Australasian Transport Research Forum 2022 Proceedings
28-30 September, Adelaide, Australia
Publication website: http://www.atrf.info

Developing a shopping destination choice model for people living in Metropolitan Melbourne

Annitta Lin[1], Yifang Chen[1], Dale Wines[1], Neema Nassir[1]

[1]Faculty of Engineering & IT, The University of Melbourne, Australia

Email for correspondence: neema.nassir@unimelb.edu.au

Abstract

One key factor to assist effective infrastructure and urban planning is to better understand the travel demands and the impact from various attributes on travellers' decisions for destination. This study develops a destination choice model for shopping trips in Melbourne. The model predicts travellers' choices for their shopping destinations from different areas in Melbourne based on demographic and trip related characteristics of individuals recorded in the Victorian Integrated Survey of Travel and Activity (VISTA) between 2012 and 2016.

One of the main challenges overcome in this research is the choice set formation. ASGS Statistical Area 1 (SA1) has been adopted as zones for categorising origins and destination locations. For each observed trip from the Victoria travel survey, nine alternative destinations within 10km travel distance from the origin are randomly generated to form the choice set together with the observed destination. The final model is calibrated and to have up to 67% accuracy in the validation process and obtain a goodness of fit of 0.65, demonstrating promising performance. The final model indicates that three main attributes, travel distance, the number of shops and supermarkets in the destination precinct has the most significant impact on the choice of shopping destination in Melbourne. In addition, trip-specific attributes such as the number of stops made during a trip, the intended dwelling time at the destination and whether the travel is made during peak time (10am – 3pm) have varying levels of impact to destination choice on top of the three main attributes.

1. Introduction

In the area of applying discrete choice model in transport demand forecasting studies, numerous contributions have been made to mode choice modelling. However, another fundamental pillar of transport decision, selecting trip destination, specifically for shopping or recreational purpose trips, has been little-studied due to its distinct challenges. The main challenge is the destination choice set formation (Hassan 2019). Also, it has often been argued that destination choice and mode choice decisions occur simultaneously (Richards and Ben-Akiva 1974). This paper aims to simplify modelling assumptions and analyse destination choices observed in the Victorian Integrated Survey of Travel and Activity (VISTA) as discrete decisions agnostic to and made after mode choice.

To identify the effective factors to the choice decision making, this research proposes to use Random Utility Maximization structure embedded in Multinomial Logit model. Based on the findings from previous literature, the variables included and tested in the systematic utility function include the accessibility of the destination in terms of travelling distance or time (Recker and Kostyniuk 1978), the size of the shopping precincts and amenities within a close proximity (Arentze and Timmermans 2001) and other sociodemographic characteristics related to travellers (Shobeirinejad et al. 2013). Empirical data surveyed between 2012 and 2016 from the VISTA dataset is adopted as the primary observed data. This is facilitated by other resources of data including TripGo API queried data, Open Street Map and Australian Bureau of Statistics Census Data (ABS) as secondary sources of information for trips.

The study generates destination choice set based on accessibility of the destination by restricting the travel distance to be within 10km from the origin. The 10km distance is determined to appropriately reflect the average choices travellers usually considered for their shopping trips and has successfully assisted to train the model to produce good prediction outcome.

The paper focuses on the methodology adopted for solving the two main challenges mentioned above and model outcomes of which as well as its implication for further study and the broader community in terms of what attributes have the most significant impact on traveller's destination choices for shopping trips in Melbourne.

2. Literature review

2.1. Destination choice modelling

Within transport modelling, trip distribution is the second step in the traditional 4-step model. The purpose of which is to model the number of trips between given origins and destinations. The most commonly used procedure is the gravity model (ATAP 2016), however as noted by Mishra et al. (2012) with the more diverse travel patterns and improvements in data collection, discrete destination choice models supported by random utility theory offer more flexibility and potentially better model outcomes.

