Re-Estimation of the Sydney Strategic Travel Model

Frank Milthorpe
Transport Data Centre
New South Wales Department of Transport
Sydney, Australia
milthorf@transport.nsw.gov.au

Andrew Daly
Hague Consulting Group
Surinamestraat 4
2586 GJ Den Haag
The Netherlands
ajd@hcg.nl

Charlene Rohr
Hague Consulting Group
36 Regent Street
Cambridge
England CB2 1DB
clr@hcg.co.uk

1. BACKGROUND

The Transport Data Centre (TDC) of the NSW Department of Transport operates the Sydney Strategic Travel Model (STM), which is used to obtain forecasts of travel demand in the Sydney region. The STM was originally developed as part of the 1971 Sydney Area Transportation Study. As new data has become available, it has been incorporated into the model; however, there has been no significant change to the structure of the model since it was first designed. Whilst the model could be considered to be "state of the art" in the 1970s, it does not represent the more complex travel behaviour incorporated in travel demand models in the 1990s. It has been recognised for some time that the STM is not appropriate for the analysis of current transport policy and planning issues and that the model requires a number of enhancements to improve its forecasts and application in the current policy environment.

In 1996 the Department commissioned a consortium led by Hague Consulting Group (HCG) to design a travel demand modelling system that would provide outputs suitable for the needs of government for relevant and timely analysis of current transport and land use policies. Two of the important criteria for the model design were firstly the specification of an implementation plan which allowed improvements to be implemented incrementally in a staged manner and secondly that the components of the model design had been proven to work in other model systems.

In 1998, the HCG consortium was commissioned to implement the first of a series of studies that will be carried out to re-estimate the STM. The first stage included the estimation and implementation of all models associated with home to work travel, including the further models that are necessary to implement these, including models of licence holding, vehicle ownership and the implementation of the prototypical sampling procedure.

1.1 Structure of the Paper

The paper describes the models estimated in the first stage of the STM re-estimation. The first section, Section 2, describes the data that were available for the modelling and the basic model definitions. Section 3 describes the structure of the mode/destination choice models and presents results obtained from the model estimation. Sections 4, 5 and 6 describe the structure of the other models required for the implementation of the model system, specifically the licence models, the car ownership models and the frequency models, respectively. Section 7 summarises the model validation that was carried out in the project. Sections 8 and 9 describe the prototypical sampling procedure and the forecasting structure. Planned future improvements are discussed in Section 10.

2. MODEL DATA

There were two major sources of data available for model estimation: household travel surveys undertaken by TDC and information from the Census of Population and Housing.

TDC has undertaken a number of household travel surveys, including one-off large-scale surveys in 1971, 1981 and 1991/92 (Home Interview Survey, HIS), and a continuous survey since 1997 (Household Travel Survey, HTS). These surveys obtain details of all travel undertaken by all members of the household on their allocated travel day. The geographic scope of the surveys is the Greater Sydney Metropolitan Region (shown in Figure 1). All days of the week were surveyed. For estimating models of commuting travel a subset of this data was relevant. The 1991 travel survey includes surveys of approximately 7,000 households, 20,000 people and 5,000 trips comprising travel from home to work on weekdays in the STM model area. The first wave (1 July 1997 to 30 June 1998) of the continuous travel survey includes surveys of approximately 2,200 households, 5,800 people and 1,700 trips comprising travel to work on weekdays in the STM model area. Data from the first two quarters of the second wave of the continuous survey (1 July 1998 to 31 December 1998) were also available for model estimation. As the same sampling fraction for the second wave of the continuous survey is the same as the first year, the available data is approximately half of that of the first wave.

The most recent five yearly Census of Housing and Population was conducted on 6 August 1996 by the Australian Bureau of Statistics (ABS). This is a complete enumeration of all people within Australia on Census day. As part of the Census, employed people are asked to provide their workplace address. These addresses are coded to the zones that are used in the STM. Also, employed people indicated the mode(s) used to travel to work on Census day. This data is referred to as Journey To Work (JTW) data. In general, the Census JTW data source is more reliable than the home interview data, because the sample size is larger. The home interview data, however, are more suited to the estimation of disaggregate models, because they offer detailed information for each traveller. There were further complications in the use of the Census JTW data for the estimation of the mode/destination choice models, specifically that a number of key variables are missing, for example information on licence holding and time of travel, and secondly that the JTW information is available only in a restrictive set of cross-tabulations. Finally, small numbers in the JTW table cells have been randomised by ABS to preserve confidentiality.

Initially, it had been considered that the JTW would be used jointly for estimation of the mode/destination choice models. However, because of the aggregate structure of the data and its limitations, and because the model parameters were accurately estimated without it, it was decided to use the JTW data for validation purposes only.

Because behavioural data existed for 1991/92, 1996 and 1997+, network information was required to explain these choices for the subsequent modelling. These data were supplied by TDC.

2.1 The Model Scope

The model area is the Sydney Statistical Division, shown in Figure 1. This area includes Sydney, the Blue Mountains and the Central Coast (Gosford and Wyong). This is an area covering 12,100 square kilometres (4,700 square miles) and in 1999 had an estimated population of 4,040,000. The population growth is currently just over 1% per annum.

Figure 1: Model Area

It was recommended that the basis of the modelling be tours. A simple tour – out and return – is the amount of travel that is required in order to take part in an activity outside the home. It is therefore logical in connection with activity modelling to model the tour as a unit rather than to split it into loosely connected outbound and homebound trips. Whilst the travel surveys provide information on intermediate travel to and from work, the JTW does not contain this information. The intention is thus that the traveller is modelled as taking a decision about the activity and the total required travel together. Decisions about travel frequency, destination, mode and time of day then relate to the entire tour, ensuring consistency in the predictions regarding outbound and homebound sections. The key advantage in the use of tours is the coherence imposed on the modelling.

