

Developing the 'Ds' – Predicting mode share from urban structure

Kym Norley¹

¹Faculty of the Built Environment, UNSW Australia, Sydney NSW 2052 AUSTRALIA

Email for correspondence: kym.norley@unsw.edu.au

Abstract

The literature on Transit-oriented Development suggests that five Ds – Density, Diversity, Design, Destination accessibility and Distance to Transit are key to its success. Transit-orientation in urban planning relates to empirical or qualitative relationships between urban structure and use of public transport (transit) and to development that encourages transit use. Studies have usually been at a local level, or have utilised overly simple parameters such as population density. This paper presents the development of a suite of models to examine transit-orientation at a metropolitan scale – measured primarily as mode share – with the Ds as key variables. The research used geospatial and regression techniques to estimate transit-orientation based on demographics, urban structure, transit infrastructure and transit service at a detailed (travel zone) metropolitan-wide level. The core of this comprises a two stage logistic model suite to first predict car ownership and secondly to use this to predict rail, transit and active mode share and travel. The models utilise factorised variables to limit co-variance, and they exhibit strong r-square values for most trip purposes. They can be substituted for the mode share step in the four-step model and have been used in this way.

1. Introduction

Transport demand, economics and environment remain the significant elements of transport infrastructure assessment in planning practice; however there is a positive new emphasis on sustainability, urban form and accessibility – the subjects of smart growth and transit-oriented development. The research presented in this paper has been undertaken with this more recent emphasis, but at a metropolitan, rather than local, scale. It addresses the question:

How can the relationship between urban structure (land use) and the impacts of investing, or not investing, in transit infrastructure best be modelled?

The research described in this paper used geospatial modelling (GIS and GIS-T) and logit regression to develop and apply metrics of demographics, urban development structure (land use) and urban rail transit infrastructure to predict mode share. The research was part of a more extensive study, using quantitative methodology contextualised by literature review, archival research, and understanding of transport systems and management. The research is significant because:

- It offers a means to analyse transit-orientation at a metropolitan rather than local level
- It provides models based on urban structure, demographics and transit service as a means to predict transit mode share. This is an alternative to the usual behavioural logit models used in travel demand modelling.
- The models can be applied to calculate mode share in the context of a four step travel model – they produce outputs in the form of travel matrices that can be used for loading networks for trip assignment.
- They have been successfully applied to produce public transport forecasts and road travel demand, as well as program-based estimation of economic benefits (Norley 2013, 2015).

2. The literature

Transport is an enabler. The full benefit of an urban transport system is not achievable by the transport system in its own right. Transit must be seen as the means by which the urban structure can be connected, made accessible, and enable it to create opportunity, vitality and wealth . and be made sustainable. The importance of transport land use interaction has been understood for many years. Even so, the extent of its application in the complex, context sensitive, wicked problem of urban development remains on a selective, rather than holistic, basis.

2.1 The 'Ds'

The 3Ds of ToD have become a headline in the literature on built environment-based transit modelling. Cervero and Kockelman (1997) used the term 3Ds to refer to three dimensions of the built environment . Density, Diversity and Design . that they examined to assess the relative impact of each on travel behaviour. They noted that, before then, only density had been given serious attention. A later paper by Ewing and Cervero (2010) expanded the 3Ds to 5Ds with the addition of Distance to Transit and Destination, and suggested that Demographics and Demand Management might also be added.

Ewing and Cervero's 2010 paper was a meta-analysis of the results of a range of studies, of which they noted some 200 had been undertaken over the previous 10 years. Most of these were based on small samples, reflecting the interest in individual transit-oriented development projects over that time. Few encompassed the metropolitan scale. Based in part on this paper, the 5Ds variables are defined as follows:

- *Density*. Most usually population or residential density, measured as persons per area of land (hectare, square kilometre or acre). Employment density has occasionally been used but this is unusual. Ewing and Cervero note that activity density . the sum of population and employment densities (urbanisation) . has been used in some references and is valuable in that it references the origin and destination of home-based work trip.
- *Diversity*. This relates to the mix of population and employment or the variety of land uses in a given area. Ratios (for example jobs to population) are sometimes used, but a concept of entropy, similar to a logit transformation, is widely used in transport studies. This was used in this research and is discussed in depth later.
- *Design*. Design relates to the street characteristics of a given area. A wide variety of measures are evident, generally related to ease of walking. The term permeability is also used, primarily in the new urbanism literature
- *Destination Accessibility*. This measures the accessibility of trip destinations, which may be as simple as distance to the Central Business District or number of jobs within a certain time or distance, or it may be based on more complex measures such as that applied in the gravity model of trip attraction.
- *Distance to Transit*. This can be the distance to the nearest transit route (station) or the density of transit routes.

Some consideration had been given to this issue by others about the same time as Cervero's work. For example Handy (1996) documented various studies, mostly from the 1990s, that looked at variables such as diversity at an aggregate level. She concluded that the high level studies masked some of the behaviours and were less useful for this reason. She also documented disaggregate studies dating back to the 1980s; however noted that r-square values as low as 0.25 or less suggested much was unexplained by the variables chosen. She concluded that better application of activity-based methodologies may prove useful.

