New appraisal values of travel time saving and reliability in Great Britain

Richard Batley[1] · John Bates[2] · Michiel Bliemer[3] ·
Maria Börjesson[4] · Jeremy Bourdon[5] · Manuel Ojeda Cabral[1] ·
Phani Kumar Chintakayala[1] · Charisma Choudhury[1] ·
Andrew Daly[1] · Thijs Dekker[1] · Efie Drivyla[5] ·
Tony Fowkes[1] · Stephane Hess[1] · Chris Heywood[6] ·
Daniel Johnson[1] · James Laird[1] · Peter Mackie[1] ·
John Parkin[7] · Stefan Sanders[5] · Rob Sheldon[6] ·
Mark Wardman[1] · Tom Worsley[1]

Abstract This paper provides an overview of the study 'Provision of market research for value of time savings and reliability' undertaken by the Arup/ITS Leeds/Accent consortium for the UK Department for Transport (DfT). The paper summarises recommendations for revised national average values of in-vehicle travel time savings, reliability and time-related quality (e.g. crowding and congestion), which were developed using willingness-to-pay (WTP) methods, for a range of modes, and covering both business and non-work travel purposes. The paper examines variation in these values by characteristics of the traveller and trip, and offers insights into the uncertainties around the values, especially through the calculation of confidence intervals. With regards to non-work, our recommendations entail an increase of around 50% in values for commute, but a reduction of around 25% for other non-work—relative to previous DfT 'WebTAG' guidance. With regards to business, our recommendations are based on WTP, and thus represent a methodological shift away from the cost saving approach (CSA) traditionally used in WebTAG. These WTP-based business values show marked variation by distance; for trips of less than 20 miles, values are around 75% lower than previous WebTAG values; for trips of around 100 miles, WTP-based values are comparable to previous WebTAG; and for longer trips still, WTP-based values exceed those previously in WebTAG.

✉ Richard Batley
R.P.Batley@its.leeds.ac.uk

[1] Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

[2] John Bates Services, Abingdon, UK

[3] Institute of Transport and Logistics Studies, University of Sydney, Sydney, Australia

[4] Centre for Transport Studies, Royal Institute of Technology, Stockholm, Sweden

[5] Arup, London, UK

[6] Accent, London, UK

[7] Centre for Transport and Society, University of West of England, Bristol, UK

🌥 Springer

Introduction

This paper provides an overview of the study 'Provision of market research for value of time savings and reliability' undertaken by the Arup/ITS Leeds/Accent consortium for the UK Department for Transport (referred to henceforth as the 'Department'). Whilst the full technical reports of the study (Arup, ITS Leeds and Accent 2015a, b) are already in the public domain, and the study team's recommendations have largely been accepted and implemented by the Department (DfT 2015, 2016, 2017), the present paper seeks to present a digestible summary that is accessible to a broad academic readership.

In the context of transport appraisal, one of the most important concepts is that conventionally referred to as the 'value of time'. This does not refer to the value that might be placed on time spent in travel, but should be seen as shorthand for the 'value of changes in travel time', relative to a reference case when investment takes place. These changes may be positive or negative, but historically have been referred to as 'savings'. Travel time savings are usually the largest single component of the monetised benefits of transport infrastructure projects and policies. Furthermore, time-related benefits such as reliability and relief of overcrowding on public transport (PT) are conventionally valued through multipliers on the 'value of time'. In this paper we have chosen to refer to the 'value of travel time' (VTT) to convey this concept.

There have been three waves of national studies of VTT in Britain. First, a series of research studies during the 1960s, the results of which were synthesised and adopted by the Department in appraisal guidance. Second, the MVA, ITS Leeds and TSU Oxford (1987) study, which led to updated guidance. Third, the 1994 study by Accent and Hague Consulting Group (published some years later as AHCG 1999), which was re-analysed by ITS Leeds (Mackie et al. 2003) before again being committed to guidance. Between 2003 and 2016, appraisal guidance was intermittently revised and updated by the Department, to reflect changes in incomes and travel patterns [e.g. as documented in WebTAG[1] Unit A1.3 (DfT 2014)]. The underpinning behavioural estimates of VTT were not however re-surveyed. In other words, between 2003 and 2016, appraisal guidance on VTT was based on survey data collected in 1994, and analysed using methods considered best-practice in 2003.

Over the subsequent 20 years, incomes, prices, demography and the mix of travel by purpose and trip length have all changed. Possibly more significant is that the world has moved on in other ways—the internet revolution, the quality and comfort of vehicles, working practices and, perhaps most fundamentally, the ways in which people perceive time spent travelling. It does not seem credible to suggest that such phenomena can be accommodated simply through updating historical behavioural values for changes in incomes and travel patterns. Also, over the same period, there have been developments in methods for collecting survey data and estimating VTT measures. Contrasting 2016 best-practice against 2003, we are now better equipped to understand various empirical phenomena such as variability in VTT (e.g. 'deterministic' variation across different travel

---

[1] WebTAG is an acronym for Web-based Transport Analysis Guidance. This resource is developed and published by the Department, and contains official "*guidance on the conduct of transport studies*" in the UK.

conditions, travellers and trip types, as well as inherently 'random' or 'unobserved' variation), discontinuities in VTT (e.g. 'size' and 'sign' effects (de Borger and Fosgerau 2008), as well as the phenomenon of 'cost damping' Daly 2010), and any residual 'uncertainty' in the resultant values (which would contribute to confidence intervals on the VTTs used in appraisal Daly et al. 2012a).

In response to these challenges, the Department has, since 2009, taken steps to review the theoretical, methodological and evidential basis of its VTT guidance. Among the key actions have been the Department's commissioning of scoping studies concerning the valuation of travel time, for both non-work and business. The ITS Leeds, John Bates and DTU (2010) study 'Values of travel time savings: updating the values for non-work travel' scoped out the research activities that would be required to update the values for non-work travel, and issued recommendations on which packages of activities should be commissioned. In a similar fashion, the ITS Leeds, John Bates and KTH (2013) study 'Values of travel time savings for business travellers' reviewed the feasibility and theoretical accuracy of different methods for estimating VTT for business travellers, as well as evidence from the UK and overseas on the values emanating from these different methods [see also the companion journal paper Wardman et al. (2015)].

Informed by these scoping studies, the Department commissioned new market research to deliver updated evidence on values of travel time and reliability (DfT 2013), and the resulting tender was awarded to the Arup/ITS Leeds/Accent consortium. The study was conducted in two phases, across a challenging timeframe of 11 months. Phase 1 of the study, which was undertaken from June to September 2014, involved the development and testing of methods for undertaking the requisite market research. Phase 2 involved a substantial field survey and detailed modelling to complete estimation of the values of travel time using the collected data.

Study aims, scope, and delivery

The Department specified the following aims for the research:

- To provide recommended, up-to-date national average values of in-vehicle travel time savings, covering business and non-work travel, and based on primary research using "*modern, innovative methods*".
- To investigate the factors which cause variation in the values (e.g. by mode, purpose, income, trip distance or duration, productive use of travel time etc.) and use this to inform recommended segmentation of the values.
- To improve our understanding of the uncertainties around the values, including estimating confidence intervals around the recommended values.
- To consistently estimate values for other trip characteristics for which values are derived from the values of in-vehicle time savings.

In pursuit of these aims, we employed an analysis framework based upon the primary dimensions of trip purpose and mode of travel (see Table 1). Within this framework, key features of the present paper include the following:

- We focus on the mechanised modes of car, bus, rail and 'other PT'.[2] The walk and cycle research encountered significant methodological challenges, and was eventually

---

[2] 'Other PT' refers to 'other public transport', namely trams, light rail and London Underground. By 'rail' we mean heavy rail.

Table 1 Summary of survey design

Mode of travel	Trip purpose				SP experiment	Covariate
	Commute	Other non-work	Employees' business	Employers' business		
Car	SP	SP	SP	SP	SP1: Time	(a) Income
Bus	SP	SP	N/A	N/A	SP2: Time and Reliability	(b) Distance/duration
Rail	SP and RP	SP and RP	SP and RP	SP	SP3: Time and Quality (e.g. crowding, congestion and other types of time)	(c) Productive time
'Other' PT	SP	SP	SP	SP		(d) Trip type etc.
Walk and cycle	SP	SP	N/A	N/A		

N/A deemed not to be applicable on the grounds that trip rates are relatively low, *SP* Stated Preference, *RP* Revealed Preference

reported to the Department separately from the mechanised modes, and with only tentative recommendations (Arup, ITS Leeds and Accent 2015b).

- Informed by the scoping studies, the Department directed us to value travel time savings using willingness-to-pay (WTP) methods—for both non-work and business.
- The latter directive in respect of business reflected the Department's interest in replacing the long-standing Cost Saving Approach (CSA) (e.g. Harrison 1974) for valuing business travel time savings with WTP—if the evidence base was adequate to support such a change.
- Whilst the direction was to implement WTP methods primarily through Stated Preference (SP) data, the Department encouraged us to validate the SP with Revealed Preference (RP) data.[3]
- The Department directed us to examine business travel from two alternative perspectives, namely those of the employee and employer. With regards to the latter, the employee is effectively 'spending' the business's time and money, and it is important therefore that the employee reports a WTP representative of his/her employer's interests. Whilst directed to examine business using WTP, the Department specifically excluded the so-called 'Hensher' equation (Hensher 1977) from our scope.

Design and implementation of the market research

This section sets out the process followed for designing and implementing the market research. As noted in the "Introduction" section, the market research was focussed around SP, but complemented by RP as a validation device. These methods were designed and developed in a systematic fashion, involving the following steps:

1. Qualitative research was conducted in certain areas of the brief that were considered to involve particular challenges; these areas included the valuation of business travel time savings, the presentation of reliability, and the presentation of car use costs.
2. The prior qualitative research informed the design of the SP and RP experiments, as well as the development of the questionnaires more generally.
3. Cognitive depth interviews tested the flow, comprehensibility and wording of the questionnaires.
4. Pilot surveys were administered in two waves, involving testing of all data collection and analysis methods.
5. The field survey involved a full 'roll-out' of the data collection and analysis methods, exploiting lessons learned from the pilot surveys.

Whilst the aims of the study required us to conduct primary research using 'modern, innovative methods', it is worth remarking that, from a theoretical perspective, we sought to ground these methods within the standard microeconomic framework underpinning both non-work and business VTT, as rationalised by Becker (1965), De Serpa (1971) and Evans (1972), and as codified in Section 3.3 of MVA et al. (1987).

---

[3] As it transpired, the RP analysis proved extremely challenging, and only limited insights could be gleaned in terms of validation of the SP; see the study report (Arup, ITS Leeds and Accent, 2015a; Chapter 5) for a full discussion. The RP survey and analysis are not discussed in this paper.

Stated Preference (SP) approach

Experimental design method

Table 2 summarises the context and content of the principal SP experiments.

Presentational considerations Informed by the prior qualitative research, together with insights gleaned from previous UK national VTT studies and the literature more generally, it was decided to define car cost in terms of fuel cost and public transport cost in terms of one-way ticket price. For car, we were conscious that the use of fuel cost could undermine the realism of SP choices between faster/cheaper versus slower/dearer journeys, in the sense that longer journeys might in practice consume more fuel and therefore be more costly. On balance, however, it was judged that fuel costs represented the best (or least worst, perhaps) available representation of costs, especially in a British context with very few toll roads/bridges. For public transport travelcard users, an appropriate one way ticket price was derived from the monthly or annual cost. The SP experiments were explicit concerning the definition of cost for car and public transport, and the preamble instructed

Table 2 Summary of principal SP formats by game and mode

Game and mode	Description of SP format
SP1	SP1 used a generic format across all modes, presenting respondents with an 'abstract' choice between two options described only on the basis of travel time and travel cost, where one option was cheaper, but the other option was faster
SP2	SP2 also presented respondents with an abstract binary choice, still focussing on travel cost and travel time but where, for travel time, five different typical trip outcomes were presented for each alternative as a representation of travel time variability
SP3	SP3 used somewhat different presentations across modes, whilst nevertheless retaining an abstract binary choice context, as described for each mode below
SP3 car	For car, the two options were described in terms of travel cost for each trip and the amount of time that each trip spends in three types of driving conditions (free-flow, light traffic, heavy traffic)
SP3 rail	For rail, two different experiments were used:
	(a) For the first group, we presented a choice similar to SP1, with the difference that for each alternative we additionally defined the level of crowding applying to the trip
	(b) For the second group, we presented a choice between up to three operators, described in terms of travel time, fare and headway
SP3 bus	For bus, two different experiments were also used:
	(a) For the first group, we presented a crowding game analogous to the rail game, albeit with different crowding definitions
	(b) For the second group, we presented a choice between two bus routes described in terms of free-flow time, slowed down time, dwell time, headway and fare
SP3 'other PT'	For 'other PT', two different experiments were again used:
	(a) For the first group, we presented a crowding game analogous to the bus game
	(b) For the second group, we presented a mode choice game ('other PT' against either bus or rail) using time, headway and cost as attributes

respondents to assume that the offered alternatives were identical in terms of other facets of cost.

