Task 8a Distribution and Mode Choice

Scope

The distribution and mode choice models are the most demanding to calibrate for many reasons:

- it can be difficult to identify the appropriate hierarchy, and this may be sensitive to different ways in which the models are specified;
- it is very sensitive to errors and biases in the input generalised costs;
- a good fit to the data is important for achieving a satisfactory overall model validation, but these models are generally insensitive to variations in travel behaviour across a region;
- it takes time to understand how best to achieve a suitable, well-fitting model because of the large data bases and the difficulty of gaining appreciation of the underlying structure of the trip matrices;
- the calibration requirements are particularly demanding on calibration software, and can be hampered by software limitations.

Because of these difficulties, there are large risks of budget and time over-runs. Consequently, to keep this task within manageable limits, a carefully planned work programme is needed, with the aim of quickly identifying the appropriate overall model specification (essentially the model hierarchy) as early as possible.

Following an initial description of the model specification in the next section, the sequence of tasks is:

- Task 8.1 Create Analysis Files/Set up Statistical Procedures
- Task 8.2 Assignment of Observed Private Vehicle Matrices
- Task 8.3 Calibration
- Task 8.4 Distribution and Mode Choice Report

Specification

Key specification issues are discussed in turn below.

Level of Detail – Segmentation

The basis of the segmentation proposed is given in Table 1; the precise structure will be developed in the preliminary studies on behavioural and sample size grounds and refined in model estimation. In particular, the preliminary studies will reach a view as to whether an income segmentation would be feasible at this level. Our preliminary view is that the car availability segmentation is considered to be best practice and cannot be dropped, in which case the additional income segmentation would increase the number of segments more than can be supported on the data and would significantly complicate model estimations. If we hold this view after the preliminary studies, then any income segmentation is likely to be brought in after the demand models.

▪ Table 1 Segmentation in DMS

Purpose	Sub-Purpose	Car Availability				Comments
		PT captive	Car captive	Compe-tition	Choice	
HBW	SEG1 SEG2					
HBEd	-					Subject to preliminary studies, Primary Education trips will not be modelled (except for vehicle escorts)
HBSh	-					
HBSo	-					
HBO	-					
NHBO	-					
HBEB	-					
NHBEB	-					

NB: cell grouping indicates possible car availability aggregations.

Cost Issues

Generalised Costs

In previous models, to improve run times and make the best use of the combined data, we have developed 24 hour distribution/mode choice models and used matrix factors to generate the trip matrices for each time period required for the assignment.

For this to be effective, the network data used in model calibration and application should reflect the times at which different trip types occur; for example, peak period network data is relevant for those purposes occurring primarily in the peaks (e.g. commuting) while interpeak network data may be more appropriate for shopping and recreational trips.

The generalised costs which will be used in these sub-models are discussed in the time period model section (refer Section **Error! Reference source not found.**). There also we discuss the option of moving away from 24 hour average costs.

Parking Costs

A description of the patterns of parking in the CBDs will be established, in terms of the type of parking, duration of parking, and parking rates, from which average parking costs will be determined and input to the model in the base and future scenarios.

Modes

Active Modes

In discussion with the peer reviewer we have agreed some simplifying approaches to short distance trips, as follows. The general principle is to remove these trips prior to the distribution and mode choice modelling (DMS) and deal with them separately.

The DMS will be developed for mechanised modes only (refer also to intrazonal trips task).

Private transport

The models will be calibrated on all light vehicle users combined (car drivers and passengers, and LCVs), with average occupancy factors used to compute the vehicle trip matrices.

Car Passengers and HOVs

We have rejected the alternative of a car passenger mode in the mode choice model for the following reasons. For most purposes, this is inappropriate where groups of family or friends pursuing the same activity travel together, and thus independent travel is not an option. It is probably more relevant for commuting where, in our view, the most appropriate approach to predicting occupancy changes is that generally used in US practice, of modelling the choice between sole driver and shared car (possibly further subdivided into groups of 2, 3 etc). Such models require additional data to that which has been collected in the ART3 surveys (for commuting car drivers, the trip purpose of non-household car passengers) but, in any case, in our experience this type of model cannot be guaranteed to provide sensible forecasts of changes in car occupancy without additional modifications (e.g. by incorporating the probable saturation level for car occupancy).

The design of a future HOV add-on module is given in a separate specification, compatible with the ART3 specification. The development of the required relationships and the software implementation is outside our present remit.

