Project Title:	Provision of market research for value of travel time savings and reliability
Subject:	Outline of QA procedures relating to behavioural modelling

1 Introduction

The purpose of this document is to outline the provisions that have been made and procedures that have been followed in order to facilitate quality assurance (QA) of the models employed on this project.

The aim of the Institute for Transport Studies (ITS) is to advance the understanding of transport systems throughout the world. This aim is achieved by teaching and research activities that develop the necessary skills and best practice in the planning, design, operation and use of transport systems. Teaching, which has its own Quality Assurance Scheme, is informed by the latest research. The Institute's research activities are the focus of the Quality Management System. In 1995 the Institute was the first University department in the UK to achieve certification for Quality Assurance procedures to ISO 9001 for research. The quality of project management and research is endorsed by the continued certification under the scheme.

2 The models which have been employed

It is appropriate to clarify what forms of model have been employed, and for what purpose, in the context of this project.

By far the primary modelling activity has been the use of discrete choice models to estimate 'behavioural' values of time, reliability and quality using revealed and stated preference choice data. This modelling has involved the use of model code written in Ox. Various procedures within this code were developed outside of the present project, and IPR for these procedures rests with the research team/University of Leeds. These procedures have previously been implemented in a number of academic papers which have been subject to expert peer review, for example:

- Ehreke, I., Hess, S., Weis, C. & Axhausen, K.W. (2015), Reliability in the German value of time study, Transportation Research Record, accepted for publication, April 2015.
- Kløjgaard, M. & Hess, S. (2014), Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: testing the benefits of a hybrid choice model approach, Social Science & Medicine 114, pp. 138-150.
- Hess, S. & Giergiczny, M. (2014), Intra-respondent heterogeneity in a stated choice survey on wetland conservation in Belarus: first steps towards creating a link with uncertainty in contingent valuation, Environmental & Resource Economics, accepted for publication, February 2014.
- Hess, S., Beck, M. & Chorus, C. (2014), Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives, Transportation Research Part A 66, pp. 1-12.
- Hess, S. & Stathopoulos, A. (2013), A mixed random utility - random regret model linking the choice of decision rule to latent character traits, Journal of Choice Modelling 9, pp. 27-38.

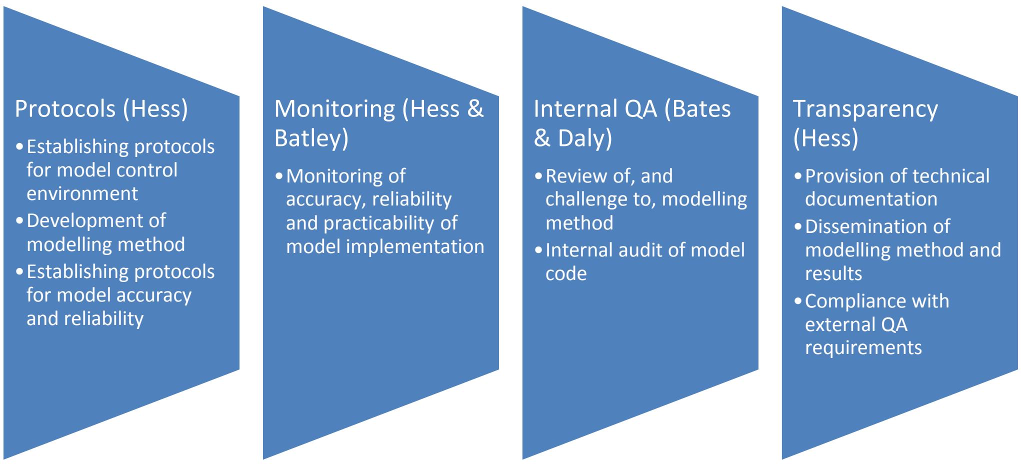
- Hess, S. & Stathopoulos, A. (2013), Linking response quality to survey engagement: a combined random scale and latent variable approach, Journal of Choice Modelling 7, pp. 1-12.
- Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V. & Caussade, S. (2013), It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity, Transportation 40(3), pp. 583-607.