Under utility maximisation theory, individuals will select alternatives (destinations) which have the greatest utility/value for them. The observable or deterministic portion of utility U, estimated by the analyst is its systematic utility V.

$$V_{ik} = \Sigma_{n=1}^{N} \beta_n X_{in} + \Sigma_{c=1}^{C} \beta_c S_{kic} \quad (1)$$

V_{ik}, the systematic utility of an alternative destination i, for a given individual k
β_n, the weight or importance of the attribute n (coefficient)
β_c, the weight or importance of the product c (coefficient)
X_{in}, the value of a trip or destination attribute n
$\boldsymbol{S_{kic}}$, the value of the product of a personal attribute dummy/binary variable and trip or destination attribute c, for a given individual k
The set of combinations C are mutually exclusive and collectively exhaustive.

$$U_{ik} = V_{ik} + \epsilon_{ik} \quad (2)$$

U_{ik}, the total utility of an alternative destination i, for a given individual k
ϵ_{ik}, the random components of the utility (error term)

Popularised by McFadden (1974), the multinomial logit (MNL) model is a commonly used logit model that assumes that the error term ϵ is Gumbel distributed with "the error components identically and independently distributed across alternatives, and identically and independently distributed across observations/individuals" (IID property) (Koppelman and Bhat 2006, p.26).

The basic MNL model, has a strict independence from irrelevant alternatives (IIA) property, which stipulates "for any individual, the ratio of the probabilities of choosing two alternatives is independent of the presence or attributes of any other alternative" (Koppelman and Bhat 2006, p.38)

The probability $Pr(i)$ of the individual choosing alternative i from a set of J alternatives, where V_j is the systematic component of the utility of alternative j.

$$\boldsymbol{Pr(i)} = \frac{\boldsymbol{exp(V_i)}}{\boldsymbol{\Sigma_{j=1}^{J} exp(V_j)}} \tag{3}$$

More structurally complex logit models have also been used in destination choice modelling and joint destination-mode choice modelling, with the strict IIA assumptions relaxed in nested logit, latent class (Wu et al. 2011), mixed logit and multinomial probit models (Dahlberg and Eklöf 2003).

2.2. Attractiveness of destinations

One of the key drivers of understanding the attractiveness of destinations is to understand where people typically travel to, depending on their purpose. Useful insights can be drawn by considering existing shopping precincts and key points of interest in Melbourne and identifying patterns between these key destinations. For example, there are a handful of very large shopping precincts in Melbourne which are contained in commercial/business only planning overlays preventing other development. Larger shopping centres are ‘one-stop’ centres of activity which have been strategically designed to maximise the amenity of them as destinations (Chatzopoulou and Tsimonis 2010), while smaller shopping precincts are convenience based.

Kristoffersson et al. (2018) found that people tend to choose from a broader choice set of destinations when the activity duration at the destination is longer. In other words, people tend to choose from a smaller set of destinations if they are likely to spend a shorter amount of time at the destination. The same paper concludes that grouping attractions together/within a close proximity of each other adds to the measured attraction of the destination. This notion is also reflected in Arentze and Timmermans (2001). This means that travellers will consider a destination more attractive if there are secondary attractions within a 2-5km proximity, regardless of the length of time spent at the primary destination. It is insightful that a potential destination-based variable to incorporate into the model is the availability of shops at a destination.

Bernardin et al. (2009) reciprocate this notion in their examination of trip-chaining, a phenomenon where travellers choose destinations such that they can combine multiple nearby stops into one convenient trip. They found that incorporating agglomeration effects into a destination choice model allowed the model to outperform models which did not consider agglomeration, similarly, suggesting that availability of alternatives at a primary destination improves its amenity. It has been understood for a while that trip chaining is a growing phenomenon, with Strathman and Dueker (1995) examining non-work stops taken on work-related trips and concluding that the complexity of trip-chains was increasing. Currie and Delbosc (2011) also concluded this for public transport trip chains in Melbourne, suggesting that travellers are resorting to trip-chaining if their needs cannot be met at one destination. Overall, it is anticipated that destination-based attributes such as availability of shops and overall destination amenity will be governing factors in a person’s destination choice, in addition to trip-based attributes such as number of stops required.

According to Recker and Kostyniuk (1978), accessibility to a destination in the form of travel time, is considered a primary factor in determining destination choice. For zonal destination choice alternatives, Barnard (1987) identifies the following destination attributes which MNL models of shopping destination choice have included: retail employment figures; retail area; zonal area; dummy variable for CBD destinations capturing urban/suburban differences.

