A workplace choice model accounting for spatial competition and agglomeration effects

Chinh Q. Ho[1], David A. Hensher[1]

[1]Institute of Transport and Logistic Studies, The University of Sydney Business School, NSW 2006 Australia

Chinh.Ho@sydney.edu.au; David.Hensher@sydney.edu.au

Abstract

This paper develops a new model of workplace choice for the Sydney Greater Metropolitan Area (SGMA) and describes the way in which this model is integrated into a general modelling framework of MetroScan, an improved version of the Transportation and Environment Strategy Impact Simulator Transportation (TRESIS). The developed model accounts for spatial competition of alternative workplaces via accessibility variables measured to attractions of both the same and different types. The new model also has two new refinements. First, a much finer geographical level is used for modelling worker's choice of workplace given the location of firms and the distribution of jobs. Second, an employment agglomeration effect is incorporated by the inclusion of jobs in the industry class relevant to the worker and two accessibility measures. Modelling analysis on data collected from a survey conducted in Sydney in 2013 identifies highly significant spatial competition and employment agglomeration effects explaining workplace choice. The application of this model to analyse policy relating to the redistribution or growth of jobs within a general framework of MetroScan is discussed.

Keywords: MetroScan, work location choice, agglomeration, spatial competition, land use models

1. Introduction

Economies of agglomeration is usually used to describe the benefits that firms obtain by locating closer to each other, whilst spatial competition arises from the fact that nearby firms in the same industry are generally competing against each other more than against distant ones. In the presence of spatial competition, the agglomeration effect may still be observed as a cluster of businesses that attract more labour and material suppliers as well as customers. From the consumer's perspective, it is convenient to have a cluster of destinations to do multiple activities with less effort of travel in between, and thus a cluster of businesses may attract more customers than dispersed ones. From the worker's perspective, the agglomeration effect on their choice of workplace is less clear, as most workers only have one place of work and thus, the benefit of less travel between work locations does not apply. However, firms having better access to other businesses to undertake work-related activities may still obtain an advantage. In addition, with agglomerated employers, workers can find an employer who wants a particular skill set that matches more closely their own. This advantage may be observed in the market via the individual's choice of workplace as firm locations translate into job locations. This paper describes the development and application of a new workplace choice model that is capable of capturing economies of agglomeration and spatial competition effects, given the location of firms and the distribution of jobs across the study area.

There is increasing interest amongst transport planners, modellers and economists in measuring agglomeration and spatial competition effects; however, very few regional travel demand models in use today are able to quantify the spatial competition and agglomeration

effects. Specifically, gravity models of destination choice (e.g., Anas, 1983) ignore the agglomeration effect as they treat location choice as independent. By contrast, competing destinations models (e.g., Fotheringham, 1986) can detect the dominating effect only, be it the agglomeration or the spatial competition effect. Activity-based models (e.g., Shiftan, 1998) have incorporated the economies of agglomeration through trip chaining but generally ignore the heterogeneity in spatial competition as defined by Tobler's first law of geography which states that: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p.236). The only exception is the agglomerating and competing destination choice model for non-work activities developed by Bernardin et al. (2009). Their approach is adapted in this paper to develop a work location choice model for implementation within a general modelling framework of MetroScan, an improved version of the 'Transportation and Environment Strategy Impact Simulator' or more commonly known in its abbreviation as TRESIS (Hensher and Ton, 2002).

This paper is motivated by the desire to equip MetroScan with the ability to analyse the agglomeration effect with respect to job concentration through an increased sensitivity of its workplace choice model to the spatial distribution of jobs in different industries. The proposed model is fully operational and compatible with trip-based and activity-based concepts of modelling regional travel demand. The next section describes the placement of the workplace choice model in the MetroScan modelling hierarchy. This serves as a precursor to the development of the workplace choice model in terms of data and inputs required from other modules. The paper then provides a brief discussion of the modelling approach to incorporate agglomeration and spatial competition and presents the estimation results. This paper concludes with a summary of the main findings and the way in which the workplace model is applied to analyse policies relating to the redistribution or growth of jobs in the study area.

2. Placement of workplace choice model in MetroScan

Figure 1 shows an overall structure of the MetroScan passenger travel and location choice model system. MetroScan is an integrated package of travel demand (and supply) models for the passenger and freight sectors that are structured in a certain way to reflect the interdependencies of travel decisions. We focus herein on the passenger model system. These models are applied sequentially at the household, the individual and the network levels (in the OmniTRANS platform) with feedbacks and links between modules. MetroScan enjoys several refinements over TRESIS. These include a wider coverage area (Sydney Greater Metropolitan Area in MetroScan vs. Sydney Statistical Division in TRESIS, see Figure 4), a finer spatial resolution of travel zones, the ability to select different zoning systems for a quick scan or a detailed analysis, a real road and public transport networks and network assignment models implemented via OmniTRANS, and a number of new/enhanced models relating to non-work and freight activities (the freight component of MetroScan is not included in Figure 1).

As for the work location choice model developed in this paper, it is applied conditional on the residential location choice. That is, residential location is treated as exogenous in this model sequence as the individual's home is already known by the time the workplace choice model is executed. On the other hand, this model is linked to the mode and time of day choice model via the logsum measure (also known as an accessibility index or maximum expected utility). Thus, mode and time of day joint choice models have to be estimated first to obtain model parameters for the imputation of maximum expected utility which is fed into the workplace choice model. Inputs from an innovative arrival time flexibility model, which describes how much flexibility a worker has in terms of the time they have to be at work, and the network assignment models (i.e., skim matrix) are also required for the estimation of the workplace choice model as shown in Figure 1. The next section describes the process of collecting the necessary data for model development, and the choice of a modelling approach to incorporate agglomeration and spatial competition effects in workplace choice.