A useful application of this concept appears in an Australian paper by McKibbin (2011). McKibbin used the same dataset as the research reported in this present paper. McKibbin's work was limited to journey to work and focussed only on the Δ s. He controlled for certain important other variables such as car ownership, and he identified incorporation of these as potentially important further research. Unlike most of the North American studies, it is similar to the study reported here in that it utilises disaggregate metropolitan-wide data.

Tsai, Mulley and Clifton (2012) used geographically weighted regression (GWR) in an attempt to improve on the multivariate analysis; however their work explains significantly less of the variance (r-square of 0.26 for their global model and 0.41 for their local model) than the multivariate analysis reported in this paper. This may be in part explained by the relatively basic land use variables they used, which were based on the early 3D work (density, diversity and design) plus accessibility, measured as road distance to the CBD.

2.2 Use of geospatial techniques

Geographical information systems (GIS) readily facilitate simple visual exploration of spatial relationships. With a detailed geospatial database such as that available for Sydney (discussed below), it is evident that the areas surrounding rail lines, which are the primary high quality transit links in Sydney, enjoyed higher levels of transit mode share than other parts of the metropolis (Norley 2010). This is not just rail share; it is higher use all non-car modes collectively. Moreover, there are other determinants; notably orientation to the cities of Sydney and North Sydney (Sydney's CBD, known as the Δ Harbour Cities and the Δ Global Corridor of high end service and professional employment (the Destination Δ d). This established the viability of the hypothesis that there was a relationship between the use and influence of transit, urban structure and the rail transit system.

It is evident that the body of ToD research would be enhanced by examination of structure at the metropolitan level, particularly the role of centres within metropolitan regions (Black 2008), a metropolitan-level concept also promoted for its sustainability advantages (Bernick and Cervero 1996; Goodman and Moloney 2004; Black 2008; Litman 2009d). This is not unlike the classic bid-rent and central place theories (O'Sullivan 2009, Chapter 6, Alonso 1964). Gonçalves, Portugal et al. (2009) use this theory to support arguments that interventions to enhance the integration of land use and transport, with particular emphasis on rail nodes, can be beneficial. They present a useful visual depiction of the problems, causal factors, solutions, procedure patterns and centrality indicators on which their study is based.

3. Data preparation and correlation

For the purposes of the research reported in this paper the data used were for Sydney Australia. The principal data used fall into four groups:

- 1) Travel data, principally in the form of trip matrices. For the mode share modelling these form the dependent variables.
- 2) Demographics, comprising data on the makeup of Sydney's past, current and forecast population, employment, households and dwellings.
- 3) Built environment or urban structure measures, represented by location of population, urbanisation, diversity of employment and population and the related entropy measure, and various measures of accessibility.
- 4) Transit service, principally comprising measures of service by rail and the type of public transport provided in the area.

The data sought were by year and both forward looking and backcast . that is previous years were required. Hence as well as the base year (2006) it was necessary to generate a data set for 1986 and 1996, and for future years at 5 yearly intervals based on the BTS population and employment forecasts.