Public Transport Sub-Mode Choice

A recurring issue with these models is deciding whether to create a logit model of the choice between the public transport sub-modes (in this case, bus and rail). While there are obvious advantages in terms of model outputs (matrices by sub-mode can be extracted), there are disadvantages to this approach where there are many multi-leg public transport trips using both modes; there are also significant extra model running costs, arising from the process of identifying the single-mode routes through the public transport network, so that the generalised costs by sub-mode can be skimmed off and input to the logit model.

For ART3, because of the disadvantages of the alternative approach, we propose to remain with the existing ART approach of sub-mode choice within the assignment path-building process. In the proposed future update of the APT, ARC may review whether to retain the APT logit modelling of sub-model choice or switch to a path-building approach.

Model Development

Estimation Principles

A database combining the HTS and intercept surveys will be used for model calibration.

The in-principle approach is to estimate the model on the unexpanded survey data[1], but this is not so easy when the data source is a mixture of surveys each having different sampling rates. We therefore propose to estimate the models on expanded data, but adjust the reported model significance tests to allow for the actual sample sizes.

As for all other aspects of demand modelling, the distribution and mode choice models will be based on P/A data and output P/A matrices (with the exception of NHB trips).

[1] As doing otherwise undermines the statistical measures. A common practice is to estimate the cost coefficients on unexpanded data and then do a final fit to the expanded data in which the alternative specific constants are estimated.

Intrazonal Trips

Consistent with the discussion of active mode trips, intrazonal trips will be excluded from the estimation. Following the estimation, intrazonal costs or zonal constants will be estimated such that the proportions of intrazonal trips are maintained when the intrazonal trips are included in DMS. In application the trip ends, including intrazonals, from all mechanised trips are used in DMS, with the interzonal costs used in conjunction with the estimated intrazonal costs or constants.

RGS Mixed Use Issues

ARC had raised the objective of attempting to reflect the impacts on travel patterns of RGS ideas on mixed developments (including transit oriented development, TOD), although it has been recognised that there is little evidence to go on in Auckland. We propose, for the purposes of the distribution and mode choice analysis, to identify such areas and test whether there are significant behavioural differences that may be incorporated in these models.

The definition of 'mixed use' is to be agreed with ARC but could be something like: any residential zone also having within its local hinterland a minimum ratio (to residents) of employment (for work) and retail employment (for shopping); both the hinterland and the ratios would be defined on the data in liaison with ARC. Input from the ARC on the location of any TOD areas will also be sought.

The mixed use and TOD areas could then, for example, be tested for different PT shares/trip rates, and different average trip lengths. While we suspect that the data may not in the end support this analysis, perhaps simply because of the paucity of mixed use areas, we suggest that it is worthwhile at least investigating. Given the data issues, this is likely to involve a k-factor structure at best.

Hierarchy

The model hierarchy options are pre-, post- and simultaneous distribution/mode choice.

Other Detailed Issues

There are other detailed model specification issues to consider, including:

- how residents' external trips[2] and non-residents' internal trips[3] are handled (initially at least these will not be distinguished from other trips in the estimation);
- we expect to model HBW and HBEd using a doubly-constrained distribution model and all other segments would be production-constrained, work and education trips being constrained to the number of work and school/student places, but no such attraction constraints exist for other purposes; this affects implementation but not estimation.

[2] Trips from outside the study area.

[3] Trips in the study area by people living outside the study area.

Task 8.1 Create Analysis Files/Set up Statistical Procedures

Important Note

This task runs concurrently with Task 8.2, providing the trip matrices which are an input to Task 8.2, and processing the network data which is an output of Task 8.2.

Inputs

Survey Data Base

Networks by time period

Estimation software: LIMDEP and SKM's max likelihood package (SKMAXL)

Processing

Specify requirements: the calibration process requires a host of files of varied structures, a wide range of diagnostic outputs and a complex sequence of model estimations using LIMDEP. The number of data files required is substantial and the pre-assembly of the data is therefore a major task.