Shobeirinejad et al. (2013) has also found significance in the sociodemographic attributes of travellers, age; gender; and income, in their Brisbane, Australia retail destination choice modelling.

2.3. Choice set formation

The problem of choice set formation for destination choice models exists because of the very large set of possible alternatives that are available to the trip maker, of which, in reality, only a small proportion is within the decision maker's consideration. In the choice set formation framework, decisions are made in two stages, the choice of a choice set and the choice of the best alternative (Spiridonov 2010). The reliability of the predictions of a model and the estimation of model parameters is dependent on the choice set provided to the model, the size and quality of the data used for model estimation (Pagliara and Timmermans 2009). Where the choice set is mis-specified, it is expected that parameter estimates are inconsistent and produce unintuitive explanations for destination choice (Thill 1992). The subset of the universal choice set of which a trip maker realistically takes into consideration can be defined through behavioural, time-space prism, time budget and spatial constraints (Scott and He 2012). Other commonly used, less intensive methods for choice set specification include predetermined alternatives which are endogenous, including only destinations chosen by individuals in the same geographic area, on the assumption that they are spatially constrained in a similar way (Thill 1992); alternatives within a set distance or travel time from each trip's origin (Pagliara and Timmermans 2009); and a simple random sampling of the unchosen alternatives without need for sampling corrections due to the IID property (Scott and He 2012).

3. Methodology

3.1. Data

The following datasets have been used in this study:

Victorian Integrated Survey of Travel and Activity (VISTA): VISTA household travel survey data from 2012 – 2016 (DEDJTR 2017) which provides records for a total of 10,625 shopping trips within the Greater Melbourne Greater Capital City Statistical Area or "Metropolitan Melbourne" and is the primary dataset contributing to the model. At their finest level of detail, trip origins and destinations are reported and adopted in the model as 2011 ASGS Statistical Areas Level 1 (SA1) zones (ABS 2010). Statistical Area 1 is categorised with average population of 400 people. In VISTA, four main tables characterised four main aspects of that can be incorporated in the model as listed following:

- H-table (Household data) & P-table (Person or trip-maker data): socio-demographic information about both the trip maker and their household includes income level, gender, employment status, children in household, car ownership level, licence ownership, age, and duration planned to stay/actual stay at the destination.
- T-table (Trip data) & S-table (stop data): information related to the trip observed in the survey, including the travel date, time of the day, length of stay at destination, traveling alone, mode chosen, and number of stops made.

OpenStreetMap (OSM): Geographic point of interest crowdsourcing data relating to the location and size of shopping centres, stores, supermarkets and public transport stops/stations (OpenStreetMap contributors 2015).

Australian Urban Research Infrastructure Network – AURIN Portal: This study used the NCRIS-enabled Australian Urban Research Infrastructure Network (AURIN) Portal e-Infrastructure (Sinnott et al. 2014) to access the following data sources in 2021:

- **ABS Census 2011:** Census night responses including SA1-based datasets for counts of employed persons.
- **VIC DSDBI – Industry Atlas:** Datasets providing information about business size by turnover, number of employees and industry sector in Victoria at SA2 level. Statistical Areas Level 2 zones consist of multiple SA1 zones and with a total population of 3,000-25,000, which represents individual communities socially and economically interacting with each other (ABS 2016).

These datasets are consolidated in three phases. From the VISTA dataset, selected trip observations and their household, person and trip attributes are extracted using PostgreSQL. Categorical variables are also converted into binary values, these are largely personal or trip attributes which are incorporated into the model by interacting with destination attributes. This forms the initial subset personal and trip attribute database.

For each selected trip, the observed destination SA1 and simple random sampling of nine unchosen alternative SA1s within a fixed 10km radius of the origin is included in the choice set, as outlined in Section 2.3.

After the choice set of alternative destinations is prepared, travel distance is calculated using the straight-line distance between the centroid of each origin and destination zone. Tripgo API (SkedGo Pty Limited, n.d.) is used to estimate travel times for each alternative destination using query package built upon the one written by Leong (2020).

OSM data is extracted using Overpass API and overlaid with 2011 ASGS SA1 boundaries in QGIS to transform and match locations with their destination SA1 zones. This is combined with other destination specific attributes exported through the AURIN portal, linked using their SA1 or SA2 code. This forms the subset destination attribute database.

