Technical University of Denmark

DTU

# Value of travel time savings

**Le Masurier, P.; Polak, J.; Pawlak, Janet; Fosgerau, Mogens; Hjorth, Katrine**

*Publication date:*
2015

Link back to DTU Orbit

*Citation (APA):*
Le Masurier, P., Polak, J., Pawlak, J., Fosgerau, M., & Hjorth, K. (2015). Value of travel time savings.

# DTU Library
## Technical Information Center of Denmark

# VALUE OF TRAVEL TIME SAVINGS - PEER REVIEW & AUDIT

# VALUE OF TRAVEL TIME SAVINGS

## VALUE OF TRAVEL TIME SAVINGS - PEER REVIEW & AUDIT

| IDENTIFICATION TABLE | |
|---|---|
| Client/Project owner | Department for Transport |
| Project | Value of Travel Time Savings |
| Study | Value of Travel Time Savings - Peer Review & Audit |
| Type of document | Final Report |
| Date | 30/07/2015 |
| File name | VTTS Peer Review - Final Report v3.0 |
| Reference number | 103217 |
| Number of pages | 48 |

**APPROVAL**

| Version | Name | | Company | Date | Comments |
|---|---|---|---|---|---|
| 1 | Authors | P Le Masurier<br>J Polak<br>J Pawlak<br>M Fosgerau<br>K Hjorth | SYSTRA<br>ICL<br>ICL<br>DTU<br>DTU | 25/06/2015 | Review Team's Draft Peer Review Report (inc. audit of Implementation Tool) |
| | Checked by | C Pownall<br>R Maldonado | SYSTRA | 26/06/2015 | |
| 2 | Authors | P Le Masurier<br>J Polak<br>J Pawlak<br>M Fosgerau<br>K Hjorth | SYSTRA<br>ICL<br>ICL<br>DTU<br>DTU | 14/07/2015 | Revised draft to consider additional model runs |
| | Checked by | C Pownall<br>R Maldonado | SYSTRA | 15/07/2015 | |
| 3 | Authors | P Le Masurier<br>J Polak<br>J Pawlak<br>M Fosgerau<br>K Hjorth | SYSTRA<br>ICL<br>ICL<br>DTU<br>DTU | 29/07/2015 | Final version |
| | Checked by | C Pownall<br>R Maldonado | SYSTRA | 30/07/2015 | |

# TABLE OF CONTENTS

Value of Travel Time Savings

| Value of Travel Time Savings - Peer Review & Audit | 103217 | |
|---|---|---|
| Final Report | 30/07/2015 | **Page** 3/48 |

# SUMMARY

**The Audit and Review**

A team of specialist market researchers and Value of Time experts comprising members from SYSTRA, Imperial College London and the Technical University of Denmark has conducted a formal audit and peer review of research undertaken by Arup/ITS Leeds/Accent to derive Value of Travel Time Savings on behalf of the UK Department for Transport (DfT).

The research team has completed their analysis and modelling of the collected data, and has delivered their Final Report to the DfT. The research team has provided the peer reviewers with access to data collection processes for on-site auditing, and some interim reports and technical notes covering interim modelling activities; and provided clarification on some points when requested by the reviewers. The review team has taken the opportunity to audit and review all aspects of the survey data collected; and the methods of analysis and modelling of the data, as reported by the research team. The review team has also been provided with some meta-code for some key model estimations but were not provided with the actual code used in the final choice model estimation, such as the syntax used to develop the final Stated Preference (SP) and Revealed Preference (RP) models that were used to derive final Values of Travel Time (VTT).

This report contains the findings of our audit and peer review of the procedures adopted by the research team during data collection of the three surveys (SP, RP and Employers Surveys); a peer review of the reported approach and implementation to data modelling; and an audit of the research team's Implementation Tool that weighted the modelled parameters to give overall national VTT estimates for use in transport scheme appraisals.

**Conclusions**

The review team recognises the complex and high-profile nature of the study; and the range of technical and logistical issues involved in designing and conducting a large-scale, national study of travellers' attitudes and preferences. Our review has uncovered some issues that potentially affect the quality of data from a **minority** of respondents; and we have raised questions about the cognitive understanding of respondents <u>and</u> the suitability of the interview environment in some cases. We believe that some of the attention, and resource, that has been invested in elaborations in the model specification and estimation stage of the project would have been better deployed on (a) undertaking a more thorough investigation of the implications for final VTT of the distinct sources of error associated with specification and implementation and (b) improving data collection, cleaning and data imputation within the research team's implementation tool.

The research team has successfully tackled many of the challenges of collecting and cleaning a large quantity of complex data, by putting in place a series of rigorous and systematic procedures for:

Value of Travel Time Savings

| Value of Travel Time Savings - Peer Review & Audit | 103217 | |
|---|---|---|
| Final Report | 30/07/2015 | **Page** 4/48 |

- recruiting willing in-scope research participants;
- collecting information on travellers' and employers' values of time savings; and
- securely transferring, cleaning and processing the raw data for subsequent analysis.

The review team is also in general content with the SP design work and technical skills demonstrated in choice modelling. In addition, our audit of the Implementation Tool found that it correctly implemented the study team's approach to the synthesis national VTTs.

However, we have five main concerns that we think could potentially impact on the quality, and reliability, of the final VTTs.

First, during data collection, a minority of respondents were asked to trade between complex hypothetical choices without the benefit of visual stimuli. In these cases, the interviewer simply read out the SP options to respondents over the phone. This insight was only obtained following interviews that the review team conducted with interviewing staff, and survey respondents. There is no evidence that the research team knows which respondents were interviewed without the aid of stimuli; nor how many respondents were surveyed in this manner; nor therefore their impact on final VTTs. Other examples of poor quality responses [affecting circa 10% of the overall sample, we believe] include those respondents who undertook the full interview in less than 5 minutes (when it is expected to take 25-40 minutes); and where respondents answered diagnostic questions confirming that they did not understand the exercise. We think, therefore, that the modelled data includes some responses that are of very poor quality but we have only limited information on whether their exclusion would lead to different VTTs as a result. With reference to the RP survey, we think that the research team's intentions to develop an RP approach to validate the SP work were laudable, but we were surprised by the number of responses discarded (with just 20% remaining) which inevitably undermines confidence in the conclusions of the validation exercise carried out.

Our second main concern regards respondents' attitudes to the context in which the SP experiments were set. The validity of the derived values of time relies on respondents considering the SP options to be plausible and meaningful to them, and for their responses to reflect (solely) the monetary value they attach to time savings. The research team designed three SP experiments, all of which featured journey options that were quicker and more expensive than other journeys. However, there is a theoretical problem with this if the cost variable is based around fuel (as is the case in this study) since the slower a journey is the more fuel would be expected to be used for the same distance and, hence, the more expensive the journey would be. [In contrast, parking or toll charges would not have this direct association with journey duration – but these alternative cost components have other SP design/presentation implications]. There is evidence to suggest that some respondents considered a journey that is quicker and more expensive than another journey to be implausible.

To explain away such implausibility, the research team provided additional context to respondents – that the travel cost and time between two options may be different because of

a change in the cost of fuel, congestion levels, breakdowns or unplanned roadworks. This introduces confounding factors into the time/cost trade-off – so we do not know whether some respondents chose a quicker and more expensive journey because (in SP2) they wanted to avoid the breakdowns/roadworks associated with the longer journey option. Nor do we know whether some respondents chose a cheaper and longer duration journey because they want the cheaper fuel costs (for all their journeys) implied by the 'cheaper fuel' option. The research team conducted 4 months of qualitative study that examined research participants' cognitive assessment of different types of SP context and wider issues surrounding time/cost trade-offs but there is no indication that respondents considered only time and cost differences and no other influences. We are reassured, to some degree, by the fact that the study team report that most respondents self-reported that they understood all aspects of the SP exercises however, it is not clear *what* they understood. Therefore, there remains a concern that some SP responses have been affected by a confounding element that has not been taken account of in the analysis (this mainly applies to car respondents only).

Third, and also related to presentation of the SP exercise, we are concerned that in the SP exercises investigating the effect of uncertainty in travel times, respondents were given no information regarding what probability they should attach to each of the five travel times shown. We regard this as a fundamental issue; any uncertain choice situation comprises outcomes and associated probabilities. Only presenting outcomes does not adequately characterise the uncertain choice. There is a vast body of literature that indicates that if respondents are not explicitly informed regarding probabilities, they are prone to infer all manner of spurious patters ("…it seems to be getting worse…") and/or make heuristic assumptions (e.g., good travel times are less likely than bad ones) which may be at variants with these intended or assumed by the analyst. In this case, the analysis seems to be based on the assumption that each of the five travel times was equally likely, which is a simple message that could have been conveyed to respondents during the SP, but it seems was not. This inevitably casts doubt on the validity of this part of the analysis (i.e. SP2).

Fourth, some members of the review team have concern that the implications of the study team's use of a choice modelling approach that depends fundamentally on notions of reference dependence have not been sufficiently investigated or thought through. There are two main aspects to this. Firstly, there appears to have been no attempt whatsoever in the qualitative work to investigate any aspect of reference dependence e.g., whether and under what circumstances respondents recognise it as a real phenomenon and if so, what constitute the relevant reference points. Secondly, and more specifically, whilst the study team present strong statistical evidence from the SP data for the presence of reference dependence with respect to the current trip (which is the only form of reference dependence they tested for) neither in the qualitative work nor in the subsequent modelling was any attempt made to investigate to what degree such reference-dependence might be an artefact of the SP approach as opposed to a substantive effect in real behaviour. This is of substantial significance since these two potential explanations for reference dependence have very different implications for the estimation of nationally representative VTTs. This point appears not to have been considered by the study team.

Our fifth and final major concern, which relates to the overall balance of effort of the study, especially as seen in the choice model specification/estimation and the implementation of the choice model estimates to produce nationally representative VTTs. Both are challenging modelling tasks, subject to various sources of model-related and data-related error. The final error in VTTs will depend on the combination of all these errors and will entail trade-offs between errors in different stages of the process. It appears that the nature of this fundamental trade-off has not been explicitly considered by the research team. Instead, their implicit strategy seems to have been to estimate the most comprehensive choice model specification possible, consistent with the SP data, and then to make a series of simplifying and sometimes rather simplistic assumptions in the implementation stage (several of which are necessarily rather *ad hoc* in nature) to recover from these models estimates of VTTs appropriate for national use. Whilst it is possible that this is the overall optimal strategy for the estimation of VTTs, we think that more sensitivity and uncertainty analysis should be undertaken to consider different conditioning variables and imputation methods. For example, a Monte Carlo analysis to investigate the implications of various alternatives at different stages in the analysis chain would have been necessary.

Overall, we are satisfied with the way the research team has implemented most of their planned programme of surveys and modelling, but we have reservations over the quality of some of the data collected and the considerable emphasis on what one might term 'more ambitious modelling' (at the specification and estimation stages) - perhaps due to the research brief's requirement for *'innovative, up-to-date analytical techniques'* – as it seems to have been at the expense of a more thorough investigation of data quality issues and more measured development of hypotheses and model testing. However, we do understand that there are always budget and time constraints for research studies such as this, and the concerns that we have do not necessarily invalidate the work carried out to date, but clearly identify areas where further work would be merited. Furthermore, it must be noted that though the research team has provided a series of working papers and reports and some model syntax, the review team has not been provided with key documents such as the core syntax to reproduce their final model and the Ox script used to test different model options that would have provided essential insights into how the final models have been derived.

# 1. INTRODUCTION

## 1.1 VTT Background

1.1.1 The Transport Appraisal and Strategic Modelling (TASM) team within the Department for Transport is responsible for maintaining and updating WebTAG[1] guidance on the modelling and appraisal of transport interventions. Values of travel time savings and reliability are often central to these appraisals, which are an important part of the Transport Business Case[2]. Therefore, TASM commissioned a research project (Value of Travel Time Saving Research) to provide fully up-to-date values for use in transport appraisal.

1.1.2 In May 2014, the Department awarded a contract to a consortium including Arup, Institute for Transport Studies (University of Leeds) and Accent to undertake research on 'Values of travel time savings and reliability' (hereafter referred to as the 'research team'). The research project was to estimate values of travel time savings (and other related factors such as crowding and reliability) based on estimates of willingness-to-pay derived from stated and revealed preference (SP and RP) exercises. These willingness-to-pay methods have been applied to business, as well as personal and commuting, travel looking at employees, employers and the self-employed.

1.1.3 The research team has collected RP and SP data using survey-based methods. Travellers for car, rail and bus (for business, commuting and other non-work journey purposes) were recruited through a mixture of street intercept and telephone contact methods. All travellers on other public transport modes (light rail, tram and London Underground) were recruited by intercept.

1.1.4 To identify potential respondents during on-street intercept surveys, a '1 in n' approach was intended, making use of a short screening interview to establish whether the traveller was in scope and they would be willing to take part in the research. The telephone recruitment survey used a Random Digital Dialling (RDD) sample that geographically represented the population of England by region. Again, potential respondents first undertook a short recruitment interview to establish whether they were in scope and willing to participate in the research.

1.1.5 Willing participants were offered the choice to complete the Main survey questionnaire online or by computer assisted telephone interview (CATI).

1.1.6 The recruitment and surveying of employers was undertaken by telephone. The sample of business phone numbers was purchased from an established provider and numbers randomly assigned to interviewers. The target respondent was intended to be the person within the company responsible for making decisions about how employees travel for business purposes.

1.1.7 As well as covering a range of journey purposes, the research project also covered active modes (walking and cycling) as well as car and public transport modes.

