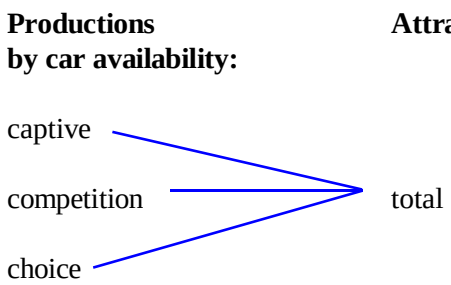
Task 8.3 Create Analysis Files/Set up Statistical Procedures

Inputs
Survey data base
Networks
Statistical calibration software

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.

Calibration software:
- statistical software up to the demand calibration task is required;
- there are a few good packages available commercially;
- we used LIMDEP for the mode choice model calibration and wrote our own maximum likelihood estimation software for the distribution model (with the advice of John Bates).

It is particularly important to have calibration software with the appropriate capabilities for distribution model calibration. For example, for our model we needed:
- maximum likelihood calibration methods;
- the ability to estimate both distribution models and simultaneous distribution/mode choice models;
- the ability to cope with model specifications where the trip productions are segmented but the trip attractions are not; eg::



- it is also important to be able to segment the models by geographical area, providing the opportunity to tune the model fit to geographical variations in behaviour.

Files:
- generalised cost matrices by mode and time period;
- trip matrices and trip ends by mode, time period and segment.

Diagnostics:
So complicated is model calibration that it is very important to set up diagnostics outputs which help the analyst understand model performance and help him 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 (eg 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 you will need a good representation of trip kms;
- 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' to smooth out variations and enable the underlying average relationship to be perceived; also observed vs modelled mode shares at a sector level;
- statistical tests.

Estimation:
- software calibration set-ups to be established, and diagnostic outputs;
- sequence of estimation steps and decision points to be specified.

Other detailed technical issues are discussed in the appendix.

Outputs
Files.
Calibration software set-up.
Output diagnostic procedures.

Appendix Detailed Technical Considerations

Segments

The expected segment aggregations for the Wellington model calibration are noted in the table. Where data is expected to be sparse, segments have been combined.

Purpose	**Segment**		
	captive	**competition**	**choice**
HBW	X	X	X
HBEd	X	X	
HBSh	X	X	
HBO (HBSo + HBO)	X	X	
NHBO	X	X	
EB (cars + vans/utes)	X		

Inputs

Daily Segment-Specific Generalised Costs
From the time period matrices we need to obtain generalised costs for calibration. As discussed in the technical specification, an averaging process of some sort is required. We propose simply averaging the peak and interpeak costs, using a weighting based on the overall proportions of trips in these time periods.

The weighting is complicated by having 3 (or even 4) time periods: am peak, interpeak and pm peak, and the issue of directionality in these periods. We propose the following as the simplest possible approach:

- for each purpose, we shall calculate the proportion 'p' of daily trips in the peak periods (am and pm combined);
- the am peak network costs will be weighted by 'p' and the interpeak costs by (1-p).

Some implications of this are noted:

- pm peak costs are ignored;
- for HBEd trips this is likely to give a 50:50 weighting because all the outbound trips are made in the am peak while the return trips are all in the midday period.

Intrazonal Costs
We tested various options for setting intrazonal costs:

- as 0.5* smallest interzonal cost for each zone;
- as some preset value for all zones, eg 5 mins;
- or use values which are calibrated to ensure that the model reproduces intrazonal trips.

Understanding the Trip and Cost Data

To gain an appreciation of the nature of our travel data, the travel behaviour in our surveys, and at the same time get a first insight into the key calibration issues, initial data analyses should look at:

- mean trip costs by purpose, segment and mode;
- trip cost distributions by purpose, segment and mode;
- mode shares by segment and distance, probably distinguishing the Wellington CBD from other areas;
- mode shares by modal cost difference (car-PT generalised cost), segment and distance, probably distinguishing the Wellington CBD from other areas;
- number and % of intrazonal trips by zone.