Four periods of the day were distinguished in the modelling: morning peak, evening peak, daytime (between the peaks) and evening/night. 'Shoulder' periods were also defined to represent the transition between congested peaks and the less congested day and evening/night periods. For model estimation, it was required that each tour be assigned network level-of-service data that best fitted with the times of day at which the tour was made. This follows from the basic assumption in the modelling that time-of-day is fixed by decisions exogenous to the transport sector.

3. MODE-DESTINATION CHOICE MODEL

The models of mode and destination choice were estimated jointly. This was not because it is believed that the decisions on these aspects of the work trip are in fact taken simultaneously, in fact no assumption on this point is necessary. The requirement for simultaneous modelling and in particular for simultaneous model estimation arises from considerations of efficiency in the modelling.

The alternatives in the mode-destination model are the full interaction of the modes and destinations, i.e. each mode is potentially available for each destination.

Seven mode alternatives were defined in the models:

- Car driver;
- Car passenger;
- Rail (possibly with bus access);
- Bus only;
- Bike;
- Walk;
- Taxi.

Mode alternatives were 'available' to specific observations depending on a number of conditions. For example the 'car-driver' alternative was available to respondents if they had a licence and if there was at least one car in their household. A constant was also applied to the car-driver utility for cases where there are fewer cars than persons with licences, in the household, to reflect the lower probability of driving a car in this situation.

Different car passenger availability conditions were tested in the model estimation procedure. In the final model formulation, the car passenger alternative was specified to be available to all respondents. A constant was also applied to the car passenger utility if there was a car in the household and if there were other persons in the household with licences, to reflect the higher probability of being able to travel as a car passenger in these cases.

Train and bus were available to respondents if train (or light rail or ferry) or bus services were available between the respondent's origin and destination. Taxi and non-motorised modes were available to all travellers, except that walking was not considered to be available for one-way journeys of over 10 kilometres.

There are between 800 and 900 zones in the model area: rather more zones being defined for the 1996/97/98 JTW/HTS data than for the 1991/92 HIS. The combination of full destination choice for each mode presented substantial difficulty for effective processing and model estimation, because of the long computer run times (models with full mode-destination choice took about 21 hours to run on Pentium II 350MHz). For this reason a strategy of sampling was undertaken for the model estimation. For simplicity the structure of maintaining a full set of modes for each sampled destination was maintained and the sampling was undertaken across destination alternatives only. The sampling strategy took account of the relative importance of different destinations, measured by the manufacturing and non-manufacturing employment in the zone, and the distance of the zone from the workers' home. The final models were estimated without sampling, for reasons explained below.

As mentioned earlier, the modelling was done using two separate data sources: the HIS and HTS data. While there were many similarities between these sources, for example the questionnaire forms were quite similar, there was reason also to question whether the same model could be applied directly to both data sets. In particular, the level-of-service data that was available was more appropriate to the HTS data; while adjustments were made to construct networks that were reasonably representative of the 1991/92 situation there was inevitably greater error in this case.

For these reasons, a scaling factor was applied to the HIS data to take account of the potential for greater error in modelling the HIS than the HTS data. This procedure has become standard practice when dealing with multiple data sources, although it is perhaps best known when merging stated preference data with revealed preferences. In the modelling results, it was found that a factor of about 0.9 was needed to reduce the HIS error to the level of the HTS. In application in the ALOGIT software, this scaling factor is applied by setting up an artificial tree structure.

Further tree structures were investigated for the modelling of mode and destination choice. The unobserved variables affecting these two choices are quite different and it is natural to expect that the net error brought about by their omission is different. In a model of commuter behaviour, in particular, it is natural to expect that the unobserved variables affecting destination choice, i.e. whether the traveller has a job in a particular zone, would introduce a larger variance into the model than unobserved influences on mode choice. However, there is no guarantee that the model will work out in this way. For this reason, two tree structures were tested:

'Tree' 1: mode choice under destination choice;
'Tree' 2: destination choice under mode choice.

The 'Tree '1 structure implies that the destination choice error is greater than the mode choice error; the 'Tree' 2 structure implies that the mode choice error is greater. It was assumed that the same tree structure is applied to both data sources.

The tests of model structure were done without sampling destination alternatives, for two reasons. First, the structural tests can prove to be quite sensitive and a small change in the model such as eliminating destination sampling, particularly since it is likely to impact destination choice more than mode choice and since it is known to introduce a small bias, might cause different results to be obtained on the structural tests. Second, the theory of destination sampling, due to McFadden, is applicable to multinomial logit models only and although it seems likely to apply to nested models, at least approximately, the abandonment of sampling at this point eliminates the theoretical problem entirely.

The structural coefficients for the two tree structures were found to be:

'Tree' 1: modes under destinations: 0.9188 (t= 63.6);
'Tree' 2: destinations under modes: 0.8006 (t= 23.1).

Both parameters significantly improved the fit of the model and both were significantly less than 1. Because the Tree 1 structure was more intuitively plausible, it was selected as the final model for implementation.

During the model estimation procedure a non-linear cost term, logarithm cost, was tested. The logarithmic cost formulation has proved to give better results in several major modelling systems, for example in the Netherlands National Model, and this formulation led once again to an improved model fit.

Two caveats must therefore be made when discussing the model 'values of time'[1]. Firstly, it is emphasised that the resulting coefficient values, particularly the cost coefficients, are highly dependent on assumptions made about appropriate travel costs, including driving costs, parking costs and fare, which are subject to substantial error. Secondly, because of the logarithmic cost formulation the concept of a single average value of time is not supported. For validation purposes, however, an estimate of the average value of time was calculated based on the average cost for the chosen mode; this, however, is an estimate only.