---

[1] https://www.gov.uk/transport-analysis-guidance-webtag

[2] https://www.gov.uk/government/publications/transport-business-case

## 1.2 Scope of the Independent Peer Review

1.2.1 On 26th January 2015, SYSTRA Ltd and partners were commissioned to verify and audit the collected data; audit the modelling processes that have been undertaken; and provide a peer review of the overall approach taken and recommendations being made for the values of travel time savings and reliability. The aim of the review was to provide external and independent verification of the methodology and interpretation of results, to establish if the survey data has been collected in the way that it has been reported, and the analysis was accurate and unbiased. Due to survey challenges that have delayed progress in the SP research with walkers and cyclists, the remit of the audit and peer review is limited to the data collection and modelling of SP data from car, bus, rail and other PT users only. The RP survey, and survey of employers, were also in scope for review.

1.2.2 The Department envisaged that the peer review and audit team would work concurrently with the research team. The primary point of contact remained the Department's project manager for the research project but the peer review team also took the opportunity to liaise directly with members of the research team.

1.2.3 For each of the surveys in scope, our review was to comprise:

- an ex-post audit of the data collected;
- an audit of modelling and peer review of recommendations in the interim report on "emerging values and segmentation"; and
- an audit of modelling and peer review of recommendations in the final report.

1.2.4 The review team submitted our Interim Report on 25th February and was based on an audit and peer review of the procedures implemented in data collection; and a peer review of the modelling methods and models as reported in the research team's Emerging Values Report. The review team submitted our Draft Report on 17th April and was based on an audit of modelling processes made available by the research team, and a peer review of the modelling methods and models as reported in the research team's Draft Final Report.

1.2.5 This Final Report documents the review team's observations and recommendations following a final audit and review of the data collected; review of the modelling as reported by the research team in their Final Report, dated 29 May 2015; and audit of the research team's Implementation Tool. Our audit of the data collection methods adopted by the research team included an exercise to obtain feedback from those directly involved in data capture - i.e. interviewers and respondents – to obtain first-hand views of how the survey was implemented and, in the latter case, how well the survey (and, in particular, the SP choice-sets) were understood. An audit of the actual modelling undertaken by the research team to create the reported models was not possible as, although it was requested, the review team did not receive details of the core modelling code, such as the syntax used to exclude respondents and modelling commands to derive the estimates and variance-covariance matrices in Ox[3].

---

[3] Meta- code for the car commute and rail employees' business models were received, and audited, by the review team but this cannot be the basis for an audit of the modelling approach adopted for the final models. This element included basic commands to exclude a list of respondents, create new variables, recode existing variables and build the single utility function for each SP exercise. Intellectual Property Rights were given as the reason for the modelling code being withheld. The review team offered to sign a Non-disclosure agreement or similar, but this did not help facilitate the transfer of information.

# 2. INDEPENDENT PEER REVIEW

## 2.1 The Peer Review Team

2.1.1 The peer review was conducted by two senior staff at SYSTRA, in collaboration with specialists at Imperial College London and the Technical University of Denmark (DTU), and whose qualifications and skills are summarised below. None of the peer review team are part of the commissioned VTTS research project.

2.1.2 **Paul Le Masurier** is Project Director Social and Market Research at SYSTRA. He has more than 25 years' experience managing social and market research studies in the transport sector, and has led the review of the research team's data collection activities. He has managed numerous SP research studies on behalf of government and the private sector, and in a range of contexts: road user charging and tolling, workplace parking levies, evaluating the cost to society of aircraft noise, and developing business cases for new transport schemes. Most of the SP surveys to inform mode-choice modelling have involved estimations of values of travel time (VTT), by traveller segment – including high-profile applications such as: Multi-modal studies in the UK (e.g. SWYMMS); the Sydney - Canberra - Melbourne High Speed rail link; Berlin – Warsaw – Minsk – Moscow train service for the EBRD; and the High Speed Rail link between Kuala Lumpur - Singapore. He has conducted audits and reviews of a range of different consultant's SP work both in the UK and abroad.

2.1.3 **Rafael Maldonado** is an experienced statistician and project manager at SYSTRA, with experience in using a variety of statistical and programming packages (including Biogeme and Gauss). He has developed advanced modelling with rules-based approaches, and is currently researching logit models to understand travellers' behaviour using advanced modelling techniques. He has been involved in all aspects of this audit and peer review.

2.1.4 **Professor John Polak** is Professor of Transport Demand, Chairman of the Centre for Transport Studies and Director of Research in the Department of Civil and Environmental Engineering at Imperial College London. He is a mathematician by background with over 30 years' experience in transport research and teaching, specialising in the areas of mathematical and statistical transport modelling and analysis. He is a past President of the International Association for Travel Behaviour Research and a past Council Member of the Association for European Transport and has served as an advisor to central and local government and industry. He has an extensive record of high quality publication on issues related to discrete choice model theory and estimation relevant to VTT. His particular expertise encompasses: VTT estimation (including valuation of travel time reliability and productive use of travel time); SP design and the treatment of SP errors; and in the translation of model estimation results to appraisal guidance (including contributions to a number of elements of WebTAG).

2.1.5 **Professor Mogens Fosgerau** of the DTU is among the leading transport economists in the world. He has published extensively in the leading journals in transportation, transportation economics and urban economics. He is guest professor at Royal Institute of Technology, Sweden, and chief co-editor of Economics of Transportation. He has considerable experience undertaking value of time studies across Europe including the Danish value of time study, for which he developed econometric methods specifically for the estimation of distributions of the value of time. These methods are based on microeconomic theory and in particular on

the use of nonparametric techniques, which he has developed in the context of discrete choice. Together with Nobel laureate Dan McFadden, Mogens Fosgerau has also contributed to the theory underlying discrete choice modelling.

## 2.2 Peer Review Methodology

2.2.1 The review was principally conducted through reading documentary evidence, and discussions with key research team personnel involved in data collection and modelling.

2.2.2 The documents scrutinised were as follows:

<u>Reports</u>

Phase 1 Report and Appendices
Initial Values Report and Emerging Values Report
Draft Final Report and Appendices
Final Report and Appendices

<u>Notes</u>

Values of Time Savings
ACT Phase 1 Report Review
Comments Phase 1 Report
Various technical notes on research processes provided by the research team

<u>Other</u>

Final questionnaires [General Travellers SP Survey; RP Survey; Employers Survey]
Raw data [General Travellers SP Survey; RP Survey; Employers Survey]
On-site review of data and documents
Partial Meta-code used for car commute and rail models (i.e. Ox syntax used for manipulation of the raw data prior to model estimation, and the code/commands defining the model estimation process in terms of identifying the utility functions but not the code that attempts to find the optimum for the joint multiplicative models)
Implementation Tool (v1.1)
The research team's written responses to some of the concerns expressed by the review team during the final stages of the peer review process.

2.2.3 The review team also undertook the following:

- Meetings with key staff from Accent to better understand the processes undertaken for data collection and cleaning;
- A review of a sample of responses to see how responses are recorded, cleaned and edited at each step of data processing (where possible) prior to modelling;
- Interviews with survey staff directly involved in the recruitment and interviewing of respondents, and a sample of survey respondents;
- Meetings with key staff from ITS-Leeds to better understand the processes and rationale for data analysis and modelling;
- Specification of a number of model runs to test the reliability of the reported results, and statistical comparison of results for the subset which converged
- Meetings with the DfT project management team to discuss the peer review remit and facilitate information provision between the research and review teams.

2.2.4　The audit of the Implementation Tool addressed seven key objectives:

1. **Model re-run**: This involved the audit team attempting to re-run the provided codes, and investigate whether the tool executes as expected.

2. **Integrity between the approach and report**: This involved the audit team checking whether the input (data files, parameters) and output files (VTT estimates) conform to those presented in the report.

3. **Inspect the overall structure of the code**: This involved the audit team scrutinising the implementation tool R code at a high (pseudo-code) level to investigate whether its operation conforms to the logic and capabilities described in the report.

4. **Clarity of documentation and ease of use**: This involved the audit team inspecting the extent to which the available documentation enables employment of the tool in the future ('ease of use') without consulting the original research team.

5. **Inspection of the assumptions on which the implementation is based and sensitivity of the output to their violation**: The research team make a number of explicit or implicit assumptions concerning the way that the final VTTs are generated.  This element of the audit therefore involved the audit team testing, wherever feasible, whether these assumptions are justifiable on the grounds of the available data and/or wider knowledge and good practice.  In addition, sensitivity of the final output (VTT values) to such assumptions was also inspected.

6. **Inspection of the way in which compatibility issues were handled**: The research team acknowledged the presence of a number of compatibility issues, especially in regards of the way in which variables were reported in the study, and in NTS.  The audit team reviewed how these compatibility issues were resolved, and whether the adopted approaches were justifiable and followed the best practice in the field.

7. **Detailed inspection of the code and routines applied:** The audit team undertook a detailed inspection of the R code, i.e. beyond the high (pseudo-code) level to investigate whether (a) the proposed formulae have been implemented correctly and (b) the R routines have been used appropriately and correctly, and in the spirit of best-practice coding.

2.2.5　For each of these objectives, the audit team developed an explicit set of processes that were applied to the relevant data and codebase.

# 3. THE RESEARCH FRAMEWORK

## 3.1 Overview

3.1.1　This section provides the review team's findings regarding the research team's chosen approach to data collection, processing and modelling; and the rationale underlying the approach.

3.1.2　Anyone designing a VTT survey has to decide how best to engage with a suitable sample of travellers; create an acceptable environment through which they can provide valuable information on VTT; and develop analytical models that satisfactorily fit the collected data.  In

addition to these technical requirements, the research team often has to consider other pragmatic issues, such as affordability and providing the client with value for money; the fact that some travellers simply will not be willing to participate in the research (regardless of incentives offered); and the existence of individual idiosyncrasies that mean that responses to SP exercises are prone to error – in how travellers' perceive the options presented to them, their logic in deciding their preferred option and relaying that information to the interviewer, etc.

3.1.3    The design and analysis of quantitative data collected from a large sample of travellers to determine values of travel time savings and reliability presents considerable challenges, requiring a clear understanding of the type and level of information that is needed in a final report; and the technical and logistical issues that inevitably arise when striving to derive individual, and aggregate, valuations from people who, themselves, do not know their inherent values.  There are complexities to designing a survey that enables travellers to reveal their value of time saving, and analysing the data such that the subsequent valuations accurately reflects the views and values of respondents.

**Review Team Conclusion**

**The VTT survey framework presents such challenges as those described above.  On balance, we believe that the research team developed an approach to data collection, cleaning and modelling that would meet these logistical and technical challenges.**

**However, as noted in the subsequent sections of this chapter, we have concerns over the intended approach in some areas.  In particular, the feasibility of collection of high quality SP data over the telephone as part of the General Traveller SP Survey and the Employers SP Survey; the plausibility of the time/cost trade-off context; and the decision to analyse the SP data in reference dependence form but present the SP alternatives to respondents in alternative comparison form.**

## 3.2    Sample design and Recruitment

*Is the theoretical basis for the research team's approach to sampling and representativeness reasonable, and fit-for-purpose?*

3.2.1    Three surveys were undertaken by the research team:

- a survey of general travellers that included Stated Preference (SP) questions;
- a survey of rail travellers that included Revealed Preference (RP) questions; and
- a survey of employers that included SP questions.

3.2.2    All three surveys adopted quotas as a means of obtaining sufficient sub-samples for each of the target traveller segments.  Quota-based sampling approaches are unlikely to obtain a fully-representative sample as they only ensure representativeness according to the stratifications designed for (for the VTT, quotas were set according to mode x journey purpose; and by region).  The research team has developed an Implementation Tool to weight

the VTT parameters derived from the sampled data to give overall national values, so a fully representative sample was not sought.

3.2.3   The research team has also recognised that covariates, beyond those defined by the quotas, will be deployed in the modelling and are likely to examine whether different types of traveller have different values of time savings compared with others (such as high and low income travellers). We are content with the research team's plan to weight the data to reflect known UK traveller population profiles in order to achieve overall national values of time savings.

> ### Review Team Conclusion
>
> **The review team is satisfied that the approach to sampling adopted by the research team meets the needs of the study, especially bearing in mind that the modelled VTTs are subsequently weighted using NTS data to reflect the UK traveller population.**

> ### Is the theoretical basis for the research team's approach to recruitment and sample size reasonable, and fit-for-purpose?

3.2.4   For the general travellers SP research, a mixed method was used comprising an intercept approach (i.e. contacting people as they travelled) and cold-call recruitment by telephone (i.e. typically contacting people at home to see if they had made an in-scope trip).

3.2.5   The research team usefully explain the rationale for their approach to recruitment. They rejected using a person- or household-based sampling frame, as they identified a bias towards predominantly short distance sample of trips. The Phase 1 Report highlights that, according to NTS, 80% of all trips made are less than 10 miles; and only 2% of trips are 'long distance' (i.e. greater than 50 miles). Yet, most SP and/or RP studies have found a strong relationship between VTT and distance, so it is important to be able to estimate VTT for longer trips accurately. Hence, the research team have adopted an alternative recruitment approach primarily based on intercepting people as they travel – as it means that the sample is skewed in favour of longer distance movements, since these have a higher probability of being intercepted. Recruiting via an intercept approach also benefits the quality of the main survey since there is an obvious trip to focus on as context to the SP exercises.

3.2.6   The research team also explain that they planned to complement the intercept approach by using person- or household-based methods after all – "to strengthen the sample in the shorter distances". The research team suggest that this natural focus on the intercepted trip would make it difficult to ask the respondent to focus on any other trip (within research scope), and the need to get some respondents to think about other trips (that are otherwise difficult to cover) is another reason for the complementary telephone recruitment approach.