Examples of SP1-3 for car are presented in Figs. 1, 2 and 3 respectively. Given the policy context of this research, it was judged important to remain faithful to established UK practices for valuing time savings, reliability and quality, and this motivated the presentations shown. Thus, Fig. 1 is essentially AHCG's (1999) presentation which underpinned previous WebTAG guidance on VTT, Fig. 2 is a variant of Hollander's (2006) presentational approach which was originally developed in the context of UK bus, and Fig. 3 is a simplified version of the presentation employed in ITS Leeds' (2008) after-study of the M6 Toll. Whilst similar approaches were generally followed for public transport, an example of SP3a for rail is also provided (Fig. 4), to illustrate the manner in which we addressed the PT-specific issue of crowding. This is a simplification of MVA's (2008) presentation which underpins Passenger Demand Forecasting Handbook (ATOC 2012) guidance on crowding penalties.

We should acknowledge that, whilst commonplace in most UK and many European studies (though not the Swiss and German studies), the presentational approaches shown in Figs. 1, 2, 3 and 4 have been supplanted by alternative approaches in some other countries (e.g. Australia). The approaches used here may not, therefore, be considered best practice elsewhere, especially in terms of: (1) separating out different time components across different SP games; (2) the means of presenting reliability; and (3) the reliance on binary choice tasks only. However, a strong reason for using these approaches is that they facilitated objective comparison—on a like-for-like basis—of updated values against previous WebTAG and PDFH values.

Exceptions to the formats described in Table 2 included SP experiments focussed upon public transport mode choice (involving separate games for concessionary and non-concessionary travellers) and rail operator choice (corresponding to the RP context). In most cases (exceptions being walk and cycle and bus concessions), respondents received all three SP games (i.e. SP1, SP2 and SP3). SP1 was presented always presented first, whilst the ordering of SP2 and SP3 was randomised in order to mitigate 'order effects'.

Statistical considerations The SP designs for this study were based upon the concept of Bayesian D-efficiency, which has the potential to give more precise (in terms of reduced standard errors) parameter estimates when used appropriately (Rose and Bliemer 2014). Whilst it was clear that different designs would be needed for different games (e.g. SP1–3), we also recognised that efficient designs needed to be optimised for the specific values of attributes and priors of interest. This had two separate dimensions in the present context.

Firstly, as substantial differences in values of time (and other valuations) were expected to exist between business and non-business travellers, separate designs were produced for these two purpose segments. In essence, this means that the trade-offs presented to business

	Option A	Option B
One way fuel cost	£33.30	£35.00
One way travel time by car	4 hours 23 minutes	3 hours 30 minutes

Option A   Option B

Fig. 1 Time versus cost experiment (SP1) for car

Imagine that on five occasions that you make the car journey **departing at the same time and on the same day of the week**, the actual travel time varies for the reasons suggested previously. We want you to think about that car journey and look at the two options below, each of which show five possible travel times that could arise.
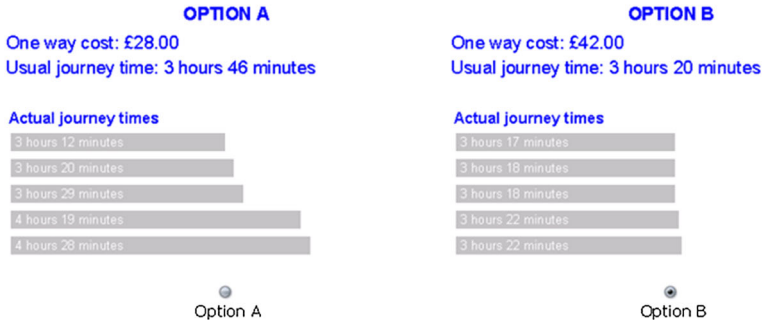
OPTION A	OPTION B
One way cost: £28.00	One way cost: £42.00
Usual journey time: 3 hours 46 minutes	Usual journey time: 3 hours 20 minutes

Actual journey times

Option A	Option B
3 hours 12 minutes	3 hours 17 minutes
3 hours 20 minutes	3 hours 18 minutes
3 hours 29 minutes	3 hours 18 minutes
4 hours 19 minutes	3 hours 22 minutes
4 hours 28 minutes	3 hours 22 minutes

Option A   Option B

Fig. 2 Time versus cost versus reliability experiment (SP2) for car

	Option A	**Option B**
One way travel cost	£42.00	£37.60
Traffic conditions	1 hour 45 minutes in heavy traffic 11 minutes in light traffic 2 hours 53 minutes in free flowing traffic	2 hours 11 minutes in heavy traffic 2 hours 36 minutes in light traffic 57 minutes in free flowing traffic

Option A   Option B

Fig. 3 Time versus cost versus quality experiment (SP3) for car

	Option A	**Option B**
One way travel time	3 hours 54 minutes	3 hours 18 minutes
One way travel fare	£18.00	£24.00
Crowding level when you boarded	Seated, 100% of seats occupied, eight people stood around each door	Standing, 100% of seats occupied, one person stood around door

Option A   Option B

Fig. 4 Time versus cost versus quality experiment (SP3) for rail

travellers were geared towards their likely higher willingness-to-pay, thereby giving us more robust estimates in the analysis.

Secondly, the surveys presented respondents with trips framed (or 'pivoted' in experimental design terms) around the travel time and (monetary) cost of a recent trip they had made. Additionally, the percentage variations in travel time and cost around the reference trip were varied with trip characteristics. Simply using a generic design—in terms of percentage variations—across all trip types, could have incited a major loss of efficiency.

Separate designs were produced for a set of representative trips. Each respondent was then given a design based on the trip closest to their reference trip (in terms of the smallest percentage difference between the reference values for the design and that respondent's values for time and cost), with percentage variations (or pivots) applied to the specific reference trip for that person, where these pivots were obtained from the design.

The number of reference trips used varied by mode, with the lowest number for bus (2) and the highest number for rail (20). Each SP game presented a respondent with five separate choice scenarios. The actual designs made use of a number of rows that was larger than the number of tasks assigned to a single respondent, to ensure sufficient richness in the variations in the data. As an example, for car SP1, the boundary value of times ranged from £0.15/hour to £372/hr. This was of course partly a result of some very cheap and very expensive reference trips, but even when looking at the reference trips used in the design process, the range extended from £0.45/hr to £90/hr. The overall design was then split into a number of distinct blocks at the design stage, minimising correlation between attributes and blocks, and each block was used as closely as possible a uniform number of times across the sample of respondents. The number of rows for designs was set to 25 after extensive testing. In total, 315 designs were produced for this study.

The subsequent discussion details the inputs used in all designs and explains how the design outputs were used to compute the values presented to respondents. We also look at any additional constraints imposed on the designs, where it should be noted that, by default, the design approach already avoided scenarios in which one alternative was dominated, e.g. there was no possibility in the simple time-money trade-offs that one option was both faster and cheaper than the other.

Car games

- No additional constraints were imposed on SP1 given the above mentioned avoidance of dominance.
- For SP2, and for reasons of realism, we excluded cases where either the shortest travel time or the highest travel time was combined with the highest level for travel time variability.
- For SP3, we imposed additional constraints which guaranteed that the implied distances for the two trips differed by no more than 25%, again for reasons of realism (assuming that light traffic speed would be 80% of free-flow, and heavy traffic speed would be 60% of free-flow). The design allowed for both increases and decreases around reference values, and the design process sought to achieve attribute level balance, i.e. guaranteeing that increases were as likely as decreases. We introduced some flexibility into this process, by allowing the share of increases and decreases to be different from 50/50, which in turn allowed the constraints on the total sum to be met.
- For all games, we defined an adjusted free-flow reference time (AFF) on the basis of the current free-flow time (CFF) and current total time (CTT) as:

$$\text{AFF} = \text{CFF} + \max(0, -(\text{CFF} + \min(\Delta\text{FF}) \times \text{CTT}))$$

Here, min (ΔFF) was the smallest (i.e. most negative) additive percentage shift used for free-flow time across both alternatives and across all five tasks for the respondents. This meant that the adjusted free-flow was shifted upwards for all tasks and all alternatives in those cases where any of the alternatives in any of the tasks would have required censoring. We then used the definition:

$$FF = AFF + \Delta FF \times CTT$$

to compute the values to be presented.

Rail games

- No additional constraints were imposed on any of the designs.

Bus games

- No additional constraints were imposed on SP1 or SP2.
- For SP3a, we imposed additional constraints which guaranteed that the implied distances for the two trips differed by no more than a third, again for reasons of realism (assuming that slowed down time speed would be 50% of free-flow).
- No additional constraints were imposed on SP3b.

'Other PT' games

- No additional constraints were imposed on any of the designs.

General public SP market research method

The core research method for the SP survey was intercept recruitment (80% of recruits) followed by on-line or telephone interviews; this was supplemented by telephone recruitment (20%) again with on-line or telephone completion.

Taking the NTS 2010-12 dataset as the benchmark for representativeness, 80% of all recorded trips cover less than 10 miles. However, many studies have found a strong relationship between VTT and distance (e.g. Mackie et al. 2003), and as we subsequently use distance weighting to derive appraisal values (see the "Appraisal values" section), it is important to be able to estimate VTT for longer trips accurately. Using the NTS definition of 'long distance' (i.e. greater than 50 miles), only 2% of trips in the NTS fall into this category. A telephone sampling approach predominantly samples short distance trips. On the other hand, an intercept sampling approach, as has been conventional in most VTT studies, favours longer distance trips, since these have a higher probability of being intercepted. Balancing these considerations, the chosen approach was two-pronged: predominantly using intercepts to ensure an adequate sample of the longer distance movements, and more generally business trips, but using telephone sampling to strengthen the sample in the shorter distances. An exception to this approach was 'other PT', which was recruited through intercept only, given the limited usage of this mode outside of London.

Another attraction of the intercept approach is that interviewers can be located where the target respondents are (e.g. at bus stops, rail stations and motorway service areas). This was particularly important to be able to recruit adequate samples of specific groups of target respondents who would otherwise be extremely difficult to recruit through other sampling approaches (e.g. those making specific 'other PT' and bus trips on corridors where there was a rail alternative (required for the operator choice SP exercise); those making trips on specific rail routes (to provide comparisons with the RP sample); long

distance car and rail travellers; employees' business travellers; or those no longer having a landline phone).

Intercept recruitment The intercept CAPI[4] survey was administered face-to-face using Android tablets. Interviewers approached a random sample of adults (typically 1 in 3) and asked scoping questions to check whether each respondent was in-scope and matched required quotas. If in-scope, the respondent was invited to undertake a follow-up survey either on-line or by phone. The interviewer collected their contact details (name and telephone number for follow-up telephone interview, and name and e-mail address for follow-up on-line survey). All intercept fieldwork took place on weekdays with fieldwork shifts either 07:00–13:00 or 13:00–19:00. Figure 5 shows the intercept locations, which were designed to cover car, rail, bus and 'other PT' users across the country.

The survey locations were selected to reflect:

- Coverage of the key trip purposes.
- A reasonable geographical spread across England, some coverage in Scotland, as well as some cross-border flows into Wales.
- A reasonable spread of the key market segmentations relevant to each mode.[5]
- Specific locations where travellers had a real opportunity to choose between different trips with different times, costs, reliability and/or quality features.

Bus concessionary passholders were only sampled in Sheffield, Leeds, Bristol and Brighton, since these areas provide bus travel at zero cost to passholders and rail travel at non-zero cost, and on bus routes with a parallel rail route, so that a bus versus rail SP exercise could be undertaken. In the case of 'other PT', most respondents (except those sampled in London Underground central locations) were sampled on routes towards the centre (along a rail or bus route), so that a bus or rail versus 'other PT' SP exercise could be undertaken.