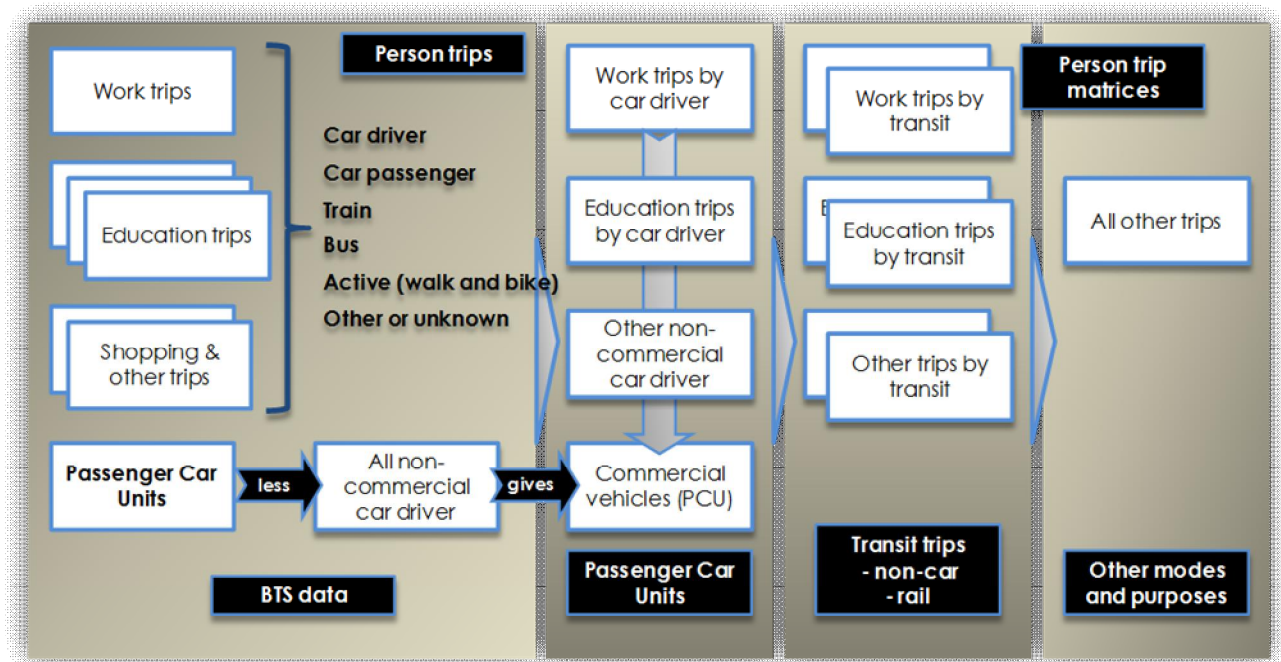
3.1 Travel Data

An extensive array of measures in dataset format is available from the NSW Bureau of Transport Statistics (BTS, formerly known as the Transport Data Centre . TDC) and the Australian Bureau of Statistics (ABS). The BTS data are generally more convenient than ABS data as, where appropriate, they have been pre-processed by the BTS from the ABS Census and structured to suit the Sydney Strategic Travel Model travel zones (Tzs) used by the Bureau. The BTS also models travel for purposes other than work on the basis of its Household Travel Survey. At the time of the modelling the 2011 data were not available, hence the analysis uses the 2006 data set, and previous years for backcasting.

The BTS data are by mode, and hence readily permit construction of models for estimation of mode share. The data are highly detailed; based on a core of 3788 travels zones, plus sundry other zones. As a matrix, used for traffic assignment and similar purposes, these form an array of in excess of 14 million cells. The source data are supplied in either comma separated files, Excel workbooks or in some cases .pdf documents. Multiple scenarios, modes and trip purposes had to be analysed. Manipulating and analysing this amount of data required considerable care and rigour.

The travel data (trip matrices) were first converted from tours (out and back) to trips (single legs by direction), transposing the tour matrix, multiplying the resultant matrices by the appropriate time of day factors and merging each time of day matrix with its transpose. The matrices were then restructured as shown in Figure 1.

Figure 1: Travel matrix structure



3.2 Other data

The other data used comprise demographic, built environment and transit service data, all of which are location based. The BTS and ABS data were supplemented with data on train services and by geospatial data (mainly proximity measures) developed using the ArcGIS® geographic information system.

The demographic data

The demographic independent variables were also largely sourced from the Bureau of Transport Statistics, with some data direct from ABS sources. The data for population, households and employment for current and forecast years (TDC 2009a, 2009b) are conveniently presented by the BTS in spreadsheet form. Backcast year data were calculated from ABS sources (Australian Bureau of Statistics 2008, 2009). The 1996 ABS data are available in Excel files by CD (Collector District) which can be converted to travel zone using equivalence data. This is not the case with the 1986 data, which are only available in ASCII file or hard copy (or PDF). Hence the travel zone populations for 1986 were estimated from the SLA data applied to the 1996 Tz distributions within the SLAs.

ABS and BTS data were both accessed to provide dwellings and household data. Transforming the ABS data to travel zones produces slightly different results than presented by the BTS. Accordingly both have been included in the database; however the TDC dwelling data are the more relevant because TDC occupied dwellings are available for future years.

The base data for the car ownership measure transformed ABS data for the 2006 year to a travel zone level. These data form the basis of the car ownership model.

Built environment and transit service

The other data that were assembled for this aspect of the research relate to built environment and transit service factors. The built environment data have been principally derived from BTS sources, in some cases using the ArcGIS proximity and feature tools. Transit service measures are largely rail service related, and are taken from published RailCorp (the Sydney rail service provider until June 2013) timetables and other sources.

Some key points associated with the various measures used are discussed below:

- Preliminary studies (Norley 2010) and professional practice recognise that a major determinant of public transport use is work trips to the CBD. A Δ % CBD tripsq parameter was used for this.
- Access to employment was based on a nominal 15 kilometre radius and is perhaps a less useful measure than time-based measures and indices that have been developed specifically for this purpose elsewhere; however it served present purposes and is relatively easy to calculate.
- Distance to the nearest rail station reflects the preliminary study (Norley op.cit.) that showed transit use to be higher in zones close to rail. It is best measured as a log (or ln) transformation to reflect the shape of the relationship.
- Urbanisation was represented by the sum of jobs and residents. As noted earlier, this measure reflects the origins and destinations for work trips. Other measures of density were applied including basic normalised measures of population and employment per hectare separately.
- Diversity has been represented as the mix of jobs and residents. Three categories were used; viz. residential population, production-based employment and service-based employment. The use of internal quantities, rather than adjoining land uses,

simplifies the calculation. The formulation was otherwise the same as that used by Cervero and Kockelman (1997); that is:

$$\text{Entropy index} = \sum \left[P_i * \frac{\ln(P_i)}{\ln(J)} \right]$$

where:

