and in particular there might be some "tapering" at longer distances. While there are no doubt other data sources which could be used to investigate this (note that NTS does not have journey costs, except for public transport tickets, and even there the response is not always available), the 2014 VTTS is convenient. Door-to-door road distances were calculated automatically from the origin and destination locations marked on the Google maps within the questionnaire. Travel time was asked directly and could also be inferred from the start and end times of the journey. The car cost was estimated (using the TAG formula current at the time) in the questionnaire, but could be corrected by the respondent. For public transport the ticket cost was asked, and an estimate was made in the case of season tickets which, again, could be corrected.

Based on this data, a straightforward statistical analysis could be carried out to deduce the shape of the functions of time and cost with distance. In respect of rail fares, it has also been noted that fares have not been based on a x pence per mile basis since the early 1960s, and a 'distance taper' gradually came into use reflecting the relative difficulty of attracting travel over longer distances. The distance taper is still embedded in Off Peak Return fares which are still offered practically everywhere and the price of which is regulated directly by Government, though Advance Fares/Yield Management/Reduction in times when off-peak return fares can be used have largely taken over. A more detailed investigation would be useful.

6.3 Variations in highway assignment route choice

While various models could be used for this, the NoHAM model is perhaps the most readily accessible. The SATURN assignment program contains options to save the routes, and given the different purposes currently treated in the model with different assumptions about the vehicle VTTS, a first investigation would be to see how much the routes differ by purpose. This could be done for all O-Ds or for a selected set.

If the variation is considered limited, a further time-only assignment could be tried for additional comparison. In this case total (light) vehicle demand could be combined to a single user class.

Tests should also be carried out on the skimmed time matrices to investigate the source of variation.

7 Conclusions

We have shown that there is strong evidence for a "framing effect": however, it is less clear how best to formulate it. Effects have been found on both time and cost, but they vary between studies. The cost effect is often stronger, but not always.

The balance of evidence suggests that VTTS increases with distance, but the mechanism is unclear: both logarithmic transformations and elasticities have been used. The theoretical conditions for increasing VTTS are potentially dependent on possible non-linearities between time, cost and distance: a proposal has been made for investigating this further using existing data.

In principle, the framing effect seems much less appropriate for destination choice than for mode choice, and much of the evidence comes from mode choice or SP "route choice". The other significant source is the "RAND models" which use the log cost term: however, in these models there is no evidence available to separate the effects on destination versus mode choice. It is proposed that some separation could be achieved by means of the CNL model.

The key theory from Small & Rosen seems to imply that any transformation of utility from the standard linear Generalised Cost should be conveyed into the benefit calculation, but it is currently unclear how important this would be in practice: a proposal has been made for testing this for the various kinds of Cost Damping currently recommended.

The framing effect should in principle also apply to route choice (assignment), but the normally applied non-stochastic nature of this procedure means that framing could only be achieved using variations in (or distributed) VTTS. Nonetheless, doubts have been raised as to how important cost is for route choice: again, a proposal has been made for testing this.

8 References

Abrantes P A L & Wardman M R (2011), Meta-analysis of UK values of travel time: An update, Transportation Research Part A: Volume 45, Issue 1, Pages 1-17

AHCG (1999) Hague Consulting Group, Accent Marketing and Research and DETR, The Value of Travel Time on UK Roads, The Hague, Netherlands

Arup, ITS, Accent (2015) Provision of market research for value of travel time savings and reliability, Phase 2 Report, 14 August 2015

Bates J J (1998), Destination Choice and the "Need to Travel", PTRC European Transport Forum, Transportation Planning Methods Seminar (Vol 1), Loughborough

Bates J J (2003, 2006), Economic evaluation and transport modelling: theory and practice. In: Axhausen, K. (Ed.), Moving Through Nets: The Physical and Social Dimensions of Travel. Pergamon, Oxford.

Bates J J (2017), Forecasting the demand for transport, Chapter 11 in J Cowie and S Ison (eds), The Routledge Handbook of Transport Economics, Routledge

Bates J J (2024), Pivoting from a known base when predicting choices using logit models, Transportation Research Part A **179** 103869, January 2024, Elsevier

Ben-Akiva, M. (1974), Structure of Passenger Travel Demand Models, Transportation Research Record 526, pp 26-42

Ben-Akiva, M., A. Daly, and H. Gunn (1987). Destination Choice Models: Design and Appraisal, presented to PTRC Summer Annual Meeting.