Telephone recruitment For the general public telephone sample, Random Digit Dialling[6] (RDD) sample was purchased that geographically represented the population of England as shown in the 2011 Census by region. Given that mobile numbers are not geographically specific; the requirement for geographical representativeness effectively forced us to survey residential landline numbers. We acknowledge that the exclusion of mobile numbers could have introduced a degree of socio-economic and/or demographic bias, since landline users tend to be older and more affluent, all else equal. That said, these biases would have been partly mitigated by the intercept recruitment.

Adult respondents were contacted and screened using a recruitment questionnaire and, if in-scope, they were invited to participate in the research either on-line or by phone. The former were sent a web-link to the customised survey using the same e-mail invite as for the intercept survey. For those who undertook the whole interview by phone, the SP

---

[4] Computer aided personal interview.

[5] Ensuring that, for example, rail included flows such as London long, non-London long, South East outer, South East inner, car included inter-urban, urban and rural, and bus included London, Metropolitan/PTE, freestanding large urban areas and market towns/rural hinterland.

[6] RDD sample is created by selecting a known existing number, and randomising the last couple of digits to generate a new telephone number that may or may not exist, and may or may not be a residential number. This is checked using a pulsing machine to dial the resulting phone numbers, and tell at the exchange whether the number is valid or not.
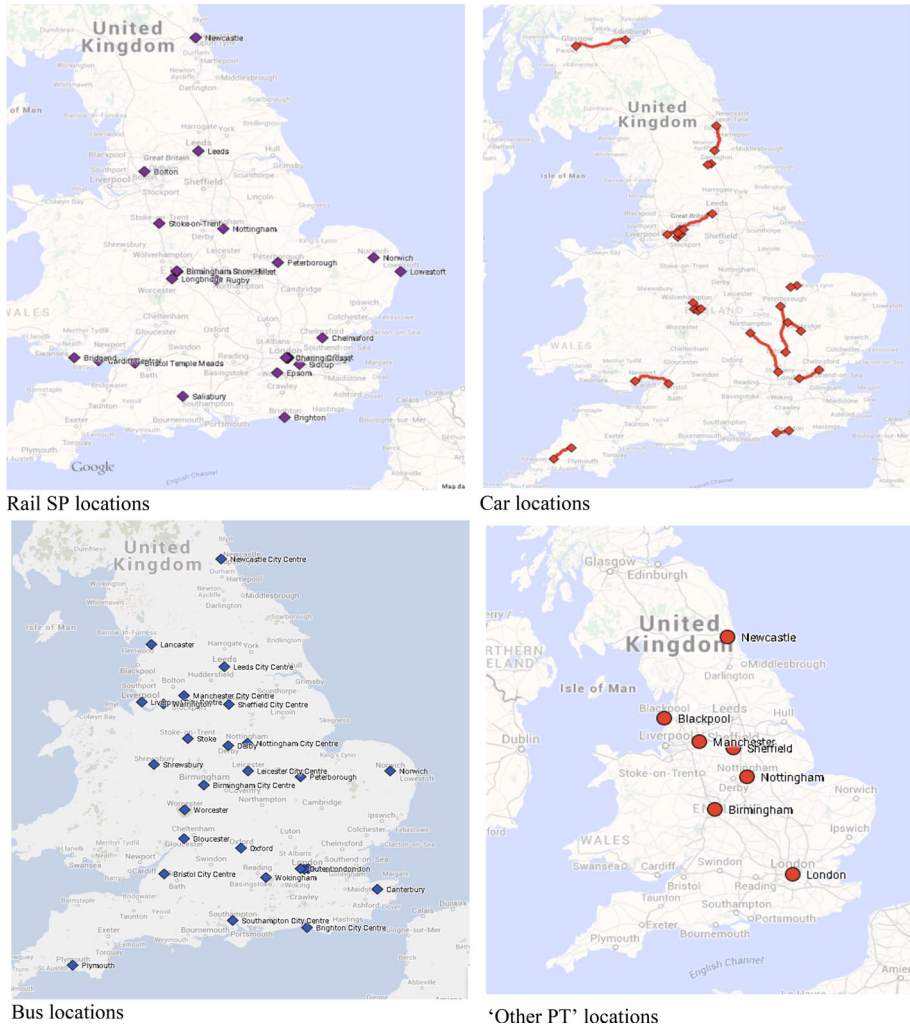
Rail SP locations

Car locations

Bus locations

'Other PT' locations

Fig. 5 Maps of the sampling locations for rail, car, bus and 'other PT'

options for the three exercises (i.e. SP1–3) were sent electronically or in hard copy. Since these exercises were customised to the responses from preliminary questions, the practical implication was that around a third of the CATI[7] interviews needed to be paused and reconvened at a later time and/or date, once the SP options had been dispatched to the respondent. The telephone fieldwork was undertaken between 14:00 and 21:00 Monday to Friday, between 10:00 and 18:00 Saturday, and between 11:00 and 19:00 Sunday, to help ensure that those in employment could be recruited.

---

[7] Computer aided telephone interview.

Employers' business SP approach

One of the aims of this study, which distinguished it from previous UK VTT studies, was to gather comprehensive evidence on VTT for business trips. This required evidence from the perspectives of both employees and employers. For the latter, we focussed upon so-called 'briefcase' travel,[8] and deliberately omitted operational functions undertaken by the likes of service engineers, travelling sales forces, delivery agents etc. Mindful that employers' business data can be difficult and costly to collect, we judged that survey resources were best directed at this key business traveller segment (see later discussion in the "Reconciling different sources of evidence on business values" section).

Experimental design method

The SP design method was exactly the same as for the general public survey, with the exception that experiment was administered to the employer, but framed around a hypothetical trip undertaken by an employee using car, train or 'other PT'.

Employers' business SP market research method

The surveys were administered by telephone, and the target respondent was *"the person within the company who was responsible for making decisions about how employees travel for business purposes, for example when travelling to meet clients, customers or suppliers or when travelling between different offices within their organisation"*. In smaller companies this could be the owner, managing director, finance director, operations manager, procurement manager or HR manager. In larger companies, there are often many such people, provoking the concern that responses may be dependent on the specific person interviewed. However, we drew reassurance from the fact that definitive travel policies were more prevalent in larger companies; in our sample, 77% of companies with over 250 employees had formal travel policies, as compared to 23% for companies with less than 20 employees. Respondents were sent the three SP exercises (i.e. SP1–3), which were customised based on various answers within the questionnaire.

The telephone sample was supplied by Sample Answers and used LBM Direct Marketing and Experian Business Files, which in turn were based on data from Thomson Directories and Companies House. Telephone numbers were randomly drawn from this sample.

In addition to quotas on the mode used by the employee for the hypothetical trip (133 car, 133 rail and 133 'other PT'), there were quotas on company size, industry grouping and region. Company size quotas were determined in discussion with the Department, with a view to ensuring adequate coverage of all industry groupings, whilst also focussing survey effort on larger companies undertaking 'briefcase' travel.

Incentives

All participants were offered a £10 incentive (an Amazon or Boots voucher or a donation to a charity) on completion of the main questionnaire. Towards the end of the fieldwork period, some participants were offered a £20 incentive to help meet certain quotas. In total,

---

[8] For the purposes of this study, briefcase travel was defined as *"trips made by office-based staff travelling to conduct meetings and similar business activities but not to provide trade services"*.

3% of the general public sample and 25% of the employers sample received £20. For employers, these participants were more likely to be rail users and from larger companies, as these were the quota groups that were being targeted at this stage of the survey.

Implementation of field surveys

Fieldwork took place between 24th October and 15th December 2014. The latter date was a 'hard' deadline agreed with the Department, so as to avoid conducting survey work during the Christmas and New Year period, when travel behaviour might be atypical. Whilst the final report of the study (Arup, ITS Leeds and Accent 2015a; Chapter 3) provided a comprehensive description of the features of the travellers and trips surveyed, the following sub-sections focus discussion on the success of the fieldwork against the target sample sizes.

General public SP survey

With reference to Table 3, 8623 SP interviews were undertaken with the general public, against an overall target of 8500. The number of interviews exceeded both the overall target, and most of the mode/purpose segment targets. The shortfall for some targets, particularly 'other PT' employees' business and bus commuting, were due to a shortage of business/commute travellers at the survey locations identified for those modes.

89% of the car, rail and bus SP interviews were intercept-recruited and 11% CATI-recruited. 'Other PT' interviews were all intercept-recruited. The proportion recruited by phone was rather lower than the 20% target, mainly because bus and rail commute and employees' business respondents were found to be relatively scarce. The car sample was 19% CATI-recruited (predominantly commute and non-work). It should be noted that any residual bias in trips/travellers in the sample was corrected at the implementation stage (as discussed in the "Appraisal values" section).

84% of the SP interviews were undertaken on-line and 16% by telephone. 45% of the SP interviews were completed within a day of recruitment and a further 33% two to seven days after recruitment. Of those who were intercept-recruited, 91% completed the questionnaire on-line and 9% undertook the interview by telephone. Conversely, of those who were CATI-recruited, 90% completed the questionnaire by telephone and 10% completed the questionnaire on-line.

Table 3 Total completed SP interviews (on-line and CATI) by mode and purpose (targets in parentheses)

Mode	Commute		Other non-work		Employees' business		Total	
Car	(1000)	1032	(1000)	1037	(1000)	956	(3000)	3025
Bus	(500)	371	(500)	672	(0)	N/A	(1000)	1043
Rail	(1000)	998	(1000)	1128	(1000)	1010	(3000)	3136
'Other PT'	(500)	614	(500)	540	(500)	265*	(1500)	1419
Total	(3000)	3015	(3000)	3377	(2500)	2231	(8500)	8623

* Includes 22 bus

Employers' business SP survey

With reference to Table 4, the target of 400 employers' business interviews was achieved, although there was a shortfall on the largest businesses.

Recruitment and response rates

To give an indication of the success of the recruitment approach, Tables 5 and 6 show the total number of 'contacts' for the general public SP survey, with breakdown by those contacts recruited and those 'lost' for one reason or another. As might be expected, the intercept-based approach—which targeted existing users of specified modes—was considerably more successful in recruiting respondents (71% on average) as compared with the telephone-based approach (6%)—which simply entailed random sampling of residential landlines.

Table 4 Total SP interviews (CATI completion) by mode and number of employees (targets in parentheses)

	Target	Actual
Mode		
Car	(194–257)*	244
Train	(130–194)*	143
'Other PT'	N/A	13
Number of employees		
1–19	(67)	74
20–49	(67)	73
50–249	(133)	149
250+	(133)	104
Total	(400)	400

* 'Other PT' dropped, remaining interviews split between rail and car (agreed revised minimum for rail of 130)

Table 5 General public SP survey intercept recruitment

	Total %	Bus %	Rail %	Car %	'Other PT' %
Recruited	71	72	79	62	76
Refusals	11	19	6	16	14
Drop-outs	2	3	2	3	2
Out-of-scope	15	6	13	19	7
Sample size	39,475	3757	9993	10,403	5462

Table 6 General public SP survey telephone recruitment

	Total %
Recruited	6
No reply/answerphone	50
Refusal	31
Number not recognised/fax/business etc	8
Out-of-scope	4
Sample size	31,960

For the intercept-recruited respondents as a whole (i.e. across all surveys), the overall response rate was 37%. Of those recruited, 93% supplied an e-mail address for the on-line survey, whilst 7% supplied a phone number for the follow-up telephone survey; the response rate was the same for both approaches. For the CATI-recruited respondents as a whole, the response rate was 61% for those who were in-scope and recruited.

Choice modelling

The remainder of this paper will devote particular attention to the general public SP dataset since, as we will see in the "Appraisal values" section, this dataset formed the basis of the appraisal values of travel time and reliability eventually recommended to the Department. The RP and employers' SP datasets were used principally to validate and corroborate the general public SP, and will not be discussed in any great detail.

Against this background, the core choice model specification was developed in a systematic fashion, as follows.

1. We undertook preliminary work to ensure that the data met appropriate quality standards.
2. We initially developed separate models for each mode and SP game (i.e. SP1-3).
3. Having identified the set of covariates applicable to each mode and game, we jointly modelled SP1-3 for each mode.
4. Developing the models further, we introduced additional elements of functionality (described in the following sub-sections), and identified the final specification to be taken forward to the Implementation Tool used for generating appraisal values in fourth section.