Given the above caveats, the resulting values of time were found to be reasonable, for example:

- values of time for in-vehicle car[2], bus[3] and train time vary from $AUD10[4] to $12 per hour;
- access time is valued at about twice in-vehicle time;
- (other) wait time is valued more highly than access time.

Two wait time coefficients were defined in the mode/destination choice models. The 'first wait time' reflected the valuation of waiting time for the first public transport vehicle of the trip. The value of this coefficient was much lower than that of the subsequent wait time coefficient, because, with knowledge of the timetable, this time can be spent at home or at work, which is perceived to be less onerous by travellers than 'other wait time', i.e. the time spent waiting time for all other public transport vehicles used in the journey.

Separate cost coefficients were estimated for different income groups, which showed the expected pattern of higher income travellers having lower sensitivity to cost and higher values of time. The pattern for the value of car time, by income group, was clear: the value of time was monotonically increasing, but the increase was not directly proportional to the increases in income. This is consistent with the findings from other studies conducted by HCG. The same pattern across income groups was observed for the other time components.

---

1 The phrase 'value of time' is used to mean the monetary value of marginal savings or losses of travel time.
2 'Car' includes drivers, passengers and taxi users.
3 'Bus' includes bus and ferry.
4 All monetary values referred to in the paper are in $AUD.

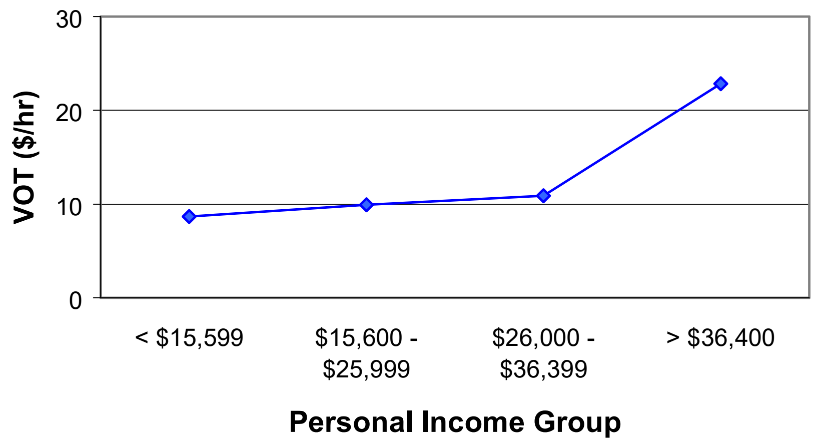

Figure 2: Values of Car Time by Personal Income Group ($/hr)

Travel segments with different behavioural characteristics within the home to work segment were identified by applying the base models to specific segments of the population, to see how the model predicted their behaviour. On the basis of this analysis, a number of socio-economic variables were incorporated into the model:

– a dummy variable for male car drivers and cyclists, expressing differential mode preferences by sex;

– a high personal income (greater than $26,000 per year) dummy variable for train use, reflecting the greater preference for trains by high-income travellers;

– a dummy variable for young (under 25) car drivers, reflecting the lower level of driving by this group;

– a dummy variable for full-time workers, reflecting their increased use of trains;

– a dummy variable for *households* with company cars, reflecting their increased driving of cars; of course it would be better to identify more precisely the *person* with the household who has use of the company car, but this was not possible in the data files currently available.

Two distance adjustment were also implemented, for manufacturing employees and for persons making journeys to part-time and second jobs, reflecting the fact that these commuters tended to make shorter journeys.

3.1 Mode/Destination Segmentation

The model structure contains a number of socio-economic segmentations.

Car Ownership and availability

The model contains 8 car ownership and car availability categories as summarised in Table 1.

	No Company Cars in Household	One or more company Cars in Household
No cars	(a)	
No personal licence but at least one household car	(b)	
Competition for car in household	(c)	(f)
Free car use, one licence in household	(d)	(g)
Free car use, several licences in household	(e)	(h)

Table 1: Car Ownership and Car Availability Segmentations

These categories are used to set a number of variables in the mode-choice model, i.e.:

- availability of car driver, requires at least one car to be owned and that the traveller holds a licence (categories c, d, e, f, g, h);
- company car dummy (f, g, h), which is applied to the car driver utility, whose availability condition excludes category b, it does not matter whether or not vehicles in category b are company cars;
- car competition dummy (c, f), which is applied to the car driver utility;
- car passenger/household car dummy (b, c, e, f, h), which is applied to the car-passenger utility; note that if there is competition for car use there must be another licensed driver in the household.

Employment Type

There are two employment type variables in the model:

- manufacturing/other industry specifications are used to define the destination attraction variable; a manufacturing distance variable is also defined to reflect that manufacturing employees tend to make shorter commuting trips;
- part-time and other job commuters/full time specifications are used in the specification of parking costs; a part-time/other job distance variable is also defined to reflect that part-time employees and employees travelling to second jobs tend to make shorter commuting trips and a dummy variable is included on rail travel for full-time workers.

Income Categories

The model contains the following 4 categories for personal income and an income dummy for rail travel for travellers with an annual income greater than $26,000:

- Income less than $15,599 per year;
- Income between $15,600 and $25,999 per year;
- Income between $26,000 and $36,399 per year;
- Income greater than $36,400 per year.

Other Socio-economic Effects

The model contains two other dummy variables for other socio-economic effects:

- travellers under 25, applied to car-driver alternatives;

- male travellers, applied to car-driver and bike alternatives.

The variables described above, if used directly for implementation, would imply a segmentation with 8 * 4 * 4 * 4 = 512 categories. The 'other' socio-economic effects segmentation was less important in explaining individual behavioural variation and were considered less important for policy than the first three segmentations. Therefore in this model the number of categories was reduced to 128 by eliminating this segmentation. The ***proportions*** of travellers under 25 and male travellers were used instead to determine the utility value for an 'average' individual.