3.2.7   The RP survey was recruited solely via an intercept survey on rail routes carefully selected and identified as having competitive choices between specific station origin-destination pairs. Only travellers who indicated that they were familiar with the route and (crucially) the times and costs of the alternative services on offer, were included for interview. We agree with this approach to include only those travellers who are fully aware of the options when deriving RP-based VTT. However, it should be noted that there are some potential issues too. By

screening-in only those that were aware of the different levels of service of the different alternatives, the sample will likely be skewed towards more frequent travellers. Moreover, it becomes a study of market behaviour only where there is 'perfect information'.

3.2.8 We are unclear as to the extent of instruction to recruiters at rail stations regarding whether they had any control over which platform they stood on to recruit passengers of interest (which would, presumably, dictate whether the passengers recruited were using the fast or slow service to/from London).

3.2.9 There is limited information provided by the research team as to why they deemed road-based RP sources such as: the M6 Toll road, Dartford Bridge Toll and the London congestion charge to be inappropriate.

3.2.10 We agree with the research team in their consideration that their target sample of 1,250 observations for business travel is appropriate – on the proviso that the requirements to derive valuations at a more segmented level is limited. The research team's rationale being that this target is higher than is customary for RP surveys (which they estimate at 1000 observations).

3.2.11 The research team chose to recruit for the Employers' Survey by telephoning businesses sourced from a commercial database. The business representative to be recruited was the person within the company who is 'the decision maker with regard to travel policy'. During the recruitment process, this was further clarified as:

> *"the correct person to speak to on behalf of your organisation with regards to any changes that might be made to your company's travel policy relating to new developments in the transport infrastructure, for example when new roads or railway systems are introduced."*

3.2.12 In the view of the peer reviewers, this definition will identify the policy-maker within the company but not necessarily the person who makes a decision as to which train an employee will catch between, say, the three operators running trains between Birmingham and London [Virgin West Coast Mainline vs London Midland vs Chiltern (that runs between Moor St/Snow Hill and Marylebone)]. <u>If</u> the DfT want valuations that will reflect actual decisions in the marketplace then only employers who would make these trip-specific decisions on behalf of their employees should have participated in the survey. This is discussed further under Questionnaire Design and Survey Environment (Section 3.4).

3.2.13 All SP, RP and Employers Survey respondents were offered a financial incentive - £10 voucher or a donation to charity) to complete the interview; and this was increased to £20 to achieve some hard-to-reach quotas. Those offered the higher incentive were typically those making business trips using car or rail. The need for a higher incentive implies that such traveller segments comprise people who have a higher value of time than £30/hour (since they were not willing to give up 20 minutes of their time – the interview length as stated to would-be respondents – for £10). It would have been useful to test whether, within each traveller segment of interest, those receiving a £20 incentive have a higher VTT than those receiving a £10 incentive.

3.2.14 Despite the above complication, we consider that the research team was right to have some form of 'thank you' as a reward for participant's time. Without any incentive, the likelihood is that response rates would have fallen significantly (increased cost) and that the residual sample of respondents would predominantly have a low VTT since (arguably) only people with a fair amount of free time would be prepared to give up half-an-hour for free. Our overall view is that the research team should have adopted non-monetary incentives [e.g. a holiday prize draw] – though we accept there is an still an indirect monetary association with most non-monetary items – and should investigate the effects of the offered financial incentive on responses and results.

**Review Team Conclusion**

**The review team supports the decision to recruit for the general traveller survey via intercept. However, we find the rationale for introducing a second component to the recruitment (i.e. telephone recruitment) less convincing and counter the rationale for intercept recruitment, but it should not adversely affect the representativeness or quality of responses.**

**We are satisfied that the approach to recruitment for the RP survey is fit-for-purpose, though consider some important survey design details to be missing.**

**We have a concern over the recruitment of some of the respondents to the Employers Survey as we consider that the views of some will not reflect decisions in the market-place and, as such, are of theoretical interest only and risk confounding the results when compared with the values expressed by business travellers, and with observations in the market-place (i.e. the RP findings).**

***Has the research team adopted a suitable approach to non-response?***

3.2.15 The non-response rates for the SP survey are not formally reported but are likely to be quite high. (A technical note of refusals and other drop-outs has been provided but the ultimate response rate and non-response rate remains uncertain). It is likely that some potentially in-scope travellers will have declined/ignored the 'on-street' or telephone cold contact by the recruiter. There are also a number of systematic exemptions, such as drivers who tend not to use the main roads/motorways; people who only have a mobile telephone or are ex-directory or on other lists excluding cold contact.

3.2.16 The research team has compared the gender, age and journey purpose profile of the final SP (and RP) sample of completed interviews against the profile of people who were initially recruited but dropped out due to attrition. The differences were small. The number of people dropping out after initially indicating a willingness to participate seems to have been quite high (circa 70% of the initially-willing SP sample and 60% of the initially-willing RP sample by our calculations) and the reasons for this are unclear. This may be important as those dropping out have revealed a relatively low tolerance to continue with the process (and secure their £10 reward) and may have different values of time compared with the final sample, despite having a similar gender and age profile.

3.2.17　It is also a reasonable assumption that many people who expressed unwillingness to participate in the survey would have high values of time.  Other profile information (especially on income) on non-responders would, therefore, have been very useful and would inform subsequent data weighting procedures.

3.2.18　No information has been made available regarding non-respondents to the Employers Survey.

**Review Team Conclusion**

**Little information seems to have been sought regarding in-scope travellers who opted out of participating in the survey; and no attempt at understanding the profile of others who have not been included in the SP, RP and Employers Survey samples.  We cannot agree with the research team's judgement that the values expressed by their survey sample must be representative of the population.  They provide no evidence to support this (only that the sample profile is similar to that of the profile who originally expressed interest in the survey).   Any differences between the survey sample profile and the population of UK travellers will be addressed by the weighting process using NTS – but any differences in valuations between those who found time/inclination to participate in the survey and those who chose not to is unknown, and could be significant.**

## 3.3　SP design

*Has a reasonable approach been taken regarding the design of the SP exercises?*

3.3.1　Efficient designs were used instead of orthogonal design techniques. They offer a benefit of generating better scenarios with more meaningful trade-offs and reducing standard errors by using prior information. The designs produced for the research team are based on Bayesian D-efficient designs.

3.3.2　We appreciate the discussion in Section 2.12.3 about the benefits and potential problems of D-efficient designs, and alternative design approaches.  The suitability of D-efficient designs hinges critically on knowing the true choice model in advance of data collection.  This is particularly a problem for SP2 where there are still unresolved issues about how respondents interpret and process the information about uncertain travel times.

3.3.3　It remains unclear why the SP designs were optimised for additive models when multiplicative models were planned.  Whilst it is true that behaviours do not depend on the model used to describe it, we do not find the discussion in page 35 to be sufficiently convincing.

**Review Team Conclusion**

**The review team is satisfied with the design from the information received, though we would have expected greater consistency between SP design, model estimation and change in travel time for which national VTTs were ultimately derived.**

| Value of Travel Time Savings | |
| --- | --- |
| Value of Travel Time Savings - Peer Review & Audit | 103217 |
| Final Report | 30/07/2015 |

**Page** 17/48

## 3.4 Nature of Data Collected and the Survey Environment

*Is the environment in which the survey is carried out, and the means of responding to the survey questions, fit-for-purpose?*

3.4.1 The research team decided to adopt a dual approach to surveying general travellers – offering the option of a CATI survey or an online survey to all willing participants. The main reasons for this mixed mode of data collection seems to be for respondent convenience, and to provide a mechanism for response from people who do not have access to online facilities.

3.4.2 Using different modes of data collection increases the risk of measurement differences. This particularly applies when combining online and telephone responses due to the fundamentally different forms of question and answer presentations. That is, the telephone respondents have the benefit of having the opportunity to be assisted by the interviewer but have the disadvantage of, generally, not being able to see the questions and answer lists. Overall though, we consider that the disadvantage of mixed mode is outweighed by the advantage of being more inclusive. That is, by offering more than one option to participate, the research team ensured that those with limited/no internet access (who are likely also to have lower incomes and/or live more remotely, and may have lower values of time) are able to respond; and hence help to increase response rates and representativeness.

3.4.3 However we have reservations about the nature of the CATI interview because of the complexity involved in ensuring that each respondent can 'see' the correct SP choice-set when providing their response to each SP choice-set. Potentially, the undertaking of three different SP exercises – overlapping in much of the context and variables – could make the interview quite complicated, repetitive or tedious.

3.4.4 The telephone survey required customised SP show material to be sent part-way through the interview, by email or post, and for the interview to be completed once the respondent had received the materials. Further explanation from the research team suggested that they had considered trying to avoid the two-part approach to the CATI survey but could not see how to do so – as, in their view, it would have taken too much time at the recruitment stage to have collected all the necessary information. However, respondents who have no internet access, or could not gain access to their emails when they received the phone call from the interviewer, would have been unable to complete the interview in one go. A two-part main interview like this risked problems with attrition rates - as respondents might lose interest in the exercise part-way through if they have to do the second part several days after doing the first part - and because some might end up doing the interview with the interviewer reading-out the SP options over the phone (to avoid delaying the second part of the interview). We discuss how the research team implemented this aspect of the interview, and the implications, in the next chapter (Section 4.4), which highlights the complexity of the SP survey and, therefore, the inappropriateness of administering it by CATI.

3.4.5 Administering the survey online has the benefit of immediate customising of SP choices, but has no interviewer interaction. This has two potential disadvantages – first, that any respondent finding it difficult to understand how to interpret the different SP options has no-one to ask and will be forced to guess answers; and second, there is nobody else present to ensure that the respondent concentrates during the interview. [Note, it is possible that some respondents may have considered the completion of the interview to be a way of securing

easy money and rush through the interview]. The interview duration for completing the online (and CATI) interviews has been recorded, but was not used to check or clean the data in any way. Our review of the data showed that some respondents took less than 5 minutes to complete the survey, that comprised 15 separate SP choice-sets as well numerous other questions before and after the three SP exercises.

3.4.6   Overall, we are surprised that the SP interview survey with general travellers has been conducted via online and CATI methods. Such administration channels are not recommended in most Best Practice guides unless budgets are severely restricted, indeed guidance for conducting SP Willingness to Pay Surveys[4] in the water industry states in its recommendations that SP surveys should be *"delivered to respondents under interviewer guidance",* but four in every five respondents (i.e. the online survey respondents) had no such interviewer guidance. [The above referenced recommendations go on to say that *"online surveys could also be considered"* – arguably leaving pretty much all options still available, though we know that water companies stopped proceeding with their planned online surveys once this 'guidance' became available]. Online surveys are usually adopted when a large quantity of data is required (such as with this study) and when the technically superior face-to-face approach would have significant cost and time implications. Hence, we understand the possible reasoning for adopting online methods but consider the complementary CATI survey to be inappropriate, especially given the problems in the field (as detailed in the following chapter).

3.4.7   It will have been important to have conducted pilot and cognitive interviews, and we note that the research team conducted a series of such interviews until, presumably, they were satisfied that the survey could be answered by the vast majority of respondents. However, the qualitative findings highlighted some problems with the perceived implausibility of the cost/time trade-off context but there is little indication that the research team subsequently addressed the issue – an issue that could undermine the usability of some of the collected data. The research team needed to demonstrate (first and foremost to themselves) that respondents thought solely about time and cost differences (with no externalities) when trading between SP options within the context actually used in the Main SP interview. Despite the report of the 4-month qualitative phase of work covering a range of different theoretical SP issues and trialled SP contexts, it fails to give clear cognitive evidence as to whether respondents associated the presented contexts (of fuel price changes, different levels of congestion, breakdowns, etc) when comparing one SP option with another. The Phase 1 Report (and Appendix A) also fails to explain why the research team seemingly tried to incorporate such externalities into the exercise. It should be noted that this is not the first VTT study to adopt time and fuel costs as the main SP variables, but a different context will often have applied and, presumably, cognitive testing will have confirmed the basis in which respondents traded-off options. How the SP exercise was implemented in the field is discussed in Section 4.4.

3.4.8   Almost all RP survey data was collected via online means (2% captured by phone). Given the information sought was of a predominantly factual nature and based on recall of a recent trip (rather than the subjective decision-making form of data that applies to SP surveys), we consider the method of survey administration to be less crucial since most forms would be

---

[4] "Carrying out Willingness to Pay Surveys", UKWIR (2011) NERA in association with Accent.

satisfactory. There is a risk that the attempts to clarify the RP time versus cost trade-offs (described in pages 24-25) may have amounted to 'leading the witness' in some cases.

3.4.9 The Employers Survey was conducted over the telephone. For reasons given above, we are concerned about having to assume that the respondent was always looking at the appropriate showcard etc, but we think the risk is lower for a survey with business representatives (who are likely to have online facilities at their fingertips, and a computer screen easily accessible when they received the call from the interviewer) compared with interviewing the general public.

3.4.10 All Employers Survey respondents will be able to respond to the SP exercises by indicating, for each choice-set presented, the option that they think is best for their business, hopefully taking into account: productive time for the employee, the cost to the business and other factors of relevance to the business' operation and cost. The implication is that using this methodology, the research team have valued the pure (but 'theoretical) employer value of time, that does not take into account the employee's welfare. Though of research interest, this is unlikely to be what is wanted for appraisal, and is not going to be consistent with RP data (i.e. actual behaviour in the market).

3.4.11 Such responses are likely to **differ**, sometimes, from the preferred responses of the business-travelling employee to the same SP choice-sets, as the latter are likely to consider (to at least some extent) impacts on their personal time as well as: productive time for the employee, the cost to the business and other factors of relevance to the business' operation.

3.4.12 There are differences in VTT estimates based on responses from employers and those from employees travelling on business. We do not consider this finding surprising, and make further observations and recommendations on the issue in the next chapter (Section 4.4).