The field of choice modelling has evolved substantially since the 2003 national VTT study in the UK (Mackie et al. 2003). The present study exploited many of these developments, relating to the error structure of the models, the treatment of reference dependence (size and sign effects) and the incorporation of unobserved preference heterogeneity in valuations. These developments are summarised in the following sub-sections, but interested readers may wish to refer to the fuller technical discussion in the final report of the study (Arup, ITS Leeds and Accent 2015a; Chapter 4) or the companion journal paper (Hess et al. 2017).

Multiplicative versus additive error structures

As is well-established, the utility in a choice model is decomposed into deterministic and random components, where the latter is referred to as the 'error' term.

The 2003 national study employed standard additive error structures, used in most VTT studies worldwide since the pioneering UK work (Daly and Zachary 1975), where $U = V + \varepsilon$, with V and ε giving the deterministic and random components of utility, respectively. In the present study, we diverged from this assumption by employing models based upon a multiplicative formulation (Harris and Tanner 1974; Fosgerau and Bierlaire 2009). In a multiplicative formulation, we replace the typical additive specification of the utility of an alternative $U = V + \varepsilon$ by $U = V \cdot \varepsilon$, where V and ε are still defined as the deterministic and random components of utility, respectively. That is, the random (error)

component of utility is taken to multiply the deterministic component, rather than be added to it.

The multiplicative formulation represents the state-of-the-art in VTT estimation for experiments of the SP1 type, and its advantages were further verified in tests conducted for the present study; again see the final report for further details (Arup, ITS Leeds and Accent 2015a; Section 4.4). A corresponding approach for SP2 and SP3 is also possible, and was used here, in common with the most recent Danish national VTT study (Fosgerau et al. 2007), but with additional development to accommodate reference dependence.

The practical advantage given by the multiplicative approach is that it becomes much easier to make an assumption of constant variance for ε. In general, it is found that utility variance increases as utility increases and this is handled automatically in the multiplicative form of the model. This benefit is confirmed by the improved model fit given by multiplicative models in this context.

In the multiplicative model, it is practical[9] to work with $\log U = \log V + \log \varepsilon$, the log function having no impact on the ranking of utilities, since it is a monotonic transformation. Technically, the assumptions regarding the distributions of ε are different in these cases. In practice, it is assumed that ε follows a log-extreme value distribution in the multiplicative model, so that the simple logit model can be used to calculate probabilities.

Size and sign effects

Many SP-based VTT studies, including the 2003 study, have found that the values obtained depend on the sign and size of time and cost changes *relative* to a 'reference' value. These findings can be related to Prospect Theory, e.g. that gains are attributed a lower absolute value than equivalent losses (Kahneman and Tversky 1979). When travellers are interviewed relative to a specific trip, the reference value for a given attribute is often and reasonably taken to be the corresponding value on that trip.

The present study tested reference-dependent effects for all SP models, but not for the RP models, since in the latter case it was unclear what constituted the reference value. Similarly, reference-dependent effects could not be included for those attributes in the SP survey where no immediate reference values were available, such as reliability, or where the number of possible effects due to reference dependence were too large to test efficiently and too difficult to implement in model application, such as crowding. In particular, we adopted the principles of de Borger and Fosgerau's (DBF's) (2008) approach to modelling reference dependence—which is arguably the most sophisticated practical approach to date—and further developed it for present purposes.

In essence, this approach specifies a value function $v(.)$ between the 'target' and 'reference' trips. Along the lines of equation (5) in DBF we used $S(\Delta x) = \Delta x/(|\Delta x|)$ and Δx as measures of sign and size inside the value function respectively, where $\Delta x = x - x_0$ captures the size difference from the reference trip. Illustrating for SP1, the implication of this specification is that respondents value the SP cost differences by $(v(\Delta c_1) - v(\Delta c_2))$ and time differences by $(v(\theta \Delta t_1) - v(\theta \Delta t_2))$, where v denotes the value of a time change from the reference and θ is the 'underlying' VTT. It is then 'rational' to choose the slower/ cheaper alternative if $|v(\Delta c_1) - v(\Delta c_2)| > |v(\theta \Delta t_1) - v(\theta \Delta t_2)|$. Estimation of DBF's value function identifies three parameters, representing: (a) differences in gain value and loss value from the 'underlying' value; (b) non-linearity in the impacts of gains and losses; (c) differences in the non-linearity of value for gains and losses. The key methodological

---

[9] It is not difficult in practice to arrange that $V > 0$ and $\varepsilon > 0$.

development here was to extend the DBF approach to attributes other than time and cost, thereby accommodating SP2 and SP3; fuller discussion of this can be found in the final report (Arup, ITS Leeds and Accent 2015a, Section 4.6.2).

Joint modelling of SP1-3

Whilst initial tests were conducted separately on the individual SP games (mentioned above when comparing the additive and multiplicative specifications), the majority of our work made use of models jointly estimated on multiple games. This was in line with our decision to present each respondent with three games of five choice tasks each. The main benefit of joint estimation was increased robustness for those parameters shared across games, which in the present case encompassed the set of covariates explaining deterministic heterogeneity in valuations, as well as the random heterogeneity parameters.

In the joint estimation, we allowed for differences in valuations across games by using separate multipliers for each valuation in our models (across the three games), relating it to a base VTT. This allowed us to capture differences in valuations that clearly related to different components (e.g. reliability as opposed to travel time), but also to test for differences in interpretation for attributes common to different games, such as generic travel time in SP1 and SP2.

Deterministic variation in modelled values

Another area of interest is the extent to which estimates of VTT are influenced by features of the traveller and/or trip—such as the traveller's income or the length of the trip. We conducted an extensive search for factors causing variation in the values, involving a large number of traveller/trip features collected in the course of the RP and SP surveys.

As has consistently been found in other national studies, we found significant evidence of VTT increasing with income. This relationship was found in all mode/purpose segments except for bus and 'other PT' commuting. We also found that VTT varied with the travel time and cost of the trip. Given that both of these factors are closely related to distance, the implication of these results was that VTT increased with trip distance.

Having tested the influence of a wide range of factors on VTT, it is interesting to note that, all else equal, time use (i.e. the traveller's ability to do something else whilst travelling, to work or surf the net), geography (i.e. area, urban/rural), current travel conditions (i.e. congestion and crowding) and current road types had little or no impact on VTT. This could be an indication that travellers, when completing hypothetical choice tasks, do not necessarily relate these back to the real world journey which these choice tasks relate to.

The result relating to time use is of particular interest, and it is perhaps useful to digress slightly and provide additional background concerning the survey approach in this regard. Focussing here upon employees' business travel, since this represents a key segment potentially affected by the productivity of travel time, respondents were reminded of their reported one-way trip time and asked approximately how much of that time was spent undertaking work and non-work related activities.

With reference to Fig. 6, the main activities undertaken by mode were:

- *Car* listening to music (53 min on average) and driving (17 min).
- *Train* using smart phone/eBook/tablet/computer (33 min, of which 17 min was work related), work related use of laptop/tablet (26 min), doing nothing/relaxing/looking out
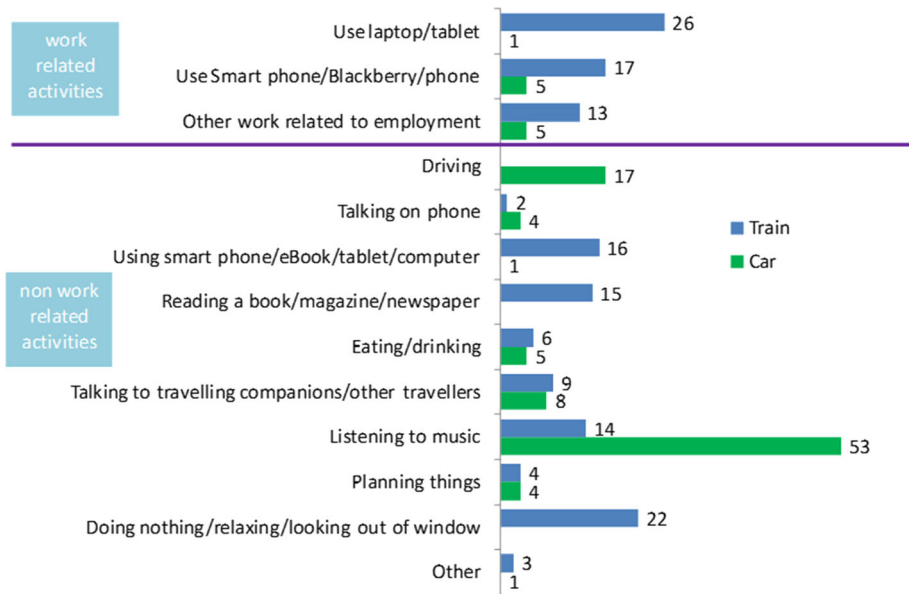
Fig. 6 Activities undertaken by employees during trip (average minutes) by mode

of window (22 min), listening to music (14 min) and other work related to employment (13 min).

The 2009 report on 'Productive use of rail travel time and the valuation of travel time savings for rail business travellers' (Mott MacDonald et al. 2009) used slightly different categories, and showed the proportion of travellers undertaking work-related activities as follows: preparing for a meeting (38%), making/receiving calls (43%), talking to colleagues/other (12%), use of a laptop (23%), use of a PDA/Blackberry (25%), other work related to employment (36%).

By comparison, in our own SP survey, 35% used a laptop, 56% used a Smartphone/Blackberry and 29% did other work related to employment. There has been a notable increase in the use of electronic devices for work-related activities on-train since 2009. Nonetheless, it is clear that a large proportion of rail travel time is spent on non-work activities.

One of the main criticisms of the Department's CSA-based values for business travel was that they failed to reflect the increasing opportunities for people to work whilst travelling. However, an attraction of WTP is that valuations should in principle reflect how travel time is used, given current travel conditions and opportunities to use that time. Whilst our results indicated that VTT did not vary with time use, this is not to say that time use is unimportant. It is possible that the results could have been different if the opportunities to use travel time productively had been substantively different for the trips being made when surveyed. Another possibility is that, despite our best efforts, the importance of time use was not fully captured by the SP exercises employed.

Returning to our more general interest in deterministic variation of VTT, we found evidence of size and sign effects, although these varied in their nature and strength across modes, games and attributes (i.e. time and cost). It is conceivable that such effects were an artefact of the SP exercises, but even if they were not, a given reference point could

become less relevant as travellers and travel conditions change over time. Notwithstanding such considerations, we ideally require 'reference free' estimates of VTT for appraisal purposes. To this end, the modelling work sought to identify the prevalence of size and sign effects, before eliciting VTTs which 'neutralised' these effects.

For the majority of the aforementioned sources of variation, multipliers of the VTT were estimated, with one category of a given attribute being used as the base—in which case the multiplier was set to a value of one. This means that the base estimate of the VTT related to an individual and trip at the base values for these covariates. We will return to this point in the discussion of the actual model results (see the "Appraisal values" section to follow).

Finally, it is worth mentioning that we tested for any significant differences in valuations between online and CATI sub-samples. While this revealed statistically significant differences in model scale between the two samples, the VTT differences were not significant at usual levels of confidence.

Random taste heterogeneity

In accordance with current best practice, the final area of model development involved allowing for random heterogeneity in the VTT. For the present study, after testing popular alternatives such as the log-normal distribution, we settled on the log-uniform distribution. This has a somewhat shorter tail than the log-normal, and any differences in fit were found to be very small, with the log-uniform avoiding problems with extreme values. Unlike some recent national VTT studies (e.g. Börjesson et al. 2012), the log-uniform also avoided the need for censoring of the tails (see the discussion in Hess et al. 2017).

Appraisal values

The models described in the "Choice modelling" section, which supply the formulae for VTT as a function of covariates, were estimated on an unrepresentative sample of travellers. Whilst we would not expect this to bias the coefficients, the models cannot, without further information, provide appropriate values for selected aggregations of the travelling population, as would be required for establishing recommended values for appraisal.

In the 2003 study (Mackie et al. 2003), the only covariates were distance[10] and income, whilst separate models were derived for the commuting and other non-work purposes (given the adherence to the CSA, the business models were not taken forward to guidance). This meant that representative values could be calculated by providing a representative matrix of trips for each 'cell' representing a combination of distance and income, applying the formula to each cell, and calculating a weighted average. In the present study, by contrast, the scope of the model was much wider in that (a) it contained many more covariates and (b) valuations were generated for a number of quantities in addition to travel 'time', such that a matrix-based approach would have been unwieldy.

---

[10] In fact, trip distance was not collected in the 1994 data, so that the variable appearing under this name is a construct derived from the reported cost (the 1994/2003 study covered car travel only).