4. LICENCE HOLDING MODEL

The modelling of licence holding was included in the system for two main reasons.

- A better model of mode choice, and ultimately of destination choice, is obtained when account is taken of the availability of car driver as an alternative for travel to work (in Sydney it is the dominant alternative). Car availability depends not only on car ownership but also on licence holding.

- Car ownership typically shows an increase over time even when account is taken of increasing income. Many car ownership models include a 'trend' variable but this is unsatisfactory because there is no basis for projecting the trend into the future. At least part of the increase in car ownership can be attributed to an increase in licence holding which can be projected on a rational basis into the future.

The model for predicting licence holding was based on methods which have previously been applied in several European contexts but that were adapted after some original research based on Sydney data. It was found that in Sydney, as in Europe, the main changes in licence holding in the last 20 years have been, and probably in the next 20 years also will be, that women's licence-holding is 'catching up' with that of men. This occurs not through the general acquisition of licences by women of all ages but by the gradual replacement of generations of women who had low licence holding by younger women whose licence holding has always been very similar to that of men. However, unlike the contexts studied in Europe, the situation in Sydney is complicated by substantial immigration from countries where women's licence holding is substantially lower than that of men. If this immigration continues, as can be expected, the difference between men's and women's licence holding will survive for longer than would otherwise be the case. The basic effects and those of immigration can be incorporated in the cohort methods that have been applied in previous European studies; a cohort forecasting system was therefore set up to predict licence holding rates.

This cohort system applies to the whole of the model area and distinguishes only sex and (5-year) age cohorts. To obtain more geographically and socio-economically specific information, disaggregate models were estimated from the HIS and HTS data, to allocate the total licences to households and persons in specific zones and socio-economic groups. Two separate allocation models were estimated: one for the joint licence holding of the first two people in the household (or for people living alone), the other for any other adults. These models contained variables on household structure (age, sex, number of children, number of adults), employment and income. The disaggregate and cohort models were implemented together to form the forecasting system.

5. CAR OWNERSHIP MODELS

The car ownership modelling is based on two linked models. The first predicts the holding of company cars by households; the second predicts the total number of cars that will be owned, *conditional on the company car ownership*. This structure was found to work better than alternative approaches.

Both models predict car ownership conditional on the household income; a logarithmic income term was included in both models, which was found to significantly improve the fit of the models relative to the use of a linear income variable. In the total car model, the income term was reduced by the average car ownership costs for the cars owned by the household. The average costs were derived from model tests, which indicated that the optimal model fit occurred assuming average car ownership costs of about $10,000 per annum.[5]

Both models take into account the number of driving licences in the household, reflecting increased car ownership with increasing numbers of household licences. A further variable in the total car model is the accessibility benefit that households derive from different levels of car ownership.

The car ownership models are therefore able to respond to a wide range of policy variables and exogenous developments and appear to be of a higher quality than most models used for this type of forecasting in current practice.

6. TRAVEL FREQUENCY MODELLING

The travel frequency model predicts the number of tours made to work, by ‘employed’ individuals. In this study, the definition of employed persons includes those who may work part-time or who may undertake casual or voluntary work.

The frequency model is made up of two models:

- the first model predicts whether ***any*** work tours will be made;

---

[5] This ‘cost’ is compared with pre-tax income; it can therefore be seen as the additional pre-tax income necessary to pay for a car, substantially higher than the amount actually paid.

- the second model predicts the extent to which repeat tours will be made, given that at least one work tour is made.

The first model is a simple binary choice, for which a logit model is used. The second model is set up in a repeating structure, to allow account to be taken of the small number of cases in which more than two tours are made. Thus the first model is applied once for each employed person; the second model is applied to each traveller, in principle, an indefinite number of times. The structure of the model is shown in Figure 3 below.

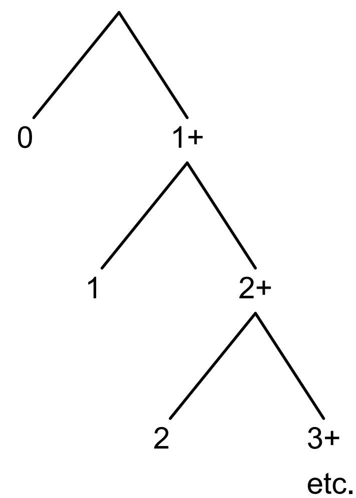

Figure 3: Structure of the Frequency Model

The first model thus predicts the choice between 0 and 1+ tours. The choices between 1 and 2+, 2 and 3+, etc. could in principle be made by separate models. However, in practice there are so few observations of multiple tours that it is more convenient to apply the same model to each of the successive choices.

In principle, it could be argued that the above model system should be linked by ‘logsum’ variables, each choice being influenced by the possibility for further choices. In practice, however, data limitations imply that the second model is very simple and the contribution of the logsum to the quality of the model would not be worth the considerable additional complexity that it would involve.

In contrast, a logsum linkage is provided from the mode-destination models. This linkage allows changes in the quality of transport connections to have an influence on the frequency of travel. While it may be surprising to see the frequency of travel to work modelled as dependent on transport quality, accessibility effects were found in the Sydney models and have been found in a number of previous studies. The findings indicate that there is evidence to support that workers are able to organise themselves to visit their workplace more flexibly.

7. MODEL VALIDATION

Model validation was carried out at three levels. Firstly, validation tests of the disaggregate models were conducted through examination of validation tables (comparing observed and predicted choices), examination of coefficient ratios, e.g. values of time and production of elasticities. Secondly, the mode/destination choice results were compared with the census Journey-to-Work data. Thirdly, validation tests were undertaken by running the model system, in its aggregate implementation, to derive base year matrices and global statistics and to extract the elasticities with respect to a number of key variables at this level. Partly this was a check on the correct programming of the separate models, but also it formed a check on the extent to which the estimation data set was fully representative of the travelling population.