Daly A J and Zachary S (1978), Improved Multiple Choice Models, in: Hensher D A & Dalvi M Q (eds), Determinants of Travel Choice, Saxon House, Sussex

Daly A J (1982), Estimating Choice Models Containing Attraction Variables, Transportation Research B, Vol. 16B, No. 1. pp. 5-15

Daly, A J and Carrasco J (2009), The Influence of Trip Length on Marginal Time and Money Values: An Alternative Explanation, Published in: Kitamura, R., Yoshii T., and Yamamoto T., eds., The Expanding Sphere of Travel Behaviour Research: Selected Papers from the Proceedings of the 11th International Conference on Travel Behaviour Research, Emerald Group Publishing

Deaton, A. and J. Muellbauer (1980) *Economics and Consumer Behaviour*, Cambridge University Press, Cambridge.

DfT (2016), Understanding and Valuing Impacts of Transport Investment: Value of Travel Time Savings Consultation Response

Fox, J., Daly, A. and Patruni, B., (2009). 'Improving the treatment of cost in large-scale models'. Paper presented at the European Transport Conference, Noordwijkerhout, October 5-7, 2009.

Fox, J and Daly, A (2013), Manchester Motorway Box: Post-Survey Research of Induced Traffic Effects - Model Estimation, RAND Europe Technical Report Prepared for the UK Department for Transport

Fox, J., Patruni, B., Daly, A and Patil, S (2014), PRISM 2011 Base: Mode-Destination Model Estimation, RAND Europe Report Prepared for Mott MacDonald

Gunn H F (1983), “Generalised Cost” or “Generalised Time”: A Note on the Forecasting Implications. Journal of Transport Economics and Policy Vol. 17; No. 1, pp 91-94

Harrison, A. J. (1969) Some Results From Using Interview Methods in Time Research., Dept. of the Environment, Highway Economics Unit, Time Research Note 6

de Jong, G, Daly A.J , Pieters M and van der Hoorn A.I.J.M. (2007), The logsum as an evaluation measure: Review of the literature and new results, Transportation Research Part A Vol 41 (9), pp 874-89

König, A., Axhausen, K.W. Abay, G. (2004), Zeitkostenansätze im Personenverkehr, Forschungsauftrag Nr. 2001/534 auf Antrag der Vereinigung Schweizerischer Verkehrsingenieure (SVI), Januar 2004

LGORU (1968) [Local Government Operational Research Unit], Modal Split: Factors Determining the Choice of Transport for the Journey to Work, Rept. C32

Mackie, P.J., Wardman, M.R., Fowkes, A.S., Whelan, G.A., Nellthorp, J., and Bates, J.J. (2003), Value of Travel Time Savings in the UK. Report to Department for Transport. Available at: http://eprints.whiterose.ac.uk/2079/2/Value_of_travel_time_savings_in_the_UK_protected.pdf

McFadden D (1978), Modelling the Choice of Residential Location, in: Karlqvist A, Lundqvist L, Snickars F & Weibull J W (eds), Spatial Interaction Theory and Planning Models, North-Holland, Amsterdam

McIntosh, P.T. and D.A. Quarmby (1970) Generalized costs and the estimation of movement costs and benefits in transport planning, *MAU Note*, **179**, Department of Transport, London.

MVA Consultancy, Institute for Transport Studies University of Leeds and Transport Studies Unit University of Oxford (1987) *The Value of Travel Time Savings*, Policy Journals, Newbury.

MVA Consultancy, Oscar Faber TPA and Institute for Transport Studies University of Leeds (1994), Common Appraisal Framework for Urban Transport Projects – Final Report

Nellthorp, J. and Hyman, G. (2001): Alternatives to the rule of a half in matrix-based appraisal. Proceedings of European Transport Conference, Methodological Innovations Seminar, Cambridge.