Figure 1: Overall framework of the MetroScan passenger travel demand model system
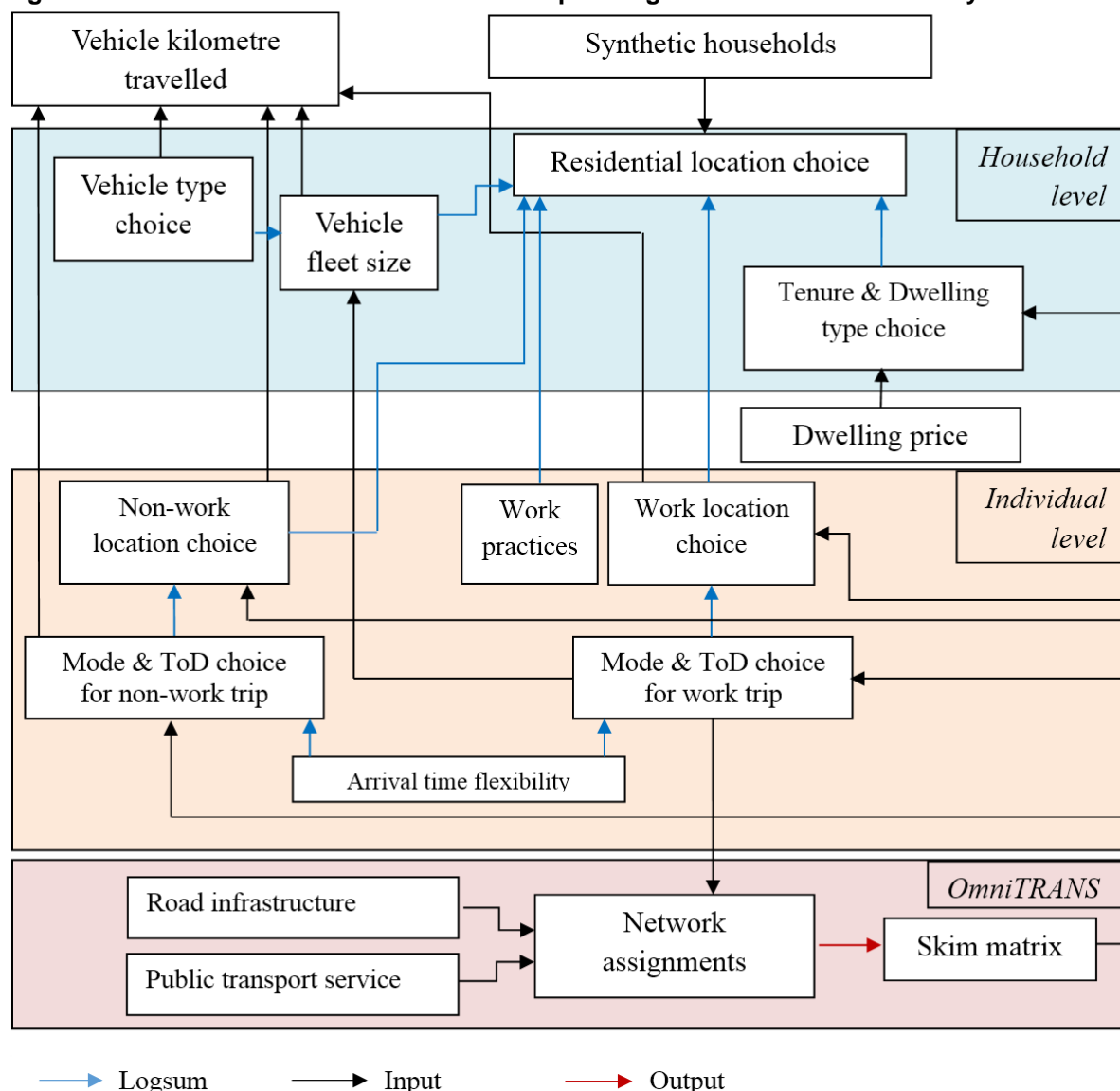
3. Data collection and modelling approach

This section describes the main survey and supplementary data from various sources that are used in developing the workplace choice model for MetroScan. The approach to modelling the agglomeration and spatial competition effects involved in the worker's choice of workplace is also described. To this end, a brief review of the literature is provided in which different modelling approaches are discussed and compared before the adopted modelling method is detailed.

3.1 The survey and data

An online survey was purposely designed to collect data for the development of the workplace choice model. A pilot survey was conducted from 25th to 26th July 2013 on a sample of 36 workers to test the comprehendability of the questionnaire and the workability of the database at the back-end. Minor edits were made as a result of the pilot survey and the main survey was conducted using *pureprofile* panel (www.pureprofile.com) from 19th August to 06th September 2013. A preset sample of 2,000 valid respondents was contracted and sampling quotas were applied based on the proportion of workers (part-time and full-time) living in each residential postcode to the total workers in the Sydney Greater

Metropolitan Area (SGMA).[1] This aims to increase the geographical representativeness of the sample. The survey data were analysed on a daily basis when the survey was in progress to ensure that the geographical quotas were closely matched.

An invitation e-mail with the survey link was sent by *pureprofile* to a total of 3,519 subscribed respondents living in the SGMA including Sydney, Illawarra and Newcastle (see Figure 2). A sample of 2,031 respondents was obtained, resulting in a response rate of 58%. However, 23 respondents live outside the SGMA and 43 respondents provided inconsistent information; these were excluded from the sample. Figure 2 shows the sample distribution of the respondents by postcodes of residence and workplace. As can be seen from the SGMA map on the left, respondents living across the SGMA were successfully recruited via the stratification and targeting strategy described above (postcodes with no respondents shown in Figure 2 are mostly bush areas). Although respondents to the survey live across the SGMA, most of them work in the Sydney Central Business District (CBD), North Sydney, Parramatta (the second biggest city in Sydney) and Macquarie Park (the fourth largest business hub in Sydney) as shown in Figure 2. This indicates that the sample is well representative of the population in terms of where workers live and work.

In terms of the information collected, the questionnaire consisted of three parts. The first part included questions relating to work location. In this part, workers were asked to describe their work location (in terms of suburb and postcode), the number of days per week they typically work at this location, their working patterns during a typical 2-week period with four selectable options: fixed working hour 5 days per week, flexible working hour 5 days per week, compressed work week 9 days per fortnight, and telecommuting/work at home one or more days per week. In the first part, respondents were also asked to describe their place of residence (in terms of suburb and postcode) and the type of dwelling (detached, semi-detached including town house, or unit including apartment) they live in. The second part included questions relating to their daily commute including the commuting mode, the typical times they leave for work and arrive at a workplace, the availability of each public transport mode (bus, train, ferry, light rail) for commuting regardless of whether they use it or not, the distance from home to work and the typical travel time during peak and off-peak hours by car and by public transport, the availability of parking at their workplace and parking cost per day. The perceived travel times and the availability of different travel modes are used for cross-checking and linking with the mode of time of day models in the system The final part included a survey of individual and household characteristics such as working industry, age, gender, employment status, working hours, occupation, annual income, driving licence status, access to car and the number of household adults and children.

Most of the data collected by this survey are included directly in the work location choice (WLC) model, but some are used for developing a complementary work practices model describing working patterns, while other data (such as the respondent's perceived travel time and the availability of different travel modes for commuting) are required for linking with the mode and time of day model which uses the actual travel times and travel costs from the network assignments via OmniTRANS (see Figure 1). On the one hand, information on the working patterns is used to estimate the work practices model that describes the spatial and temporal dimensions of individuals' work patterns over a typical 2-week working period. This model has important roles in influencing the levels of commuter traffic for each O–D pair via adjustment factors that account for the location of the workplace being at home or out-of-home (see Hensher and Ton, 2002 for more details). On the other hand, data on the availability of each PT mode for commuting as well as access to a car is used to compute the maximum expected utility from the choice of mode and time of day (ModeToD) for commuting between each and every pair of home and work zones. This is described in the next section.

---

[1] It would have been better to stratify the sample by the place of work, but this information is rarely reported by *pureprofile* panellists.