A secondary modelling activity has been the translation of 'behavioural' values of time, reliability and quality into 'appraisal' values of time – this is referred to in the draft Final Report as the 'Implementation Tool'. Preparatory work for this task was undertaken in Excel, with the final implementation programmed in R. QA for the Tool is documented in a separate note and the Implementation Tool has been provided in full to DfT's external auditor for review.

3 General modelling principles

An overview of the QA process which has been followed is given in **Figure 1** below.

Figure 1: Schematic diagram of stages in modelling and analysis QA, with responsible individual



More specifically, modelling work on this project has been conducted in accordance with the principles detailed in the Department's guidance document *'Quality Assurance of Analytical Modelling'*. We reproduce these principles below, and explain what measures have been taken in order to ensure compliance.

Principle I: The Model Control Environment

All finalised models and associated data are stored on a secure 'sharepoint' at the University of Leeds. This sharepoint is backed up on a daily basis, in accordance with standard University of Leeds data protection procedures.

The location of the sharepoint is: *N:\Earth&Environment\ Research\ITS\Research-1\VoT2014*

On the sharepoint, separate folders have been established for analysis undertaken in the course of Wave 1 pilot, Wave 2 pilot, and field data analyses.

Within each such folder, there is storage of the raw data received from our data collection partner (Accent), the 'finalised' data following reformatting, the relevant model decks, and model output.

The model estimations themselves have largely been performed on individual team member's local machines, before being transferred to the sharepoint (which is of course only a storage site rather than an environment in which models can be run). However, in the latter stages of model development, the increasingly complex nature of the models meant that run times on standard computers became burdensome, and we therefore estimated a number of the final models via remote access to high performance computers owned and managed by the University. Independently of whether models were estimated on standard computers or high performance computers, the results were then copied to the sharepoint.

More generally, staff working on the modelling are governed by the University's IT policy and practices (http://it.leeds.ac.uk/info/116/policies).

Principle II: Model Accuracy and Reliability

In order to ensure the accuracy and reliability of the models which underpin the valuations presented in the Final Report, the discrete choice models were developed through a careful and systematic process involving the following stages.

Stage 1: Data checking, cleaning and diagnostics

The 'raw' behavioural data from the Stated Preference (SP) and Revealed Preference (RP) surveys, which was collected by Accent, was subject to initial checking. This involved verifying that the correct data had been dispatched to us and was complete (i.e. the correct files had been sent, and each file contained the correct number of variables for the correct number of respondents).

Following these checks, there was some reformatting and reorganisation of the datasets, but this involved no substantive changes to the data itself. This essentially involved replacing some text entries by appropriate numerical codes and removing columns with pure text entries (e.g. direct respondent feedback) which could not be used in the modelling process which is described below.

In order to ensure the quality of the data taken forward to the modelling stage, the data was subjected to certain 'cleaning' and exclusion criteria. Detailed discussion of these procedures and the precise cleaning/exclusion criteria employed are reported in sections 4.2.1 and 4.2.2 of the Final Report. It is important to note that we conducted comparative testing of alternative criteria, so as to demonstrate the impact of different levels of cleaning/exclusion on the modelling.

For reasons of consistency, where a given data point was removed (e.g. for a given SP choice task) – in accordance with the final cleaning/exclusion criteria – all other data points for that respondent (e.g. for all other choice tasks and games) were also removed. A record was kept of which data points/respondents were removed, and the whole process was subject to internal review and audit[1]; this audit included a check that the exclusion criteria were correctly programmed.

Furthermore, DfT's external auditor was given the complete (but checked) dataset, a list of the exclusion criteria, and a file listing the ID of each respondent who was 'cleaned out' on the basis of the exclusion criteria; this gave the auditor the opportunity to replicate the cleaning process if he/she so wished.