<u>Preliminary Processing Tasks</u>

- create private transport and public transport trip matrices (refer Task 1.11)
- based on economic inputs, specify and extract generalised cost components from networks (Task 8.2)
- set up parking costs
- time period processing to give time classifications and average tour generalised cost matrices by mode and purpose (compatible with Task 12.3)
- identify "mixed use zones"

<u>Files</u>

- average generalised tour cost matrices by mode and purpose
- trip matrices by mode and segment

<u>Diagnostics</u>

So complicated is model calibration that it is very important to set up diagnostic outputs which help the analyst understand model performance, identify deficiencies and possible solutions. All the following examples of diagnostics are aggregations of the data which simplify it and make it possible to appreciate underlying behavioural relationships:

- aggregate the trip matrices and trip ends to sectors (typically it is good to have at least 2 levels of aggregation, for example a simple 4*4 structure and perhaps another with 8-10 sectors); procedures should be set up to produce these aggregations for the each calibrated model, together with a comparison with the observed data (e.g. predicted versus observed trips at a sector level);
- for the distribution models: trip length/cost distributions; it is always worth doing both as the fit to each is usually different and, although calibrations are based on trip costs, to achieve a good model validation we need a good representation of trip km;
- for the mode choice models, plots of mode shares versus difference in generalised costs between public and private transport; for this to be useful, the costs should be aggregated into relatively few 'bins' (defined in relation to the sizes available for each trip purpose) to be perceived; also observed vs modelled mode shares at a sector level;

- statistical tests (compute average expansion factors).

<u>Estimation</u>

- LIMDEP and SKMAXL set-ups to be established for all 3 model structures, and diagnostic outputs
- sequence of estimation steps and decision points to be specified

Outputs

Calibration files

Statistical package set-ups

Output diagnostic procedures

Task 8.2 Assignment of Observed Private Vehicle Matrices

Purpose

There are 2 objectives of this task:

- to obtain an insight into under-reporting in the survey data;
- to generate realistic road and PT costs for use in model calibration.

The second point overlaps with Task 5, in which the roading and PT networks are implemented and tested. As such they will need to be carried out in conjunction with this task, which is concerned with obtaining generalised costs for the model calibration. Having no prior information that we can use in relation to the ART3 zone system, we will establish initial values by assigning the observed matrices from the survey data (HTS and external cordon interviews) and undertaking initial comparisons with observed data (counts on screenlines, travel times). Overall adjustments will cover under-reporting or traffic (trucks) omitted from the surveys.

The distinction between the private and commercial vans/utes will be made in the household and road surveys on the basis of trip purpose and in the traffic count surveys on the basis of logos on the vehicles. For the latter a sample of manual counts will be used and then applied to the tube count sites.

In the final model tuning any errors arising from the use of interim generalised cost estimates will be removed.

Inputs

Observed car and public transport trip matrices by time period

Public and private networks by time period

Screenline counts by time period, including proportions of EB LCVs

Road speed surveys by time period

Processing

<u>Road</u>

- Expand car matrices uniformly to make a broad allowance for other (commercial) vehicles not included in the observed matrices
- Assign car matrices and compare with counts; note evidence of under-reporting;
- Extract the EB LCV matrices from the HTS and compare against count data: across screenlines, and around town, commercial and sub-regional centres and estimate survey under-reporting factors (if needed);
- Using the HTS data develop factors for splitting LCVs on EB trips from the light vehicles EB trip matrices;
- Reassign vehicle matrices and compare with counts and observed speeds in each time period
- Adjust vehicle matrices by matrix scaling to obtain a reasonable representation of road journey times in each time period, in comparison with the observed times;

<u>Public Transport</u>

- Assign PT matrices and compare with counts; note evidence of under-reporting by sub-mode.

Outputs

Factors for adjusting EB LCV demands to account for bias or under-reporting

Factors for splitting out LCV trips from the EB light vehicle matrix

Road and PT networks with realistic journey times and costs for input to model estimation

An appreciation of matrix biases by mode and sub-mode, which may affect later tasks

Task 8.3 Calibration

Inputs

Calibration files

Processing

Understanding the Trip and Cost Data

We need an appreciation of the nature of our travel data and at the same time get a first insight into the key calibration issues. Data analyses should look at:

- mean trip costs by purpose, segment and mode (car and PT)
- trip cost distributions by purpose, segment and mode
- mode shares by segment by distance (or car generalised cost), probably distinguishing the CBD from other areas
- mode shares by cost difference (car-PT) as above

Initial Calibration Programme

The initial calibrations will use wholly generic models to investigate model structure, and later work will explore:

- geographical segmentations
- non-generic specifications

Develop initial post-distribution mode choice models

Tasks are:

- calibrate lower level: a PT:car mode choice model
- compute composite modal costs
- calibrate higher level: a distribution model
- do the hierarchy test

Calibrate a PT:car mode choice model

For each purpose, calibrate using LIMDEP based on:

- car and PT generalised costs
- joint calibration for car availability segments with generic cost parameter but separate ASCs
- ASCs attached to PT mode for captive/competition (choice to be the base segment, car to be the base mode, because these will probably have the largest sample).