Figure 2: Distribution of sampled respondents by postcode of home and workplace
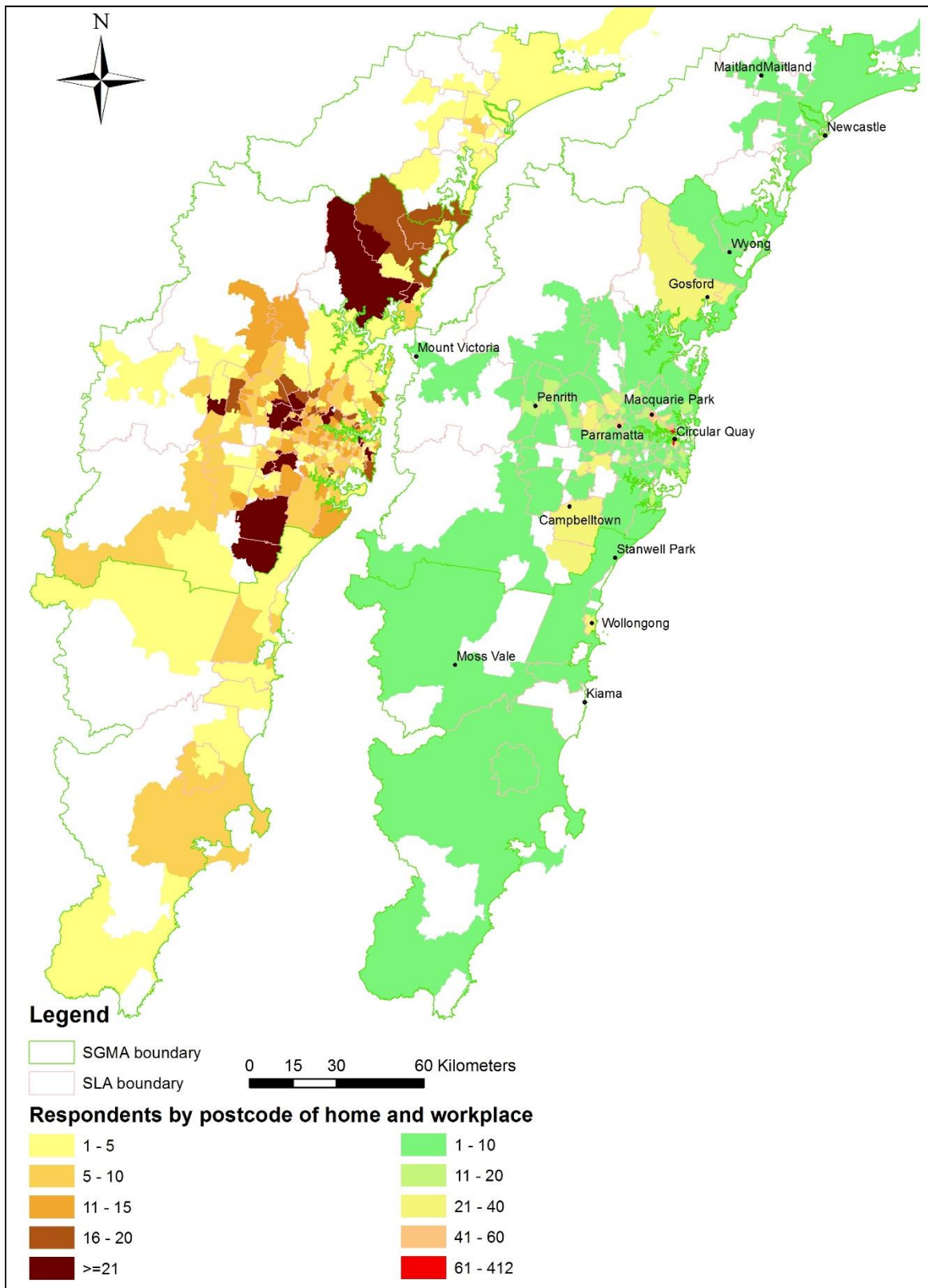
3.2 Supplementary data and input from the mode and time of day choice model

The logsum of the ModeToD choice model is required for the estimation of the WLC model. A substantial effort has been invested into computing the maximum expected utility that each

worker derives from their joint choice of mode and time of day for commuting. To this end, supplementary data from the Census journey to work (JTW), the Sydney Household Travel Survey (HTS) and the Sydney Strategic Travel Model (i.e., skim matrices) were used. Supplementary data are required given that the ModeToD choice model and the WLC model are based on two different datasets with significant variables in the ModeToD choice model such as access and egress modes, access and egress times, toll cost, and public transport fare which were not collected in the WLC survey in order to reduce the survey burden on respondents. In addition, given the residential location, the logsum measures are required for all alternative workplaces considered in the WLC model. This includes not only the location that the workers were observed to choose but also the locations that they did not choose (i.e., non-chosen alternatives). Thus, supplementary data are required for the estimation of WLC model even if both the WLC and the ModeToD choice models were based on the same dataset.

To compute the logsum of ModeToD choice model, its parameters were applied to the corresponding variables derived from the Census JTW, the Sydney HTS and the skim network matrices. The ModeToD choice model has been estimated in which a day is discretised into 6 time-of-day periods and Stated Preference (SP) data are used to enrich model behaviour. The results of the ModeTOD choice model will be reported in the near future (Ho and Hensher, forthcoming). For the purpose of this paper, a description of variables entering the ModeToD model is sufficient. These variables are classified into three groups. The first group includes origin and destination based variables (in-vehicle travel time, wait time, toll cost, public transport fare, fuel cost, and travel time reliability that is represented by the standard deviation of travel time between each O-D pair). These were derived from the STM skims and then matched with the WLC data based on the worker's home and all alternative workplace locations. The second group includes origin-based variables (access mode and access time). These variables were derived from the Census JTW and the Sydney HTS conditioned on the main mode of travel being train, bus or car (as driver or as a passenger). Origin-based variables were then matched with the WLC data based on the respondents' residential location. The final group includes destination-based variables (egress mode, egress time) which were also derived from the Census JTW and the Sydney HTS and matched with the WLC dataset based on alternative work locations. The final group also includes parking availability and parking cost per day that were derived from the WLC survey itself. In sourcing all necessary variables for the computation of the ModeTOD logsum, it is important to reproduce the population distribution. This was done by matching the supplementary data sources with the WLC data at the postcode level. By doing so, we are certain that the distribution is reproduced at any geographical level larger than postcode.

Figure 3 shows the distribution of ModeToD logsums across all alternative workplaces for workers living in Camden (left) and Ryde (right), both locations highlighted in the map with the green boundary lines. In deriving the logsum of ModeToD, the availability of each travel mode for commuting is taken into account. Specifically, the logsum or inclusive value (IV) of ModeToD (*IVMDT*) can be written as in equation (1):

$$IVMDT = \ln\left[\sum_{m \in M_i} \sum_{t=1}^{T} \exp(U_{imt})\right] \tag{1}$$

where U_{imt} is the utility that worker *i* derives from departure time *t* and mode *m* available in their choice set M_i. Thus, the maximum expected utility from the choice of mode and time of day for commuting varies across O-D pairs (see Figure 3) and across workers with any difference in available modes or access/egress modes or access/egress times. Compared to the use of an average logsum by O-D pair for all workers that most modellers adopt, the method proposed in this paper has an advantage of maintaining the variation across

individuals even if they commute between the same O-D pair. This variation would help increasing the sensitivity of the WLC model to network changes that affect commuting.