Figure 4 shows the validation tables comparing the observed and predicted trip length distributions for the tours (outward and return legs). The zero distance category reflects intrazonal trips. The results indicate that the model fit is very good. There was also good fit observed in other dimensions.

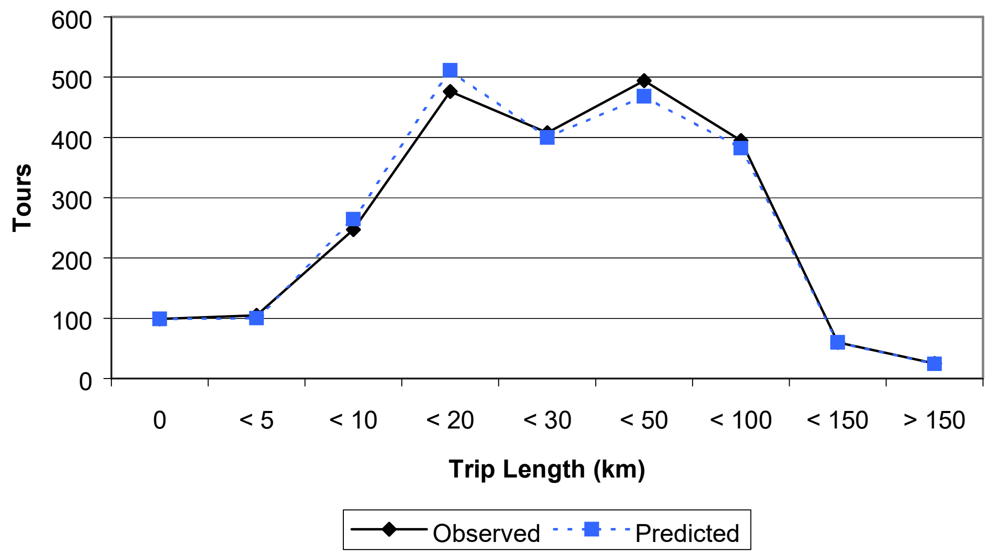

Figure 4: Observed and Predicted Trip Length Distributions, for HTS

Five elasticity tests were made on the aggregate model system. Firstly, in order to test income elasticity of the licence and car ownership models, a test was carried out to examine the effect of increasing incomes by 10%. In this test, no change was made to the total number of licences, but instead the income was increased in the models that allocate licences to individuals in the household. This is somewhat artificial in that real income changes would occur simultaneously with general increases in licence holding over time. The model test reflects the changes in licences due to income only.

The results indicated that car ownership levels increased as a result of the increased household income and the increased licence-holding predicted from the licence-holding models. The car ownership elasticities, by number of workers per household, are presented in Table 2.

Workers	Base Model		Household Income Increased by 10%		Elasticity	
	Mean Company Cars	Mean Total Cars	Mean Company Cars	Mean Total Cars	Company Cars	Total Cars
0	0.0000	0.7605	0.0000	0.7775	n/a	0.224
1	0.1840	1.3051	0.1975	1.3256	0.734	0.157
2	0.3252	1.7984	0.3474	1.8211	0.683	0.126
3	0.3825	2.5309	0.4090	2.5594	0.693	0.113
4	0.4489	2.9816	0.4803	2.9970	0.700	0.052
Total	0.1925	1.3955	0.2059	1.4160	0.696	0.147

Table 2: Car Ownership changes with Increased Household Income (10%)

The elasticity for company car ownership is much higher than that for total cars: 0.70 compared with 0.15, respectively. Company car ownership elasticities also increase as the number of workers per household increases. The total car ownership elasticities are lower than those observed in Europe. They are, however, consistent with the pattern of car ownership observed in Sydney, which is illustrated in Figure 5.

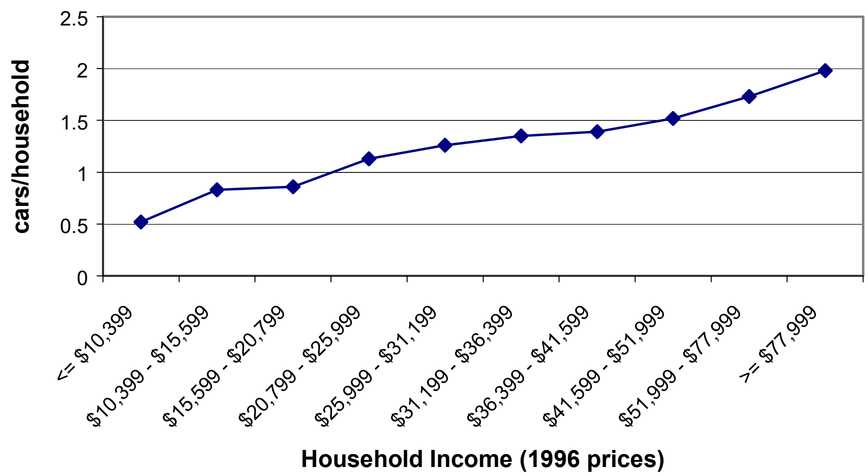

Figure 5: Mean Cars per Household, by Household Income Group (HTS)

The graph showing the relationship between cars and income has a surprisingly shallow slope: an income increase of a factor of 7.5 (from $A 10,400 to 78,000) produces an increase in car ownership of a factor of 2.6 (from about 0.7 to about 1.8), an apparent elasticity of 0.35. However, when the other relevant variables, such as numbers of workers, numbers of licences and numbers of adults, all of which are strongly correlated with income, are taken into account, it is clear that the elasticity relating to income alone must be substantially less.