Sample enumeration approach

Whilst the principles are essentially the same, it was more convenient to make use of a 'sample enumeration' approach. This involved the calculation of appropriate valuations (of time, etc.) for each observation in the sample, making use of the relevant covariates, followed by the calculation of weighted averages over the sample to ensure national representativeness. We can represent this mathematically as follows:

$$\bar{v} = \frac{\sum_n w_n v(z_n)}{\sum_n w_n}$$

where $\bar{v}$ is the weighted average, n represents an observation in the sample with w_n the necessary weighting to obtain representativeness, $v(.)$ is the time value formula derived from the model as a function of a vector of covariates $\mathbf{z}$, and $\mathbf{z}_n$ is the set of covariates relating to observation n.

As the best way of ensuring national representativeness, it was agreed with the Department that we would use the NTS sample of trips made by persons over 16 years of age by all motorised modes collected during the years 2010–2012. It was judged that a three-year period was appropriate for giving a representative picture of current (as opposed to historical) travel behaviour, and 2012 was the most recent year at our disposal.[11] At the time of undertaking this work, a further update of NTS to 2013 was pending, but it was not anticipated that this update would introduce significant differences. While NTS is—in common with the RP and SP datasets—a sample, it contains a set of weights aimed at achieving a representative picture of national travel.

For each trip in the NTS sample, the recommended choice model arising from the "Choice Modelling" section was used to calculate appropriate valuations, taking account of the covariates of the NTS record. This calculation made use of the same 'code' used in the model estimation procedure, to ensure complete compatibility. In addition, the estimated standard errors (etc.) were transferred, such that each NTS trip generated information about the statistical reliability of its valuations, obviating the need for a special subsequent step to calculate the confidence intervals associated with the recommended values.

Note that it is possible to restrict the calculation of the quantity to summations over the NTS sample observations with particular characteristics, whether or not these characteristics are within the set of covariates defining the valuation formula. In this way, it is possible to derive separate valuations for e.g. geographical breakdowns or for income bands, as well as mode, purpose etc. In order to provide maximum flexibility for the Department, an 'Implementation Tool' was developed in 'R', which permits the calculation of valuations for different segments and based on a variety of weighting options.

Conceptual issues

Aside from the above considerations, the translation of behavioural values of travel time, reliability and associated quality factors into values suitable for use in appraisal provokes a number of technical considerations; the following sections will highlight the principal such considerations.

---

[11] The NTS data available in the 'special licence version' has been used.

Trip versus distance weighting

Invariably, practical demand modelling does not occur at a level of disaggregation consistent with the covariates used in the choice models, because of constraints in the data underlying the demand model. This has implications for appraisal, since it creates a need for some form of average VTT and—following from the discussion in the "Sample enumeration approach" section—a need to re-weight this average for representativeness.

The two principal candidates for re-weighting are trip rates and distance travelled. The previous WebTAG values were distance-weighted averages. There are two arguments in support of this position. Firstly, the probability that a trip will experience a travel time change is a function of trip distance. When transport interventions are targeted on specific links in the network, long trips have more chance of experiencing a travel time improvement than short trips—as they travel over more 'links'.[12] Secondly, at a more conceptual level it was argued in Section 7.3 of Mackie et al. (2003) that if individuals j have different values of time v_j and travel time t_j, it would seem sensible to try to ensure that $\bar{v} \sum_j t_j = \sum_j t_j v_j$, such that the value of the total time disutility is correct. Since time is reasonably correlated with distance, a distance weighting would be one way of achieving this—although the presence of congestion can reduce the degree of correlation.

Income effects

Another technical consideration is whether the VTT used in appraisal should be a pure WTP value or should be adjusted to a standard value reflecting the income distribution of the travelling population. Discussions with the Department identified essentially four possible ways of dealing with this issue:

Option (1): Averaging over income, but not segmenting by income This was the Department's approach in previous WebTAG guidance, where the effect of income is included when calculating the value for each trip VTT. These VTTs are then averaged to give single, 'standard' values for commuting and other non-work (and other levels of segmentation).

Option (2): Calculating values at 'average' income This approach is similar to option (1) but treats all trips in the weighting process as having 'average' income. The user specifies household income (for non-work VTT) and personal income (for business VTT).

Option (3): Removing the income covariate from the choice model, thus allowing the effect to be picked up by other covariates This is a pragmatic option, but introduces model misspecification.

Option (4): Applying distributional weights from the 'Green Book'[13] Values are again calculated for each trip, using the same parameter values as under option (1). The resulting values are then weighted, according to the income quintile which the income band falls in, using the weights given in the Green Book (HM Treasury 2013). This would apply to non-work trips only.

---

[12] Notwithstanding the theoretical argument for distance-weighting, we drew reassurance from the empirical finding that, on segmenting VTT by distance, we found little difference between distance- and trip-weighted VTTs across modes and purposes. For further discussion, see Section 7.6 of the study report (Arup, ITS Leeds and Accent 2015a).

[13] The Green Book details the HM Treasury's *"guidance for public sector bodies on how to appraise proposals before committing funds to a policy, programme or project"*.

Size effects

The modelling analysis in third section indicated the clear presence of 'design' effects in the SP data. The specific modelling approach used (based on DBF) meant that sign effects cancelled out in the VTT calculations, but the same did not apply to size effects. In other words, the VTT calculations were dependent upon the size of the time change from the reference (denoted Δt). Therefore, in translating these behavioural values to appraisal values, it was necessary to make a definitive assumption concerning Δt. Informed by analysis of the sensitivity of VTT to different Δt, together with review of the corresponding assumptions employed in the most recent Danish (Fosgearu et al. 2007) and Swedish (Börjesson and Eliasson 2014) national studies, we eventually settled upon a Δt of 10 minutes.

Mode-specific values

Mackie et al. (2003) in Sections 7.3 and 8.3 noted a number of reasons why values might vary by mode:

(1)   The income and socio-economic characteristics of travellers might vary systematically by mode. Low income users with low average VTT might gravitate to mode A while high income users with high average VTT might tend to choose mode B.

(2)   The composition of trips and purposes might vary systematically by mode. Mode A might have a strong market share in short distance trips, while mode B might be stronger at longer distances.[14]

(3)   A cross-section of people with given income and socio characteristics making a given trip will have a distribution of values of time (and individual values may vary according to the constraints faced). People with low VTT for that trip will self-select into relatively low cost/high time modes and vice versa.

(4)   For any individual, VTT by mode may vary due to the different characteristics of the modes in terms of comfort, cleanliness, reliability, level of personal control, and other quality attributes.

Mackie et al. argued that, point (4) aside, individuals should, from a theoretical perspective, have the same VTT for a given trip regardless of mode used, hence favouring an approach which picks up (1) to (3) through the income, socio-economic characteristics and trip and purpose characteristics of the traffic modelled to the various sub-markets. Any remaining variation in VTT should then reflect 'comfort' effects (note that these effects could conceivably be related to 'time use', i.e. the extent to which travel time is used productively or otherwise enjoyably, as discussed in the "Deterministic variation in modelled values" section above).

In the present study we found that, even with income option (2)—where the effect of income was neutralised—substantial differences by mode remained. The variation in VTT by mode and person for a 'typical' person/trip combination is given in Table 7.

If the modal variation related solely to 'comfort', then we would expect the lowest values for rail and the highest for bus, with car and 'other PT' intermediate. For commuting, we can see such an effect for car, rail and 'other PT', but the bus values are low rather than high. For business, where bus is not of relevance, there does appear to be some relation to comfort (especially in terms of the possibility of working on the train). For other non-work, bus values are much higher, in line with the 'comfort' hypothesis, but the values

---

[14]   Given the much lower distance elasticities being found in the current study, this example has less force.

Table 7 Ratio of modal VTT by trip purpose for an average person based on income option (2) (car = base)

	Commute	Other non-work	Employees' business
Car	1.00	1.00	1.00
Bus	0.51	2.14	N/A
'Other PT'	0.99	3.19	0.69
Rail	0.73	2.29	0.39

Table 8 All mode non-work VTT by treatment of income option (£/hr, perceived prices)

	Commute	Other non-work
Using income option (1)	11.21	5.12
Using income option (2)	10.03	5.49

Fixed income trip-weighted (option 2) is £49,684 for household (non-work) income. For personal income (EB) it is £35,070 for Car, £20,219 for Bus, £45,019 for 'Other PT', £55,319 for Rail. Ratio of distance-weighted VTT. Modelling used imputation for zero reported PT costs and employers paying for EB trips. $\Delta t = 10$. Tool version 1.1

for rail and 'other PT' go in the opposite direction to what we would expect. All in all, we considered that these differences between modes could not be explained solely by comfort.

Given the above considerations and evidence, and in line with the argument in Mackie et al., our preference for non-work was to retain mode-free values, by averaging the values over the sample of trips for all (motorised) modes, maintaining the distance weighting. Implementing this, but also maintaining (in the short term, at least) the approach of using the SP1 values, the values given in Table 8 were obtained.

It can be seen that the two income options do not produce very different results, but option (2) narrows the gap between commute and other non-work values. As might be expected, the car values dominate, given that well over 80% of distance travelled is by car.

Reconciling different sources of evidence on business values

Much of the background thinking to the recommended approach to valuing business travel time savings was undertaken in the scoping study which preceded the present study (ITS Leeds, John Bates and KTH 2013), and what is reported here is essentially the implementation of that thinking. However, relative to the analysis of non-work trip purposes where people are making their own travel decisions involving their own time and money, business trip-making is inherently more complex and models require a greater degree of interpretation and judgement. Whilst interested readers may wish to refer to the detailed discussion in the report of the scoping study, it is perhaps useful to briefly summarise the conclusions of that study, before reporting the new analysis which has been undertaken here.

In the scoping study, we expressed reservations regarding the CSA traditionally employed by the Department, essentially reiterating long-standing and well-rehearsed concerns that not all travel time is unproductive and not all time savings would be converted into productive use to the benefit of the company. In particular, the digital

revolution has increased the potential for using travel time productively, and indeed can be expected to have increased the productivity of any such time spent working while travelling. Other arguments against the CSA surround difficulties in estimating the value of the marginal productivity of labour (which underpins the approach), the benefits of spending more time at the destination (say with a client or at a sales pitch), and the benefits of avoiding overnight accommodation and travel in unsocial hours. By contrast, these effects should in principle work themselves through into a WTP-based valuation, thereby eliciting a reliable representation of what the company would pay.

We felt that an intuitively appealing approach would be to survey employers about how much they would be prepared to pay to reduce their employees' travel time. From a conceptual point of view, it might be argued that employers should be the focus, since it is they who will actually be purchasing the time savings. After all, if the CSA is a valid representation of the value of business travel time savings, then the employer should simply express a WTP in line with the CSA. Nonetheless, the difficulties of, and uncertainties surrounding, a valuation approach based on surveying employers were recognised. For example, the data collection costs are high, there are challenges involved in identifying the appropriate employer agent, and even then the agent may not be entirely familiar with specific kinds of business trips. Furthermore, another challenge is to achieve a representative sample of travel-using employers.

A potentially complementary approach is to undertake employee surveys, using either RP or SP approaches, which are couched within an awareness of company travel policy. Compared to collecting employers' surveys, obtaining large samples of employees travelling on company business is relatively straightforward, and indeed the business scoping study demonstrated that SP studies along these lines tend to be the norm. The concern here is whether employees are able to make choices in response to hypothetical scenarios that accurately represent the company's willingness-to-pay, or worse still simply represent their own willingness-to-pay. If the employee is to be an acceptable proxy for the employer, then we need employees to respond in accordance with the company's interests as opposed to their own private interests. An interesting special case here is self-employed business travellers, where it might be presumed that company and private interests are one-and-the-same, and that SP responses would therefore reflect what the company would pay.

In the scoping study, we expressed a preference for WTP-based approaches, using different methods for corroborative and interpretive reasons. This reflected a proposition that well designed and conducted quantitative research can provide a coherent 'story' as to how business travel time savings are valued, or better still, can elicit direct estimates of WTP that lend themselves to comparison against the CSA.

Against this background, the business travel component of the present study was informed by three sources of survey evidence, namely employer SP, employee SP, and employee RP. The information collected on income and working hours in the course of the survey also allowed comparison with the CSA. In order to reconcile these different perspectives on business VTT, we pursued two lines of enquiry. First, we were interested in the degree of similarity between SP-based estimates of VTT and the CSA (the RP-based estimates were largely used as a corroborative device and will not be discussed in any detail here). Second, we were interested in the degree of consistency between various properties of the VTTs emanating from the different surveys.