Figure 3: Average logsum of mode and time of day choice by workplace for two example home locations
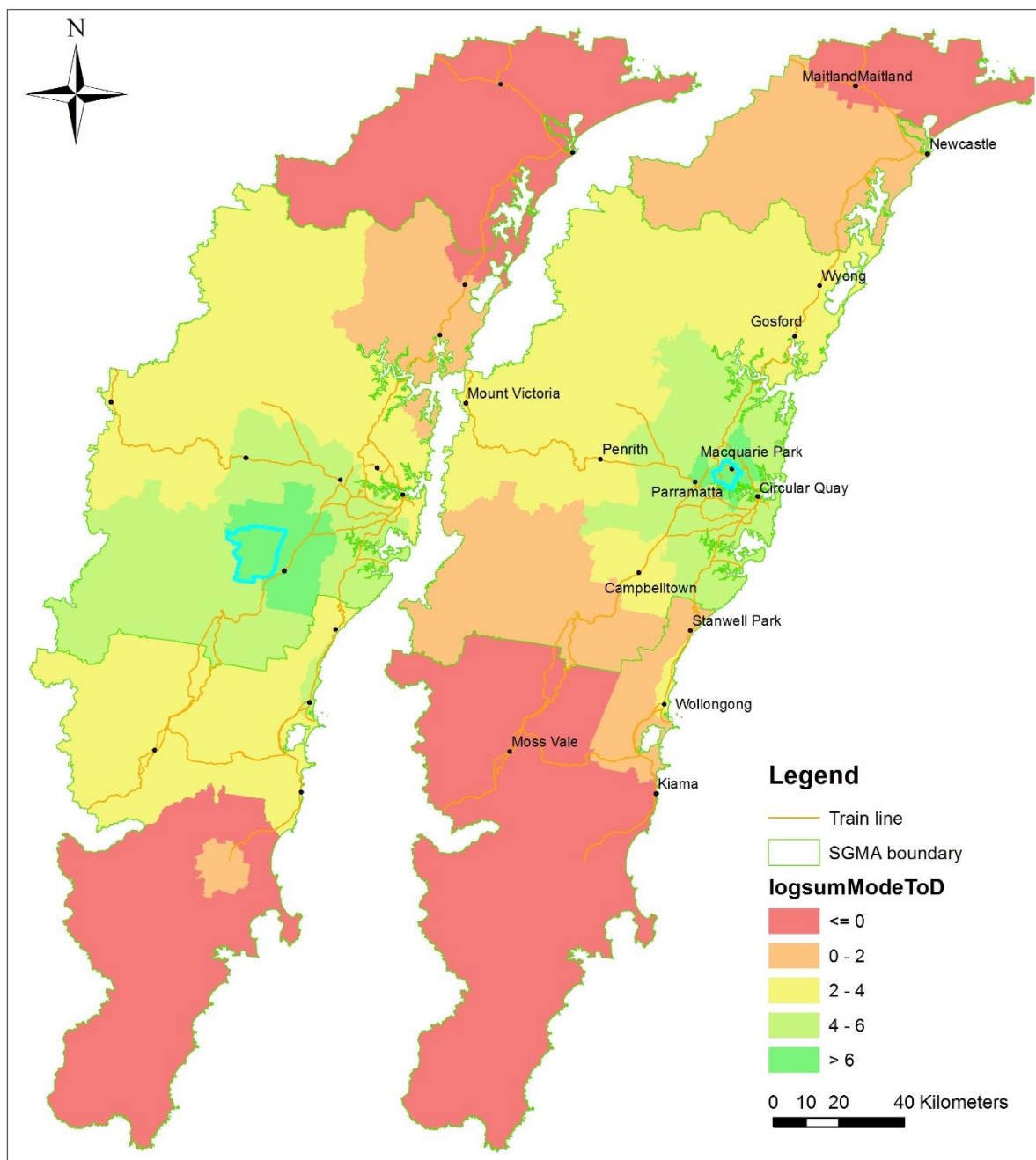
3.3 Approach to modelling agglomeration and spatial competition

Two modelling techniques have been applied in the literature to spatial choice as a means of allowing spatial competition and economies of agglomeration. The first approach uses the generalised extreme value (GEV) family of models, usually in a form of Nested Logit (see Hensher *et al.*, 2015), and a priori knowledge to group alternatives into different groups to have differential competition amongst alternatives. In the current application of modelling workplace choice, the alternative work locations can be grouped spatially into different regions so that the GEV models can be used as a means of allowing for differential spatial competition. This modelling technique can also be used as a means of capturing the

agglomeration effect, but the modelling framework must be tour-based or activity-based rather than trip-based. Following Shiftan (1998), this approach to agglomeration has been used in a few tour-based or activity-based models, but the adoption has been limited due to the costs relating to data requirements, model development and application.

The second approach to spatial competition and agglomeration uses the traditional multinominal logit (MNL) models but including accessibility measures that encapsulate information about other alternative destinations. Through these accessibility measures, the popular independence from irrelevant alternatives (IIA) assumption of the MNL model will not apply. Thus, destination choice models with accessibility measures allow for heterogeneity in spatial effect, be it the competition or the agglomeration. This approach was first introduced by Fotheringham (1986) and adapted in many subsequent studies (see Bernardin *et al.*, 2009 for a review). Generally, the accessibility index of a destination (A_j) is measured to a single attraction variable (B_k), such as employment, in other destinations (k), and the travel cost (c_{jk}) between the origin (j) and each potential destination (k):

$$A_j = \ln \sum_{k \neq j} \frac{B_k}{c_{jk}} \qquad (2)$$

The utility associated with a destination j is specified as a linear function of this accessibility measure and its parameter estimate (known as spatial structure parameter) and is used to identify the agglomeration or the spatial competition effect. Specifically, if the spatial structure parameter is negative, zones in close proximity to other opportunities (approximated by employment in nearby destinations B_k) have a lower utility than zones in spatial isolation, and thus the spatial competition effect dominates the agglomeration effect (Bhat *et al.*, 1998). However, if the spatial structure parameter is positive, the agglomeration effect dominates the spatial competition effect, whilst a zero parameter may indicate either the absence of both effects or equally strong effects that cancel each other.

A limitation of accessibility indices measured to a single attraction described above is that only the net effect of agglomeration and spatial competition can be detected while they can co-exist. To remedy for this limitation, Bernardin et al. (2009) introduced agglomeration and competing destination choice models where two accessibility measures are used to model the destination choice for non-work trips. This paper follows their approach and defines two accessibility measures: one to attractions of the same type (that they called accessibility to substitutes A_j^S) and one to attractions of different type (that they called accessibility to supplements A_j^C). These two measures emerge from an assumption that attractions of the same type (B_{jk}^S) are substitutes and those of different types (B_{jk}^C) are complements. Bernardin *et al.* (2009) proposed the use of Lieberson's D dissimilarity statistic (Lieberson, 1969) for the estimation of the number of substitutes and complements to one zone (j) to another zone (k). Given a classification scheme, Lieberson's D dissimilarity statistic can be defined as the probability that two items (e.g., activities) selected at random will belong to the same category. In the current application, we use 2-digit standard industrial classification, and Lieberson's D dissimilarity statistic can be estimated as:

$D_{jk} = P(\text{randomly visiting different work industries in zone } j \text{ and zone } k)$

$\quad = 1 - P(\text{randomly visiting same work industry } g \text{ in zone } j \text{ and zone } k)$

$\quad = 1 - \sum_g P_{jg} P_{kg} \qquad (3)$

$\quad = 1 - \sum_g \frac{B_{jg}}{B_j} \frac{B_{kg}}{B_k}$

where B_{jg} and B_{kg} are the number of jobs in industry g in zones j and k, respectively.

