Task 8.1 Develop Model Specification

Inputs
Household survey tabulations.

Processing
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;
- they are 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 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 trip matrices;
- the calibration requirements are particularly demanding on calibration software, and calibration will be severely hampered by the use of software with limited capabilities.

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.

The preliminary specifications considered for Wellington are identified in the following tables and figures. By way of introducing the key issues:

- we adopted a process whereby:
  - (1) on the basis of a review of other models in this field and our own experience, we set out our expected alternative model structures, as the first step in limited the number of alternative model specifications which we needed to explore
  - (2) on a simplified data base, we then sought to establish the appropriate model hierarchy for each trip purpose, ie pre- or post-distribution mode choice or simultaneous distribution and mode choice, and
  - (3) finally, on the full data base, we tuned the preferred model structures to gain a good model fit; in practice, this sequence only partially succeeded and for some segments we only finally established the model hierarchy in this last stage;
- 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 of and input to the logit model;
- the approach to time periods needs to be considered; we favour developing a 24 hour distribution/mode choice model and using matrix factors to generate the trip matrices for each time period required for the assignment; but, 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 (eg commuting) while interpeak network data may be more appropriate for shopping and recreational trips;
- the combination of trip data to be used for model calibration should be defined; it may include more than the household travel survey - in the case of Wellington, for example,

because public transport trips in the household survey were few in number, we had supplemented the data collection with a rail passenger intercept survey, a survey of trips to school; we has also done an external cordon survey to intercept trips from outside the study area; we combined all of these data bases for model estimation;

- the complex statistical estimation process must be specified and set-up on appropriate software; in particular, 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; in the end, we estimated the models on expanded data, but adjusted the reported model significance tests to allow for this;
- there are many other detailed model specification issues to consider, including:
  - how external trips[2] & non-residents' trips[3] are handled;
  - how to deal with the difficult modes: car passengers and slow (or active) modes (ie walk and cycle);
  - issues relating to the representation of costs, for example intrazonal costs and parking costs;
- because the data is sparse[4], the level of aggregation and segmentation of the estimation process needs careful design;
- the number of data files required is substantial and the automated pre-assembly of the data is therefore a major task.

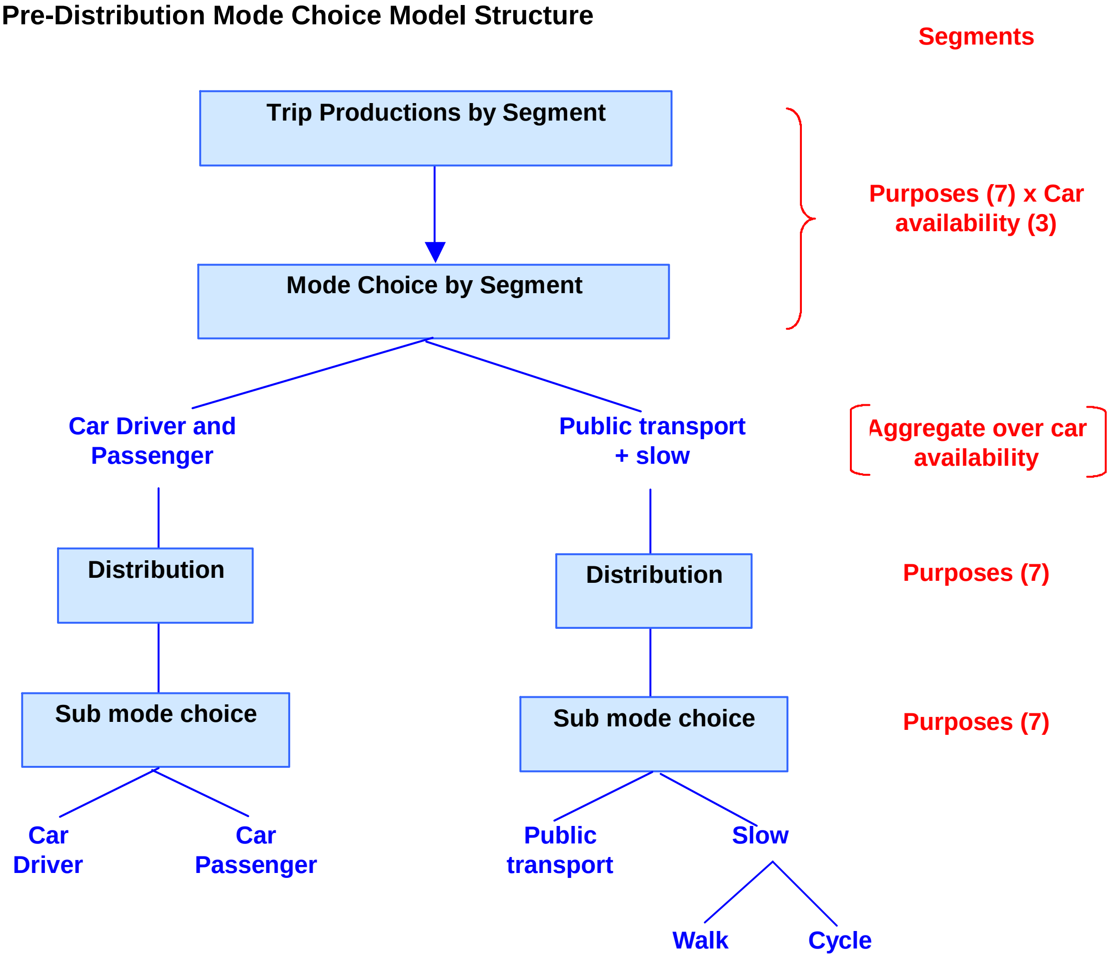
Outputs
Preferred Specification.