P = proportion of total population plus employment in category i

J = total number of categories

- Service measures used reflect the importance of specific destinations and the transit service levels to these. First, the number of train services to each of the CBD, to North Sydney and to Parramatta were included, both for all travel zones and then by zones within 2 kilometres of stations.
- A more detailed measure of rail service was developed for CBD services. This measured the service to the CBD (based on the Sydney CBD Town Hall station) in terms of its components, as follows:
 - Average in vehicle rail time for the nearest station to the travel zone
 - Average wait time
 - Walk time from the travel zone to the station
 - Transfer time between trains if required to access Town Hall.
- For the initial analysis a weighting of two (2 x) was adopted for the wait penalty factor plus 30 seconds boarding time, which is also allowed for cross platform transfer. This value is very conservative. A value of one (1) minute was used for suburban transfers other than cross platform and four (4) minutes for transfer at Central to and from the Terminal (Intercity) platforms, based on timings by the author.
- Finally a classification - Transit Category . was used to identify the type of public transport services (and hence the quality of service) in the travel zone concerned. Four categories are used in hierarchical order as follows:
 - 1) Rail-served, within 2,000 metres of a railway station with a direct service to the CBD.
 - 2) Areas otherwise serviced by Sydney Buses
 - 3) Areas serviced by contract bus services, offering commuter services in developed areas
 - 4) Other areas, typically serviced by infrequent rural services, if at all.

4. Correlation

Correlation between the variables was comprehensively examined in order to identify a suitable set on which to build the models in the next stage. This was done statistically, but the results were examined pragmatically in order to ensure that they stood up to common sense.

The analysis was undertaken with the following as the important dependent variables:

- Car ownership
- Work mode share
- Education (tertiary) mode share

- Business mode share
- Shop and other mode share

The mode share evaluation was undertaken on a logit basis for each of non-car and rail share of total trips. Car ownership examined predictors for total cars in each travel zone and cars per person, adult, worker and dwelling or household.

4.1 Car ownership

Car ownership is often used as an independent variable in transport studies. In the context of the built environment it may be argued that it is in fact a response to population and the availability of other resources, and hence is dependent.

Ordering the correlations by number of cars in each zone is of little value - the number of cars in each travel zone relates to the population in that zone (r -square > 0.7). Ordering by car ownership rates (per dwelling, adult or person) is rather more informative. In table 1 shading indicates the significance of the r -square value shown. The darkest shading shows correlations that are significant at the 0.1 level and the lighter shading at the 0.5 level. Only the highest r -square value correlations for the cars variable being examined are shown.

Table 1: Correlation (r-square) – Work trip non-car share sorted by cars per dwelling

	Cars per dwelling	Cars per adult
Population Density(P/Hectr)	0.48	0.49
Log of distance to rail	0.32	0.34
Percent work trips to CBD	0.29	0.28
Access to employment	0.27	0.35
Percentage of trips to CBD	0.27	0.24
Transit Category	0.27	0.26
Total time	0.22	0.28

This table shows that car ownership per dwelling correlates with measures such as density, proximity to rail, and CBD orientation. There is a strong (negative) correlation between the percentage of CBD workers and cars per dwelling. The distance to rail station also strongly affects car ownership per dwelling, again suggesting that a household will limit its car ownership if rail access is available, more so if its workers work in the CBD. Earlier work had suggested that the relationship to distance to station was in fact a log curve and this was adopted for analysis.

4.2 Mode Share

Table 2 shows the part of the correlation matrix (r -square values) for the logit of work trip non-car share with the independent variable candidates arranged in descending order of the r -square value. The Pearson correlation r values are also shown.

Table 2: Correlation – Work trip non-car share

Dependent Independent	Logit Work non car		Logit work rail
	Pearson	r-square	
Cars per adult	-0.803	0.645	0.179
Cars per dwelling	-0.798	0.637	0.158
Cars per work age	-0.764	0.584	0.190
Percent work trips to CBD	0.752	0.566	0.058
% Total trips to CBD	0.688	0.473	0.014
Log rail	-0.665	0.442	0.632
Population Density (p/hectr)	0.663	0.440	0.074

The most significant correlation is that between the non-car share and car ownership expressed on a per unit basis (per adult, dwelling or work age person), with r-square values of 0.6. The negative Pearson correlation shows that the greater the car ownership the less likely people are to use means other than car to travel to work, which makes logical sense. The next most significant determinant is the CBD orientation (percentage of work and other trips from the zone to the CBD, and the trip rate to the CBD). This puts great importance on the central business district as a destination. The Pearson correlation is positive in this case.

Noncar mode share . that is the share of modes other than car driver of car passenger, including train, bus and active modes . is also very highly correlated with the parameters that relate to rail service. These include the distance to the nearest rail station (expressed as the log of the distance), the total and in-transit (i.e. excluding access) travel times by rail to the CBD, and the transit category. In each case the sign of the Pearson correlation is logical - the greater the times or distances to rail the less non-car travel (negative sign). The negative sign for the transit category reflects the ordinal value of 1 given to rail service and higher values for the various bus categories. Access to employment is also seen as reasonably important with an r-square value of 0.37; however this may be simply a reflection of proximity to the large employment base in the CBD and Global Arc.

The various density parameters represent the next group; however the literature suggests that there may be considerable covariance between density and other parameters that lead to high non-car mode choice. The inclusion of school age children in this high ranking selection with a negative Pearson sign suggests that tours that involve child drop-off may mitigate against non-car work trips. The last parameter on this list is diversity, expressed as entropy at the SLA level, where areas more diverse have higher non-car work travel.

Correlations between these independent variables and rail mode share are less pronounced, but follow the same pattern.