Ortúzar J de D. (2001), On the Development of the Nested Logit Model, Transportation Research 35 B (2), pp 213-216

Pearce D. W. and C A Nash (1981), The Social Appraisal of Projects: A Text in Cost-Benefit Analysis, MacMillan

Quarmby, D. A. (1967) Choice of Travel Mode for the Journey to Work: Some Findings. Jour. of Transport Economics and Policy, Vol. 1, No. 3

Significance, VU University, John Bates Services, TNO, NEA, TNS NIPO, & PanelClix (2013). Values of time and reliability in passenger and freight transport in The Netherlands. Report for the Ministry of Infrastructure and the Environment, Significance, The Hague.

Significance (2023), Values of Time, Reliability and Comfort in the Netherlands 2022: New values for passenger travel and freight transport, Technical report | 21 November 2023

Small, K.A. and H.S. Rosen (1981) Applied welfare economics with discrete choice models, *Econometrica*, **49** (1) 105-130

Sugden R (1999), Developing a consistent cost-benefit framework for multi-modal transport appraisal: A report for the Department of the Environment, Transport and the Regions

Wardman M, Lythgoe W & Whelan G (2007), Rail Passenger Demand Forecasting: Cross-Sectional Models Revisited, Railroad Economics: Research in Transportation Economics, Volume 20, 119–152

Williams, H.C.W.L. (1977), On the formation of travel demand models and economic evaluation measures of user benefits, Environment and Planning A, 9 (issue) 289–344.

Appendix A Terms of Reference

Consistency between transport modelling and economic theory: think-piece

Purpose and background

Cost damping

The Department for Transport's (DfT) Transport Analysis Guidance (TAG) Unit M2.1 allows the use of cost damping in modelling. There is strong empirical evidence that the sensitivity of demand responses to changes in generalised cost reduces with increasing trip length (see, for example, Daly (2008, 2010)). The mechanisms by which this may be achieved are generally referred to as 'cost damping'.

Cost damping is part of our current best understanding of travel behaviour and would be expected to be incorporated into models. There are, however, some contexts where the range of travel distances that need to be represented in a transport model are limited.

If cost damping is employed, TAG recommends it should apply to all person demand responses. The same cost damping function should be applied to both car (private) and public transport costs. While the starting position should be that the same cost damping parameter values are used for both modes, it may be necessary to vary the cost damping parameters between the modes in order to achieve satisfactory realism test results. It may also be necessary to vary cost damping parameters by trip purpose. However, these variations by mode and purpose should be avoided unless it is essential to achieve acceptable model performance (and always reported). The use of cost + log(cost) has also been popular with RAND, and there may be scope to also use time + log(time). This should be explored further.

Research undertaken for the Department has demonstrated that for all trip purposes there is a relationship between travel distance and the value of travel time savings. This evidence likely reflects that travellers' sensitivity to cost declines with trip length more rapidly with distance than their sensitivity to time. The implication is that this ideally should be expressed in the utility function. This was explored in Annex C of the 2016 VTTS consultation response by Hanson & Whittaker. Increasing VTTS with trip distance is one of the recommended methods of cost damping within TAG and departs from linear utility functions. There is therefore a close association here between the appraisal and modelling guidance. Within the usual generalised time based formulation of demand models, this approach to cost damping is equivalent to decreasing the marginal utility of cost with trip length.

From an economic perspective, cost damping is a modification of utility functions, and ideally the same utility functions should be used for both modelling and appraisal of behavioural changes. A justified exception to this could be where behaviour is governed by constraints which are not explicitly represented within the transport model. In these cases, there could be a muted behavioural response but a large appraisal value (an example is the treatment of rail reliability (i.e. lateness) in TAG, where a relatively modest demand elasticity is combined with a large appraisal multiplier on late time).

Recent discussions in relation to major rail projects have raised the question of whether cost damping should be applied to the generalised costs which are fed into WITA. This also raises an analogous question about TUBA. In one such study reviewed by a TASM panel, use of the VDM cost damped matrices in WITA calculations caused significant changes to outturn agglomeration estimates.