[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.
[2] Trips from outside the study area.
[3] Trips in the study area by people living outside the study area.
[4] A daily trip sample drawn from 2,500 households spread across 40,000 matrix cells, and further split by purpose, segment and transport mode.

Table 1 General Specification Issues

Issue	Approach
Structure	See possible alternative structures in following figures and tables. The decision was taken to model public transport trips together, subsequently allocating them between rail and bus modes in the assignment procedure - because bus is widely used as an access mode to rail in Wellington.
Time period	24 hours
Network level-of-service inputs	Although this is a 24 hour model, the level-of-service data used for each trip purpose is for the time period in which most of the trips occur: ☐ HBW & HBEd purposes: use the AM peak networks ☐ all other purposes: use the interpeak networks Alternatively, consider the feasibility of combining level-of-service for each time period in proportions appropriate to each trip purpose.
Statistics of estimation	Specify requirements
Singly/doubly constrained distribution?	Suggest: ☐ HBW & HBEd doubly-constrained: these should be constrained by workplaces and schoolplaces ☐ other purposes production-constrained: the trips attracted to shops, recreational and social destinations are not necessarily constrained by the destination characteristics
Other model specification issues which were considered	☐ model all residents and non-residents trips (ie including externals) ☐ pre- or post-distribution mode choice? ☐ should car passengers be combined with car drivers or treated as a separate mode in these models? should car passengers be forecast using a sub-mode choice model or by using a car occupancy factor? ☐ how should slow mode trips be handled (being mainly intrazonal and thus poorly represented in the relatively coarse networks)? ☐ representation of car parking costs ☐ geographic segmentation of model parameters – identification of an appropriate geographical structure ☐ calculation of intrazonal trip costs