The quantities measuring attractions to jobs in the same industry (B_{jk}^S) and to jobs in different industries (B_{jk}^C) can then be estimated as:

$$B_{jk}^S = (1 - D_{jk})B_k$$
$$B_{jk}^C = D_{jk}B_k$$

(4)

and the accessibility indices of zone j to jobs in the same and in different industries can be computed using a unit distance decay function (i.e., using a parameter of 1 for cost c_{jk}):

$$A_j^S = \ln \sum_{k \neq j} \frac{B_{jk}^S}{c_{jk}} = \ln \sum_{k \neq j} \frac{(1 - D_{jk})B_k}{c_{jk}}$$

$$A_j^C = \ln \sum_{k \neq j} \frac{B_{jk}^C}{c_{jk}} = \ln \sum_{k \neq j} \frac{D_{jk}B_k}{c_{jk}}$$

(5)

Figure 4 shows the accessibility to jobs in the same industries and in different industries for all statistical local areas (SLA) in the SGMA. Both accessibility measures are high in Inner Sydney where most of the jobs are, and gradually decrease as distance from the CBD increases. This reflects the influence of distance on the attractiveness of activities that are spatially separated from the zone for which the accessibility is measured.

Figure 4: Accessibility to jobs in the same and in different industries
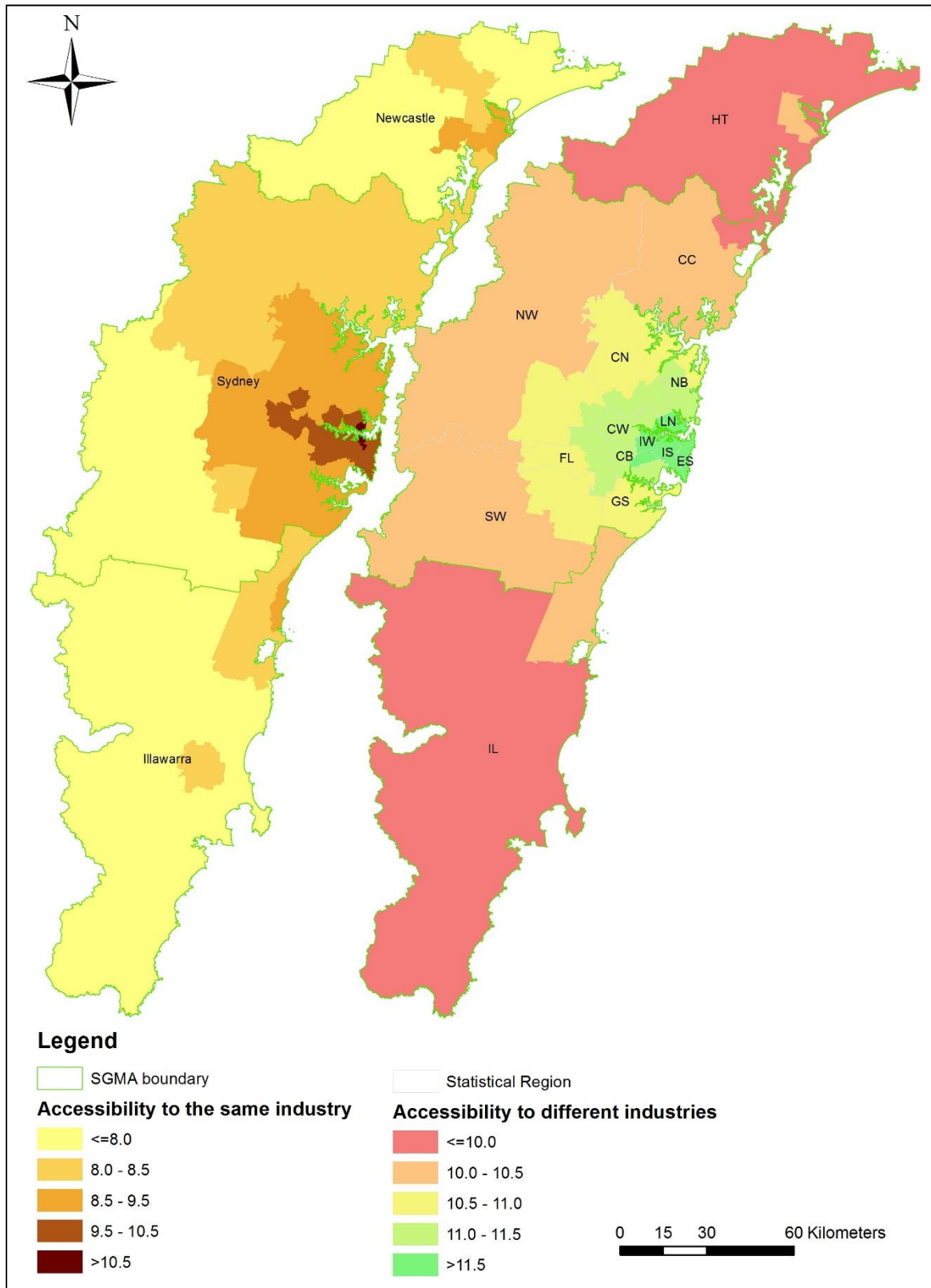

Data sources: Developed from GIS layers with employment by industry data from ABS Census 2011.

4. Model specification

Figure 5 shows the structure of the WLC model and how it is linked to the ModeToD choice. The WLC model has a MNL form and is estimated at the Statistical Local Area (SLA) level. The SGMA is divided into 80 SLAs (see Figure 2), labelled as SLA1 to SLA80 in Figure 5. The WLC model therefore will predict the probability that a worker with given personal characteristics and home location will choose to work in each of the 80 SLAs in the SGMA.