The final combined dataset consists of over 50 attributes including travel time by chosen mode, aerial distance and personal, trip, and destination attributes (number of shops, number of supermarkets, destination within the CBD, number of public transport stations, total employed persons, retail business turnover, retail business employees) for consideration.

3.2. Model calibration and validation

3.2.1. Calibration

The destination choice model in this discrete choice analysis is a basic multinomial logit model which provides a starting point in explaining shopping destination choice behaviour of people in Metropolitan Melbourne and to inform the development of future more flexible models. The model assumes the mode choice is determined before or is unaffected by destination choice. Using maximum likelihood estimation (MLE), as described in Section 2.1, Biogeme (Bierlaire 2020) is used to estimate the model coefficients.

Explanatory variables that remain in the MNL model are significant, improve the model goodness of fit and have coefficients/signs that have intuitive explanations. A simplified and a complex model are calibrated, which exclude and include personal or trip-specific attributes interacting with trip origin-destination specific and destination attributes respectively.

3.2.2. Validation

80% of the data obtained is used for model calibration and 20% of the data is reserved for validation. Two methods will be adopted to validate the results. The first one is to fit the utility function produced by the training set to the data in the validation set, comparing the calibrated coefficients to ensure choice sets are not mis-specified. In the second method of validation the validation data and choice sets are applied to the final calibrated utility function. The accuracy of the model is indicated by the ratio between the sum of probabilities for the chosen destination and the alternative destinations, described by Parady et al. (2020) as the ‘fitting factor’.

4. Results

This section summarises the findings of the study into shopping destination choice modelling for Metropolitan Melbourne

4.1. Model iteration: Base Model

To develop the model, various destination relevant or trip specific attributes were tested to determine the most important variables, being travel time (TT), linear aerial distance (Distance or Dist), and the number of shops (Shops) and supermarkets (Supermarkets or Superm) within an SA1.

Table 1 below is a summary of the iteration process to determine the optimal destination choice model based on destination specific attributes. Each key attribute is presented individually, as well as the best model from combinations of 2, 3 and 4 attributes.

Table 1: Iteration summary

Variables	Goodness of Fit	Fitting Factor	Travel Time (mins) Coefficient	Distance (km) Coefficient	Shops (#) Coefficient	Supermarkets (#) Coefficient	Coefficient p-value
Travel Time	0.347	35.46%	-0.169				0
Distance	0.353	33.79%		-0.586			0
Shops	0.297	37.26%			0.135		0
Supermarkets	0.381	44.66%				1.69	0
Base Model	0.626	63.82%		-0.596	0.0451	1.27	0 (All)

Table 1 indicates that the presence of a supermarket has the most significant goodness of fit of all models (0.381). When the utility functions are applied to the 20% of validation data it is found that the supermarkets attribute has the greatest model fit of 44.66%, with the other models performing at around 33-38% each, with the base model combining distance, shops and supermarkets having a fitting factor of 63.82%.

As expected for these single attribute utility functions, the coefficients for distance and travel time are both negative (-0.169 and –0.586 respectively), where the utility of a destination decreases by 1 for approximately every 6 minutes of additional travel time or every 1.71km of additional travel distance. Supermarkets have a significant effect on the utility function, with every additional supermarket at a location providing an additional utility of 1.69. This is anticipated due to the small number of supermarkets in any one destination SA1 area, considering shopping precincts typically only have around 1-3 supermarkets depending on the size of the precinct. The utility increases by 1 for approximately every 7.4 shops, which indicated that a greater quantity (and arguably variety) of shops is required to increase attractiveness. It should be noted that the p-value for all coefficients is 0, which indicates a strong statistical significance and a low likelihood of particular coefficients demonstrating extreme results.

The base model will be assessed further in subsequent sections. As travel time and distance are too similar to be appropriately considered together in this model, the distance model is deemed as more reliable in terms of the initial data sourcing and the results obtained in the preliminary model testing.