Generally speaking, we found similar values for the two different SP analyses (employer and employee), and for some occupational types we also observed similar values for the SP analyses and the CSA. This was particularly so for blue collar workers, who would be expected to have relatively low productivity whilst travelling. For briefcase travellers,

who are more likely to be productive, SP-based values appeared to be lower than the CSA. Moreover, the degree of similarity between the SP-based VTTs and the CSA was partly dictated by the trip length distribution, and did not hold over all distances. The self-employed values were lower than those for employees; whilst we cannot substantiate empirically, this result would seem plausible if the time saved is taken as leisure.

Turning to the properties of the VTT estimates, the theoretically-driven CSA embodies an income elasticity of one (i.e. implying that VTT increases in direct proportion to income) and applies a constant unit value to all trips (e.g. irrespective of time, cost, distance, travel conditions, productivity, etc.). By contrast, the SP-based VTTs exhibited income elasticities within the range 0.3–0.4 (and significantly less than one), and significant variability by several of the aforementioned dimensions (and notably by distance). Thus, whilst there was some correspondence between the actual estimates of VTT from the CSA and SP analyses, this correspondence did not extend to key properties of those estimates.

Having reconciled the various sources of evidence on business VTT through the lines of enquiry summarised above, it was decided that the employee SP survey should be the definitive source of evidence taken forward to the Implementation Tool. This was because it generated—with some qualifications—similar values to the employer SP survey, but offered a considerably more substantial dataset, amenable to generating statistically robust values for a range of trip and traveller segments. Furthermore, the employee dataset was more comparable to the NTS data used as the basis for the sample enumeration. That is to say, the Tool applied the choice model from the employee SP to business trips in the NTS, to derive an average value over specified segmentations, as shown in Table 9.

The average distance-weighted personal income across the NTS sample was £46,615 (2014 prices and values). This gives a business VTT in 2014 perceived prices of £28.27 using the CSA (second row, third column). This compares to the CSA-based WebTAG values which have an all modes value of £25.47 (first row, third column). If we also compare these values to the SP-based values re-weighted for NTS (third row), then we see that the VTT for employees' business across all modes is £18.23. This is 72% of the WebTAG value. We also find substantial variation by mode with 'other PT' lowest at £8.33 and rail highest at £27.61 for the all distance values. As proportions of the WebTAG values, these range from 34% ('other PT') to 92% (rail).

We have already noted that the SP-based VTTs for business are sensitive to trip distance. From Table 9 we can also see that at low distances the SP-based values are substantially less than the previous WebTAG values, but as trip distances increase the SP-based values increase to be close to the previous WebTAG values at long distances (>50 miles).

Reconciling different values of travel time savings from different SP games

The results from the simplest time/cost trade-off (SP1)—which is, of course, the game most comparable to the results on which previous WebTAG guidance, using the 1994 AHCG data, was based—gave car values of £11.70 for commute, £4.91 for other non-work, and £16.74 for employees' business. For all purposes, the values fell between the corresponding values from SP3 for light and heavy traffic, and this was in line with the underlying model, since the relativities were unaffected by NTS re-weighting.

We had to form a judgment as to whether the SP1 and SP3 values were compatible, and if not, which values to carry forward to appraisal. To do this, we reflected on how the time values were presented in the SP tasks. For SP1, the instructions were:

Table 9 Business VTT by method of calculation, mode and distance (2014 perceived prices and values, £/hr)

Source/method	Distance	All modes	Car	Bus	'Other PT'	Rail
Previous WebTAG (2014 prices and values)	All distances	25.47	24.43	15.64	24.72	30.07
CSA estimate from NTS 2010–2012 data (2014 prices and values)	All distances	28.27	27.05	13.13	26.33	36.46
Employees' business SP re-weighted to NTS 2010–2012 (2014 prices and values)	All distances	18.23	16.74	N/A	8.33	27.61
	<5 miles	5.39	5.27	N/A	8.33	N/A
	5–20 miles	8.84	8.79	N/A	8.28	10.19
	>=20 miles	21.14	19.51	N/A	N/A	28.99
	>=50 miles	24.55	22.53	N/A	N/A	32.56
	> = 100 miles	28.62	25.74	N/A	N/A	N/A

All modes. Distance-weighted, income option 1, SP1 $\Delta t = 10$; Tool version 1.1. PT cost is imputed for a trip with a zero cost, and employers paying for EB trips. WebTAG 'Other PT' is Underground passenger, WebTAG car EB is weighted average of driver and passenger (vehicle occupancy of 1.2)

"Please imagine that each situation is exactly the same as for #LEG2# actual car journey at the time you made the journey, except… The one way travel time may be different because of changes in congestion". Variations in cost were suggested in terms of changed fuel cost.

It is noteworthy that in the 1994 survey (which was for car only) the corresponding instructions were:

"Please imagine that each situation is exactly the same as for your actual journey at the time you were surveyed except… the travel time… can be different from the actual situation at that time because there is, for example, more or less congestion". Variations in cost were suggested in terms of changes in petrol price or parking charges.

Hence in terms of the background to the SP1 experiment, it seemed fair to conclude that they were identical in both surveys, with a clear suggestion to relate the changes to the conditions of the reference trip. Any changes in results could therefore be attributed to:

(a) Changes in the SP design (this should largely improve the accuracy rather than lead to different results per se).
(b) Changes in preferences and behaviour etc. (including the underlying NTS travel characteristics).
(c) Changes in the method of analysis (in particular the switch from additive to multiplicative model specifications with flexible heterogeneity patterns).

Respondents had previously been asked 'about how much' of their time was spent in each of the three following conditions, presented in words and pictures, as:

"Heavy traffic: Your speed is noticeably restricted and frequent gear changes are required".
"Light traffic: You can travel close to the speed limit most of the time, but you have to slow down every so often".
"Free-flowing: You can travel at your own speed with no problems over-taking".

Table 10 Average shares of congested driving conditions in the SP sample

	Commute	Other non-work	Employees' business
Free-flow	0.33	0.36	0.34
Light traffic	0.36	0.41	0.40
Heavy traffic	0.31	0.23	0.26

Attempts were made to relate the valuations from SP1 to these implied proportions of congestion on their real life trip. However, the effects were generally weak,[15] suggesting that respondents related the valuations in SP1 not to their actual trip but to some imagined congestion level, perhaps based on the changes in travel time from their current time.

Turning to the results from SP3, while these choice scenarios were again framed in relation to the reference trip, the hypothetical journey times were now explicitly split into different levels of congestion, and major differences were retrieved in how respondents reacted to these. The SP3 values of time for commuters displayed ratios relative to free-flow time (ff) of 1.4 for light traffic (lc) and 2.66 for heavy traffic (hc). For business the ratios were 1.61 and 2.99, and for other non-work they were 1.76 and 3.98.

As already noted, NTS does not carry information on conditions for individual trips. The Tool was therefore set up to calculate an average of the three types of time in SP3, using the average shares in the SP sample for different purposes, as shown in Table 10. Using these overall proportions, we obtained SP3-based values of £9.98 for commute, £12.44 for business and £4.62 for other non-work. These are lower than the SP1-based values reported at the beginning of this section: the commute value is 15% lower, the business value 26% lower and the other non-work value 6% lower. It is arguable as to whether this comparison is appropriate: the proportions are likely to be related to the kind of trip (urban vs. inter-urban etc.), and the SP sample was certainly not representative in terms of distance travelled. Nevertheless, it is also important to acknowledge that a game of the SP1 type may overstate the value of time as respondents need to trade between only one time and one cost component. As far as the implied variation in valuations for the three levels is concerned, the levels presented are three possible positions within a continuum related to the 'volume/capacity' ratio, and could be difficult to apply in practice, since they would need to be aligned with actual traffic conditions in relation to volume-delay functions.

On balance, we recommend that the Department should adopt the SP1 values in the short term, but conduct further work to build confidence in the SP3 values and explore their practical applicability in appraisal. Indeed, some progress towards these ends has already been achieved, with Hess et al. (2017) demonstrating the greater robustness of SP3—relative to SP1—towards reference dependence in the SP games. The future availability of data on differing levels of congestion may permit the use of SP3 results in the medium term, though there would still be the need to resolve how model output (in terms of volume/capacity ratios) relates to the levels of congestion presented in the SP, and how to interpolate between the free-flow, light and heavy traffic points.

---

[15] The multipliers (relative to free-flow) were reported for commuting as 1.40 for light traffic and 1.56 for heavy traffic: however, neither of these were significantly different from 1 at the 95% level. The corresponding values for other non-work were 1.36 and 1.46, again with neither being significantly different from 1.

VTT values for use in appraisal

Bringing the discussions presented in this section together, we can draw out some recommendations regarding the basis for VTTs for use in appraisal:

- Whilst each of the three SP games (i.e. SP1–3) could potentially be used to elicit the 'headline' estimate of VTT, we recommend that in the short term this should be based on SP1 with $\Delta t = 10$. In the medium term, pending further development, there could be a case for replacing SP1 with SP3.
- There are material differences between the three trip purposes, and we should therefore continue to disaggregate VTT by trip purpose.
- VTT should continue to be distance-weighted, but should be disaggregated into distance bands to reduce the level of approximation between the standard VTT values and the 'real' scheme level VTT value. Further work is required to determine appropriate distance bands for use in appraisal[16,17].
- For non-work, we should use an 'all modes' value due to the non-work VTTs reflecting some self-selectivity between modes. For business, we should use modal values as we interpret differences between modes to represent real differences.
- In the case of business, the choice model in the sample enumeration should be based on those employees who reported that their employers would be willing to pay for time savings.
- We should distinguish between appraisals of small and medium sized schemes (referred to as 'Level 1' in the UK), and appraisals of major schemes and policies ('Level 2') and significant 'user pays' initiatives ('Level 3'). For Level 1 appraisal, standard 'national' values of time can be used. For Level 2, the values may be amended to more accurately reflect local conditions. For Level 3, appraisal values derived from bespoke quality surveys would be appropriate.

For Level 1 and 2 appraisals, we make the additional recommendations:

- Non-work: for Level 1 appraisals with VTTs distance-banded, as recommended above, we should use income option (2)—that is, treating all non-work trips as having the same average household income (if however, distance-banding is not implemented—at least in the short term—then income option (1) should instead be used[17]). For Level 2 appraisals, we should use income option (1) applied at the appropriate regional level.
- Business: for Level 1, income option (1) should be applied using national data, whilst for Level 2, income option (1) should be applied at the appropriate regional level.

We have illustrated these recommendations for a Level 1 appraisal in Table 11. This table also presents, for the purposes of comparison, the previous WebTAG values converted to a comparable base (2014 perceived prices).

---

[16] Whilst it falls outside the scope of the present paper, it is worth noting that, in responding to consultation on our recommended values, the Department subsequently revised the three distance bands shown in Table 11 into four bands (0–50 miles, 50–100 miles, 100–200 miles, and 200+ miles) and, in the case of business, fitted a continuous function to these four bands (DfT 2017).

[17] For non-work, the Department opted not to disaggregate by distance (DfT 2017).

Table 11 Appraisal VTTs for a Level 1 appraisal (routine appraisal of small and medium sized schemes) with illustrative distance bands (2014 perceived prices, £/hr)

Source	Distance	Commute	Other non-work	Employees' business				
		All modes	All modes	All modes	Car	Bus	'Other PT'	Rail
WebTAG	All	7.62	6.77	25.47	24.43	15.64	24.72	30.07
Re-surveyed values	All	11.21	5.12	18.23	16.74	N/A	8.33	27.61
	<20 miles	8.27	3.62	8.31	8.21	N/A	8.33	10.11
	20 to 100 miles	12.15	6.49	16.05	15.85	N/A	8.33	28.99
	>=100 miles	12.15	9.27	28.62	25.74	N/A	8.33	28.99

Distance weighted, 'all distance' values based on income option 1, for distance-banded values non-work based on income option 2 (household income = £49,684) and business on income option 1, VTT imputed for PT trips with zero cost, SP1 VTTs, $\Delta t = 10$, employers paying for EB trips, Tool version 1.1

VTT multipliers for use in appraisal

In addition to the overall VTTs, we also make recommendations for adjustments to these values for different types of time, and we present these as multipliers. In doing this, we must take account of the different VTTs coming from the different games, as well as our general approach of using SP1 values for the overall recommendations about VTT.