Figure 5. Structure of the work location choice model in MetroScan
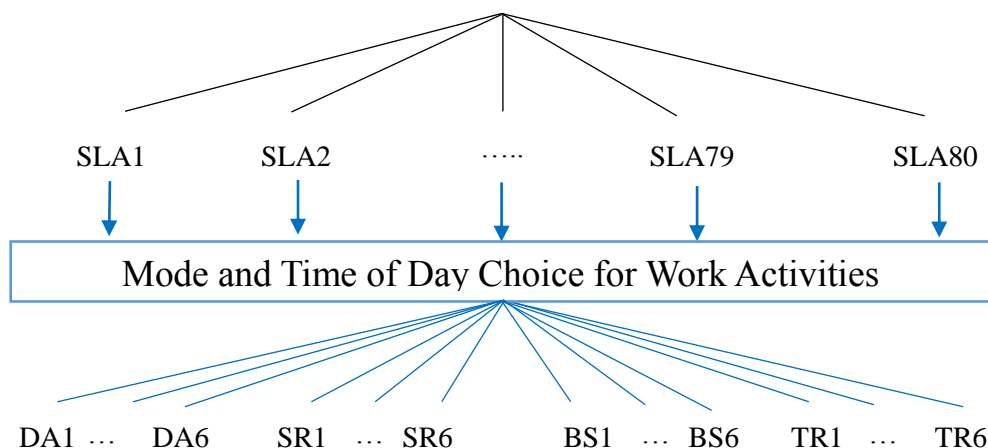
The probability that a worker living in zone *i* chooses zone *j* as a workplace is specified as a function of variables included in traditional destination choice models (such as the log of the size of the destination (S_j), the travel cost or level of services (c_{ij}) between the origin (*i*) and the destination (*j*), and interaction effects of worker's characteristics (W_o) with the size (S_j) of the destination and travel cost) and the two accessibility measures described in section 3.3 above. The size variable is approximated by the total number of jobs by industry in each work zone, while the level of service is replaced by the logsum of ModeToD choice ($IVMDT_{ij}$). The utility a worker living in an SLA *i* derived from working in an SLA *j* can be written as:

$$U_{ij} = \alpha_j + \beta_s \ln S_j + \beta_{WS} W_o \ln S_j + \beta_c IVMDT_{ij} + \beta_{wc} W_o IVMDT_{ij} + \beta_{AC} A_j^C + \beta_{AS} A_j^S + \varepsilon_{ij}$$
$$= V_{ij} + \varepsilon_{ij}$$

(6)

where the α and βs are parameters to be estimated, V_{ij} and ε_{ij} are observable and unobservable components of the utility expression, and other variables have been defined above. Assuming ε_{ij} follows an *iid* Gumbel Type I distribution, the probability a worker *i* choosing an SLA *j* as his workplace is expressed as (Hensher et al. 2015):

$$P_{ij} = \frac{\exp(V_{ij})}{\sum_k \exp(V_{ik})}$$

(7)

The model is estimated using a maximum likelihood estimator in Nlogit 5 (Econometric Sofware, www.limdep.com).

5. Estimation results

A number of models were explored to identify the best interaction effects between worker characteristics and the sizes of the destinations to be included in the utility function (terms $W_o \ln S_j$ in equation 6). Worker characteristics (W_o) tested included work industry, employment status (full-time, part-time, casual, unpaid volunteer) and worker's occupation. Measures of destination sizes tested include the number of jobs in different industries and

the statistical region to which an SLA belongs (see Figure 4). The best model fit was found with the interaction effects between a worker's work industry and the number of jobs in that industry (called 'jobs in relevant industry') and between worker's occupation and statistical regions.

Table 1 presents the estimation results of the WLC model based on a sample of 1,965 workers in the SGMA. The model fits reasonably well to the data with an adjusted McFadden R^2 of 0.328 (LL_β = -5318, LL_α = -8049, 94 parameters). Interpretation of the parameter estimates is straightforward, with the following highlighting the more interesting results.

Both accessibility measures are highly significant but have opposite signs, suggesting the presence of both agglomeration and competition effects with respect to workplace choice. Given that a negative spatial structure parameter identifies the competition effect while a positive spatial structure identifies the agglomeration effects, the results suggest that competition forces dominate the agglomeration effect (see the magnitudes of the two parameters associated with the accessibility measures). This is reinforced by the estimation results of the competing destinations (CD) model in which a combined accessibility measure (equation 2) was used. [2] This combined accessibility measure has a statistically significant parameter of –1.584, which represents the net effect of agglomeration and spatial competition (-1.837 + 0.286 = -1.551). This result is consistent with previous findings (Bernardin *et al.*, 2009) and suggests that a CD model for work location choice may be satisfactory if we are interested in the net effect only.

What is more interesting is that the accessibility to jobs in different industries has a negative parameter, while the accessibility to jobs in the same industry has a positive parameter. This is the opposite to Bernardin et al.'s findings for non-work activities, where accessibility to activities of the same type has a negative parameter (hence the name accessibility to substitutes), while accessibility to activities of different types has a positive parameter (hence the name accessibility to complements). Together, these suggest that agglomeration and spatial competition effects with respect to trip chaining in work and non-work location choices emerge from different sources, with travellers on non-work activities valuing destinations with a mix of activities, and workers preferring locations with the same type of business. This seems reasonable as workers/firms are more likely have work-related activities with businesses in the same industry than businesses in different industries, and thus the attractiveness of a zone would increase (i.e., an agglomeration effect is observed) if it is located in a cluster of businesses of the same type. By contrast, non-work travellers are more likely to see attractions of the same type as substitutes for each other (i.e., the spatial competition effect is observed) because of two reasons. First, travellers undertaking trips for non-work purposes are more likely to chain activities of different types than activities of the same type (Primerano *et al.*, 2008). Thus, attractions of different types complement each other when it comes to trip chaining. Second, where multiple activities of the same type such as shopping are chained, these are likely to be fulfilled by visiting one destination without the need to visit another zone (Ho and Mulley, 2013). Thus, attractions of the same types are likely to be substitutes for non-work location choices.

As shown in Table 1, the size of the destination, represented by jobs, strongly influences workplace choice with the parameters associated with jobs in a relevant industry (one digit ANZSIC), where statistically significant, are all positive as expected. The three size variables that do not significantly attract workers in the corresponding industries are jobs in the accommodation and food services industry, jobs in agriculture, forestry and fishing industry, and jobs in the mining industry. The non-significant parameters are likely to be due to a lack of variation in the number of jobs in these three industries across SLAs in the SGMA. Specifically, jobs in the accommodation and food services industry are everywhere, while very few SLAs have jobs in the agriculture and mining industries.

---

[2] Estimation results of the CD model are available on request.