A final stage of preliminary work to ensure and confirm the quality of the data was to undertake diagnostic testing of the data for 'irrationality' of RP/SP responses, 'non-trading' and such like. This exercise is documented in detail in section 4.2.4 of the Final Report, and our broad judgement was that the data was fit for purpose in this regard. Like the cleaning process, the diagnostic testing process was reviewed internally; see the note referred to under footnote 1.

Stage 2: General specification search

Having finalised the dataset to be modelled through the aforementioned quality checks, there then followed a process of determining the general form of the model to be employed for estimating values of time and associated factors.

Broadly speaking, this involved comparative testing of the two key candidates, namely 'additive' and 'multiplicative' models. These tests are reported in section 4.3.4 of the Final Report, and demonstrate the clear empirical preference for 'multiplicative' over 'additive' specifications for all three SP games (i.e. time vs. cost; time vs. cost vs. reliability; time vs. cost vs. quality).

Once we had identified the multiplicative specification as the preferred model, we then conducted tests to demonstrate the benefits of further elements of functionality, namely size and sign effects. This is outlined in section 4.5.6 of the Final Report. It is important to note that such testing was conducted incrementally, thereby demonstrating the distinct benefit of each additional element of functionality.

Stage 3: Detailed covariate search

Having finalised the generic model specification through the activities outlined in stage 2 above, there then followed a more detailed and exhaustive process of determining which covariates (i.e. relating to features of the traveller and/or journey) should be included in the models. This exercise was conducted separately for each mode of travel (i.e. car, rail, bus and 'other PT') and purpose, and followed a general-to-specific approach.

The process used in the specification search was iterative, gradually building up model complexity and using results from intermediary models as starting values for more complex models, leading to substantial reductions in estimation time, which was essential in the context of a tight analysis schedule. We started with models excluding covariates and first added elasticities in relation to travel time, travel cost, distance and income. This was then followed by adding in size and sign effects, before gradually adding in the remaining covariates. These were added in batches, each time removing insignificant parameters before adding additional ones, and every time

[1] See for example the 'Note on data cleaning and diagnostics' referred to under *Stage 5* below.

using the final estimates from the previous model as starting values. This process is reported in section 4.7 of the Final Report.

As part of the external audit (Stage 6 below), the auditor asked us to estimate some variants of our recommended models, in order to explore the sensitivity of the results to adjustments in the covariates. We do not believe that this exercise revealed any significant concerns regarding the rigour of the covariate search and robustness of the resulting model specification.

Stage 4: Model estimation

The estimation process can be broken down into the following steps (**Table 1** details the members of the study team involved in each step, for each model):

- Development of base Ox code, which was subject to internal model audit (see *Stage 7* below).
- Additional coding; this involved an iterative process of experimenting with restricted forms of the base model code, in conjunction with different starting values for the estimation routine. Starting values were chosen so as to avoid numerical issues (e.g. very large starting values would lead to computational failures) and also taking into account base values for given parameters, e.g. zero for elasticities. Small deviations (generally up to 50%) were used in changing starting values.
- Coding check by another member of the modelling team; this involved checking the Ox code to confirm that the additional specification work referred to above was being implemented correctly.
- Collective decision by the modelling team on the final model specification.
- Final estimation run.
- Sense check of results; this was undertaken by personnel from outside of the modelling team (namely Batley and Bates), and involved assessing the plausibility of the results (for example, assessing whether the relativities of values for different travel conditions and different purposes aligned with intuition) and making sure that results were consistent with DfT's appraisal needs (for example, making sure that values of time, reliability and crowding/congestion were presented at an appropriate level of segmentation).

Again, the final models are reported by mode in section 4.7 of the Final Report.

Stage 5: Methodological review

In parallel with stages 1-4 outlined above, methodological review and challenge was undertaken on an ongoing basis, and especially at key junctures in the modelling development process. This exercise was led by a senior study member outside of the modelling team (Bates).