With reference to the SP2 results, we found higher valuations based on the average time presented in the reliability experiment relative to the SP1 values, by a factor of 1.31 for commute, 2.17 for other non-work, and 1.52 for employees' business. Now it might be argued that by implying the possibility of unreliability, there is some suggestion of (greater) congestion. However, the questionnaire said that the situation was the same as the reference trip, while the reasons for variation in overall travel time were attributed to 'improvements in traffic control', and the variation (unreliability) was attributed to 'breakdowns, unplanned roadworks, or general traffic'. It is not obvious that this has to imply that SP2 values exceed SP1 values, particularly not at the scale seen for other non-work, where the value was well in excess of that for heavy traffic.

This also presented a problem for the 'reliability ratio', as we had to decide whether to take the value of the standard deviation relative to the SP2 VTT or the SP1 VTT. The former gave values of 0.33 (commuting), 0.42 (business) and 0.35 (other non-work): the latter gave values of 0.43, 0.64 and 0.77 respectively. The former values are low by 'received wisdom' (though the evidence base for that is not especially strong), while at least for other non-work, the SP1-based result is close to the previous WebTAG value of 0.8.

On balance, it seemed more reasonable to interpret these values relative to the SP2 time multiplier, on the grounds of internal consistency within the SP2 experiment. So, for example, the reliability ratio for car was taken from the ratio of the 'Value of sd travel time' to the value of 'Average travel time' from SP2, and multiplied by the relevant SP1 VTT to get an absolute valuation of the standard deviation (Table 12). The same approach was taken for the early and late multipliers. This is also in line with the way reliability ratios were derived in other work (e.g. Black and Towriss 1993). However, this is not a strongly-based recommendation, and we are still left with the conundrum of explaining the high SP2 time multiplier. The fact that the SP2 VTTs are rather higher than those for SP1

Table 12 VTT multipliers, with SP2 VTT taken as base

Mode	Multiplier	Commute	Other non-work	Employees' business
Car	Reliability ratio	0.33	0.35	0.42
Bus	Value of early	−2.69	−3.20	N/A
	Value of late	2.88	2.52	N/A
'Other PT'	Value of early	−2.40	−2.98	−1.66
	Value of late	1.75	2.24	1.95
Rail	Value of early	−1.77	−2.34	−1.55
	Value of late	2.86	3.21	2.76

Table 13 VTT multipliers, with SP3 VTT for 'A few seats free but had to sit next to someone/could not sit with people travelling with some standing' taken as base

Mode	Multiplier	Commute	Other non-work	Employees' business
Bus	Plenty of seats free and did not have to sit next to anyone	0.85	0.83	N/A
	A few seats free but had to sit next to someone/could not sit with people travelling with	0.89	0.84	N/A
	A few seats free but had to sit next to someone/could not sit with people travelling with Some standing	1.00	1.00	N/A
	No seats free—a few others standing	1.24	1.30	N/A
	No seats free—densely packed	2.14	2.32	N/A
'Other PT'	Plenty of seats free and did not have to sit next to anyone	0.95	1.00	1.00
	A few seats free but had to sit next to someone/could not sit with people travelling with	0.97	1.00	1.00
	A few seats free but had to sit next to someone/could not sit with people travelling with some standing	1.00	1.00	1.00
	No seats free—a few others standing	1.13	1.10	1.17
	No seats free—densely packed	1.70	1.87	1.78

in the case of car and rail does mean, of course, that the implied valuations of reliability will be lower. Without a clear understanding of the reason for the difference between SP1 and SP2 VTTs, this must remain an arbitrary judgment.

Relative to reliability, the remaining multipliers may be considered of lesser importance, but it is worth noting some additional complications which arose in the case of public transport crowding (SP3). For the public transport modes, we chose to align the results with the level of crowding closest to the SP1 VTT: for bus and other public transport this corresponds to the level *"a few seats free but had to sit next to someone/could not sit with people travelling with. Some standing"* (Table 13),[18] while for rail it corresponds to a load factor of 100% (i.e. *"all seats taken but no standing"* (Table 14). Turning to SP3 for car, we note that in all cases the SP1 VTT fell inside the range between the light and heavy traffic values from SP3, though it is difficult to justify this

---

[18] There are multiple rows of ones in some cases, since the relevant multipliers were not significantly different from one.

Table 14 VTT multipliers, with 'seated 100% load' taken as base

Mode	Multiplier	Commute	Other non-work	Employees' business
Rail	Seated 50% load	0.73	0.72	0.75
	Seated 75% load	0.79	0.72	0.76
	Seated 100% load	1.00	1.00	1.00
	Seated 1 pass per m^2	1.09	1.14	1.13
	Seated 3 pass per m^2	1.31	1.39	1.36
	Standing 0.5 pass per m^2	1.16	1.21	1.29
	Standing 1 pass per m^2	1.19	1.27	1.38
	Standing 2 pass per m^2	1.32	1.57	1.56
	Standing 3 pass per m^2	1.57	1.79	1.61
	Standing 4 pass per m^2	1.86	2.17	2.03

Table 15 VTT multipliers, with SP1 VTT taken as base

Mode	Multiplier	Commute	Other non-work	Employees' business
Car	Free-flow	0.51	0.47	0.42
	Light traffic	0.72	0.83	0.68
	Heavy traffic	1.37	1.89	1.26
Bus	Value of free-flow	0.99	1.22	N/A
	Value of slow down	1.39	1.36	N/A
	Value of dwell time	0.68	1.57	N/A
	Value of headway	1.68	1.60	N/A

apart from perhaps the commuting case. Since in any case we consider these ratios to be only indicative of the possible impact of congestion, we divided the SP3 values for the three levels by the SP1 VTT to get the multipliers in Table 15.

Variance and uncertainty around the results

As noted in the study aims in the "Study aims, scope, and delivery" section, the Implementation Tool was developed in such a way that it outputs not only the mean VTT, but also the standard error of the mean and confidence intervals. The latter takes account of both estimation error and NTS sample error, where the latter is a development beyond previous examinations of VTT uncertainty (Wheat and Batley 2015). Using this functionality, 95% confidence intervals were calculated for income option (1) based on distance-weighting.

Whilst the multiplicative models gave direct estimates of WTP measures, the additive models were estimated in preference space, hence giving estimates for the component marginal utilities. These were then used to calculate WTP measures, by taking ratios against the cost coefficients, and standard errors for the WTP measures were computed using the Delta method to allow for errors in the estimated parameters. This is known to give the exact same result as working in preference space, both for the estimates and the standard errors, using MNL models (Daly et al. 2012a, b). Additionally, using the bootstrap method, we calculated errors that allowed for the fact that the NTS is itself a sample.

Though we do not present the confidence intervals here—interested readers can find them in the final report[19] (Arup, ITS Leeds and Accent 2015a; Section 7.7)—we note that, broadly speaking, the models and associated VTTs were well-estimated. In general, the confidence intervals for the VTT from SP1 were just below ±30% of the mean; rail and 'other PT' had the narrowest confidence intervals, whilst bus had the widest intervals. Another finding was that the SP1 VTTs had slightly narrower confidence intervals (<30%) than SP2 and SP3. All of the valuations were found to be significantly different from zero, apart from the value of bus dwell time.

Summary and conclusions

In summarising and concluding this paper, it is appropriate to draw reference to the four aims introduced at the outset (see the "Study aims, scope, and delivery" section).

To provide recommended, up-to-date national average values of in-vehicle travel time savings, covering business and non-work travel, and based on primary research using modern, innovative methods.

Within the standard microeconomic framework of goods versus leisure trade-off subject to money and time constraints, we employed Stated Preference (SP) experimental methods to estimate WTP for time savings, for travellers in the course of business and non-work trips. The SP involved three games, which considered different trade-offs, namely: SP1 (time vs. money), SP2 (time vs. money vs. reliability), and SP3 (time vs. money vs. crowding/congestion).

The main field survey employed, primarily, an intercept-based recruitment approach at around 30 representative locations across England for each mode. Recruits were directed to a web-based questionnaire, and were paid a nominal reward on completion of the questionnaire. Potential gaps in this primary approach (e.g. infrequent travellers and/or those without computer access) were covered by a secondary approach of telephone recruitment and telephone and/or paper-based administration of the questionnaire. The field survey was completed successfully; most targets were achieved, and overall the pilot survey provided a good guide to the likely success, issues and response rates in the field survey.

To investigate the factors which cause variation in the values (e.g. by mode, purpose, income, trip distance or duration, productive use of travel time etc.) and use this to inform recommended segmentation of the values.

The modelling was developed in a systematic fashion, whereby alternative model specifications were tested in relation to:

- time and cost gains, losses and size effects;
- person characteristics such as age, gender, employment status, household composition and income; and
- trip characteristics such as mode, purpose, distance and geography.

All values were reported by purpose and mode—given the importance of these variables as key policy segments—and as a function of income elasticity, as well as a function of elasticities with respect to the time and cost of the 'reference' trip. Where other covariates

---

[19] The final report also discusses the potential sources of variance, which in the present case include: (1) differences in the estimation errors in the behavioural models by mode and purpose; (2) differences in the representativeness of the NTS data by mode and purpose leading to different bootstrapping errors; and (3) differences in sample size.

were identified as significant, these were retained in the preferred choice models and, in turn, fed into the process of eliciting values for implementation in appraisal.

Indeed, a key task was the progression from the values derived from best specification models to recommended values for use in appraisal. This entailed four steps:

• review and modification of the choice models in the light of interpretation;
• specification of standard values to be produced;
• re-weighting of those values for national representativeness using the National Travel Survey (NTS); and
• assumptions concerning changes in VTT over time.

To assist the Department's analysts in generating VTTs for appraisal guidance, we developed an 'Implementation Tool' which calculates VTT for different segments, based on a variety of weighting options. Rather than just apply weights from the NTS sample to the respondents in our estimation data, we made use of sample enumeration which applies the estimated models to the respondents in the NTS sample.

To improve our understanding of the uncertainties around the values, including estimating confidence intervals around the recommended values.

For any appraisal scenario of interest to the user, the Implementation Tool generates confidence intervals around the mean VTT. This estimation takes account of two potential sources of uncertainty in values, namely:

• error in the mean VTT arising from the parameters estimated in the choice model; and
• error in the NTS sample, which arises from the fact that NTS is a sample drawn from the travelling population.

More generally, we demonstrated the robustness of the estimated values through a number of exercises. First, we validated our SP values against comparable RP values. This exercise involved the design and implementation of separate SP and RP experiments around a common choice context—namely the choice between competing rail operating companies for medium distance rail trips to London. Whilst the resulting values were somewhat high, due to confounding with operator preferences, the SP and RP delivered similar results. Second, we validated our preferred choice model specification by testing against alternative specifications, including the recommended model from the 2003 national study. This demonstrated the greater functionality and superior statistical fit of the 2015 model, as compared with the 2003 model. Finally, it should be noted that any decisions in relation to model structure were not solely based on best practice from past studies and more recent developments in the academic literature, but also on extensive testing on the data at hand, such as tests comparing additive and multiplicative model structures.

To consistently estimate values for other trip characteristics for which values are derived from the values of in-vehicle time savings.

Whilst Stated Preference experiment 1 (SP1) offered the respondent trade-offs between time and cost, SP2 and SP3 offered trade-offs between travel time, cost, reliability and congestion/crowding. The objectives of this exercise were three-fold: first, to ascertain whether, and to what extent, the 'headline' VTT from SP1 was confounded with the value of reliability and congestion/crowding; second, to estimate incremental multipliers for reliability and congestion/crowding suitable for implementation in appraisal; and third, to gain insights into what the SP1 values represented in terms of type of time (i.e. trip conditions).

The use of the joint modelling framework across all games allowed us to examine the differences between individual valuations across games. This joint estimation yielded key insights, for example showing the extent of VTT increases with road congestion (for car and bus), as well as with the level of crowding (for all PT modes). It also showed differences across modes and across purposes as to what type of trip conditions in SP3 the values from SP1 relate to. The values coming out of SP2 diverge from those of SP1 and SP3, and this is likely to be a behavioural impact in terms of how respondents reacted to variability in travel time.