3.4.13 If the VTT estimate to be used for appraisal is to reflect true market decision-making, the modelling should be based on the responses of the individuals who determine actual travel behaviour. It is likely that only some of those who participated in the Employers Survey will actually have such influence. As described in 3.2.14, to be in-scope for the survey, the business representative need only have confirmed that they have some role in their company's travel policy. Though some business representatives have influence over employee travel policy, they will not necessarily have sufficient control over their employees' specific travel decisions, many of whom are likely to act autonomously. Therefore, responses of some respondents to the Employers Survey will not reflect actual behaviour and so should be omitted from the evaluation. Similarly, there will be some respondents to the general traveller SP who will have responded in the context of an employers' business trip but who will not actually have the autonomy to make the travel decision (it will be made by their employer). These responses should also be excluded from the estimation - if the desired VTT is to reflect market behaviour.

**Review Team Conclusion**

**We have some concerns over the decision to conduct the general traveller SP survey via telephone with a minority of respondents. We appreciate that it is a relatively cost-effective interviewer-administered survey approach but there are quality issues associated with it – in particular, a high likelihood of logistical difficulties when ensuring**

that respondents see each SP choice-set in timely fashion. As we report in the following chapter, there were indeed problems with the implementation of the two-stage CATI interview.

For such an important study as this, we are also surprised that an online interview has been conducted with the majority of general traveller SP survey respondents; but we recognise the approach offers good value for money; and has been the main approach to data collection for many VTT studies overseas. The research team rely on the general public's ability to comprehend the three SP exercises – without assistance from an interviewer - and continue to be motivated to concentrate throughout the 20-45 minute survey. We are reassured, to some extent, by the level of qualitative research and testing that has been reported as part of Phase 1 which, presumably, confirmed to the research team that the design, and survey channel, was workable for all types of respondent. However, the Phase 1 Report does not focus on the main issue of checking the cognitive basis for how respondents chose between the different time/cost options and whether they considered the option of a quicker but more expensive car journey to be plausible [this is discussed in more detail in the implementation Chapter 4]. We consider that more data quality checks could, and should, have been conducted; and that face-to-face interviewing would have led to far better quality of data.

We are satisfied that the chosen survey environment is suitable for both the RP and Employers Surveys. We are aware of the differences in reported VTT for employers business between those derived from employers' SP responses and those from travellers on employers business. We do have concerns on the nature of the Employers Survey sample and how such values are to be interpreted.


## 3.5 Data Cleaning & Processing

*Is the planned approach to data processing satisfactory?*

3.5.1 The research team has included a number of questions or controls to help with data quality checking post-data collection, including: upper and lower limits of values of time; interview duration; and questions for the respondent to self-report the ease, or difficulty, they had in answering each SP section.

3.5.2 The research team planned to undertake most of the main data checks that we would expect, with reasonable justification for all remedial actions by the research team.

3.5.3 We comment on the procedures undertaken by the research team in the next chapter (Section 4.5).

**Review Team Conclusion**

**We are satisfied that the research team planned to undertake the necessary data checking.**

## 3.6 SP Analysis & Modelling

*Is the planned approach to SP data modelling satisfactory?*

3.6.1 There are three elements to consider: the treatment of the underlying micro-economic theory informing the analysis of VTT; the econometric implementation of the micro-economic theory; and alternative models considered.

*Treatment of the underlying micro-economic theory informing the analysis of VTT*

3.6.2 The final report contains some discussion of the micro-economic foundations of VTT for non-work and business purposes (Appendix A) but there is insufficient explanation for why alternative approaches have been rejected. [However, we understand from DfT that the exploration of different theoretical approaches may have been specifically excluded from the project remit]. The need for clarity in respect of micro-economic considerations is perhaps most important in respect of the valuation of business travel time, where a number of strongly divergent theoretical frameworks exist and where substantial new challenges, such as the productive use of travel time, are emerging. We could not find a clear presentation of the relative performance (in terms of goodness-of-fit and diagnostics, not just VTT) of the different theoretical approaches.

*The econometric implementation of the micro-economic theory*

3.6.3 The main emphasis of the modelling work has been placed on econometric implementation. The research team shows excellent awareness and understanding of both the recent VTT literature and more widely of underlying econometric technique. The work is characterised by the research team's evident desire to push the boundaries of the state of the art (e.g. multiplicative model forms, WTP-space analysis, reference dependence, treatment of uncertainty etc.).

3.6.4 Whilst such research ambitions are in general terms laudable, some members of the review team have doubts as to whether this research project was an appropriate vehicle through which to seek to fulfil all these ambitions. The DfT client, the research team and indeed, the peer review team, share the view that the combination of the demanding scope of the DfT's requirements (in terms of the range and granularity of appraisal VTTs sought) and the aggressive timescale for the work resulted in an extremely challenging agenda. We believe that the research team may have given insufficient consideration to the implications of the compounding of the unavoidable challenges of scope and timescale with the additional challenges associated with seeking to explore more sophisticated approaches to model specification. There is evidence that the resulting emphasis on elaborate econometric specification has been at the expense of consideration of more fundamental theoretical and empirical issues.

3.6.5 One of the aspects of the econometric modelling in which these issues are most apparent is the representation of reference dependence. In the context of SP work, reference dependence can in principle arise from two quite distinct sources – as an artefact of the form of the SP exercise or as a reflection of a real phenomenon present in actual behaviour. This distinction is important because the implications of each for VTT estimation are quite different. In the case of the former, the question is one of purging these effects from the

estimation results (similar to the treatment of other SP artefacts such as layout or fatigue) whereas, in the case of the latter, the question is how best to accommodate the effects in model specification and subsequent analysis. In practice, of course, both effects might be present so there is a question of disentangling the two.

3.6.6    It seems to us to be good practice that if one takes the view that it is necessary to explore for the possibility of reference dependence, one should take steps to gain insight into what type – artefactual or substantive – one is dealing with. From this perspective, we find it surprising that there is no evidence that the qualitative work undertaken as a pre-cursor to the SP exercises sought to explore this issue, in particular whether evidence existed of substantive reference dependence and if so, of what form and whether particular forms of SP presentation might induce artefactual reference dependence. The psychological literature explores these issues at great length and makes apparent that substantive reference dependence can be complex and context dependent (sometimes the reference point might indeed be a chosen alternative, sometimes it is an aspired-to ideal alternative, sometimes it is a worst case alternative, sometimes it is an alternative chosen by others…etc.). It is, therefore, disappointing that the research team chose neither to explore the issue in the qualitative work nor to explore alternatives in the quantitative work (e.g. best-worst would have been quite easy to implement). One consequence of this is that it remains ambiguous whether the reference dependence effects detected are indeed artefactual or substantive (although in the research team's subsequent calculation of appraisal VTTs, the latter interpretation is implicitly used).

3.6.7    Only one form of reference dependence is examined – changes relative to the characteristics of the current trip – and the only testing that is done is between this form of reference dependence and none; there is no testing of alternative potential specifications of reference dependence effects. The SP exercises appear to have been presented to respondents in alternative comparison form with no explicit presentation of the assumed reference point (which respondents are, presumably, expected to recall from memory).

3.6.8    In the final report, the study team seek to justify their SP approach to reference dependence in terms of time pressure and "current best practice". The latter is a reference to a number of recent Scandinavian VTT studies that have adopted a similar approach. These studies are all exemplary pieces of work but in science, terms such as "current best practice" are usually understood to be attached practices that have demonstrated superiority over a number of years and contexts and are accepted as such by the overwhelming majority of a relevant academic (and where relevant practitioner) community. We are not aware that this is the case here. What is undoubtedly the case is that the recent Scandinavian studies (which have been in a sense a template for the current study) are by some distance the most econometrically sophisticated approaches to VTT estimation currently available in the literature. This, however, is not necessarily the same as best practice.

3.6.9    A consequence of the use of a reference dependent specification is that it significantly complicates the derivation of VTTs. This is especially the case given that the research team use a value function specification that accounts for sign, (asymmetric) curvature and scale effects. This means that (assuming, as the research team implicitly seem to have done, that the reference dependence are entirely substantive) the VTT is context specific; in particular, depending on the sign and magnitude of the changes travel time and travel cost relative to the reference levels of travel time and travel cost. This presents major complexities from the

point of view of practical implementation. In essence, the choice of an econometric model specification that makes the VTTs choice context dependent implies that, in application, future choice contexts need to be predicted up to and including future reference points and the magnitudes of the signed changes in attributes relative to these reference points. So, strictly speaking, application of the econometric results to the estimation of appraisal VTTs values, effectively involves a second modelling step and one for which no established methodology exists.

3.6.10 This forces a series of quite aggressive simplifications in the application stage. From the point of view of the estimation of final appraisal values, this is a classic case of trading off errors in different stages of a modelling exercise, in this case errors in estimation *vs* errors in application. It appears that the nature of this fundamental trade off was not explicitly considered by the research team. Instead, their implicit strategy seems to have been to estimate the most comprehensive econometric specific consistent with the SP data and then to make a series of simplifying assumptions (several of which are necessarily rather *ad hoc* in nature) to recover from these models estimates of reference free VTTs. Whilst it is possible that this is the optimal strategy for the estimation of VTTs, it is very far from being axiomatically the case. We are surprised that there appears to have been no Monte Carlo analysis undertaken to investigate the implications of simplification at different stages in the analysis chain. We do not think that "no time" and "best practice" constitute adequate justification for not investigating this issue more fully.

3.6.11 An example of the issues of this sort that arise relate to the treatment of asymmetric scale effects in the value function. The research team puts forward an argument on p108 (based on the geometric mean) to the effect that the asymmetric curvature parameters $\eta$ and $\gamma$ can in fact be ignored in the computation of appraisal VTTs, because they drop out in averaging across gains and losses. This argument, based on the dBF paper, however, is not developed as a general result but rather is simply illustrated using the special case of identical but opposite signed attribute changes. The 'dropping out' argument therefore appears to depend on the assumption that the unknown future changes in attributes, considered relative to unknown future reference points will be such that, on average the magnitude of gains and losses are similar. This is not impossible but neither is it anything more than a rather strong *ad hoc* assumption, and one that is inherently very difficult to verify, even *ex post*.

3.6.12 Essentially similar issues arise in Chapter 5 and Chapter 7 in the context of the discussion of scale dependence in the value function implying the need to fix $\Delta t$ for the estimation of appraisal VTTs. The arguments made in p208-210 are entirely plausible but are again essentially heuristic, and contrast very markedly with the theoretically driven approach used in the estimation stages of the study. Reference is made to investigations being carried out into the sensitivity of final results to this assumption but this work does not appear to constitute the type of full scale Monte Carlo design exercise we would have expected to have been undertaken.

3.6.13 Although the issues associated with reference dependence dominate the econometric work in the report, there are in fact even more challenging issues associated with the treatment of uncertainty. One source of concern is the relationship between the presentation of the five travel times to respondents and the modelling assumptions made in the analysis of these data. As far as we can see from the SP interview scripts included in the Appendix, respondents

were given no information regarding what probability they should attach to each of the travel times shown. Any uncertain choice situation comprises outcomes and associated probabilities. Just presenting outcomes does not adequately characterise the uncertain choice. There is a vast body of literature that indicates that if respondents are not informed regarding probabilities, they are prone to infer all manner of spurious patterns ("…it seems to be getting worse…") and/or make heuristic assumptions (e.g., good travel times are less likely than bad ones) which may be at variants with these intended or assumed by the analyst. In this case, the analysis seems to be based on the assumption that each of the five travel times was equally likely, which is a simple message that could have been conveyed to respondents during the SP, but it seems was not. This casts doubt on the validity of this part of the analysis (i.e. SP2).

3.6.14 The standard deviation is used as measure of travel time variation (TTV). This has a microfoundation in alpha-beta-gamma preferences (Fosgerau & Karlstrom). The research team has used the standard deviation as a measure of TTV but with no justification for doing so instead of adopting a factor that accounts for the shape of the travel time distribution.

3.6.15 Given there are still many unresolved issues about how respondents interpret and process the information about uncertain travel times in SP choices, we are not fully convinced that it is suitable to include SP2 within a combined analysis of the SP exercises (though we appreciate why it is desirable to do so). Evidence from both Swedish and Norwegian studies suggests that preferences inferred from choices with and without uncertainty may be inherently different, and we do not yet understand fully what drives people's choices. So we are concerned about how large the effect is of SP2 on the overall results (through the common covariate parameters). We did not find a test of the parameter restrictions entailed in merging the three models. It would have been easy to perform a likelihood ratio test of the joint model against separate models.

3.6.16 We did not find a presentation of the range of VTT offered to respondents. This is critical information for the purpose of evaluating the estimates of VTT distributions. Specifically, it is important to know whether final results rely on VTT values outside the range presented in the SPs.

3.6.17 On the positive side, the rate of time and cost non-trading is low.

3.6.18 Though consistent with the NTS data used for weighting, using gross annual income (as reported in Table 3.9) is problematic for the estimation of income elasticities, since it is net income that is relevant for behaviour.

3.6.19 Although we list above a number of concerns and uncertainties, we consider most of the theoretical basis for the final model to be sound. An impressive amount of work has been done to fit the three SP games into one single model, which allows a single individual specific base VTT to enter into the three models. The single individual specific base VTT varies with observed characteristics of respondents and trips, and includes a random component.

*Alternative Models Considered*

3.6.20 In Chapter 4, comparisons are presented between additive and multiplicative model forms for SP1-3 for the case without covariates but similar comparative results are not presented

for models involving covariates or for models involving more complex functional transformations, such as those introduced to represent reference dependence.

3.6.21 Some members of the review team have concerns with the derivation of the final best specification for the following reasons:

(a) Part of the multiplicative effect comes via scaling and this could also enter via distance, time, cost covariates, in either V or ε or both.