An elasticity test was carried out to determine the sensitivity of accessibility in the frequency model. Specifically, a test was made where the car travel time was reduced by 10%. The resulting elasticities for this accessibility change were small:

'Main' model:	–0.032
'Stop-Go' model:	–0.091
Total Tours:	–0.034

The credibility of the model is enhanced, however, by including these small but properly estimated elasticities rather than by assuming they are zero.

Four further 'transport' elasticity tests were conducted, incorporating all components of the demand forecast system (all adjustments were applied to the complete journey tour):

- 10% reduction in car costs;
- 10% reduction in car times;
- 10% reduction in pt costs;
- 10% reduction in pt times.

The results of the tests are set out in the following tables. For simplicity, taxi has been grouped with car passenger as 'passenger' and walk and cycle have been grouped as 'non-motorised'. Note that the passenger/taxi shares increase when car time is reduced, a negative elasticity, because their travel time is also set equal to the car travel time. The first table shows the resulting mode change elasticities.

	Car costs	Car time	PT costs	PT time
Car-Driver	–0.11	–0.23	+0.07	+0.14
Passenger/Taxi	+0.24	–0.36	+0.18	+0.28
Train etc.	+0.20	+0.74	–0.32	–0.59
Bus	+0.18	+0.61	–0.35	–0.60
Non-motorised	+0.19	+0.59	+0.20	+0.27

Table 3: Tour Elasticities

In general, the elasticities appear reasonable. Comparison with European values (Netherlands National Model and the 'Antonin' model of Paris) suggest that they are certainly comparable.

Because the model reflects both mode and destination choice changes, the policy tests will result in both mode and destination choice changes, the latter leading to changes in trip lengths. Table 4 presents the resulting elasticities taking into account changes in kilometrage, by mode.

	Car costs	Car time	PT costs	PT time
Car-Driver	–0.12	–0.93	+0.06	+0.14
Passenger/Taxi	+0.25	–0.85	+0.18	+0.28
Train etc.	+0.21	+0.86	–0.33	–0.84
Bus	+0.21	+0.73	–0.36	–0.99
Non-motorised	+0.19	+0.58	+0.22	+0.29

Table 4: Kilometrage Elasticities

8. PROTOTYPICAL SAMPLING

The objective of the prototypical sampling procedure is to provide a description of the population of Sydney that is representative of the future socio-economic and spatial distribution of the population. The level of detail of information in that description must be sufficient to provide the inputs to the transport demand models that have been estimated. However, the forecasts that can be made by the responsible planning and demographic agencies fall well short of this level of detail. The solution that is offered by the prototypical sampling procedure is to assume that the future population will be generally like the present population, apart from the changes indicated by the specific forecasts that the responsible agencies are able to make.

Thus the prototypical sample represents a balance between a base-year description of the population in detail and a future-year sketch of some of the key characteristics. The most convenient way to obtain the base-year description is from detailed household interview survey files and the most convenient way to present the output of the prototypical sampling process is as adjusted files from such an interview. In a sense, output of this kind can be seen as what might be the product of a future home interview.

These methods involve the adjustment of the weights which expand the survey records to represent the population. By changing these weights, a set of records can be made to be representative of a different area and/or a different time. The specific method used to find the new weights is to minimise an objective function which contains terms measuring both the change from the base population (i.e. the change in the weights) and the divergence from the future summary statistics. This objective function thus makes explicit the balance between the base and the future.

The function that is chosen for minimisation is simply the weighted sum of squares of changes in the base weights plus the weighted sum of squares of divergences from the future population description. This is a quadratic function, so the minimisation process is a quadratic optimisation and the program that performs it is called QUAD. The process is complicated only by the fact that it is necessary to constrain the weights to be positive, so that an iterative procedure is needed to find the optimum. The use of weights in the procedure allows the balance between base and future to be adjusted; additionally the weights can be used to give more priority to matching some of the future statistics – known as ‘targets’ – than others. Differential weights for different targets would be appropriate, for example, when some targets were more important or more accurately forecast than others. Weights can also be set to zero when *no* forecasts can be made for specific targets items.

Two inputs are therefore required for the prototypical sampling procedure. The first is a representative base sample of data. The most natural base-year sample to use for this procedure was the HTS sample. This gives a data set of adequate size to represent the important sub-groups of the population and it contains exactly the variables that are needed to exercise the new STM models. However, because this sample is representative of the model area as a whole, when geographically specific information is required, *even in the base year*, the prototypical sampling procedure has to be applied to adjust the weights on the HTS records to make the sample representative of the population of a zone in the base or in a future year.

Secondly, targets describing the future population characteristics are required. One of the key advantages of the prototypical sampling procedure is its flexibility with respect to the data inputs that can be accommodated. For the Stage 1 Sydney Travel Model, the prototypical sampling produces forecasts for each zone to match the following 15 targets:

- population, age-sex cohorts in four age groups: 0-20, 20-40, 40-60, 60+;
- households, subdivided into single persons, couples with children, couples without children, single parent families and other households;
- workforce, manufacturing and non-manufacturing.
- income is *not* be used as a target.

Income is excluded as a target because no information is available on the income distribution; instead an overall factor will be applied to incomes. This factor is applied to all of the incomes in the prototypical sample, *after* the recalculation of the weights is done; this order of calculation is necessary because the recalculation of the weights will of itself change the incomes by a small amount.

The output of the prototypical procedure is used in two ways. First, at model area level, factors are calculated giving the change in the population in each category for use in forecasting licence-holding and car ownership. Second, for each zone, the totals in each of the segments required by the travel demand modelling are calculated from the QUAD output and the output of the licence and car forecasting modules.

9. MODEL IMPLEMENTATION

The model implementation system comprises two separate components:

- the ‘Population Model’;
- the ‘Travel Demand Model’.

The Population Model contains the prototypical sampling procedure (PROTOSAM), the licence-cohort projection module and the licence-holding and car ownership models. All models within the Population Model use sample enumeration procedures and are run outside the EMME/2 environment. The car ownership forecasts made at this stage do not take account of accessibility changes.