References

Accent and Hague Consulting Group (AHCG): The value of travel time on UK roads. Report to Department for Transport (1999)

Arup, ITS Leeds, Accent: Provision of market research for value of time savings and reliability. Phase 2 report to the Department for Transport (2015a). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470231/vtts-phase-2-report-issue-august-2015.pdf

Arup, ITS Leeds, Accent: Provision of market research for value of travel time savings and reliability. Walk and Cycle report to the Department for Transport (2015b). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470237/vtts-walk-and-cycle-report-issue-14-aug-2015.pdf

Association of Train Operating Companies (ATOC): Passenger Demand Forecasting Handbook. Association of Train Operating Companies, London (2012)

Becker, G.: A theory of the allocation of time. Econ. J. **75**(299), 493–517 (1965)

Black, I.G., Towriss, J.G.: Demand Effects of Travel Time Reliability. Centre for Logistics and Transportation, Cranfield Institute of Technology, Cranfield (1993)

Börjesson, M., Fosgerau, M., Algers, S.: Catching the tail: empirical identification of the distribution of the value of travel time. Transp. Res. Part A **46**(2), 378–391 (2012)

Börjesson, M., Eliasson, J.: Experiences from the Swedish Value of time study. Transp. Res. Part A **59**, 144–158 (2014)

Daly, A.: Cost damping in travel demand models. Report of a study for the Department for Transport (2010). http://assets.dft.gov.uk/publications/pgr-economics-rdg-costdamping-pdf/costdamping.pdf

Daly, A.J., Hess, S., de Jong, G.: Calculating errors for measures derived from choice modelling estimates. Transp. Res. Part B **46**(2), 333–341 (2012a)

Daly, A.J., Hess, S., Train, K.E.: Assuring finite moments for willingness-to-pay in random coefficients models. Transportation **39**(1), 19–31 (2012b)

Daly, A., Zachary, S.: Commuters' values of time. Report to Department of the Environment, LGORU report T55, January (1975). http://www.alogit.com/General_papers.htm

De Borger, B., Fosgerau, M.: The trade-off between money and travel time: a test of the theory of reference-dependent preferences. J. Urban Econ. **64**(1), 101–115 (2008)

Department for Transport (DfT): Understanding and valuing the impacts of transport investment: latest DfT technical research and next steps in transport appraisal (2013). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253485/technical-research-next-steps-appraisal.pdf

Department for Transport (DfT): TAG unit A1.3: User and provider impacts, November (2014). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/539338/webtag-tag-unit-a1-3-user-and-provider-impacts-november-2014.pdf

Department for Transport (DfT): Understanding and valuing impacts of transport investment: values of travel time saving. Moving Britain ahead' (2015). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470998/Understanding_and_Valuing_Impacts_of_Transport_Investment.pdf

Department for Transport (DfT): Understanding and valuing impacts of transport investment: value of travel time saving consultation response. Moving Britain ahead (2016). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/544165/understanding-and-valuing-the-impacts-of-transport-investment-values-of-travel-time-savings-consultation-response.pdf

Department for Transport (DfT): TAG unit A1.3: User and provider impacts, February (2017) (in press)

De Serpa, A.: A theory of the economics of time. Econ. J. **81**(324), 828–846 (1971)

Evans, A.: On the theory of the valuation and allocation of time. Scott. J. Polit. Econ. **19**(1), 1–17 (1972)

Fosgerau, M., Hjorth, K., Lyk-Jensen, S.: The Danish Value of Time Study: Results for Experiment 1. Danish Transport Research Institute (2007)

Fosgerau, M., Bierlaire, M.: Discrete choice models with multiplicative error terms. Transp. Res. Part B **43**(5), 494–505 (2009)

Fowkes, T., Wardman, M.: The design of stated preference travel choice experiments. J. Transp. Econ. Policy **22**(1), 27–44 (1988)

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

Harrison, A.J.: The Economics of Transport Appraisal. Croom Helm, London (1974)

Hensher, D.A.: Value of Business Travel Time. Pergamon Press, Oxford (1977)

Hess, S., Daly, A., Dekker, T., Ojeda Cabral, M., Batley, R.: A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. Transp. Res. Part B **96**, 126–149 (2017)

Treasury, H.M.: The Green Book: Appraisal and Evaluation in Central Government. Treasury Guidance, TSO, London (2013)

Hollander, Y.: Direct versus indirect models for the effects of unreliability. Transp. Res. Part A **40**(9), 699–711 (2006)

International Transport Forum: Valuing Convenience in Public Transport. Roundtable report 156. OECD, Paris, France (2015)

ITS Leeds: M6 Toll study: modelling of passenger choices. Final report to the Department for Transport (2008). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/194011/passenger-surveys-report.pdf

ITS Leeds, John Bates, DTU: Updating appraisal values for travel time savings Phase 1 study. Report to the Department for Transport (2010). https://www.gov.uk/government/publications/values-of-travel-time-savings-updating-the-values-for-non-work-travel

ITS Leeds, John Bates, KTH: Valuation of travel time savings for business travellers. Main report, prepared for the Department for Transport (2013). https://www.gov.uk/government/publications/values-of-travel-time-savings-for-business-travellers

Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979)

MVA, ITS, TSU: The Value of Travel Time Savings. Policy Journals, (1987)

MVA: Valuation of overcrowding on rail services. Prepared for Department for Transport (2008)

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

Mott MacDonald, Hugh Gunn Associates, TRI Napier University, Accent and Mark Bradley Research, Consulting: Productive use of rail travel time and the valuation of travel time savings for rail business travellers. Final report to the Department for Transport (2009)

Rose, J.M., Bliemer, M.C.J.: Survey artefacts in stated choice experiments. In: Proceedings of the International Conference on Transport Survey Methods, Leura, Australia (2014)

Wardman, M., Batley, R.P., Laird, J.J., Mackie, P.J., Bates, J.: How should business travel time savings be valued? Econ. Transp. **4**(4), 200–214 (2015)

Wheat, P., Batley, R.: Quantifying and decomposing the uncertainty in appraisal values of travel time savings. Transp. Policy **44**, 134–142 (2015)

Richard Batley is Professor of Transport Demand and Valuation at the Institute for Transport Studies (ITS), University of Leeds, UK. He is also Director of ITS Leeds. He is a transport economist by background, and has particular expertise in the valuation of non-market goods and welfare economics.

John Bates is an Independent Consultant, now based in Berlin, Germany. He specialises in mathematical modelling, and has been at the forefront of transportation model development and evaluation methodology, in particular the valuation of time savings and reliability. In addition, he has been a leading figure in the development of stated preference techniques within the transport field.

Michiel Bliemer is Professor of Transport Network Modelling at the Institute of Transport and Logistics Studies (ITLS), University of Sydney, Australia. He is a leading expert in transport planning and traffic management methodologies in order to forecast traffic volumes and conditions, as well as methodology around stated choice experiments and choice analysis.

Maria Börjesson is Associate Professor of Transport Systems Analysis at the Centre for Transport Studies (CTS), Royal Institute of Technology, Stockholm, Sweden. Her research interests include transport cost-benefit analysis, appraisal and sustainability, transport policy and pricing, and econometrics for valuation of non-market goods such as travel time, reliability and security.

Jeremy Bourdon is Consultant at CB&S Limited, London, UK; he was previously at Arup. He established his own consultancy in order to work freelance whilst transitioning his career from programme and project risk management of infrastructure to sustainable finance.

Manuel Ojeda Cabral is Research Fellow in Transport Economics and Appraisal at ITS, University of Leeds, UK. He has experience in various areas of transport economics, including valuation studies, choice modelling, cost-benefit analysis and efficiency analysis. His main research has focused on the economic theory and the estimation of individuals' valuation of travel time savings.

Phani Kumar Chintakayala is Senior Research Fellow in Big Data at the Leeds Institute for Data Analytics (LIDA), University of Leeds, UK. He has experience using stated preference methods and discrete choice models to draw consumer insights for both public and private sector organisations.

Charisma Choudhury is Lecturer in Transport Engineering and Emerging Economies at ITS, University of Leeds, UK. She is also Deputy Director of the Choice Modelling Centre (CMC) at ITS. Her key research interests include behavioural modelling and discrete choice analysis, traffic microsimulation, transport modelling using big data sources, and transportation in developing countries.

Andrew Daly is Emeritus Professor at ITS, University of Leeds, UK; also Senior Research Fellow at RAND Europe and Visiting Professor at KTH, Stockholm. He is proprietor of ALOGIT Software & Analysis Ltd. His pioneering work on the value of time in the 1970s introduced Random Utility to value of time estimation, and he was recognised internationally as an expert in the field by 1980. In the 1980s and 1990s he contributed to national value of time studies in The Netherlands and the UK.

Thijs Dekker is Lecturer in Transport Economics at ITS, University of Leeds, UK. In both environmental and transport economics, Thijs has a proven track record of publishing in the field's leading journals. His research in the field of discrete choice modelling is on the cutting edge between econometrics, microeconomics and psychology.

Efie Drivyla is Associate Director at Arup, London, UK. Efie has directed and managed a broad range of projects varying from research projects and multi-modal and urban/inter-urban studies to local transportation

studies. She led Arup's transport planning input into a number of major projects, such as the M1 Widening (junctions 21–30).

Tony Fowkes is Visiting Reader in Transport Econometrics at ITS, University of Leeds, UK. Tony's background is in economics and statistics, working briefly in labour economics before joining ITS in 1976, where he worked initially on car ownership forecasting. In 1982 he joined the DfT/BR value of time study, and the associated ESRC study of business travellers value of time (including groundbreaking RP and SP surveys of employees and a WTP survey of employers).

Stephane Hess is Professor of Choice Modelling and Director of the Choice Modelling Centre (CMC), both at ITS, University of Leeds, UK. He is also Honorary Professor in Choice Modelling at ITLS, University of Sydney. He is a leading international researcher in the development of advanced discrete choice models for the analysis of travel behaviour, primarily with the use of stated preference data.

Chris Heywood is Technical Director at Accent, London, UK. With over 30 years' market research experience, Chris is a highly skilled manager of complex survey logistics. He has particular expertise in quantitative methods in the transportation market. He is a specialist in the use of stated preference techniques and other innovative research approaches.

Daniel Johnson is Senior Research Fellow at ITS, University of Leeds, UK. He is an economist with 24 years' experience specialising in labour market analysis and transport-economy linkages. Dan has a wide range of modelling experience covering productivity analysis, transport networks and wider economic impacts of bus provision.

James Laird is Visiting Research Fellow at ITS, University of Leeds, UK. He has 23 years' experience within both academic and consultancy environments. His primary research and consultancy interest lies in the economic appraisal of transport infrastructure. He has a strong interest in cost-benefit analysis, economic impacts of transport projects, valuation of travel time, wider economic benefits and demand forecasting.

Peter Mackie is Emeritus Professor at ITS, University of Leeds, UK. He is a leading UK academic figure in transport policy and economics and has specialised in issues relating to transport appraisal and to regulation and deregulation. He has a well-established relationship with the Department for Transport, having been a member of SACTRA for its influential 1992, 1994 and 1999 reports, and also one of the analytical Friends advising Sir Rod Eddington on his 2006 report.

John Parkin is Professor of Transport Engineering at the Centre for Transport and Society (CTS), University of West of England, Bristol, UK. His research interests include the experimental analysis of factors that affect infrastructure design, the assessment and modelling of perceptions of risk and effort as they influence mode and route choice, and the longitudinal analysis of transport datasets for monitoring purposes.

Stefan Sanders is UK, Middle East and Africa Rail Business Leader for Arup, London, UK. He has more than 25 years' international experience in transport consulting and management, particularly in rail. Stefan is highly experienced at leading major strategy projects, preparing major business cases, modelling railway organisations, advising on rail restructuring (he led Arup's contribution to the McNulty Rail Value for Money study) and understanding/advising on the rail regulatory framework.

Rob Sheldon is Managing Director at Accent, London, UK. He has a background in econometrics and over 35 years' experience of market research consultancy. Rob is an internationally recognised expert in the field of stated preference. He was instrumental in introducing the technique to the UK, for a project for British Rail in the 1970s. He has published extensively and lectured widely on the technique and has managed or directed more than 250 stated preference projects.

Mark Wardman is Visiting Professor at ITS, University of Leeds, UK. He has over 30 years' experience of research into the quantitative analysis of travel behaviour. His first research position was on the UK Department for Transport's pioneering national value of time study in the 1980s. More recently, Mark was responsible for the most extensive meta-analyses ever conducted of price elasticities, time elasticities and values of time and time related attributes.

Tom Worsley is Visiting Fellow in Transport Policy at ITS, University of Leeds, UK. Tom joined the UK Government Economic Service in 1969 and worked as an economist in a number of different government departments. Most of his time was spent in the Department for Transport and its predecessors. He held various senior economist posts in the Department between 1988 and May 2011, leading teams responsible for developing and improving economic methods.