As evidence of this process, we have supplied technical notes produced by Bates alongside this document, as follows:

1. General scoping note on modelling, dated 30th October 2014
2. Note on data cleaning and diagnostics, undated
3. Note on additive vs. multiplicative specifications and Δt effects, dated 17th November 2014

4. Note on additive vs. multiplicative specifications, dated 19th December 2014
5. Note on sign and size effects, undated
6. Note on distance effects, dated 16th April 2015

In addition to Bates' contributions in this regard, technical advice and challenge has been sought/invited from other experts on the team (Bliemer, Börjesson, etc.) as necessary. E-mail correspondence can be provided as evidence of this.

Stage 6: External model audit

The behavioural models developed through the process outlined above have been subject to external review and audit undertaken by contractors appointed by the Department. This has involved the following broad tasks:

- Review of modelling method as documented in the Final Report, in conjunction with a telephone interview with the auditor. *This exercise sought to confirm the appropriateness of the conceptual approach to the modelling which had been adopted.*
- Review of 'meta' Ox code for two key models – namely car commute SP and rail employee SP. The meta code outlined the key mathematical statements that drove valuations of time, reliability and quality, but – for reasons of IPR – not the optimisation routine used to estimate the model. *This exercise sought to confirm that the key mathematical statements given in the Final Report (e.g. utility functions and probability statements) had been correctly implemented in the model code.*
- Site visit by the auditor to observe models being estimated. *This exercise sought to confirm that the model estimation had been correctly initiated, and that the results presented in the Final Report were indeed those emanating from the data collected and models developed in the course of this study.*
- Additional model runs to explore sensitivity of the results to variants of the final model specifications and alternative starting values of the estimation routine. *This exercise sought to confirm that the covariate search had been conducted in a rigorous and systematic fashion.*

To the best of our knowledge, the aforementioned audit tasks have not revealed any significant concerns regarding the rigour of the modelling process and the robustness of the resulting outputs.

Stage 7: Internal model audit

In order to provide further assurance concerning the accuracy and reliability of the behavioural modelling, a senior team member (Daly) not directly involved in the development of the Ox code was commissioned internally to audit check two key SP models – namely car commute and rail employee. *This exercise is outlined in a separate note alongside this document, and was intended to check that elements of the model code (such as the estimation routine) not covered by the meta code were correctly programmed.*

The modelling lead (Hess) and supply chain leader (Batley) have carefully studied Daly's note and are satisfied that this raises no significant areas of concern regarding the Ox code. Daly summarises his audit findings by noting: *'The issues I have found* *were minor and have largely been resolved by the project team, so that, as far as I can see, nothing of consequence remains to be corrected in the code'*.

Principle III: Model Practicability

An important consideration in the context of the present project was to ensure that the models developed would be fit for purpose in terms of the translation of modelled values of time to appraisal values.

In practical terms, this involved monitoring a range of issues including the following:

- Does the modelling framework cover the necessary modes and journey purposes, and permit breakdown by mode/purpose as necessary?
- Do the models specify the necessary covariates (e.g. income and distance)?
- In the case of such covariates, is the data recorded in appropriate units (e.g. household income before tax)?
- More generally, does the modelling framework allow interface with key external data sources such as the NTS?

Model practicability was monitored on an ongoing basis by the supply chain leader (Batley), and especially at key milestones in the modelling work.

Principle IV: Governance and Transparency

Whilst Hess was responsible for all aspects of the modelling suite, the internal review of method (Bates) and internal audit of model code (Daly, Appendix 1) were undertaken by experienced individuals not directly involved in the estimation of the models or production of model code.

Whilst observing the necessary provisions to ensure the security of models and data, we have sought to promote the transparency of the modelling and analysis work through various initiatives, such as ongoing dialogue and liaison with the DfT project officer Adam Spencer. This will also ensure no loss of information value at point of transfer from consultant to DfT.

Furthermore, we have made all data and the theoretical model specifications available for external audit.

When data are delivered to the client and/or auditor in electronic format, the following have been checked:

- The file format is compatible with the software specification agreed with the client/auditor.
- An up-to-date anti-virus check has been made.
- Completeness – that the correct number of files and records are in each file.
- Any file is accompanied by a structural description.
- Any file is labelled with the contents and where applicable accompanied with instructions on limitations of use.
- That the file is password protected if the file contains personal data (suitable data encryption is used where the client requests this).