(b) A further question that could have been addressed is whether reference dependence is the true underlying mechanism or not – e.g. is the distance/time scaling driven indirectly by reference dependence or directly by distance/time?

3.6.22 In Section 4.3.4, the research team justify the use of the multiplicative model on the basis of goodness of fit, and this rationale raises a subtle question. Whilst we have no fundamental objection to the multiplicative model per se (it is in essence just another form of mixture model), in the context of VTTS estimation, the objective of the models is measurement (of VTTS) not prediction of behaviour - hence accuracy and precision in VTTS should be the key criteria.

**Review Team Conclusion**

**The review team is satisfied with many elements of the initial theoretical framework and design. We can confirm that the research team has demonstrated considerable effort and skill in combining the different data and in considering different approaches to modelling and interpreting the data. Though the approach adopted by the research team is essentially sound, we have a concern that there is an inconsistency between assumptions made at the design stage and in application that focus on reference dependence.**

## 3.7 RP Analysis & Modelling

### *Is the planned approach to RP data modelling satisfactory?*

3.7.1 Overall, the description of the RP data analysis reveals some serious problems:

- It is a well-known disadvantage of RP data that it is often impossible to identify effects separately, because data does not display independent variation of different effects. As an example, the research team mentions that one operator is always cheaper while another is faster. If this is the case, it does not make sense to include operator dummies in the model, as they cannot be identified separately from the VTT. So we suggest excluding operator dummies from the models and checking if the resulting estimates are comparable between RP and SP.

- The fact that the SP design (SP3a operator choice model used to compare with the RP model) provides a poor explanation because attribute values were designed to mimic real world conditions rather than display independent variation of different factors

deserves to be highlighted as a lesson we can learn from the study. There is an ongoing debate among practitioners whether SP designs should prioritise realistic scenarios over sufficient attribute variation, and the research team's experience here would provide a contribution in that regard.

○ We remain unconvinced by the supportive evidence described in Section 8.1, which reports a validation of the SP analysis by means of the RP data analysis. Since the validation consists of comparing two models where the VTT is poorly identified due to confounding with the operator dummies, we are not fully convinced that the results serve as a validation of the SP data.

3.7.2 We are also unclear as to how the collected RP data was intended to be treated, in terms of the representativeness of the data and any need for weighting to reflect population flows along the rail corridors of interest. Given the problems encountered in both the RP and operator choice SP (which provides the closest comparison for validation purposes, both in terms of sample and choice context) we note the limited scope for the RP to validate the SP results. We are, at least a little, encouraged that the "very sure" RP responses are comparable to those from the SP.

**Review Team Conclusion**

**We believe that the research team's intentions to developing an RP approach to validate the SP work was reasonable, but we raise some technical questions regarding their treatment of the data and use of operator dummies in the estimation.**

## 3.8 The Implementation Tool and Deriving National Values

3.8.1 The procedure of obtaining appraisal values from the estimation results is very well documented in Section 7 and Appendix I. Many important issues that are crucial for the computation of appraisal values are openly discussed and laid out – indeed many choices in the process are policy decisions that should be taken by the DfT.

**Review Team Conclusion**

**The review team agrees with the research team's intention to weight the SP derived parameters according to the NTS-based national profile.**

# 4. IMPLEMENTATION

## 4.1 Overview

4.1.1 Anyone implementing a VTT survey has to contend with a range of issues in the field and back at HQ. In this section, the review team reports its findings on how the research team has

conducted the data collection and processing.  In particular, we have assessed whether the research team has implemented the survey in the way that was originally intended, and reported.

**Review Team Conclusion**

**Overall, we are satisfied with the way the research team has implemented most of their planned programme of surveys and modelling – whilst noting that the review team has not been provided with the full code/syntax used in modelling that would have provided essential insights into how the final modelling has been carried out.**

**Whilst there is evidence that respondents reported that they found the SP exercise realistic, there is little evidence that respondents traded-off time and cost without also considering external, emotive factors that could have confounded the cost/time trade-off.  The research team has not explored the possible impact of these externalities, that they deliberately introduced into the context of the SP exercises, on the derived VTTs. Some SP responses (possibly only a minority) were collected without the use of any visual aids, and we think these responses should have been excluded from the model estimation.**

**The modelling approaches adopted are of an advanced technical nature, but we have some reservations about the disjoin between SP design, presentation in the field, and model specification.  We would have liked to have seen further segmentation of the Employers Survey data according to level of influence over actual travel on employers' business.**

**The Implementation Tool applies the assumed formulae correctly and provides the desired capabilities.  However, the confidence intervals around the VTTs are large, making reliable appraisal difficult; it relies on simplistic imputation procedures which may influence the final estimates; and the evidence for some recommendations is not convincing (e.g. the claim that $\Delta T=10$ mins is the range for which the reported appraisal values are most reliable/suitable).**

## 4.2      Sampling & Recruitment

*Did the research team implement the recruitment approach satisfactorily?*

4.2.1      The recruitment of the SP survey with the general public was administered face-to-face using CAPI.  Interviewers approached adults and asked scoping questions to check whether the respondent was in scope and matched required quotas.  If in scope, the respondent was invited to undertake a follow-up survey online or by telephone.  Those providing email addresses were sent an email with a unique web-link to the survey at the end of the shift. The names and telephone numbers of those preferring to undertake the interview by telephone were loaded into Accent's telephone unit sample on a daily basis.

4.2.2      The intercept locations covered car, rail, bus and 'other' PT users across the country.  The survey locations selected reflect a reasonable geographical spread across England and Wales. The recruitment sites were Motorway and A-road service stations and petrol stations in towns

or in local high streets nearby for car; and bus, Metro and tram stops and rail stations for public transport users. Interviewers approached 1 in *n* adults to reduce the chance of recruiting people from the same travel group, and to increase the chances recruiting a varied distribution of travellers across time segments.

4.2.3 All 'on-street' intercept fieldwork took place on weekdays with fieldwork shifts either 07:00-13:00 or 13:00-19:00, reflecting the TASM's focus on typical weekday effects for transport appraisals.

4.2.4 We understand that the '1 in *n*' approach to initial contact was generally applied consistently. Mostly, contact was with 1 in every 3 people but, for potential employees, this was 1 in 1. Given the challenging nature of securing willing participation from a large sub-sample of such travellers, and the low risk of interaction/association with business travellers, we think this a reasonable approach (assuming that only one person from a travel group was recruited).

4.2.5 The telephone recruitment for the general travellers survey was done via random digit dialling of a commercial database of domestic household telephone numbers, that was broadly geographically representative by region. The telephone recruitment was undertaken between 2pm-9pm Monday – Friday; between 10am – 6pm on Saturday; and between 11am – 7pm on Sunday. Adult respondents were contacted and screened using a recruitment questionnaire and, if in scope, they were invited to participate in the research either online or through an interviewer-administered telephone interview. The former were sent a web-link to the customised survey using the same email invite as for the intercept recruitment approach.

4.2.6 All recruitment for the RP survey was conducted, as planned, at seven rail stations intercepting people as they travelled between the station (one of Birmingham New Street; Birmingham Moor Street; Birmingham Snow Hill; Stoke; Stafford; Rugby; Peterborough) and London. All in-scope travellers had to be making a rail journey originating at the surveyed station and ending at the London terminus station; and familiar with the times and costs of alternative operators.

4.2.7 Recruitment for the Employers Survey was undertaken 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.

4.2.8 Employers Survey respondents were asked to respond to an SP exercise, involving different journeys with varying times and costs, in the context of a recent trip made by an employee travelling on business:

> *"We would now like you to think about the #LEG# part of the #MODE# trip your employee made … and indicate what you think they should have done if some of the travel conditions changed"*

4.2.9 Hence, if the desired VTT estimation is to reflect the values of actual decision-makers in the real-world, and likely actual behaviour, then we consider that the research team should obtain the values direct from the actual decision-makers. In some organisations, this might be a senior policy-maker within the company who can dictate which rail service their

colleagues will use, etc; but in other cases the decision will ultimately remain with the business traveller.

4.2.10 In order to derive the 'true' business VTT, we recommend that a weighted composite value is derived from combining true decision-makers only. From the Employers Survey sample, this might only include the sub-sample that have indicated (from Q1b) that they 'make most/all decisions regarding company travel' and 'I arrange travel' or 'I authorise travel decisions'. Our analysis of the raw Employers Survey data suggests that around 15% of respondents fall into this category. They represent a very small sample in this study (circa 60 respondents) which indicates the need for a better sample design criteria for employers business in future studies on VTT.

**Review Team Conclusion**

**The review team is satisfied with the research team's implementation of the sampling and recruitment strategy overall.**

*Is the achieved final sample size, and breakdown, satisfactory?*

4.2.11 A sample of 8,623 SP interviews were undertaken with general travellers, exceeding the target of 8,500.

4.2.12 There was a shortfall of data for some segments (e.g. employee's business for car and 'other' PT, and commute for bus), but we agree with the research team that these shortfalls (with sub-samples sizes of at least N=265) are unlikely to significantly compromise the modelling analysis.

4.2.13 As reported in Section 3.2.2 of the Final Report, a sample of 2,646 RP interviews were obtained with rail travellers, exceeding the target of 2,500, and split fairly evenly across employers' business / non- employers' business. However, Table 5.3 suggests a usable final sample of 725 RP responses (discarding almost 3,000 responses from the original raw data file, suggesting a considerable amount of wasted effort) which raises a significant concern over the suitability of the collected RP data and its use for 'validation' purposes.

4.2.14 The Employers' Survey sample comprised 400 interviews.

**Review Team Conclusion**

**The review team is satisfied with the overall sample size and sub-samples.**

## 4.3 SP Design

*Is the implemented SP design satisfactory?*

4.3.1 We are content with the D-efficient designs and the way they have been implemented. In principle, efficient designs demand complete knowledge of not just coefficients, but also of

Value of Travel Time Savings

| Value of Travel Time Savings - Peer Review & Audit | 103217 | |
| --- | --- | --- |
| Final Report | 30/07/2015 | **Page** 30/48 |

the model: If wrong assumptions are made regarding the appropriate choice model, the resulting design is not optimal, and it may even do a much worse job identifying valuations than a simple orthogonal design.

**Review Team Conclusion**

**The review team is satisfied with the application of the SP design.**

## 4.4      Nature of Data Collected and the Survey Environment

***Was the context to the SP experiments considered plausible and meaningful by respondents?***

4.4.1      This is a rather fundamental issue which could have far-reaching implications for the derived values if respondents had considered the experiments to be unrealistic or only realistic if they assumed other extenuating (ie external) circumstances also applied over and above the journey time and cost differences (e.g. petrol price policies or breakdowns only applying to some SP options and not others).

*4.4.2*      The SP1 experiment comprises a series of binary choices, with each option described in just two variables – time and cost.  Each choice-set basically consists of one option being both quicker and more expensive (in terms of fuel, for car users) than the other option.  However, *the implicit correlation between travel in slow, congested traffic and fuel economy means that such a situation may be difficult to believe for some respondents.*

4.4.3      The research team's qualitative research report only provided insight into respondents'' cognitive understanding for trialled SP contexts (but not, as far as we can see, the actual context that the research team proceeded with in the Main Survey).  However, the qualitative report indicates that it was difficult for some respondents to understand the justification for a car journey that was quicker and more expensive – e.g. as reported in research team's Phase 1 Report, Appendix A (within a route-choice SP context):

> *"I'm not too sure here why there would be such a high fuel premium."*

> *"It seems illogical. Motorways tend to be a fixed speed, no slowing down, stopping, starting. I'd expect to use less fuel on a motorway."*

4.4.4      Presumably, to overcome this potential problem, the research team gave respondents a reason for why the time and cost may vary, as part of the SP context adopted for the Main Survey, as follows [for SP1].

- *"The one-way travel time may be different because of changes in congestion"*
- *"The one-way travel cost may be different because of a change in the cost of fuel"*

4.4.5      The first bullet explains that the slower of the two journeys is more affected by slow moving traffic.  For there to be useful trade-off, the slower option must also be the cheaper option, and the second bullet states that it is cheaper because of a change in the cost of fuel (ie in unit petrol prices or government tax levels on petrol).

4.4.6 It appears that the research team has tried to introduce a disconnect between the two variables by introducing the concept of different unit fuel charges – a policy that many motorists would be delighted by and could confound the time/cost relationship that is the intended focus of the study. So, it may be that some, or all, of the 514 car respondents (by our calculation, and 17% of the car sample) who always chose the 'cheapest' option across the five choice-sets that they were presented with, did so because they supported the implicit idea of reduced fuel charges rather than because they have a low value of time.

4.4.7 Alternatively, some respondents may have felt that the costs were unrealistic and that, given that the two options are for the same journey, the quicker option would also likely be the cheaper option. So, it may be that some, or all, of the 404 car respondents (13% of the car sample) who always chose the 'quickest' option across the five choice-sets that they were presented with, did so because they thought the stated costs were implausible.

4.4.8 The research team have pointed out that there is an inherent plausibility issue with SP1 for car which is difficult to overcome. However, there appears to have been no consideration of using other car cost variables that do not have an association with journey time, such as toll levels and parking charges. The vast majority of respondents self-reported that they understood each of the SP exercises (83% - SP1, 77% - SP2, 84% - SP3) – and this we find to be reassuring (though the responses to the diagnostic questions were excluded from the raw data that we received for review, so we have not been able to verify or investigate the diagnostic information).

4.4.9 Similarly, with regards to the context-setting for SP2, we are concerned about the introduction of significant externalities as a means of 'explaining' how variations in time may change, that may confound the intended time/cost relationship. These include "breakdowns", "unplanned roadworks" and "traffic control" – all pretty unpleasant scenarios – compared with the fourth reason "general traffic". We do not know whether respondents who have chosen the option with the shorter journey were simply willing to pay the extra cost in order to experience a shorter journey OR have chosen to pay the extra cost in order to avoid being affected by a breakdown or unplanned roadworks. It would be interesting to test for a fixed penalty for an increase in journey time (to reflect the onset of roadworks or breakdowns) as well as a continuous element for extent of delay/unreliability.