4.2. Model iteration: Composite Model

For the second stage of modelling, the best performing destination attributes from the base model were amalgamated with personal and trip attributes from the VISTA data, which will be referred to as composite variables. This is done as VISTA data does not include data for alternative destinations and only the recorded destination a person travels to or from. Therefore, if these attributes are combined with another attribute that does change based on the destination (all coefficients tested in the base model) then this can account for the effect of personal and trip attributes from VISTA data have on destination choice in the model.

A variety of composite variables were tested, and the most successful variables are presented as the 'Composite Model'. Variables that have a range (I.e. time spent at a destination) were tested to find a 'threshold' value. For example, 15 minutes was the threshold that produced the greatest goodness of fit and a distinct difference between the coefficient for more than 15 minutes and under 15 minutes spent at a destination. It is important to note that all variables are tested under the condition that they are mutually exclusive and collectively exhaustive. This ensures that the same observed trip is not tested twice against the same attribute and so that data is not missed while capturing all trips, as demonstrated in Figure 1 below.

Figure 1: Composite model tree diagram

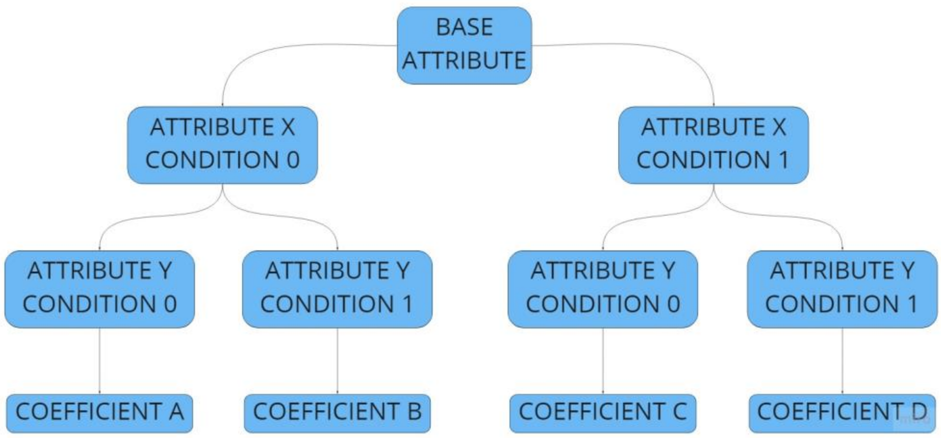
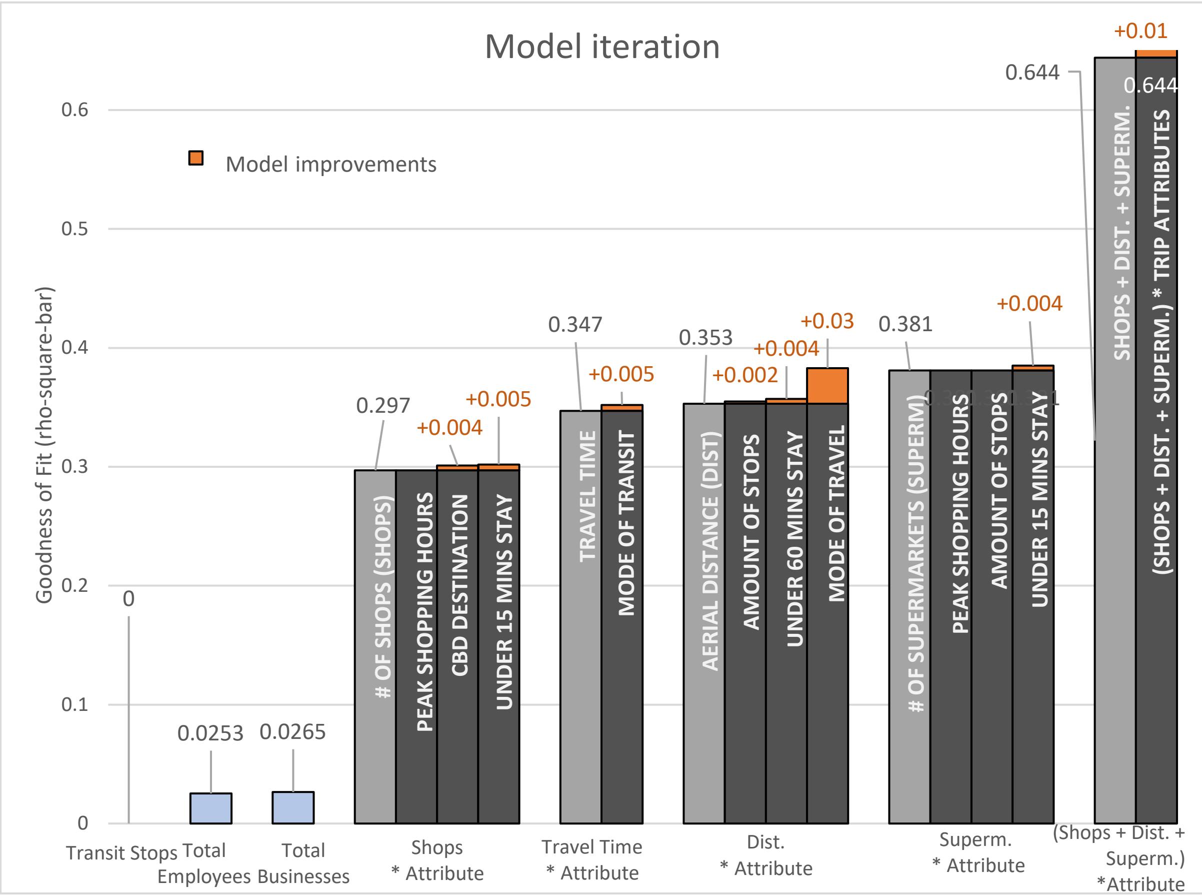