The travel demand forecasting model, implemented in EMME/2, calculates accessibility, applies an adjustment to car availability to account for accessibility changes, applies frequency and mode-destination choice models and then factors the output of these processes to obtain the matrices for assignment at different periods of the day.

A key feature of the design of the system is that the Population Forecasting component is not influenced by transport network conditions and thus does not need to be run iteratively. Thus this module can be run once only, perhaps just once for each forecast year, before the travel demand module is run repeatedly for alternative policy options, infrastructure variants etc..

9.1 Implementation of the Population Module

Within the Population Model (Figure 6), the prototypical sampling procedure is first run at model area level. Next the licence-holding models are run. The licence holding forecast requires two stages. First, using an EXCEL spreadsheet, forecasts are made of licence holding for age-sex 'cohorts' in the population. Second, forecasts are made of licence holding, using the adjusted disaggregate models, for all of the households in the prototypical sample. Two models of licence holding are applied, one for the first two adults in the household, the other, dependent on the first model and applied after it, for any other adults in the household. Car ownership, dependent on licence holding, is forecast after the licence models are applied and also requires two models. The company car model is applied first and the total car model, which depends on the number of company cars, is applied after that model.

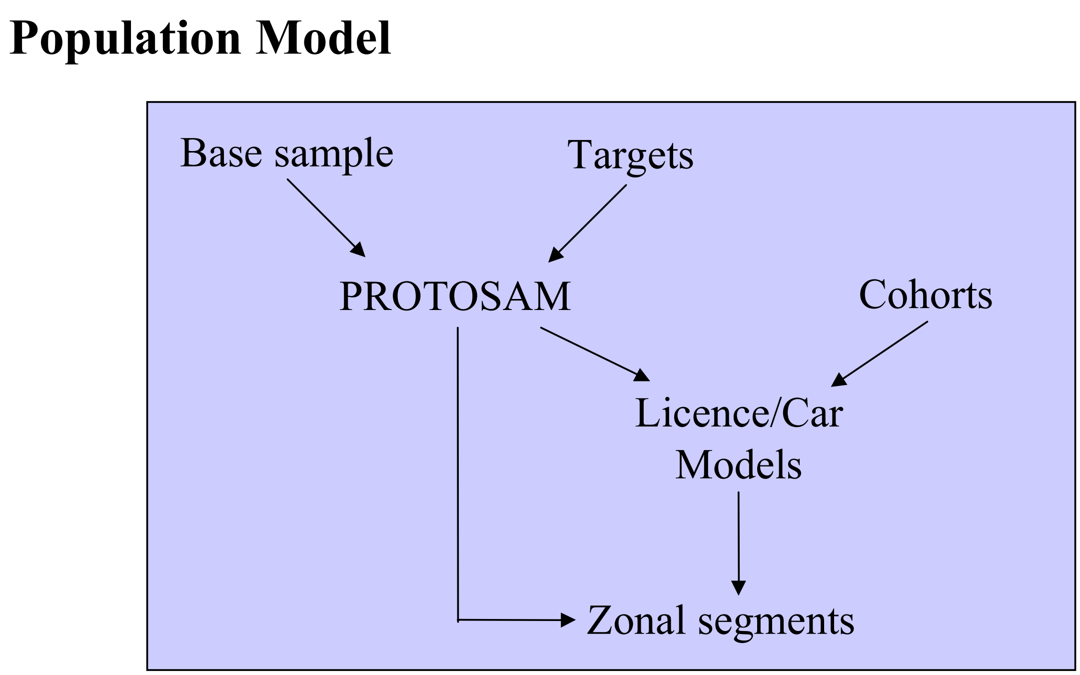

Figure 6: Population Model

All of the licence holding and the main car ownership forecasting is carried out at model area level and is based on the results of the prototypical sampling at that level. Once the forecasts from these models are available, they are allocated to travel zones using the results of the prototypical sampling at travel zone level and accumulated to form the zonal segment file which is the basis of the EMME/2 travel demand forecasting.

9.2 Implementation of the Travel Demand Module

All parts of the model system other than those which comprise the 'Population Model' are implemented within the EMME/2 environment. Figure 7 shows the components of the 'Travel Demand Model'.

Fundamental to all the applications in the Travel Demand module is the calculation of the accessibilities of each origin. In this context 'accessibilities' means the levels of accessibility of each origin zone for each car availability segment. These are used in the car availability adjustment and travel frequency modules. The accessibilities used in these two modules are segmented by car availability only, a subset of the full set of 128 segments used in the mode-destination model.

Eight accessibility values, one for each car availability segment, are calculated for each zone, for input to these two modules. This reduction to 8 values is achieved by taking standard representative segments with respect to each traveller's occupation (the value for group 3, full-time, non-manufacturing employment, is used) and income (the value for income group 3, $ 26,000 to 36,400 per annum, is used).

The car availability adjustment adjusts the availability of cars at person level to take account of changes in accessibility. This is achieved by comparing the base-year accessibility measures with corresponding forecast measures. In this way, any inconsistency between aggregate and disaggregate accessibilities are eliminated.

Accessibility is not a strong variable in the car ownership model, so the impact of accessibility changes on car availability are not large. Nevertheless, it has been considered worthwhile to include the impact of such changes in a simple way to complete the interconnection of the modules of the model. The impact of accessibility is measured by changes in the mode-destination logsum variable. Simple models are then used to predict the impact of changes in this variable on the distribution of the working population over the 8 car availability segments.

The travel frequency model predicts the numbers of tours that will be made by travellers of different types (defined by population segments), using the two frequency models that were estimated. The travel frequency is dependent on accessibility, using the same measures used for the availability adjustment, i.e. for 8 different car availability segments but not distinguishing other segmentations of the population.