4.4.10 Unfortunately, the Phase 1 report (and related cognitive work) makes little reference to this plausibility issue. Moreover, there is no indication that the research team is aware of the rationale assumed by respondents when choosing between the different journey options.

4.4.11 We also have a concern over the reference times and costs for the SP experiments. Each experiment was customised to each respondent's individual journey details.

> Q16  At what time did you start the journey from … [origin]?
> Q17  And at what time did you reach … [destination][5]?
> Q18  About how long did you spend …?

---

[5] In our testing of the online survey script, the appended words on Q17 for non work trips were, incorrectly, *"… at your workplace"*, further undermining confidence in the SP context. However, the research team have confirmed that they are confident that Q17 always showed the relevant text in the main survey.

Value of Travel Time Savings

| Value of Travel Time Savings - Peer Review & Audit | 103217 | |
| --- | --- | --- |
| Final Report | 30/07/2015 | **Page** 32/48 |

4.4.12  The answer to Q18 was used as the basis for the journey times presented in the SP experiments and in the estimate of the reference (fuel) cost.  The concern we have is that the question seems to be ambiguously, and sometimes poorly, worded.   It is not clear to the review team, and possibly unclear to the respondent, why Q18 was, effectively, a repeat of information already implicitly collected.

4.4.13  The intention was for Q18 to be extended with phrases like:  "… in the car" but this does not always seem to have happened (e.g. for car non-work, when we test the online questionnaire).  In these cases, it is not clear what respondents will have been thinking of – possibly the duration at the destination rather than the duration of the journey (which they have already covered by giving the journey start and end times].  For those that were presented with the full question, it is quite possible that some respondents will have given the total time spent in the car (i.e. outward and return).

4.4.14  If the journey duration implied by questions Q16 and Q17 was the same as that recorded in Q18, our concern would disappear.  For a significant minority, however, this does not seem to be the case.  For non-work car respondents we have analysed the journey durations according to Q16 and Q17 [ie. the difference in time] and compared it with Q18.  In 62% of cases, the duration was the same.  For 38% of respondents, the stated duration was significantly different from the difference between the start and end time.  In more than 10% of occasions, the value stated in Q18 (and the basis of the customised SP exercise) was less than half of that inferred from the start and end times (which, we suspect, is the correct answer based on a clearer pair of questions).  It would be interesting to see if the derived parameters are different for the 38% sub-sample compared with the 62% sub-sample.

**Review Team Conclusion**

**The review team has some concerns with the context to the SP experiments.  There has been no demonstration that respondents understood the implicit (not explicit) presentation of the time and cost trade-offs; nor the presentation of probabilities in SP2.   Without this evidence, we remain concerned that respondents may have considered other factors (in addition to time and cost differences – such as fuel price policies, risk of breakdowns, etc) when choosing between the SP alternatives.  However, we are comforted, to some extent, by diagnostic information recently provided by the research team that showed that the majority of respondents indicated that they considered the choices to be realistic.**

*Has the data collection been implemented satisfactorily?*

4.4.15  The adopted survey administration for the general traveller SP survey was primarily via online (circa 84% we understand), but a significant minority (16%) completed the survey via CATI interview.

4.4.16  About one-third of the way into the telephone interview, respondents were informed that they needed to be sent some show material that they would need to look at when completing

the rest of the interview. Depending on respondent preference, the show material (i.e. the customised SP choice-sets) were sent by post or email. If by post, or difficulty for the respondent to easily get to a computer, the interview would be curtailed and a time on another day agreed to re-start it. For those respondents that could get to a computer, the interview could continue relatively reasonably smoothly.

4.4.17 We have conducted 13 interviews with survey recruiters, interviewers and respondents in order to get anecdotal insights into the actual interview dynamics. Some of the respondents who participated in the telephone survey explained that they had not received the SP options during the interview, and had had the options read out to them. Some indicated that this had made it difficult to remember all the SP options within a choice-set, and that, if they could have seen the options in front of them, they probably would have given a different answer.

> *"It [the show material] came through after the call, so he [the interviewer] had to read them out over the phone. … It was difficult to weigh up the scenarios just because there was so many variables" [Respondent 1]*

> *"It was ridiculously involved, it was almost impossible to remember the first one by the time they got to the last one. There were too many options and there were too many words … and it just went on for ages" [Respondent 2]*

> *"They sent it by email, and I sort of read it through and I think he asked me some questions on the material. [Were you confident that you were looking at the correct scenario for what the interviewer was asking?] No, to be honest. I thought that the material that was sent over…was ambiguous in some cases." [Respondent 4]*

4.4.18 The activity of the interviewer reading the options over the phone to the respondent was corroborated by some of the interviewers who we spoke to. They also confirmed that, sometimes, it led to respondents becoming frustrated.

> *"There was a small minority that wouldn't either do it by email or didn't have access to email, so you'd do it over the phone with them. … They sometimes got a little bit stressed with the amount of time it would take"[Interviewer 1]*

> *"If they preferred it to be sent by post, we'd do it but we'd say do you mind doing it over the phone? [How many SP surveys were sent by post?] "A small percent, I'd say like 10% maybe. [And how many read over the phone?] Probably about 40[%].*

> *"Some people didn't understand [when read over the phone] in which case I'd have to read it out again a few times so that they understood it, and I could sense the frustration in their voice." [Interviewer 3]*

4.4.19 Some of the feedback indicated other potential problems such as realism or understanding of some of the variables and/or values. However, these may be isolated instances as, presumably, such issues would have been highlighted during initial interview testing by Accent, and addressed.

*"[some of the variables] didn't really apply to me, so I did have to, you know, invent what I thought it would take. It just doesn't really apply to me as such, not for where I work and where I travel…I did adjust them [the variable values] [Respondent 3]*

*"I don't know how much use I was because I live in a rural area and I always take the same route to work. So there wasn't any being able to alter my route or traffic or anything involved in it. [Were they realistic?] Not in my particular circumstances, but I understood where it was coming from."[Respondent 5]*

4.4.20 This anecdotal evidence highlights that some of the telephone survey responses were based on options read out over the phone and, for the second and third SP exercises in particular, it would be almost impossible to carefully consider all the variables supposedly being presented. Neither the review team, nor the research team, seem to know how many respondents responded to the SP exercises without being able to see the choice-sets they were supposed to be considering.

4.4.21 The interview duration was not reported, nor does it seem to have been used for data checking. From our analysis of the CATI raw data, it seems that 30% of CATI interviews took more than half an hour (fatigue is likely to have set-in part-way through these interviews; 20-minutes is an accepted maximum telephone interview length in the market research industry). We also found that 2% took no more than 10 minutes which suggests little time was spent considering each SP choice-set. This could have happened in circumstances where the second part of the interview was scheduled for a later date giving the opportunity for the respondent to consider the SP options before-hand, but other evidence suggests that very few interviews were conducted in two parts.

4.4.22 A sample of 2,646 RP interviews were obtained; 98% of which were undertaken online with the remaining 2% by telephone. We consider that the predominantly online approach to data collection is reasonable given the mainly factual data that was involved.

4.4.23 All survey data were treated in confidence and in accordance with the Market Research Society Code of Conduct. All personal data was collected using password-protected tablets, and subsequently encrypted and finally stored in a password-protected Access database in a Restricted Access part of the server. We understand that, where necessary, files were transferred securely via FileZilla.

**Review Team Conclusion**

**The review team has some concerns with the way that the research team conducted the logistics of the CATI SP survey as feedback from respondents and interviewers suggest that, contrary to initial final reporting, some interviews have been conducted with the interviewer reading out the SP options over the phone to the respondent. This is not good practice and does introduce a substantial risk that some respondents will not have been able to fully consider all the options and variable values being presented to them. The very short duration of some online interviews is also a cause for concern, as is the sub-sample of respondents who self-reported that they did not understand facets of the**

| Value of Travel Time Savings | |
| --- | --- |
| Value of Travel Time Savings - Peer Review & Audit | 103217 |
| Final Report | 30/07/2015 |

**Page** 35/48

**SP exercises. We would have liked to have seen further data checks to ensure poor quality responses are not biasing the derived values.**

**From the evidence we have reviewed, the data collection for the Employers Survey and RP Survey have been implemented satisfactorily, and as reported, though the need to exclude most of the collected RP data (see next section) clearly casts a doubt over the suitability of the initial screening process.**

## 4.5 Data Cleaning & Processing

*Has the data processing been implemented satisfactorily?*

4.5.1 The research team report that the data has been subject to careful scrutiny to give the research team confidence regarding the quality of the data collected.

4.5.2 We can confirm that data from the different surveys (i.e. the online and CATI surveys) have been combined for the general traveller SP survey, and then cleaned. Data cleaning has been conducted systematically by Accent. The extent of this data cleaning was as follows.

- checking for any accident or test interviews and excluding;
- checking against quotas;
- checking for duplicate responses, and manually addressing any issues; and
- identifying missing values in the raw dataset and making a judgement on any issues.

4.5.3 The research team also report a number of other checks on consistency and logic of responses, as we would expect. We understand that the following checks were undertaken by ITS and we have been able to verify successfully a random sample of them, including:

- eliminating those individuals who reported a zero travel time or cost for their current journey (this will have adversely impacted upon the quality of the SP exercises and, hence, responses);
- checking for outliers (i.e. extremes) in relation to: distance, cost and revealed VTT (i.e. £ per hour) according to how each individual responded to a particular SP choice-set; and
- establishing additional exclusion conditions for particular trip circumstances and/or missing respondent details (such as gender or age).

4.5.4 We would have liked more detail on the rationale underlying the final data cleaning thresholds/conditions used to generate the estimation dataset; and more sensitivity analysis on the reported final model. This is especially important given that significant proportions of the data end up being removed (e.g. 11.5% for bus). The review team remains unconvinced over the research team's preferred approach for retaining more data over cleaning more data out.

4.5.5 It would be useful to know more about whether any work has been done to establish whether there are systematic demographic, or journey related, characteristics associated with these filtered-out respondents? And/or do such characteristics relate to data collection processes?

4.5.6    The received meta-code gives further insight to the processing carried out.  It identifies the data files that are read in to the modelling process; the file of exclusions; defines the income segmentation variable; and defines other calculated variables.   The data processing undertaken seems entirely reasonable.  However, there is no discussion as to why quality checks using the self-reported answers to the ease, or difficulty, the respondent had in responding to each SP exercise were not part of the exclusion criteria for modelling.

4.5.7    The research team's final report provides more detail of data cleaning and editing processes applied to the RP data, and we are surprised by the number of responses discarded (as previously identified) – with just 20% remaining, the whole exercise risks being undermined.

**Review Team Conclusion**

**From what we have seen, we are satisfied with how the research team has processed and edited the data, and the justification for treating the data as they have.**

## 4.6      Analysis & Model Estimation

### *Has the approach to data analysis and modelling been implemented satisfactorily?*

4.6.1    The received meta-code has shown the model specification of the utilities for each of SP1, SP2 and SP3, but only individually.  The review team has not been able to see how the final results (final VTTs and covariates) have been derived – which, we have been told, are based on a mix of separate and joint model estimations.  A parallel model build would provide further validation of the results.   Most of our peer review therefore, is based upon what has been reported by the research team and complemented, to a degree, by model runs defined by the research team and selected meta-code.

4.6.2    We are pleased to see that the VTT distribution has now been tested.  The test, on the SP1 data, rejects the maintained assumption of a log-uniform distribution, which is unfortunate.  The study has chosen to impose a VTT distribution that is rejected by tests on the SP1 data.  Some members of the review team are sceptical about this solution since the research team is then imposing a model that they have shown to be mis-specified.   In principle this invalidates all their statistical inference. The Danish study chose instead to use a distribution that described the data as well as possible, which avoids mis-specification, and then dealt with the tail problem afterwards.

4.6.3    In Table 4.10, the number of draws is very low.  We do not understand why the research team did not choose a large number here?  It is virtually costless, as there is no iteration.  The reporting of the maximum simulated is not that meaningful.  It would be better to give some information about the tails of the estimated distributions.  This could be done in a simple way by just plotting cdfs.

4.6.4    More generally, the research team argues elsewhere that the models should just describe the data.  Choices are consequently made about how to use the estimation results.  This principle applies e.g. to the discussion about \delta t.  It seems reasonable to apply the same principle to the distribution of random VTT.

4.6.5    However, as explained in Section 3.6 of this report, there is no evidence that the research team questioned some more basic model forms, with inadequate consideration of wider specifications.  Within the narrower focus than we would have liked, the research team has diligently performed tests to identify the models that best fit the data, using a simple stepwise eliminated strategy.

4.6.6    The justification for dependence on the assumption of $\Delta T=10$ as exposed on page 210 raises a question of why the SP exercise was not designed to take into account the need to provide estimates of the parameters in the 10 minutes range of values?  This reinforces the overall impression that there was a lack of an overall end-to-end modelling strategy.

4.6.7    We note the discussion regarding reference-free values (page 108).  Our concerns remain regarding the degree of arbitrary-ness of the argument that is used to justify dropping beta and eta in the estimation of 'average' VTTS, and have not been addressed in the final report.  This is all the more alarming given that the use made of the average VTTS will be in the context of the appraisal of policies that are likely to have strong asymmetric impacts (i.e. most schemes are about travel time reduction).

4.6.8    At the start of Section 4.7 there is limited explanation stating that the research team built up model complexity using results from intermediary models as starting values for more complex models.  Were different starting values tested, if so this would increase confidence that a global optimum was achieved (see below further consideration of different start values).