Other trip purposes

Other trip purposes were explored in the same way. The major conclusions were:

- There is very little correlation between any of the independent variables measured and the (logit of) education trip non-car share . mode share for education overall does not depend on the built environment or demographics. This makes intuitive sense, with the possible exception of household income which may relate to private, selective and higher education.

- In the case of tertiary education, the universities have large populations that act as focal points for trip making; conditions that suit public transport and, for residential colleges, walk in. Technical and further education facilities generally are less concentrated. Car ownership is the most significant variable in regard to tertiary trip making, with r-square values of between 0.5 and 0.6.
- Although the non-car mode share for business trips typically is small, there is a decided relationship between the built environment and the share. The results suggest that this may be due to cross correlation effects. Unlike other trip purposes, business trips tend not to be related to car ownership per person, although per dwelling is important. The most important determinant is work trips to the CBD (r-square of 0.6) which, together with other CBD orientation variables (access to employment for example) suggests that there is CBD orientation in regard to non-car business trip making.
- For shopping and other trips, car ownership, with r-square values from 0.48 to 0.63, is the most important of the independent variables, followed by population density at 0.38. Beyond that the variables are mainly associated with CBD orientation (0.22 to 0.29), which is intuitively sound. Overall urban density (population plus jobs) has an r-square value of 0.21. After that the level of rail service is important (r-square values from 0.11 through to 0.21). It appears that diversity (SLA entropy) is has some importance, albeit modest, in choice of mode for shopping and other trips. Rail share, as before, follows a similar pattern to non-car, but not as marked.6.2.2. The GASTM-transit model suite

5. The multivariate predictive models

The modelling process was initiated with the population, employment, built environment (urban structure), transit service and travel data. The travel data were restructured into trip matrices by time of day, mode and purpose to suit the modelling. These and other data were analysed to create what were labelled the GASTM-transit (Growth Area Strategic Travel Model - transit) suite of tools. This constitutes the major outcome of the research.

The platforms with which the analysis was conducted comprised ArcGIS, TransCAD® and SPSS Statistics, with Excel used to monitor, transfer and present the analysis. Some steps required a text processing capability, and Microsoft Word and WordPad were used in this role. The primary statistical tool was SPSS Statistics®. TranCAD was used only for demonstration of the mode share modelling to assign road demand, and for some data manipulation tasks.

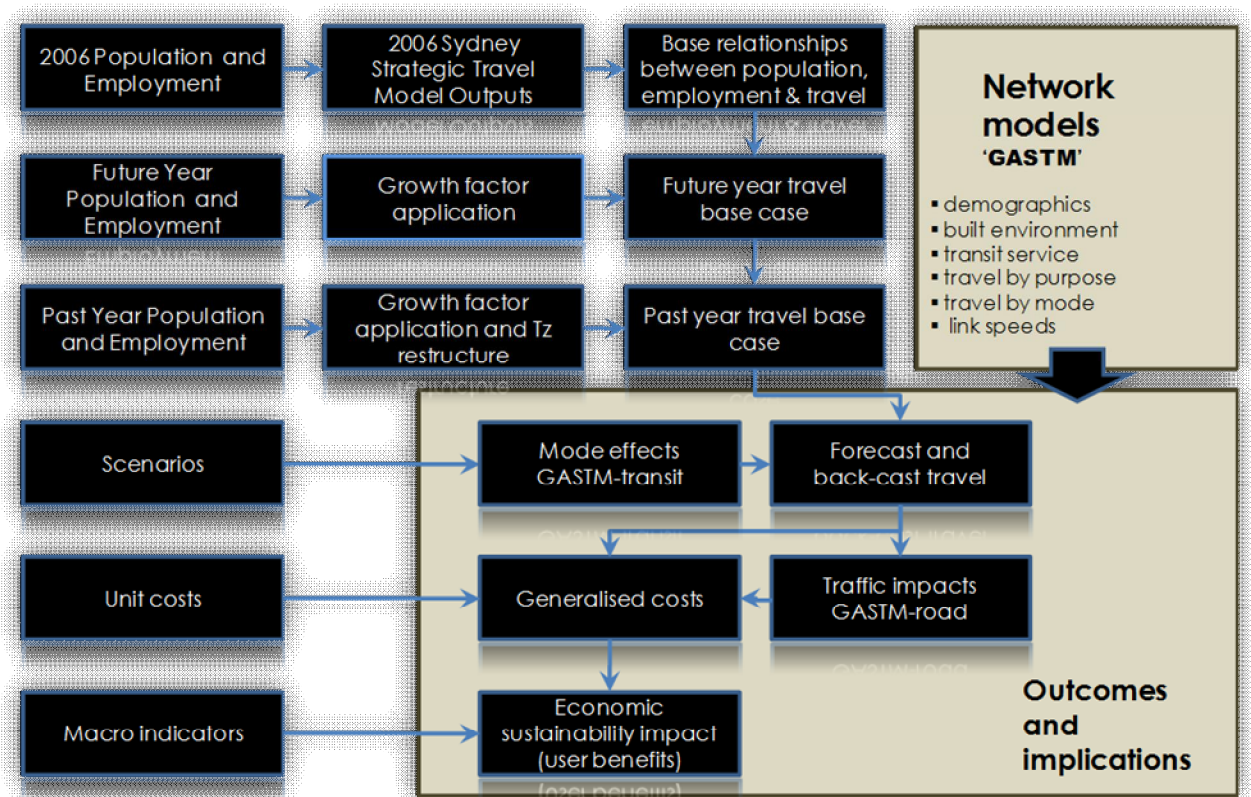
The GASTM-transit mode share models comprise a two stage regression-based model suite:

- Stage 1 . The cars models: car ownership was modelled as the dependent variable in the first instance, based on demographic and built environment variables
- Stage 2 . The mode share models: predicted car ownership was then treated as an independent variable in the development of mode share models for each trip purpose.