Figure 2 below is a summary of the iteration process to determine the optimal destination choice model based on coupled destination and trip/personal attributes.

As shown in Figure 2, the transit stops, total employees and total businesses are not used due to a poor goodness of fit value. The models with the light grey bars indicate the destination only tests from the 'simplified' model as shown above. The dark grey and orange bars indicate where an individual VISTA attribute has been applied to the destination data. These show the incremental improvement which each attribute makes to the model. As shown the following attributes from VISTA have some effect on the trip choices: the mode of choice (I.e. car, public transport, walking, cycling), how many stops are made throughout the whole trips, whether they are travelling in peak shopping periods of 10am to 3pm (this was the only attribute without a significant improvement to the model which was left in as it does have significance when coupled with other attributes later on), the length of stay at a destination (I.e. under/over 15 minutes or under/over 1 hour) and whether the destination is located in the CBD or not. These attributes all contribute to the final composite model which is presented in the subsequent sub-sections below.

Figure 2: Goodness of fit for different model iterations



4.3. Composite Model results

Table 2 (Page 9) is an overview of the coefficients used for the composite model. Including a breakdown of composite variables and the associated model results. All composite variables are mutually exclusive and collectively exhaust all different options in the data.

The utility function from Table 2 returns a model goodness of fit of 0.654. The maximum p-value for all the coefficients is 0.0185 which remains less than 0.05 (which is for number of shops * if staying for over 15 minutes * shopping during peak hours of 10am – 3pm * destination in the CBD). The correlation between attributes is also considerably lower for most coefficients when compared to the simplified model.

4.3.1. Out of sample validation

The predictability rate increases significantly from the base model to the composite model. Where the base model is 0.6382 and the composite model has a fitting factor of 0.674. It should be noted that the 'composite model' which includes VISTA attributes does not demonstrate a significant improvement on the predictability. Therefore, the destination choice prediction validation test has confirmed that the original training models developed perform well and are able to predict the choice of destination for individuals living in Metropolitan Melbourne with approximately 63-68% accuracy in the final models.