4.6.9    There is no evidence that potential SP-design effects such as different modes of administration and the different incentive mechanisms used at various stages in the recruitment stage were accounted for in the analysis.

4.6.10   We have not found any test for panel and fatigue effects which analyses repeated observations by the same respondent (each respondent provides 15 observations) and the sensitivity variation of the respondents across their 15 responses.  The review team captured correlation across the 3 SP experiments but not the effect on dependant observations or the ordering of choices.  Hess et al. (2012) reported[6] the following finding in relation to the Danish VOT study: "There is also evidence of increasing time sensitivity for non-commuters as the experiment progresses, while the other changes are of lesser importance."  Hence, the review team suggests that a similar investigation is important at least for non-commuters.

4.6.11   The review team requested additional runs for the car commuter and rail employees' business models to the study in order to check:

a) whether the online survey sample has different values to the CATI survey sample;
b) whether assuming different start-up values leads to different derived parameters; and
c) produce a model that includes all the covariates at the same time.

4.6.12   The separate model estimations for the online and CATI sub-samples (a) undertaken by the research team failed and standard errors were not calculated.  This fact indicates the difficulty for using subsets of the sample due to the complexity of the model.   Consequently, we

---

[6] Hess, Hensher, Daly (2012) Not Bored Yet – revisiting respondent fatigue in stated choice experiments.  Transportation Research Part A: Policy and Practice, 46 (3).  626-644

requested a car commute model run as reported but adding a dummy variable for CATI sample. The research team added two variables into the model (a CATI multiplier and its scale factor – standard deviation); and the result revealed a significantly different VTT scale factor at 95% confidence level for the two sub-samples; and a significantly different VTT multiplier (CATI was 20% lower) but only at the 90% confidence level. The finding that both variables (scaling and multiplier) are both significantly different indicates that the different survey approaches give derived values. The research team's chosen way of estimating and interpreting the coefficients differently and not following the same statistical criteria, could distort the test findings – hence why the review team's specified model required just one online/CATI dummy variable. As a consequence, the review team is not entirely confident with the reported final model.

4.6.13   The results for b) showed that the model estimations may be sensitive to the chosen starting values - which may have its source in the non-convex character of the likelihood-function and possibly the model being not identifiable. The review team has not seen evidence of the research team attempting to use any global optimisation routines [essentially estimating the model using a number of different starting points and comparing the resulting values of likelihood to conclude about global/local nature of the optimum] which could demonstrate robustness of the presented estimates. We consider it remiss of the research team not to have tested for other optimums.

4.6.14   For c) unfortunately, we have not received full model results for all the covariates as was requested. We have received only results for the car commute model using personal income instead of household income; and similarly, for the rail employees' business model but household income instead of personal income. In both cases, the reported model provided a better fit than the alternative income model.

4.6.15   We also asked for a clearer way to follow the sequence from the model with all the covariates to the reported final model. But only the car commute model with all covariates was forthcoming - and not useful enough to track the in/out covariates. It would be beneficial for the research team to keep a log file document to register a list of all model runs undertaken if that does not exist. General good practice requires the recording of all model runs by the research team to be logged so that it is possible to follow the process of gradual building up of model complexity, and to demonstrate evidence of improved quality of model fit.

4.6.16   To fully audit and peer review the modelling we would have needed access to all the Ox code.

### Cross-validation to validate the model results

4.6.17   We have not seen any cross-validation once the model has been estimated. It is particularly important to see how well the model forecasts a supplementary sample (not used for modelling). Maldonado, Sivakumar and Polak (2014) used this technique to evaluate the impacts of different cycling policies in London. We recommend using the Jack-knife technique, or similar, to produce a number of separate samples for modelling and validation and analyse the results. This would give more confidence and assurance in the modelling, in particular, about the robustness of the study's team final joint multiplicative model.

**Review Team Conclusion**

**The review team is satisfied with the modelling application from the evidence that we have been able to review, with the caveat that some of the review team would have liked the research team to have tested different modelling assumptions (in relation to starting point and reference point). The lack of access to the research team's Ox code means that we have been unable to audit the research team's final modelling procedures, so we cannot be entirely sure of the accuracy of the values derived.**

## 4.7 Implementation Tool

*Accuracy and Suitability of the Implementation Tool*

4.7.1 The audit processes generated a considerable volume of model output and information. This section summarises the key findings from this work.

4.7.2 The audit team understands that, subject to our reservations regarding the complexity of the modelling, the implementation requires a more advanced approach (i.e. the calculation of applicable values of VTTS), involving the development of a purpose-built program for sample enumeration and weighting. We also note that the final estimates provided by the tool will strongly depend on the quality and credibility of input data, both in terms of the estimated parameters, and NTS (or other) input data, as well as the assumptions made throughout the study, e.g., relying on log-normal distribution, using geometric mean, or using multiplicative effects coding to 'neutralise' the design effects.

4.7.3 We have found no inconsistencies between the underlying formulae, as outlined in Appendix I, and their implementation in the tool. This is consistent with the results of the validation exercise conducted by the research team (Section I4.3). We have also found that the report largely reflects the capabilities of the tool in terms of producing mean values, standard deviations, and bootstrap-based confidence intervals.

4.7.4 The tool embodies, due to the nature of methods it implements, a considerable level of complexity. Despite correct implementation of the basic capabilities, we have reservations regarding the ease of use of the tool beyond simple modifications of the generic parameters. The explanations provided in Appendix I1.1 describe how adjustments to weighting parameters, choice of $\Delta T$, or choice of modes can be made using the relevant Excel sheet (Master Sheet). However, adjustments to the tool beyond that, while signalled in the report, are not explained in detail.

4.7.5 Though not specified to be used in the wider transport community, the review team can imagine the tool may be used and updated on an ongoing basis. Potentially, the tool could serve as a cornerstone of VTT policy but, if so, its operation will need to be explained to a much higher detail and with much greater transparency. For example, a simple description without a diagrammatic representation of the relationship between the files does not adequately convey the intricacies of the tool's operation (Figure 7.1 offers a limited step towards such treatment) and a more picture-rich manual explaining the use of Master Sheet by the future users would be needed.

4.7.6   We note that in general, the tool presents an example of what we might term 'applied econometric programming' practice rather than professional software development.  By this we mean that, although there is no evidence of errors in the implementation of the relevant routines, there is also no evidence that the code has been developed to professional coding standards.  Given the considerable complexity of the code and high likelihood of its re-use in the future by the Department or other stakeholders, we think there is a substantial risk that the code will prove unsupportable.  We believe that if the code is to be retained and used in future as a workable tool, it must be re-written to professional standards of development and documentation.

4.7.7   Examples of poor coding practice that might hamper the supportability of the codebase include but are not limited to:

a)   Use of non-transparent variable names (e.g. mlhs, Index, Test1) and inconsistencies between variable names and the notation adopted in the report (e.g., formula I.2 uses $\alpha,\beta$ while in the code these are named 'theta_car_com_w' and 'theta_car_com_s' respectively).

b)   The embedding of key control parameters such as seed value, number of bootstrap draws, maximum VTT for mixed logit and reference values (for reference dependence)  in the code body, rather than in the Master Sheet.

c)   Inconsistency of variables categorisation (e.g. unstated income is categorised as 'traveller covariate' in the report, but among 'elasticities' in the input file).

d)   The Master Sheet File contains a number of sheets which serve as input to the tool based on the parameters provided by the user in the Master Sheet.  We believe that sheets which should not be modified by the user should be clearly marked, or even secured to prevent unintended corruption.

e)   Some sections of the code are commented out, and it is not clearly discussed (sometimes not at all) what purposes such sections serve.  For example lines 1095-1096 perform a linear regression of log(VTT/weight) with respect to trip distance.  This relationship is discussed in the report briefly but a more detailed exposition how such analysis can be performed using the implementation tool is warranted.

4.7.8   The audit team was largely able to run the code following adjustments in the working directory and installing the relevant packages:

a)   The only issue was in relation to the term: *VTT_tool_v1.exended_imputedcost_mixl_simulation.R* which displayed *'Error in eval(expr, envir, enclose): object 'NTS_data' not found'*.  It seems that line 200 in the code: *rm(NTS_Data)* appears to have been the source of error.  When the line is commented out, the code could be run correctly.

b)   Our recommendation in this sense is to re-perform a detailed de-bugging of the code.

c)   Since the tool relies on a number of packages and possibly on the present version of the language, it may be convenient to retain (together with the tool) all the relevant packages and version of R in a repository (subject to relevant conditions of use).  This can aid future use and update for compatibility.

4.7.9    The input files used in the tool are consistent with the tables reported in the files: multipliers from Tables 4.11-4.13 are used, and the input dataset is consistent with the distribution presented in Table 3.11.  We could not test for the covariance matrices similarity as these are not reported directly in the report.

4.7.10   Capturing potential impact on VTT from different time changes is achieved through Table 7.9 where appropriate multipliers for different changes in time are given.  We think that this way of handling the dependence on time change is adequate.

4.7.11   The research team states that it is possible to readily adjust the tool to make use of external data.  This applies only to a limited extent (through the Master sheet), and no detailed manual is presented regarding how such adjustments (input files, input data, format of variables, parameters) would have to be implemented.  For example, we were surprised to discover that, in its current form, the tool relies on a number of averages derived from the SP dataset.  These variables include:

- *mult_outside_city*
- *mult_company_pays*
- *mult_lc_current*
- *mult_hc_current*
- *mult_one_way*
- *mult_company_or_other_pays*
- *mult_freq_<=3_per_mth*
- *mult_dwell*
- *mult_shared_costs*
- *mult_freq_less_than_twice_per_week*
- *mult_freq_<1_per_mth_or_first*
- *mult_freq_less_than_every_day*
- *mult_no_reserved_seat*
- *mult_nightsaway_1_plus*
- *mult_nightsaway_1*
- *mult_nightsaway_2_plus*
- *mult_peak*
- *mult_cbuy_all*
- *mult_cbuy_not_stated*
- *mult_cbuy_no_for_employed*
- *mult_costs_NOT_covered_self_empl*

4.7.12   At present the 'average' values are included as fixed values in the Master Sheet, rather than variables appended to the input (NTS) data. This means that if a new variable was added to the NTS dataset, or an alternative way of imputing a variable not present in the NTS were used, substantial re-coding would be necessary.  We think that this raises doubts regarding the readiness of the tool for accommodating adjustments based on external data.  We would recommend that a more consistent description and approach to enabling adjustments in the code be developed.

4.7.13   Regarding the compatibility issues discussed in Appendix I (Section I.2.5), the audit team is largely satisfied with the approach presented by the research team.  Our only concern is with respect to dropping 10% of the sample for which PT costs were not available.  Was there any

specific reason for why imputation of these costs was not possible using the remaining data (in a similar way that cost imputation was dealt with in SP data)?

4.7.14 We have concerns regarding the simplicity in which compatibility between the variables present in the SP dataset, and absent from NTS dataset, was resolved. The variables which are not present in the NTS are introduced in the model by relying on mean (average) values from the SP dataset (see p.I5 in Appendix I), without clear and convincing justification. The variables are introduced directly in the code, e.g.:

> *VARS_VTT_car_com <-*
> *cbind(VARS_VTT_car_com,**repmat(t(SP_means[,1]),**gnObs_car_com,1))*

4.7.15 We believe that this way of introducing the imputed values has a number of weaknesses:

a) It lacks transparency -- it should be explicitly stated in the manual where the imputed values are located (presently in the Master Sheet).
b) It would be consistent and better practice to place them among the input files rather than Master file.
c) If they are to be retained in the Master Sheet file, there is a need for a clearer manual explanation of how they can be modified (Master sheet – columns N-AL).
d) Imputing 'average' values for categorical variables does not qualify as best practice. Imputing actual 0-1 values while preserving the distribution is at the very least an approach that should be followed. Otherwise any variance in the variable is essentially neglected with a potential effect on the subsequent derivation of the confidence intervals, including intervals derived from bootstrapping.
e) A single value approach to imputation (e.g. based on a mean or a regression coefficient) will distort the data covariance structure. A process of multiple imputations is required to retain biasedness in covariance structure. There is a large and well established literature on this topic, covering how to (and how not to) perform imputation. This literature appears to have been ignored, and we believe that in a critical context such as the present one, such short-cuts should be avoided or at the very least inspected and discussed. We can understand the complex nature of the study, but distorting data quality through overly naïve approaches while arguing for a complex modelling framework is an incoherent approach.

4.7.16 It is unclear how the averages were calculated. We followed the reported methodology but there we have uncertainties regarding how the adopted averages were derived: weighted? With or without excluded observations? Trip or individual-weighted? Our resulting estimates did not coincide with those used by the research team.

4.7.17 It appears that the resulting mean VTT values display some degree of sensitivity to the changes in the assumed average values. For example, using a mean value of 0.21 (which is an arithmetic average calculated from the raw data provided by the team as Audit Item 29 dataset) instead of the postulated 0.17591 for 'mult_company_or_other_pays' for rail commute leads to change in the mean VTT from £12.43 to £12.75 (SP1), £16.62 to £17.04 (SP2 on time), £47.56 to £48.76 (SP2 late). While these values fall within the estimated confidence intervals, we note, following from the discussion in 4.7.16 that using simple global means as is the case at present means that the imputation uncertainty will not be captured by the bootstrapping procedure, and that an imputation mechanism at the level of individuals

(micro-simulating values for the categorical variables) would work better. We also note, based on our understanding of the study design that NTS-based weighting had been decided at an early stage of the study. If this is the case, an appropriate set of imputation models could have been designed a priori, and internally validated, giving more credibility for the imputation mechanism, or at least more understanding on the possible degree of bias. Moreover, bootstrapping mechanism taking into account such uncertainty could be an effective mechanism for ensuring inclusion of the uncertainty in the estimated confidence intervals. It appears that the approach adopted so far is somewhat negligent of these issues (an alternative possibility could be to omit the variables and let them be picked-up by others – we understand that this deliberate mis-specification would not be appropriate, but neither is the present approach). We also note that the validation exercise as it stands now is unable to capture these issues since it only deals with the consistency of implementation of the formulae.