Principle V: Proportionality

Whilst the above provisions have been made in respect of internal QA of modelling and analysis, we have been open to discussion with DfT and/or the main contractor

(Arup) throughout the project with regards to additional provisions that might have been judged appropriate. A parallel redevelopment of the models by the external auditor would have added further reassurance, but this would have needed additional time and budget.

4 QA Programme

Table 2 below records the points at which data and models were available for auditing.

Table 2: Schedule for QA checks

Modelling task	Start	Finish	Point at which checks could begin	QA check
Pilot Wave 1	27/8/14	3/9/14	8/9/14	8/9/14
Pilot Wave 2	25/9/14	26/9/14	29/9/14	29/9/14
Field (initial values)	20/10/14	28/11/14	1/12/14	1/12/14
Field (emerging values)	1/12/14	26/12/14	29/12/14	29/12/14
Field (final values)	5/1/14	27/2/14	2/3/14	2/3/14

Table 1: Internal QA checking of models at each stage of development (to be read alongside *Stage 4* described above)

Mode	Purpose	Base code developed by…	Additional coding undertaken by…	Code checked by…	Sign-off on final model specification by…	Final models estimated by…	Model results sense-checked by…
Car SP	Commute	SH	TD	MOC, AJD	SH, AJD, MOC, TD	SH, TD	RPB, JJB
	Employee	SH	MOC	TD	SH, AJD, MOC, MRW	SH, MOC	RPB, JJB
	Employer	MOC	MOC	TD	MOC, MRW	MOC	RPB, JJB
	Other NW	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
Rail SP	Commute	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
	Employee	SH	MOC	TD, AJD	SH, AJD, MOC, MRW	SH, MOC	RPB, JJB
	Employer	MOC	MOC	TD	MOC, MRW	MOC	RPB, JJB
	Other NW	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
Bus SP	Commute	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
	Other NW	SH	MOC	TD	SH, AJD, MOC, TD	SH, TD	RPB, JJB
‘Other PT’ SP	Commute	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
	Employee	SH	MOC	TD	SH, AJD, MOC, MRW	SH, MOC	RPB, JJB
	Employer	MOC	MOC	TD	MOC, MRW	MOC	RPB, JJB
	Other NW	SH	TD	MOC	SH, AJD, MOC, TD	SH, TD	RPB, JJB
Rail operator choice RP	Commute	MRW	CC	SH	SH	SH	RPB, JJB
	Employee	MRW	CC	SH	SH	SH	RPB, JJB
	Other NW	MRW	CC	SH	SH	SH	RPB, JJB
Rail operator choice SP	Commute	MC	TD	SH	SH	SH	RPB, JJB
	Employee	MC	TD	SH	SH	SH	RPB, JJB
	Other NW	MC	TD	SH	SH	SH	RPB, JJB
Mode choice SP	Commute	MC	TD	SH	SH	TD	RPB, JJB
	Employee	MC	TD	SH	SH	TD	RPB, JJB
	Other NW	MC	TD	SH	SH	TD	RPB, JJB
Concessionary SP	Commute	MC	TD	SH	SH	TD	RPB, JJB
	Employee	MC	TD	SH	SH	TD	RPB, JJB
	Other NW	MC	TD	SH	SH	TD	RPB, JJB

Note: SH (Stephane Hess), TD (Thijs Dekker), MOC (Manuel Ojeda Cabral), CC (Charisma Choudhury), MRW (Mark Wardman), AJD (Andrew Daly), RPB (Richard Batley), JJB (John Bates)

Appendix 1: QA of Ox modelling code

Andrew Daly, ITS, 19 May 2015

This document presents the results of my review of the Ox modelling code for the VTT models. I have reviewed the models for Car Commute (CC) and for Rail Employer's Business (REB). It is likely that the models for other purposes and modes share the same properties as these two, but of course that cannot be guaranteed. The issues I have found were minor and have largely been resolved by the project team, so that, as far as I can see, nothing of consequence remains to be corrected in the code. This document summarises what I have reviewed and my findings.