A workplace choice model accounting for spatial competition and agglomeration effects

Table 1: Estimation results of the work location choice model for Sydney GMA 2014

Variable description	Parameter	Sig. level
Accessibility measures		
Accessibility to jobs in different industries	-1.837	***
Accessibility to jobs in the same industry	0.286	***
Log of jobs in industry relevant to worker		
Manufacturing	0.522	***
Accommodation and Food Services	-0.054	
Agriculture, Forestry and Fishing	0.174	
Mining	0.265	
Health Care and Social Assistance	0.297	**
Electricity, Gas, Water and Waste Services	0.649	***
Education and Training	0.573	***
Financial and Insurance Services	0.641	***
Information Media and Telecommunications	0.500	***
Construction	0.580	***
Transport, Postal and Warehousing	0.868	***
Rental, Hiring and Real Estate Services	0.722	***
Retail Trade	0.353	**
Public Administration and Safety	0.760	***
Professional, Scientific and Technical Services	0.634	***
Wholesale Trade	0.447	***
Arts and Recreation Services	0.750	***
Administrative and Support Services	0.514	***
Other Services	0.714	***
Logsum of mode and time of day choice	0.445 – 1.0	***
Interactions between worker's occupation and statistical region		
Clerical worker (1/0/-1, base = other occupations) × Central Coast	0.606	*
Managers (1/0/-1, base = other occupations) × Central Coast (1/0)	-0.466	
Professional (1/0/-1, base = other occupations)× Illawarra (1/0)	-0.420	
Clerical worker (1/0/-1, base = other occupations) × Illawarra (1/0)	0.660	
Professional (1/0/-1, base = other occupations) × Inner Sydney (1/0)	0.264	**
Clerical worker (1/0/-1, base = other occupations) × Inner Sydney	-0.600	***
Manager (1/0/-1, base = other occupations) × Inner Sydney (1/0)	0.368	***
Professional (1/0/-1, base = other occupations) × Lower North (1/0)	0.210	*
Clerical worker (1/0/-1, base = other occupations) × Lower North	-0.276	
Manager (1/0/-1, base = other occupations) × South West (1/0)	-0.302	
Region-specific constants		
Canterbury-Bankstown (1/0, base = Inner Sydney)	-2.435	***
Central Coast (1/0, base = Inner Sydney)	-2.272	***
Central Northern Sydney (1/0, base = Inner Sydney)	-1.877	***
Central Western Sydney (1/0, base = Inner Sydney)	-0.878	***
Eastern Suburbs (1/0, base = Inner Sydney)	-1.075	***
Fairfield-Liverpool (1/0, base = Inner Sydney)	-2.790	***
Hunter (1/0, base = Inner Sydney)	-3.840	***
Illawarra (1/0, base = Inner Sydney)	-2.890	***
Inner Western Sydney (1/0, base = Inner Sydney)	-1.774	***
Lower Northern Sydney (1/0, base = Inner Sydney)	-0.732	***
North Western Sydney (1/0, base = Inner Sydney)	-1.862	***
Northern Beaches (1/0, base = Inner Sydney)	-2.154	***
Outer South Western Sydney (1/0, base = Inner Sydney)	-2.285	***
St George-Sutherland (1/0, base = Inner Sydney)	-2.377	***

*Note: * significance at the 10% level, ** at the 5% level;*** at the 1% level.*

Alternative specific parameters were specified for the logsum of the ModeToD model (*IVMDT*). Values of these parameters range between 0.445 and 1.0, consistent with random utility theory. Figure 6 maps the logsum parameters by SLA. Generally, Figure 6 shows that SLAs closer to the CBD and a train line have smaller ModeTOD logsum parameters than SLAs further away. A smaller logsum parameter suggests higher substitution between alternatives under that nest (SLA), and the distribution of the logsum parameters reflects the fact that workers who work in the CBD or in SLAs close to a train station are more likely to substitute one mode of travel (e.g., car) for another (e.g., train, bus) than workers with workplaces outside the CBD or in SLAs with no quality train service. This validation lends credit to the model results.

Figure 6: Logsum parameters of the ModeToD choice model
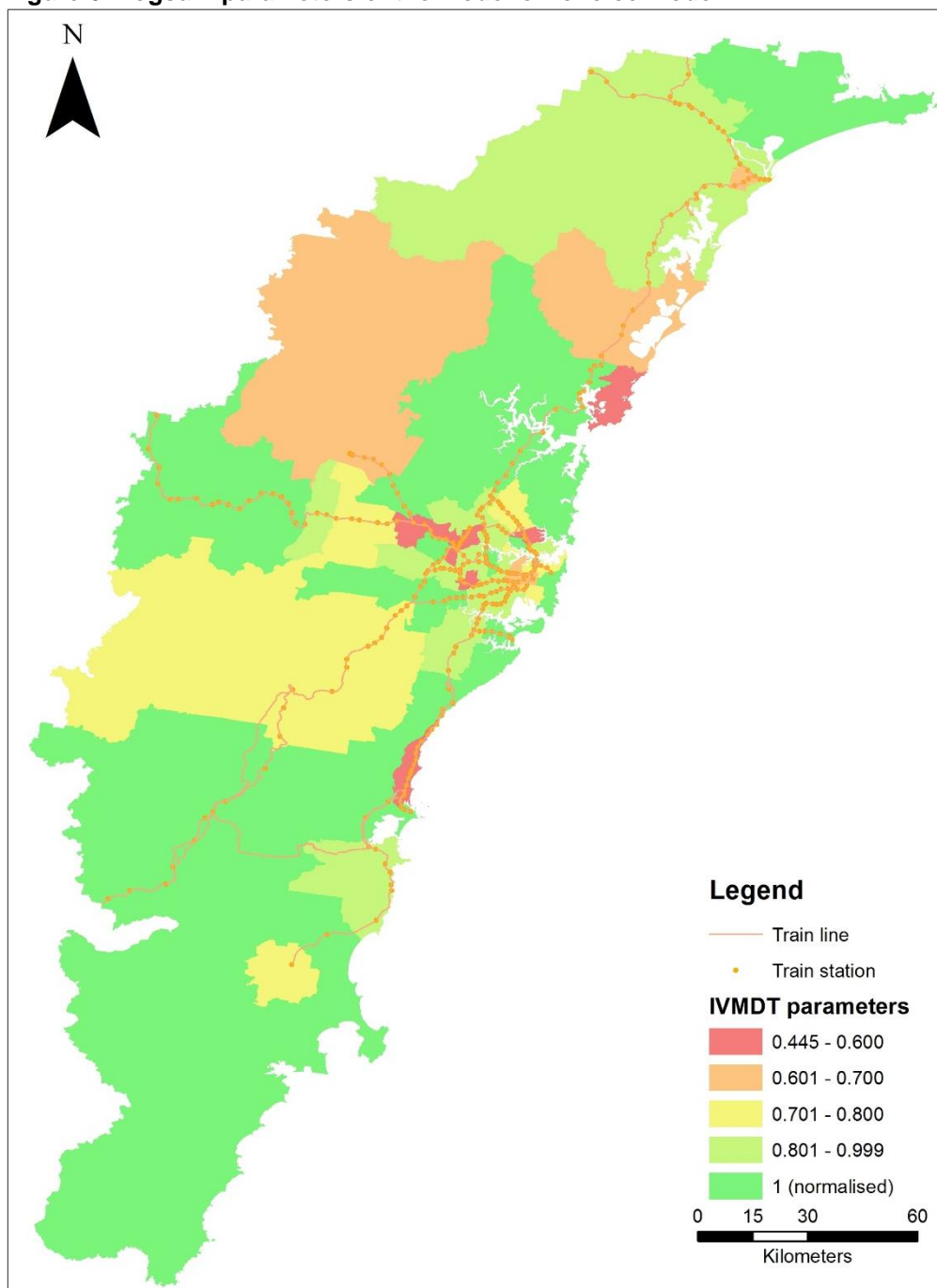
A worker's occupation also plays a role in deciding where they are likely to work, especially the propensity to work in the Inner Sydney (IS) statistical region. The worker's occupation was effects coded[3] and multiplied with each of the 15 statistical regions (see Figure 4) in the SGMA to create interaction terms; hence their parameters must be interpreted relative to an 'average worker' in each statistical region (i.e., average across all occupations). These interaction terms are highly significant in the Inner Sydney region only, with the statistically significant level being much lower in other regions. The results indicate that professionals and managers are more likely, compared to the overall average, to work in the Inner Sydney region while the opposite is true for clerical workers. The likelihood that workers with other occupations (i.e., the base) will work in the Inner Sydney region can be calculated as the negative sum of the three parameters included in the model (-0.264 + 0.600 - 0.368 = -0.033). That is, sales workers, machinery operators and drivers, labourers, technicians and trades workers, and community and personal service workers are just slightly less likely than the overall average, to work in the Inner Sydney region.