4.7.18 Even if we temporarily set aside our concerns regarding the inappropriateness of single value imputation schemes, we note that no regression diagnostics are provided by the tool for the models used to impute travel costs. Careful inspection of the models used to impute travel costs reveals that the fit of such models is poor (bus-commuting: $R^2$=0.148; rail-commuting: $R^2$=0.139). A number of other aspects such as the need for accounting for non-negativity of the variable seem to be ignored altogether, whilst the proportion for which the imputation is performed is high (e.g. 54% for bus).

4.7.19 The shares of time in different driving conditions (for SP3 values) are based on average values from the SP sample, again without clear and convincing justification, or a more elaborate means of imputing the values at a disaggregate level. The research team reports also that the observed values are relatively equal when aggregated (page 217, Table 7.10). One of the suggestions made is that the question might have been misinterpreted. We sense that this is an indication that the results from any pilot survey in which issues such as this one should have been captured were not appropriately incorporated in the actual data collection. In general, our sense is that the arguments presented in the discussion of changes in valuation under different congestion/and or reliability conditions are weak (p. 218: 'this is not a strongly-based recommendation, and we are still left with the conundrum explaining the high SP2 time multiplier').

4.7.20 We tested the result's sensitivity to different seed values, and also to different number of bootstrap rounds. The results did not deviate significantly (max +/- £0.1, typically around +/-£0.01).

4.7.21 The presented confidence intervals (Table 7.30, p. 247) include some very high values (e.g. above 70% for some non-work scenarios, while on average and for the headline SP1 values they are around 30%). Such high values of the confidence intervals present a challenge for application in appraisal since this effectively means that estimation of the benefits can vary for some groups by as much as 70% [car – other]. Moreover, since the use of single imputation methods has the effect of suppressing sources of uncertainty, the true variability in the VTT estimates are higher still. We understand that the width of the confidence intervals may be simply an artefact resulting from the underlying behaviour, yet we feel that a more detailed discussion on the potential implications for appraisal (e.g. cost-benefit ratios) is essential. We also understand (based on the notes below Table 7.30) that the confidence intervals were estimated for ΔT=10 mins. However, we feel that there is a need to comment

on whether, and if so how, such confidence intervals would need to be adapted for different assumptions on ΔT, e.g. should they be scaled in line with the multipliers from Table 7.9? We understand that such intervals can be calculated using the implementation tool, but the tool will not be made available from the WebTAG.

**Review Team Conclusion**

**Overall, the audit team have found the assumed formulae and advertised capabilities to be accurate. However, it relies on simplistic imputation procedures which may influence the final estimates in ways that are unclear.**

**We can imagine that there would be benefits of continuing to use a version of the Implementation Tool to provide further updates in future. If so, then we consider that certain revisions would be required:**

   **a) professional standards of coding and documentation would need to be developed that would ensure it can be properly maintained in the future.**

   **b) it is not readily adjustable to incorporate new data given that the code in its current form incorporates fixed values imported from the master sheet, rather than deriving the values from the input data (even if imputed).**

## 4.8 Reporting & Appraisal

*Is the Variance and uncertainty around the results reasonable?*

4.8.1 The review team finds generally that the VTT confidence intervals are high (typically ±30% of the mean and reaching ±70% (in the case of car other). That implies that the VTT values by purpose, mode or other covariates might not be statistically different within them and this fact would question their use at a segmented level for appraisal. We recommend that, in due course, a more detailed analysis of the sources of variation should be conducted. For example, a comparison with previous studies and a consideration for the aggregation of VTT values. We note some VTT aggregation in Section 8.3 of the research team's Final Report and we find this to be appropriate. In general, we think the VTT values are useful.

4.8.2 As the research team has explained, there is a bias-variance trade-off and they have opted for less bias. The results can be used to argue for aggregating VTT values for use in applications. It is good that the research team has reported the uncertainty of estimates carefully. However, we have not seen any bias/variance trade-off model assessment and selection. In general, finding an optimal bias-variance trade-off is challenging, but acceptable solutions can be found - e.g. by means of cross-validation or regularization. Cross-validation is used in nonparametric contexts for choosing smoothing parameters. We have not seen it used in this context such as the present but it is has been applied in other market research areas.

***Implications on schemes current WebTAG values vs new values***

4.8.3    The review team would like to have seen a section in the final report comparing, at least, two business case appraisals where current and new VTT values are used.  This will enable a fuller assessment of the implications of the methodology used in this study.  In essence, we would like a quantification of the impacts in terms of sensitivity testing.

***Is the conclusion on CSA versus WTP reasonable?***

4.8.4    In Sections 6.7 and 6.8, the research team have thoroughly wrestled with the CSA vs WTP issue.  On the evidence provided, we would, perhaps, have drawn a different conclusion - which is that the issue is currently undecided.  We do not think that the evidence presented here is strong enough to draw clear conclusions.

***Are the overall report conclusions reasonable?***

4.8.5    We believe that an explicit and upfront statement regarding the objective of section 7.8 (review/overview of the available evidence regarding how VTT evolves over time?) would avoid the feeling that the section is very speculative, with little supportive evidence either from data or, failing that, simulation.  This is understandable given that the research specification did not allow for more detailed analysis in this area.

4.8.6    Our concerns expressed regarding the general data and modelling process has a bearing on the reliability of the recommended values.  First, the overwhelming emphasis in the study seems to have been on what one might term 'more ambitious modelling' and that in a number of respects there has been insufficient care and attention given to data collection by the research team.  Moreover, there is a risk that the elegance and rigour of the 'innovative' work injected into the specification and estimation stages is in stark contrast to the more arbitrary nature of some of the assumptions made in the implementation (ignoring beta and eta, setting delta t to 10, etc.).  We recommend that future values of time research studies are less focused on breaking boundaries in technical modelling and more focussed on producing practically useful and reliable guidance.  It is far from clear to the review team that the strategy followed (innovative modelling + aggressive and sometimes arbitrary simplification in application) leads to better results than a strategy that uses simpler models that can be more straightforwardly and theoretically rigorously applied in practice.  This question of the balance of errors in specification, estimation and implementation appears not to have been considered at any point.

4.8.7    Recommendation 3 (page 262) suggests that the Implementation Tool be used by the Department.  We believe that the tool provides necessarily capabilities to derive the values for the present study, but these values will be subject to the limitations outlined earlier in the report.  Based on our understanding that the Department is currently using the tool to investigate different levels of aggregation, we feel obliged to inform about the implicit assumptions and possibly uncaptured uncertainties unaccounted for in the present form of the tool.  Furthermore, we consider the tool is likely to be beyond the scope of a typical Local Authority or consultant transport planner so its use will need to be limited to helping DfT to generate future values rather than being used by third parties.

4.8.8    Recommendation 6 indicates that the research team's finding was that ΔT=10 mins 'produced the most representative' values.  This is a strong statement, and the discussion and strength of arguments provided in section 7.6.1 (pp.208-210) do not provide indication regarding what is meant by 'representative' and what tests were performed in this respect.  In addition, the frequently invoked argument of 'following the best practice' seems to indicate 'current practice' which may be adopted on the grounds of feasibility rather than quality assurance.

4.8.9    The recommendations regarding business travel (section 8.3.2) are silent with regards to the emerging issues concerning the role of comfort, level of personal control, or possibility of more productive use of travel time (section 7.6.5, p 234).  We feel that a more definitive position/recommendation should be made in this respect, i.e., whether multipliers are required (or not), additional work be conducted, etc.

4.8.10   The recommendation section is also silent with regards to compatibility issues between NTS and the research team's modelling approach, and the possible biases/limitations resulting thereof.   Perhaps a recommendation towards collecting additional NTS variables, or investigating a means of more pro-active resolution of such issues, is warranted since a simple 'global average' solution seems to contradict the use of a very advanced model for which no suitable data exist.

### *Is the approach to reporting satisfactory?*

4.8.11   In general, we are impressed by the thoroughness and quality of the report.  It is good to see considerable discussion on the steps from estimation results to appraisal values.  There is inevitably a number of decisions that must be made in bridging the gap, and transparency here is important so that readers can distinguish between what is founded in the empirical evidence; and what is judgement on behalf of the DfT.

**Review Team Conclusion**

**The review team recognises the complex and high-profile nature of the study; and the range of technical and logistical issues involved in designing and conducting a large-scale, national study of travellers' attitudes and preferences.  Our review has uncovered some issues that potentially affect the quality of data from a minority of respondents; and we have raised questions about the cognitive understanding of respondents.  The modelling, and reporting, has been of a very high standard though we believe that some of the attention, and resource, that has been invested in the model estimation stage of the project would have been better deployed on improving data collection, cleaning and data imputation within the research team's implementation tool.  Nevertheless, this study has progressed our understanding of travellers' values of time savings and we recommend that the DfT undertakes more frequent, smaller-scale updating exercises in future.**

**SYSTRA provides advice on transport, to central, regional and local government, agencies, developers, operators and financiers.**

**A diverse group of results-oriented people, we are part of a strong team of professionals worldwide. Through client business planning, customer research and strategy development we create solutions that work for real people in the real world.**

**For more information visit www.systra.co.uk**

**Abu Dhabi**
AS Business Centre, First Floor, Suites 201-213,
Al Ain Road, Umm al Nar, P.O. Box 129865,
Abu Dhabi, UAE
T: +971 2 558 3809   F: +971 2 558 9961

**Birmingham**
Second Floor, 37a Waterloo Street
Birmingham B2 5TJ United Kingdom
T: +44 (0)121 233 7680  F:+44 (0)121 233 7681

**Dublin**
1st Floor, 12/13 Exchange Place,
Custom House Docks, IFSC, Dublin 1 Ireland
T: +353 (0)1 542 6000  F: +353 (0)1 542 6001

**Edinburgh**
Prospect House, 5 Thistle Street, Edinburgh EH2 1DF
United Kingdom
T: +44 (0)131 220 6966

**Glasgow**
Seventh Floor, 78 St Vincent Street
Glasgow G2 5UB United Kingdom
T: +44 (0)141 225 4400

**Lille**
86 Boulevard Carnot, 59000 Lille, France
T: +33 (0)3 74 07 00  F: +33 (0)1 53 17 36 01

**London**
Seventh Floor, 15 Old Bailey
London EC4M 7EF United Kingdom
T: +44 (0)20 7529 6500  F: +44 (0)20 3427 6274

**Lyon**
11, rue de la République, 69001 Lyon, France
T: +33 (0)4 72 10 29 29  F: +33 (0)4 72 10 29 28

**Manchester**
25th Floor, City Tower, Piccadilly Plaza
Manchester M1 4BT  United Kingdom
T: +44 (0)161 236 0282  F: +44 (0)161 236 0095

**Marseille**
76, rue de la République, 13002 Marseille, France
T: +33 (0)4 91 37 35 15  F: +33 (0)4 91 91 90 14

**Newcastle**
PO Box 438, Newcastle upon Tyne, NE3 9BT
United Kingdom
T: +44 (0)191 2136157

**Paris**
72 rue Henry Farman, 75015 Paris, France
T: +33 (0)1 53 17 36 00  F: +33 (0)1 53 17 36 01

**Woking**
Dukes Court, Duke Street
Woking, Surrey GU21 5BH  United Kingdom
T: +44 (0)1483 728051  F: +44 (0)1483 755207

**Hong Kong**
14th Floor West, Warwick House, TaiKoo Place,
979 King's Road, Island East, Hong Kong
T: +852 2529 7037  F: +852 2527 8490

**Shenzhen**
Room 905, Excellence Mansion, No.98, No.1 Fuhua Road,
Futian Central Zone, Shenzhen, PRC, Post Code：518048
T：+86 755 3336 1898  F：+86 755 3336 2060

**Shenzhen - Beijing Branch Office**
Room 1503, Block C, He Qiao Mansion, No. 8 Guanghua Road,
Chaoyang District, Beijing, PRC, Post Code：100026
T：+86 10 8557 0116  F：+86 10 8557 0126

**Beijing Joint Venture**
Room 1507, Main Building, No. 60, Nan Li Shi Road,
Xi Cheng District, Beijing, PRC, Post Code：100045
T：+86 10 8807 3718   F：+86 10 6804 3744

**Mumbai**
Antriksh, Unit no. 301, 3rd Floor, CTS Nos.
773, 773/1 to 7, Makwana Road, Marol, Andheri East ,
Mumbai 400069
T: +91 22 2647 3134
B 307, Great Eastern Summit Sector - 15, CBD Belapur Navi
Mumbai - 400 614
T: +91 22 2757 2745

**New Delhi**
5th Floor Guru Angad Bhawan, 71 Nehru Place, New Delhi
110019
T: +91 11 2641 3310

**Noida**
3/F, C-131, Sector 2, Noida-201301, U.P.
T: +91 120 432 6999

**Singapore**
25 Seah Street #04-01 Singapore 188381
T：+65 6227 3252   F：+65 6423 0178

**Thailand**
37th Floor, Unit F, Payatai Plaza Building,128/404-405 Payathai
Road, Rajthewee, Bangkok 10400, Thailand
T：+662 216 6652   F：+662 216 6651

**Vietnam**
5/F Perfect Building, Le Thi Hong Gam St, District 1,
Ho Chi Minh City, Vietnam
T：+84 8 3821 7183   F：+84 8 3821 6967

**SYSTRA**