The source material I have used is:

- the part of the report describing the models as mathematical equations, of which I was an author;
- the two documents describing these models as 'meta code' which were prepared for the audit and to which I contributed annotations; and
- the Ox code, which was explained to me in some detail by Stephane Hess and which I believe I largely understand, though I am not proficient in the use of Ox.

In these files I have focussed on the sections required for the estimation of the parameters that are later used (in the 'Application Tool') to calculate population VTT and associated statistics. I have therefore ignored:

- sections carrying out the SP sample-based enumeration, including the variables saved to make this possible, which gives preliminary VTT to help in modelling but which is then superseded by the Tool results;
- sections identifying and printing 'outliers';
- sections calculating and printing other diagnostics.

Input to the Ox code is from two spreadsheets for each model, which contain the data for the relevant mode and all purposes[2]:

- car_joint.xlsx, which contains the full car SP data or
- rail_joint.xlsx, which contains the full rail SP data; and
- Exclusions.xlsx, which lists the exclusions over all modes and purposes.

I am not in a position to check the validity of these files. I am informed that a number of checks have been made that the SP data has been properly processed and appears in the correct location and properly labelled in the data file. The exclusions file is simply a list of 679 record ID numbers but again I am unable to check that this corresponds with the data processing that identified the respondents who should be excluded.

The car SP data file contains 15045 (rail 15175) rows of data plus a header record containing the data labels. Each row of the car file contains 178 data items (rail 199),

[2] That is, those included in the main part of the study: modes car, rail, bus and other PT; and purposes commute, employer's business and leisure. As usual, there is no data for employer's business by bus. Obviously, separate files also exist for bus and other PT SP data.

of which the first is the unique respondent ID number. There are five rows of data for each respondent (i.e. car 3009, rail 3035 respondents). The k^{th} row of data for each respondent contains all of the background data for that respondent (i.e. this information is repeated five times) together with the attributes presented and the choice made for the k^{th} presentation in all three of the games (SP1, SP2 and SP3); the unusual ordering of the data is made to facilitate later processing by Ox. A check on the validity of the SP data file would include a check that the data is properly duplicated but no problems are reported in the modelling work.

CC model

The CC model is estimated using the file car_commute_final.ox, which interacts directly with the Ox software. The software proceeds in the following steps.

Technical Ox components that are needed for maximum likelihood estimation, e.g. the SQP optimisation algorithm, are loaded.

'Global' variables are defined for the basic dimensions of the modelling. Next, variables are defined for the input of data, using the header record of the SP data file; I am informed that this has been done automatically, reducing the potential for typing error, while the internal checking given by Ox would detect most errors anyway. Variables are then defined that are used in the data transformations, in four groups for SP1, SP2, SP3 and 'socio-demographics' (including income and trip variables); again the internal checking of Ox should detect most errors here, while the data transformations themselves are discussed below.

It is then necessary (for reasons of Ox syntax) to declare three internal functions, MXL, LL and haltonsequence, which make the main calculations of likelihood and quasi-random numbers; these are discussed below.

The 'main' part of the Ox code is now presented. This starts by defining a few working variables, loading the data and setting the number of tasks per game (equal to the number of rows of data per respondent, i.e. 5) and the number of games (3). Note that the number of rows and columns in the data and the number of exclusions are defined automatically from the data files and that Ox automatically ignores the header row in the data, so that errors here are improbable.

The income variable code is then set to use the household income (not the adjusted household income, nor the personal income). This does not (yet) set the income.

Exclusions are then implemented:

- for respondents whose ID value (data item 0, only the first record is checked for each respondent) appear in the exclusions data and
- for respondents interviewed for a purpose (data item 43) other than code 2, commute.