Finally, the constants associated with 14 statistical regions are all statistically significant and negative. With the base being the Inner Sydney region and all categorical variables effects coded, these constants play their true role of reproducing the sample shares. Interpretation of these constants is straightforward with the more negative parameters (e.g., Illawarra and Hunter) indicating that fewer workers in the sample are observed to work in those regions (also see Figure 2).

6. Conclusions and discussion

A major task involved in setting up a strategic travel demand model such as MetroScan relates to data collection and imputation. This is particularly true when it comes to linking discrete choice models that are estimated on separate datasets and linked together using the concept of maximum expected utility or the logsum measure. There are two main reasons why different modules of a strategic travel demand model must be estimated separately and the issue of linking models arises. First, many decisions are not made at the same time but in a certain order. Work location choice and commuting mode choice presented in this paper is a typical example in which workplace choice precedes and conditions commuting mode choice, while daily pressure of commuting provides feedback that impacts on the longer-term decisions relating to the place of work and/or the place of residence. In this case, the decisions should be broken down and modelled in this particular sequence to reflect the hierarchy of the decisions. Herein lies the need of linking these models, and in the absence of a temporal adjustment scheme, to impose heuristics to distribute the probability responses across an agreed period of time, instead of allowing for an instantaneous change (see Hensher and Ton 2002).

Second, it is inevitable that interdependencies of travel decisions have to be broken down and modelled sequentially in operational travel demand models, even if in theory these decisions can be modelled simultaneously. This is because trade-offs have to be made between the cost of model development/application and the accuracy of model parameters. For example, modelling the choices of workplace and mode and time of day for commuting simultaneously requires much larger sample sizes, and therefore a much higher cost, than modelling them sequentially (although this still requires extensive data to form the logsum terms). The cost difference between the two ways of modelling depends on the number of alternatives in the choice system. For example, for the workplace and mode and time of day choice models described in this paper, each is based on a sample of about 2,000

---

[3] Effects coding is alternative to dummy coding in which an attribute with *L* levels is transformed into *L* – *1* variables with the reference level being coded as –1 instead of 0. Each effects coded variable is set equal to 1 when the attribute is present, equal to –1 if the reference case is present, and equal to 0 otherwise. See Bech and Gyrd-Hansen (2005) for the advantage of effects coding compared to dummy coding.

observations, while an equivalent model where the two decisions are modelled simultaneously would have 1,930 alternatives (80 SLAs × 4 Modes × 6 TODs = 1920) and require a sample size of a minimum 20 times larger to obtain the same level of parameter robustness.

Reducing costs relating to model development does not come without the risk of obtaining biased parameters. One way to reduce this risk, while basing models on their own datasets, is to have some key information in common across the datasets, and to invest in the imputation of the logsums using these common variables and external data sources. This paper has demonstrated how the common variables and external data help the imputation of the logsum of mode and time of day choice for each individual in the work location dataset from which the work location choice model was developed. This approach has delivered model parameters that are behaviourally meaningful and intuitively appealing.

With respect to the drivers of work location choice, the model detects the presence of both agglomeration and spatial completion forces, with the latter effect being stronger than the former effect. This highlights the importance of using separate accessibility measures if both agglomeration and spatial competition effects are to be captured for a particular planning scenario. Competing destinations models, however, may still be satisfactory if the net effect is of interest.

Scenario analysis can be conducted using the model developed in this paper through a specification of variables being affected by a specific policy. For example, a scenario with a growth of jobs in a specific area could be simulated via changes to the number of jobs in relevant industries but also changes to the attractiveness of this and every other locations via the accessibility indices. However, scenario analysis should be performed on a representative sample such as synthetic households/workers since work industry and occupation are among the drivers of workplace choice. This task is out of the scope of the current paper, which is part of a larger project aiming to equip MetroScan with more powerful scenarios analysis and evaluation of spatial and aspatial transport, environmental and land use policies. For a complete MetroScan model system, it is necessary to establish the logsum of workplace choice and feed this into the residential location choice, together with logsums from the other three models describing non-work location choice, work practices and vehicle fleet size choice (as shown in Figure 1). All modules of MetroScan have been estimated and we plan to report the results of the remaining models and the application of MetroScan to evaluate transport projects in subsequent papers.

Acknowledgements

References

Anas, A. (1983) Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological,* 17**,** 13-23.

Bech, M. & Gyrd-Hansen, D. (2005) Effects coding in discrete choice experiments. *Health Economics,* 14**,** 1079-1083.

Bernardin, V. L., Koppelman, F. & Boyce, D. (2009) Enhanced destination choice models incorporating agglomeration related to trip chaining while controlling for spatial competition. *Transportation Research Record: Journal of the Transportation Research Board,* 2132**,** 143-151.

Bhat, C., Govindarajan, A. & Pulugurta, V. (1998) Disaggregate attraction-end choice modeling: formulation and empirical analysis. *Transportation Research Record: Journal of the Transportation Research Board***,** 60-68.

Fotheringham, A. S. (1986) Modelling hierarchical destination choice. *Environment and planning A,* 18**,** 401-418.

Hensher, D. A., Rose, J. M. & Greene, W. H. (2015) *Applied choice analysis*, Cambridge University Press.

Hensher, D. A. & Ton, T. (2002) TRESIS: A transportation, land use and environmental strategy impact simulator for urban areas. *Transportation,* 29**,** 439-457.

Ho, C. & Hensher, D. (forthcoming) Joint estimation of mode and time of day choice accounting for arrival time flexibility, travel time reliability and crowding on public transport. *World Conference on Transport Research Society*, 10 - 15 July 2016, Shanghai, China.

Ho, C. & Mulley, C. (2013) Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice. *Transportation Research Part A: Policy and Practice,* 49**,** 206-219.

Lieberson, S. (1969) Measuring population diversity. *American Sociological Review***,** 850-862.

Primerano, F., Taylor, M., Pitaksringkarn, L. & Tisato, P. (2008) Defining and understanding trip chaining behaviour. *Transportation,* 35**,** 55-72.

Shiftan, Y. (1998) Practical Approach to Model Trip Chaining. *Transportation Research Record: Journal of the Transportation Research Board,* 1645**,** 17-23.

Tobler, W. R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* 46**,** 234-240.