Other cases of inconsistent appraisal and modelling values

There is also a broader question around the consequences of divergent values between modelling and appraisal. While it is generally well understood that some discrepancies can hardly be avoided and can indeed be tolerated, as implied by TAG Unit M2.1, the scale and potential severity of any issues this could create have been less well explored by DfT. There are a few key cases/issues in point which could be considered within this think-piece.

- Where travellers face direct trade-offs between time and price, having very different appraisal VTTS compared to modelling VTTS could lead to wrong signed benefits. In other cases, there could still be a bias even if the sign remains correct.
- Within rail modelling, where it is common to use demand models based on separate elasticities to fare and GJT. This is not consistent with the generalised time (GT) (or cost (GC)) formulation used in multimodal demand modelling. In particular, while the elasticities to GJT and F can be used to infer a VTTS, we think this only holds within a GT (or GC) based model. As a result, the validity of applying VTTS to appraise the outcomes of unimodal demand models may be called into question.
- Related to the previous bullet, whether the implied reduction in forecast year price elasticities, as value of time grows, is empirically and behaviourally valid. While this follows from a GT based model of demand, there does not appear to have been an historical reduction in the elasticity. This may suggest empirical evidence in favour of the uni-modal 'constant elasticity' modelling approach. If this evidence is accepted, what can/should be done, and what are the implications of this in terms of the theoretical consistency of appraisal?
- Another recently re-emerging issue surrounds non-fuel vehicle operating costs. Research by Ricardo for DfT points to a significant upwards adjustment in these costs, which would further call into question the current TAG assumption that these costs are unperceived for non-work travel.
- The handling of combined sub-mode PT assignment. The TUBA manual currently recommends a flow weighted average costs but hints at the possible use of composite costs. If composite costs were used, however, these would embed behavioural/modelled VTTS which may differ from the desired appraisal VTTS.
- Recent work by DfT on distributional weights points to the use of more segmented VTTS (better reflecting heterogeneity), subject to application of distributional weighting as part of the appraisal process. This presents an opportunity to enhance the consistency between modelling and appraised behaviour, as a single distributional weight can be applied to a traveller's overall change in generalised cost (which can

then be permitted to be constructed with a best estimate of their behavioural VTTS, rather than an 'equity' figure as at present).

- There are also a range of features present in most contemporary / recent choice models for estimating VTTS, such as the use of reference dependent utility functions, multiplicative error terms and the use of 'random valuation' as opposed to 'random utility' models. It is currently not clear what, if any, implications these should or could have for appraisal.
- More broadly, how and whether findings from behavioural evidence, such as a smaller unit value for small time savings, or an increased aversion to £1 of toll versus £1 of fuel cost, should be carried forward into applied appraisal (and modelling). There is an argument that even if these findings are empirically valid, they should not service as guide to public policy.

Aims, objectives and scope

The Department overall aim is to further its understanding around the consistent application of cost damping within modelling and appraisal, as well as the broader issue of divergences between modelling and appraisal values. Including, specifically, advice on whether and how cost damping, where employed, should have implications for economic appraisal.

The commission has four main objectives:

1. A conceptual framework/model or formal mathematical approach should be deployed to explore the issues surrounding cost damping and inconsistent appraisal and modelling values. This should be used to tackle the and provide recommendations around the issues discussed above in the *Purpose and background* section.
    a. As part of this, precise statements should be made about where differences between modelling and appraisal values are (or are not) valid theoretically, versus where they *may* be tolerated practically.
2. Provide recommendations on whether and how 'damped' costs, where the demand model uses damping, should be carried forward to appraisal. This should cover user benefits as well as wider impacts. Consistency with underlying empirical evidence should also be considered (e.g. the measures of impedance used to derive agglomeration elasticities), as well as alignment with economic theory and key principles such as utility maximisation.
3. Set out the key considerations for cost damping under the two main demand modelling 'paradigms' in use for scheme appraisal: discrete choice based multimodal models, and unimodal elasticity-based models.
4. Explore the issues, and provide recommendations where appropriate, around practically reflecting cost damping within appraisal tools such as TUBA and WITA.

In delivering against these, the key relevant literature should be surveyed and referenced. This does not need to be a systematic literature review, but should cover all important developments highly relevant to the questions at hand.