The disutility for each mode is then simply the sum of the constants and parameters specific to that mode in the model:

$$U^s_{ijm} = \alpha^s_{m=pt} + \beta_m G_{ijm}$$

where:

G is the mode generalised cost

i,j are the zones

m is the mode

s is the segment (captive, choice etc)

α the mode constants attached to PT

Compute logsums

LIMDEP may conveniently output the composite costs. If not the formula is:

$$U^s_{ij*} = \ln(\Sigma_m \exp(U^s_{ijm}))$$

Calibrate a distribution model

For each purpose:

- the trip matrices are the observed matrices (PT & car combined) for the car availability segments;
- in the calibration using SKMAXL, trip productions are segmented by car availability, trip attractions are not;
- the relevant cost matrices, one for each segment s, are $U^s_{ij*;}$
- there will be a deterrence function for each segment s (fitting to the mean segment trip cost).

Hierarchy test

If the distribution model parameters are >1 then a post-distribution mode choice model is inappropriate.

<u>Develop Initial Pre-distribution Mode Choice Models</u>

Tasks are:

- calibrate combined PT and car distribution models;
- compute production zone composite costs by mode;
- calibrate a production mode choice model,
- do the hierarchy test.

Calibrate PT and car distribution models

As above, this is a joint calibration (because the two modes compete for the all-mode trip attractions), but in this case the segments are PT and car users and the relevant costs are the mode-specific generalised costs. There is no segmentation by car availability, this being designed specifically for explaining the choice of mode.

In detail, for each purpose, we calibrate a single distribution model with two segments (PT and car):

- the input trip matrices are the observed car and PT matrices;
- in the calibration, trip productions (accumulated from the observed matrices) are segmented by mode; the trip attractions are not segmented;
- the relevant cost matrices are the car and PT generalised costs;
- test a segmentation of the deterrence functions for trips attracted to Auckland CBD.

Compute production zone composite costs by mode

The disutility of mode m from the distribution model calibration can be expressed as:

$$U_{ijm} = \alpha_j + \beta_m G_{ijm}$$

Where:

G is the mode-specific generalised cost

i,j are the zones

m is the mode

β is the deterrence function, and

α is related to the attraction factor.

The composite cost for each production zone and each mode is then simply the same formula above, summing instead over all the destination zones i.e.

$$U_{i*m} = \ln(\Sigma_j \exp(U_{ijm}))$$

Calibrate a production mode choice model

As above, using:

- for car and PT generalised costs the production zone composite costs, and
- mode shares for production zones.

Hierarchy test

If the mode choice model parameters are >1 then a post-distribution mode choice model is inappropriate.

Simultaneous Distribution/Mode Choice

In previous work we have found that consistent pre- or post-distribution model structures cannot be found for some trip purposes (e.g. HBW) and that simultaneous models are better. If this situation recurs, then the simultaneous model can be estimated in SKMAXL.

Final Calibrations

The distribution models should be refined to improve the fit. Essentially this should focus on the geographical fit and whether we need to segment the deterrence parameter. So tabulate the fit at sector level and examine the discrepancies. Issues are likely to be: CBD trips, counter-peak direction trips for HBW and intra-zonal or intra-district trips. Look also at average trip lengths/costs by sector and distributions by TA in case we are getting things wrong in some places. It is very important to look at both trip distance and trip cost. If there are key screenlines across which we need to get the traffic flow correct, then construct these (approximately) from the matrices and check them.

For the mode choice models, again the checks concern geographical errors at district or TA levels. External and non-residents’ trips deserve attention.

In both models non-generic structures can be explored.

Outputs

Final model

Task 8.4 Intrazonal Trips

Scope

This task is to retrofit the estimated model to include intrazonal trips, by estimating the intrazonal costs or zonal constants required such that these proportions are maintained in application.

Inputs

Calibration files (specifically, mechanised mode trip matrices by purpose)

Processing

For each purpose apply the estimated DMS models including intrazonal trips and with estimates of intrazonal costs, adjust these costs aiming to achieve the same proportions of intrazonal trips as in the calibration files. The fitting will allow for the sampling confidence intervals for intrazonal trips.

Outputs

Intrazonal costs or zonal constants for application

Task 8.5 Distribution and Mode Choice Report

Write the Distribution and Mode Choice report