Note that these exclusions apply for all the five records for the respondents concerned. To check that this was operating correctly, we made a specific test of the exclusions and purpose selections in Excel, obtaining the same number of records and respondents.

The number of observations is then set to be the length of the remaining data and the number of respondents is set equal to than number divided by the number of tasks. The quasirandom draws are then made, 500 for each respondent, using haltonsequence and prime 2. Additional redundant draws are made, three sequences which would be used if more than one quasirandom sequence was used per respondent. Vectors are set up for the values of time for each game and for SD (SP2) and congestion level (SP3). These steps seem unlikely to cause error.

Then names are given to each item in the data file, using the names taken from the SP data file header (0 to 177, i.e. 178 items). Stephane Hess has shown me how this was done automatically and there is therefore unlikely to be an error here. However, the name of the first item is set to 'ID'. From this point on, these names can be used to refer to the data, reducing the possibility of error.

Transformations and renaming of the data variables is then undertaken. These transformations were provided in full (using the Ox coding) for the external audit and I made additional annotations to them at that time, questioning anything that was not clear to me. These transformations cover generic and game-specific data items.

Of particular interest are the income transformations:

- 'income' is set to household income code if transformed or untransformed household income is indicated, otherwise it is set to personal income code;
- positive income codes 1-8 are converted to continuous income, using group midpoints, in £k/year, with £5k/year for the lowest group (under £10k/year) if personal income is used, £7.5k/year for the lowest group if household income is used and £130k/year for the highest group (over £100k/year); continuous income is set to zero for missing income and income code 0;[3]
- then, if adjusted household income is used, the income is divided by the household size;
- finally, missing income is set for income codes that are negative or greater than the highest group.

Also, reference values are set for income (£40k/year), cost (£5), time (½ hour) and distance (20 miles), these being one-way values.

The following commands relate to diagnostics counting non-traders etc. and have not been reviewed.

Coefficient labels are then set for output and start values are set for the coefficients. I have checked that the labelling of the variables corresponds to the setting of the variables (in routine LL), with the note that "mult_additional_travellers" corresponds

[3] I have checked the values assigned against the survey script. Processing of income=0 when not noted as missing would apparently lead to an error in running the program, so there is effectively a check that this does not occur.

to the variable for two or more additional travellers[4]. The initial values are unlikely to cause error and have been chosen to minimise the number of iterations taken to convergence.

The following lines initiate the maximum likelihood estimation, using the SQP algorithm and routine MXL, mentioned previously but for which the code follows. A number of controls are set for the optimisation. Their function is to specify the output from the optimisation and to retain the default convergence criteria.

When the optimisation is finished, outputs are made to the log file of the converged results and an analysis is given of the results, including the 'robust' errors and a list of outliers. Results are also saved in spreadsheet format. I cannot check the Ox syntax here and it would be useful to be sure that the labels are attached to the right variables and that the errors are correctly output for use in sample enumeration.

The code of the routine MXL is presented: it loops over the respondents and calls routine LL to calculate the log likelihood contribution for each respondent. LL was mentioned previously and its code follows. MXL also maintains pointers to the first and last of the five lines of data for the current respondent and sums the likelihood over respondents. This routine is very simple and should not cause problems.

The routine LL is the core of the modelling and its code is presented. It starts by setting the variables: the checking of these commands was described above. Then the calculations are made to obtain log likelihood for a respondent. At each step I have checked as far as I can the calculations and in the following points I note what may be weaknesses.

- The exponent κ is calculated for travel time for each game and for s.d. for SP2.
- The elasticities λ are applied to obtain distance (distance elasticity is zero in this model), time and cost multipliers. The elasticities are applied to the ratio of the variable to the reference value.[5]
- An income multiplier is calculated comprising:
    - an elasticity λ is applied to obtain a multiplier, when the income is not missing;
    - separate simple multipliers are applied when income is not stated (or not applicable, both have negative codes), unknown (code 9) or refused (code 10).
- Choice-dependent position multipliers are set, probably for each record of the five for the current respondent (my knowledge of Ox is not sufficient to understand this code fully), for
    - in SP1, a multiplier when time is presented first *and* a multiplier when the cheap option is on the left;
    - in SP2 and SP3, a multiplier when the cheap option is on the left (the multiplier for SP2 is fixed to 1 in this model).