The travel frequency models are however mainly a function of the segmentation variables. At this stage, segmentation takes account not only of car availability a (8 segments), employment type b (4 groups) and income level c (4 groups), which are used by the mode-destination models, but also of age (3 groups) and work status (for other than full-time workers, 5 groups). Thus the segmentation that is input to these models, specifying number of people in each segment resident in each zone, is more detailed than the segmentation that is output, specifying numbers of tours made by travellers of a given segment originating in each zone.

The mode and destination choice model forecasts that number of tours made to any destination 'j' by mode 'm' for people of segment h (h=abc) living in zone i. The output of this stage of the model system comprises the total matrices

$$S_{ijm} = \Sigma_h T_{ijhm}$$

giving the 'synthetic' matrices used in subsequent parts of the system.

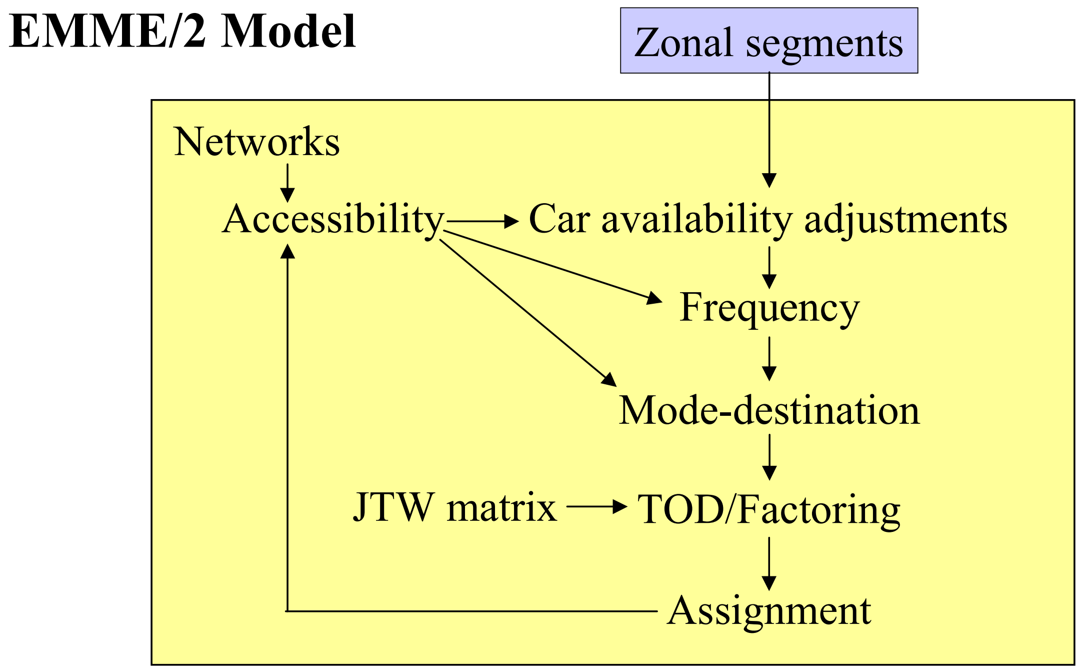

Figure 7: Travel Demand Model

The final step in the Travel Demand procedure is the splitting of predicted demand into different time periods, applying the pivot procedure to obtain best-estimate matrices for travel to work and factoring these up to reflect total demand including travel for other purposes. For the peak hours, in particular for the morning peak, this procedure should give reasonable estimates of total travel, but for day-time, evening and night travel, purposes other than work dominate and the estimates would not be adequate for use. It is noted that the base matrix used for pivoting is derived from the census Journey-to-Work data, which gives 24-hour travel. For this reason it is necessary to apply the pivoting based on this matrix *before* splitting the matrix by time-of-day.

10. FUTURE IMPROVEMENTS

This paper has described the first phase of improvements to the Sydney Strategic Travel Model. The first phase focussed on modelling home to work travel and the associated models. It is planned that additional model estimation and implementation will be undertaken in subsequent phases. These include the explicit modelling of additional purposes including home to education, home to shopping, home to other and work to work-related business travel. The model system has been implemented as a pivot point system, and it is proposed to construct better base year matrices using matrix estimation procedures.

The model system has been designed to allow flexibility in extending the scope of the model in the future as needs arise and / or practice is enhanced. Presently the allocation of the tours to specific time periods is undertaken using fixed proportions, rather than taking into account differences in travel time in the different time periods, however this allocation process could be done using a model. Also the modelling of tours could be extended to incorporate travel to intermediate destinations.

REFERENCES

Antonisse, W., Daly A., and Gunn H. (1986) The Primary Destination Tour Approach to Modelling Trip Chains. **Proceedings PTRC 14th Annual Meeting,** England.

Ben-Akiva, M. and Lerman S. R. (1985) **Discrete Choice Analysis: Theory and Application to Travel Demand.** The MIT Press, Cambridge Massachusetts.

Daly, A. J. (1998) Prototypical Sampling as a Basis for Forecasting with Disaggregate Models. **Proceedings PTRC European Transport Forum Annual Meeting**, England.

Daly, A. J. (1998) Prototypical Sampling as a Basis for Forecasting with Disaggregate Models. **Proceedings PTRC European Transport Forum Annual Meeting**, England.

Daly, A.J. (1997) Improved Methods for Trip Generation. **Proceedings PTRC European Transport Forum Annual Meeting**, England.

Domencich, T.A., and McFadden D. (1975) **Urban Travel Demand: A Behavioural Analysis.** North-Holland/American Elsevier, Amsterdam, The Netherlands.

Gunn, H. (1993) The Netherlands National Model: A Review of Seven Years of Application. **International Transactions in Operational Research, Vol. 1, No. 2**, 125-133.