Table 2: Composite model utility function coefficients

Destination Variable	Travel mode?	Multiple stops?	Peak time? (10am-3pm)	Staying for over 15 minutes?	Staying for over 60 minutes?	Destination in the CBD?	Coefficient	Standard Deviation	t-test	p-test	N
Distance	Car	Yes	N/A	N/A	N/A	N/A	-0.241	0.0788	-3.06	0.0022	82
Distance	Car	No	N/A	N/A	Yes	N/A	-0.407	0.0211	-19.2	0	1337
Distance	Car	No	N/A	N/A	No	N/A	-0.57	0.00981	-58.1	0	5326
Distance	Other	Yes	N/A	N/A	Yes	N/A	-0.518	0.112	-4.64	0	223
Distance	Other	Yes	N/A	N/A	No	N/A	-0.671	0.11	-6.08	0	106
Distance	Other	No	N/A	N/A	Yes	N/A	-1.04	0.498	-2.08	0.0373	12
Distance	Other	No	N/A	N/A	No	N/A	-0.938	0.129	-7.28	0	69
Distance	PT	N/A	N/A	N/A	Yes	N/A	-0.511	0.0944	-5.42	0	186
Distance	PT	N/A	N/A	N/A	No	N/A	-0.222	0.0521	-4.25	0	305
Distance	Walk	N/A	N/A	N/A	Yes	N/A	-2.96	0.465	-6.37	0	99
Distance	Walk	N/A	N/A	N/A	No	N/A	-2.44	0.14	-17.4	0	1210
Shops	N/A	N/A	No	Yes	N/A	Yes	0.0392	0.00552	7.1	0	670
Shops	N/A	N/A	No	Yes	N/A	No	0.0657	0.0044	14.9	0	3168
Shops	N/A	N/A	Yes	Yes	N/A	Yes	0.0608	0.0131	4.63	0	223
Shops	N/A	N/A	Yes	Yes	N/A	No	0.0385	0.00552	6.97	0	1563
Shops	N/A	N/A	N/A	No	N/A	Yes	0.0188	0.00597	3.15	0.0016	689
Shops	N/A	N/A	N/A	No	N/A	No	0.0512	0.00464	11	0	2642
Supermarket	N/A	Yes	No	Yes	N/A	N/A	1.47	0.164	8.96	0	451
Supermarket	N/A	Yes	Yes	Yes	N/A	N/A	1.65	0.229	7.22	0	264
Supermarket	N/A	Yes	N/A	No	N/A	N/A	0.953	0.168	4.49	0	190
Supermarket	N/A	No	No	Yes	N/A	N/A	1.43	0.0487	29.3	0	3387
Supermarket	N/A	No	No	No	N/A	N/A	0.969	0.0527	18.4	0	2447
Supermarket	N/A	No	Yes	Yes	N/A	N/A	1.39	0.0675	20.6	0	1522
Supermarket	N/A	No	Yes	No	N/A	N/A	0.586	0.0814	7.2	0	694

5. Discussion

5.1. Findings

5.1.1. Analysis of variables

When assessing the destination choice models to establish features that are important for the traveller, the coefficient indicates the influence of a variable on an individual's choice of destination. For example, the distance coefficient in the base model is –0.596, this means that for every additional kilometer travelled, a shopping destination of one or more shops or supermarkets loses its attractiveness by 0.596 of utility. In the base model, supermarkets have a high coefficient of 1.27. However, for shops this coefficient is only 0.0451, which indicates that a larger number of shops is required for a destination to have the same amount of attractiveness of a supermarket for a given origin-destination distance. Where there is an additional supermarket the utility would increase by 1.27, the number of additional shops to generate the same increase in utility would be 29 shops. This trade off, along with other trade offs between attributes in the base model are shown in Table 3 below.