- **Possible Pre-Distribution Mode Choice Model Structure**

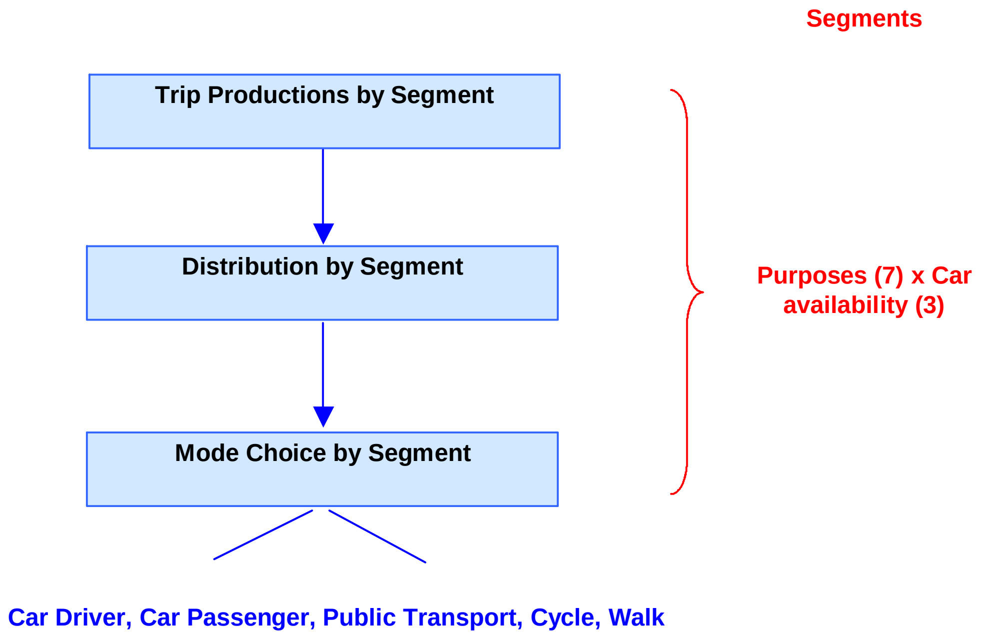


- **Specification of Pre-Distribution Mode Choice Model Structure**

Issue	Approach	Comments
Production Mode Split	Input Productions: by segment ☐ HBW: Ca, Co, Ch[5] ☐ HBEd: Ca, Co, Ch ☐ HBSh: Ca, Co, Ch ☐ HBSo: Ca, Co, Ch ☐ NHBO: Ca, Co, Ch ☐ BU: Ca, Co, Ch Modes: a 2-mode model is envisaged: (1) car passenger & driver combined and (2) public transport and slow combined.	Extent of captive/choice segmentation depends on pre-analysis. Zonal data may be too sparse for reliable estimation given low proportion of public transport trips so either aggregate zones on some basis (eg by location, or by differential accessibility) or calibrate at a disaggregate level[6] updating the alternative specific constants on aggregate data.
Distribution by mode	Segments: purpose (7) and mode (2) Inputs: ☐ productions by mode ☐ attractions by purpose (modes compete for zonal trip attractions).	With fewer segments and model parameters, zonal distribution model calibration should pose few problems.
Sub-Mode Choice	To be specified.	Decide on approach from initial data analysis.

[5] Ca: captive, Co: competition, Ch: choice. See 'segmentation' for full descriptions of the segments.
[6] Level-of-service will be at zonal level, so aggregation issues are unlikely to be significant.

- **Possible Post Distribution Mode Choice Model Structure**



- **Specification of Post-Distribution Mode Choice Model Structure**

Issue	Approach	Comments
Distribution	Input Productions: by segment ☐ HBWa: Ca, Co, Ch ☐ HBWo: Ca, Co, Ch ☐ HBEd: Ca, Co, Ch ☐ HBSh: Ca, Co, Ch ☐ HBSo: Ca, Co, Ch ☐ NHBO: Ca, Co, Ch ☐ BU: Ca, Co, Ch Input Attractions by purpose [car availability segments compete for attractions].	Extent of captive/choice segmentation depends on pre-analysis. Some concern about the sparseness of the captive and competition sub-matrices being insufficient for deterrence function determination.
Mode Choice	Segments: purpose (7) and car availability (3). Inputs ☐ matrices by segment ☐ costs by mode.	Matrices will be very sparse for the captive and competition sub-segments. At this stage it is theoretically feasible to aggregate these over purposes. Need to consider structure of mode choice model (to deal with slow modes and car passengers).