[4] The label may be misleading, but the model is clearly defined.

[5] The pointer to the first record (of five) for the current respondent is used so that the base variables are taken from that first record.

The definition of these terms depends on the OPTIONORDER variables on the SP data file, which I am unable to check. These multipliers are set using the formula $\beta\delta + {(1-\delta)}/{\beta}$, where β is the multiplier to be estimated and δ is the indicator of its applicability to the current choice, this formulation meaning that the 'null' position, i.e. ignoring the multiplier, is the geometric mean of the two possibilities.

- Further simple multipliers are set, using the first record for the current respondent, for 15 background covariates (age, sex etc.). These depend on the codes in the data file and I am unable to check in full detail, but the calculations have been supplied to the audit team previously.
- Congestion and road type multipliers are set using power-weighting averaging as a function of the base shares (for the first record) of congestion and road types.
- Scales are set for the games, taking account of the order of SP2 and SP3 and using a geometric averaging for the order multiplier.
- The covariate multipliers are combined into a single multiplier.
- The base random VTT is calculated using the log-uniform distribution and the quasi-random draws (I can't check the Ox syntax). Note that this is negative to indicate WTP for time increases.
- VTT for each game is calculated using the base VTT and the covariate and position multipliers, raised to the power of ${1}/{\kappa}$ and using the appropriate κ for each game; additionally,
    - for SP2, a value of s.d. is calculated;
    - for SP3, separate values are calculated for each congestion level.
- For SP1, for each of the five choices:
    - value functions are calculated for each of the time and choice changes presented in the two alternatives; note that these depend on the VTT for each draw;
    - from these value functions, the 'log bid' is calculated;
    - from the log bid, the logit probability is calculated;
    - the product of probabilities over the five responses is calculated.

  These calculations are made over all of the draws, which is automatically arranged by the Ox syntax, though I am unable to verify this.
- For SP2 the calculations are quite similar to SP1, except that, instead of calculating log bid, explicit utilities are calculated for the alternatives, including the base values, and these include additional terms for OPTIONORDER and for 'no variability'.
- For SP3 the calculations are quite similar to SP2, except that the three time components (depending on congestion) appear separately in the value functions and (for the base value) in the utility functions; OPTIONORDER appears again but not (of course) 'no variability';
- The overall likelihood contribution is calculated as the mean over the draws of the product of SP1, SP2 and SP3 probabilities.

Finally, the routines haltonelement and haltonsequence are defined and the code is presented. haltonelement is referenced only by haltonsequence (and so does not need to be declared earlier). haltonelement returns one element in the Halton sequence (for a given prime), while haltonsequence returns the whole sequence of a defined length for a given prime, calling haltonelement for each element of the sequence. These routines appear to be correct and have been used in many previous studies.

REB model

The REB model is estimated using the file rail_employee_final.ox, which interacts directly with the Ox software. The file is quite similar to the CC model file, with the following notable differences.

The data file contains more variables, e.g. referring to crowding and to operators, so these have to be declared.

Personal income is used, not household income.

An additional exclusion is performed to remove respondents who have code 3 for the ‘self-employed’ question. This is still not clear but I understand that people with this code were excluded from the business questionnaire.

The data transformations are different, of course, but these have been reported to the audit already, with the exception of those for the operator game, which I have checked briefly, although the operator codes and OPTIONORDER codes are complicated and I cannot check them thoroughly.

The parameter names are different, which I have checked these against the order of the variables set later, and initial values are set.

The routine MXL appears to be the same as in the CC model.

The multipliers are of course different from those in the CC model, but have been provided to the audit previously. Similarly the calculations of VTT are different to allow for early/late, crowding levels etc., but the audit team have seen these already.

Note that the operator choice game is not included in this model.