Both stages used the variables in un-factored from and utilised factorisation to reduce the number of variables and to reduce the covariance between them. In the event, the variables selected for the models comprised a mix of factors and un-transformed variables. The resultant models generally have strong ANOVA and perform well in other respects.

The modelling procedure is shown in Figure 2.

Figure 2: GASTM model structure



5.1 Stage 1 – The cars models

The model development was initiated by examination of a series of models potentially suitable for the prediction of car ownership. Throughout this text cars models refers to models to predict the number of cars. As with other aspects of the model suite, car ownership is considered to be static over time; in fact it has and will vary. There is some evidence that ownership is levelling out, and so the assumption of static car ownership into the future may not be unreasonable.

The modelling process entailed multiple linear regression, progressively testing combinations of variables based on the correlations that had been undertaken in the previous phase of analysis. The procedure utilised factorisation to reduce covariance and to reduce the number of variables. The factor analysis was undertaken using the variables from the first non-factorised series, excluding those that the earlier regression procedure indicated were of low importance to the model. IBM SPSS Statistics® dimension reduction factor analysis default parameters were used, except that varimax rotation of the solution was applied to assist interpretation of the results, and extraction was based on eigenvalues greater than 0.5 rather than the default of 1.0.

The rotated matrix provides useful information on the significance of each factor and the nature of each component. In this structure the components are:

- 1) Demography . this component is driven by the population age groups other than seniors and elderly (which nevertheless have some influence on the component). It is the most important, as it determines 37% of the variance.

- 2) Transit service . driven by train service to the CBD and by the type of public transport in the travel zone. This determines 27% of the variance.
- 3) Density . driven principally by population density, along with work trip orientation to the CBD. This determines a further 12% of the variance.
- 4) Income . income and work trip orientation to the CBD drive component 4. This determines 7% of the variance and so is marginally important but not insignificant.
- 5) Seniors . an interesting outcome, in that the last component is dominated by the seniors and elderly demographic. It explains just 6% of the variance.

Table 3: ANOVA for selected models of car ownership

Model	Predictors	r	Adjusted r-square	Standard error of estimate	F
Cars per dwelling CF5	Seniors factor, Income factor, Density factor, Transit service factor, Demographic	0.85	0.72	0	903
Cars CFL7	Household Income, Residential Density, Seniors and Elderly, Rail service factor, Transit type factor, Demographic factor, CBD orientation factor	0.95	0.91	190	2,398
Cars CFL8	Transit service factor, Demographic factor, CBD orientation factor, Household Income, Seniors and Elderly, Residential Density	0.95	0.91	190	2,787

The preferred cars models are of the following form:

$$N_{cars} = 514 + 0.83 * N_{senior} - 3.00 * N_{density} + 448.19 * F_{demog} - 97.31 * F_{cbd} + 77.34 * F_{cat} + 46.36 * F_{service} + 0.35 * N_{income}$$

Where:

F_{demog} = demographic factor

$F_{service}$ = transit service factor

F_{income} = household income factor

$N_{seniors}$ = number of seniors and elderly people

$N_{density}$ = residential density

F_{cbd} = CBD orientation factor

F_{cat} = transit type factor

The signs in the model are reflective of the manner in which the underlying variables are expressed. The model is dominated by the numbers in each age group, and so F_{demog} and $N_{seniors}$ simply reflect the population. The positive signs on the $F_{service}$ and F_{cat} components may appear counterintuitive, however in each of these cases the components with the better transit service have lower values. Service was measured by service to stations in major office (CBD) employment areas as the sum of the weighted times, as follows:

Average in vehicle rail time for the nearest station to the travel zone

Average wait time
 Walk time from the travel zone to the station
 Transfer time between trains if required to access the CBD.

Transit Category (F_{cat}) was used to identify the type of public transport services (and hence the quality of service) in the travel zone concerned. Four categories are used in hierarchical order as follows:

1. Rail-served, within 2,000 metres of a railway station with a direct service to the CBD. The distance is based on walk access (generally taken as up to 800m), and short car access trips. The direct service criterion eliminates the Carlingford and Southern Highlands lines in Sydney. Stations on the Airport Line attracting a surcharge were also excluded from this parameter.
2. Areas otherwise serviced by Sydney Buses, which offers relatively high levels of service
3. Areas serviced by contract bus services, offering commuter services in developed areas
4. Other areas, typically serviced by infrequent rural services, if at all.

Thus the lower value for F_{cat} were zones with rail access.

This model had an r-square value of 0.9.

5.2 Stage 2 – The mode share models

The mode share models were developed so as to use the car ownership results and other variables to estimate the logit of non-car mode share (1 minus car driver and passenger) and rail share. The cars results were expressed as cars per adult for this purpose. The other candidate variables were selected from the correlations in the first phase, as with the cars models. Models were developed for each trip purpose. The logit is used because mode share is a fraction, or proportion, that is not normally distributed. Accordingly it is necessary to transform it to a form that has properties suited to statistical analysis . the logit transformation. The model was developed by multiple linear regression, testing the combinations of variables and factorising them.

The preferred work purpose model is:

Logit noncar share

$$= -1.36 - 0.36 * F_{service} + 0.34 * F_{access} + 0.42 * F_{carown} + 0.38 N_{diversity}$$

Where:

$F_{service}$ = transit service factor

F_{access} = accessibility factor

F_{carown} = car ownership factor

$N_{diversity}$ = SLA entropy

Again, the negative sign for $F_{service}$ reflects the attributes of the service component, in that the lower the value (i.e. lower weighted travel time) the better the service. The positive sign of F_{carown} is curious however. The factorisation process effectively combines some variables, and this could be a result of other variables in the factor. Unfactorised models with cars per adult or cars per dwelling showed negative signs for car ownership. The models for rail, however, show different signs for different trip purposes, even with car ownership left unfactored, suggesting that car ownership is actually less important for some trip purposes if other variables are taken into account. This aspect of the models warrants further work.

This model has an r-square value of 0.6. While this is substantially lower than for the cars model and lower than for the unfactorised models, it is nonetheless sound. This model not

only substantially eliminates covariance, but it gives weight to the variables transit service levels while excluding density. This is a key finding, given the debate on density as a driver of transit use that has occurred in the literature in recent years. These results show that, while it affects car ownership, it can be omitted from the mode share calculation as it is covered by other variables.

The r-square value varies by trip purpose, with work trips and non-car share the better performers. Education trips other than tertiary do not exhibit any relationship with urban structure. The other purpose non-car models are:

$$\begin{aligned} \text{Tertiary logit non-car share} \\ = -0.15 - 0.26 * F_{\text{service}} + 0.13 * F_{\text{access}} + 0.26 * F_{\text{carown}} + 0.19 * N_{\text{diversity}} \end{aligned}$$

$$\begin{aligned} \text{Business logit non-car share} \\ = -2.95 - 0.24 * F_{\text{service}} + 0.33 * F_{\text{access}} + 0.26 * F_{\text{carown}} + 0.91 \\ * N_{\text{diversity}} \end{aligned}$$

$$\begin{aligned} \text{Shop and other logit non-car share} \\ = -1.50 - 0.07 * F_{\text{service}} + 0.06 * F_{\text{access}} + 0.15 * F_{\text{carown}} + 0.33 \\ * N_{\text{diversity}} \end{aligned}$$

Where

F_{service} = transit service factor

F_{access} = accessibility factor

F_{carown} = car ownership factor

$N_{\text{diversity}}$ = SLA entropy

The r-square values were 0.7 (tertiary education), 0.8 (business) and 0.6 (shop and other).

Rail mode share is not as predictable, on the basis of the variables used, as non-car is more generally, but acceptable results were obtained. The rail models are:

$$\begin{aligned} \text{Logit work rail share} \\ = -1.70 - 0.02 * F_{\text{cat}} + 0.51 * F_{\text{service}} - 0.37 \\ * F_{\text{access}} - 0.39 N_{\text{caradt}} \end{aligned}$$

$$\begin{aligned} \text{Logit education rail share} \\ = -3.27 + 0.004 * F_{\text{access}} - 0.29 * F_{\text{service}} - 0.16 * F_{\text{cat}} + 0.68 * N_{\text{caradt}} \end{aligned}$$

$$\begin{aligned} \text{Logit tertiary rail share} \\ = -2.02 - 0.36 * F_{\text{access}} - 0.43 * F_{\text{service}} - 0.31 * F_{\text{cat}} + 0.96 N_{\text{caradt}} \end{aligned}$$

$$\begin{aligned} \text{Logit business rail share} \\ = -3.53 + 0.18 * F_{\text{access}} - 0.52 * F_{\text{service}} \\ - 0.34 * F_{\text{cat}} [+ 0.03 N_{\text{caradt}}] \end{aligned}$$

$$\begin{aligned} \text{Logit shop rail share} \\ = -4.01 + 0.12 * F_{\text{access}} - 0.34 * F_{\text{service}} - 0.22 * F_{\text{cat}} + 0.12 N_{\text{caradt}} \end{aligned}$$

Where:

F_{access} = accessibility factor

F_{service} = transit service factor

$F_{cat} = \text{transit type factor}$

$N_{caract} = \text{cars per adult}$

The r-square vales were:

Work	0.6
Education	0.2
Tertiary	0.5
Business	0.6
Shop	0.7

The models described above were determined from the 2006 data in isolation. For the purpose of applying the models to future and past years, it was necessary to expand the factorisation to cover all years. The data for all years must be included in one dataset to retain the predictive capability. To estimate the model with the expanded factors, the selected travel zones, base case 2006 were extracted from the enlarged set and used as previously.

6. Conclusion

The research that has been described in this paper was the core of a much larger exercise. The GASTM-Transit mode share models developed as described above were used to forecast and backcast a series of scenarios that reflected the development of Sydney's rail system within the evolving land use. It was specifically designed to assess the effect of provision of rail service in the context of urban development and to estimate the use of public transport (noncar modes) that resulted. It recognised the importance of the Ds, but in the event the factors that resulted were combined to reflect accessibility, (both to transit and to destination), public transport service, car ownership and to some degree diversity. Density was effectively embodied in the car ownership and accessibility variables. The resultant models were very effective in prediction and the factorisation process served its purpose in reducing the number of variables and in reducing covariance. Factorisation is something of a trade-off in that it reduces the r-square value while improving the clarity of the model; however the models remain strong.

This work clearly supports the ability to model mode share and to predict travel demand more directly from land use and the public transport service that is offered. They reinforce the fact that good public transport such as that offered by frequent rail service to the areas of major employment and other activity are key to reducing car-dependence.

References

- Alonso, W. (1964). Location and Land Use. Cambridge MA: Harvard University Press.
- Australian Bureau of Statistics (ABS) (2008a). Basic Community Profile Travel Profile CD_NSW_B45. Canberra ACT: Australian Government.
- Australian Bureau of Statistics (ABS) (2008b). 9208.0 - Survey of Motor Vehicle Use (SMVU). Canberra ACT: Australian Government.
- Australian Bureau of Statistics (ABS) (2009a). 3218.0 - Regional Population Growth, Australia, 2007-08. Canberra ACT: Australian Government.
- Australian Bureau of Statistics (ABS) (2009a). 6291.0.55.001 - Labour Force, Australia, Jul 2009. Canberra ACT: Australian Government.
- Bernick, M. and R. Cervero, R. (1996). Transit villages in the 21st century. New York: McGraw-Hill.

- Black, J. (2008a). "Polycentric Employment Formation in Australasian Cities", 9th World Conference of Metropolis, Sydney NSW Oct 2008.
- Cervero, R. and K. Kockelman (1997). "Travel demand and the 3Ds: Density, diversity, and design", *Transportation Research D* 2(3): 199-219.
- Ewing, R. and R. Cervero (2010). "Travel and the Built Environment", *Journal of the American Planning Association* 76(3): 265-294.
- Gonçalves, J. A. M., L. d. S. Portugal, et al. (2009). "Centrality indicators as an instrument to evaluate the integration of urban equipment in the area of influence of a rail corridor", *Transportation Research A* 43: 13-25.
- Goodman, R. and S. Moloney (2004). "Activity centre planning in Melbourne revisited", *Australian Planner* 41(2): 47-54.
- Handy, S. L. (1996). "Methodologies for Exploring the Link between Urban Form and Travel Behavior", *Transportation Research D: Transport and Environment* 1(2): 151-165.
- Litman, T. (2009d). *Reevaluating the Role of Public Transit for Improving Urban Transportation*. Victoria Canada: Victoria Transport Policy Institute.
- McKibbin, M. (2011). "The Influence of the built environment on mode choice - evidence from the journey to work in Sydney", 34th Australasian Transport Research Forum, Adelaide S. Aust., Sept. 2011.
- Norley, K. (2015). "Get used to itq. the economics of Sydney's Rail Future", 37th Australasian Transport Research Forum, Sydney, Australia, Sept-Oct 2015 (forthcoming)
- Norley, K. (2010). "The Role of the Rail System in the Sydney Journey to Work . A Geospatial Analysis", 33nd Australasian Transport Research Forum, Canberra ACT, Sept.-Oct. 2010.
- Norley, K. (2013) *Applied geospatial modelling of the relationship between mode share, rail transit infrastructure and urban structure*, PhD Thesis, Sydney, Australian Graduate School of Urbanism, UNSW Australia.
- O'Sullivan, A. (2009). *Urban Economics*. New York: McGraw Hill Irwin.
- Transport Data Centre (2008c). Sydney GMA06 2006 travel data. 2006_24h_tour_csv.zip. Sydney NSW: NSW Government.
- Transport Data Centre (2008d). TblTable06 Tz Origin and Destination Tables. Sydney NSW: New South Wales Government.
- Transport Data Centre (2008g). Sydney 2001 journey to work data. 2001JTW_T09_GMA. Sydney: NSW Government.
- Transport Data Centre (2010b). Population Forecast (Population and Households) 2010 Release. Popforecast_Hhld_Rebased_Dec_2009.xls. Sydney NSW: New South Wales Government.
- Transport Data Centre (2010c). Employment Forecast 2010 Release. EmpForecast_Hhld_Rebased_Dec_2009.xls. 2010 ed. Sydney NSW: New South Wales Government.
- Transport Data Centre (2010e). GMA07 STM2 Tour Time Period Allocation. Sydney NSW: New South Wales Government.
- Tsai, C-H., C. Mulley, et al. (2012). "The Spatial Interactions between Public Transport Demand and Land Use Characteristics in the Sydney Greater Metropolitan Area", 35th Australasian Transport Research Forum, Perth W. Aust., Sept. 2012.