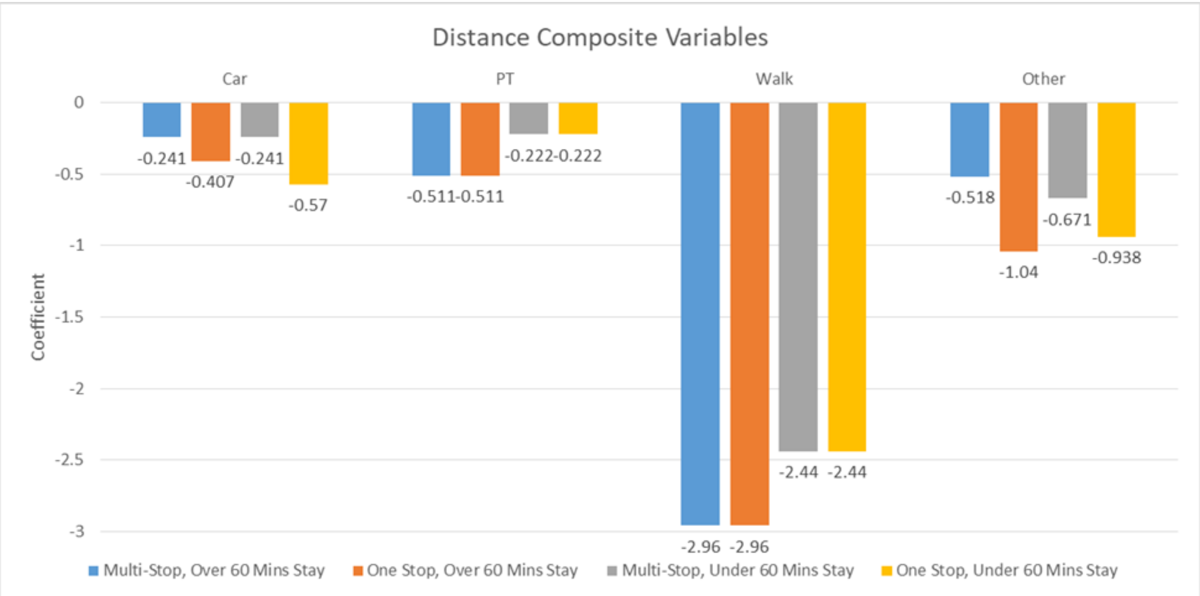
Table 3: Marginal rate of substitution (MRS) of attribute X for attribute Y

X	Y	MRS
Supermarkets	Shops	28.160
Supermarkets	Distance less travelled (km)	2.131
Shops	Distance less travelled (km)	0.0757
Distance less travelled (km)	Shops	13.215

The marginal rate of substitution (MRS) calculated as the ratio of the respective model coefficients in this simple case with linear utility functions, indicates the distance people would be willing to travel to shopping destinations. A small local shopping village of a dozen retail businesses of various industries and goods not including a supermarket, may attract residents from a modest 1 km radius. However, a single supermarket twice as far away, is just as attractive, a retail catchment area four times larger.

Whilst the distance, number of shops and supermarkets form a basic derivation of destination attraction, further nuances in travel behaviour can be found using composite variables.

Figure 3: Comparison of distance composite variables



Three variables are part of a composite with the base attribute travel distance: mode choice, whether the trip consists of one or multiple stops and whether the dwelling time at the shopping destination exceeds 60 minutes. Figure 3 above summarises the impact these attributes have on the destination choice, grouping by trip mode choice.

Overall, we can see there are distinct differences in the coefficient between car & PT, walking and 'other' modes, which signifies that these variables have a distinct influence on the distance travelled. This is intuitive and expected given variation in the speed of travel and perceived effort for each travel mode. Car and public transport are the preferred modes of travel with higher marginal utility. This is particularly noticeable when the traveller intends to stay under 60 minutes at a retail destination. The MRS of distance travelled by public transport for distance travelled by walking is 0.0910. This demonstrates the strong effect walking has on an individual's willingness to travel with walking trips typically performed in a shorter range.

The number of stops in a trip is not significant factor for shopping trips by public transport or walking. On the other hand, motorised modes such as Car and those included in 'Other' are more favoured in multi-stop trips compared to single stop trips by the same mode. When an individual intends to make multiple stops travelling by car, the distance is less of an impendence than for a single stop trip. With an MRS of 2.37, an individual is equally happy to travel more than twice as far from their origin to their final shopping destination when making a multi legged trip by car, than when making a one stop trip to a destination with the same shops and supermarkets composite attributes.

Similarly, number of shops at a destination is combined with three variables, whether the dwelling time at the shopping destination exceeds 15 minutes; whether shopping occurs during the busiest (peak) hours of shopping between 10am and 3pm; and whether the shopping location is within the Melbourne CBD. Number of supermarkets at a destination is combined with three variables, whether the dwelling time at the shopping destination exceeds 15 minutes; whether shopping occurs during the busiest (peak) hours of shopping between 10am and 3pm; and whether the trip consists of one or multiple stops. Figure 4 shows the influence of these variables by grouping coefficients by whether the dwelling time at destination exceeds 15 mins.

Figure 4: Comparison of other destination & trip-specific coefficients for Composite Model

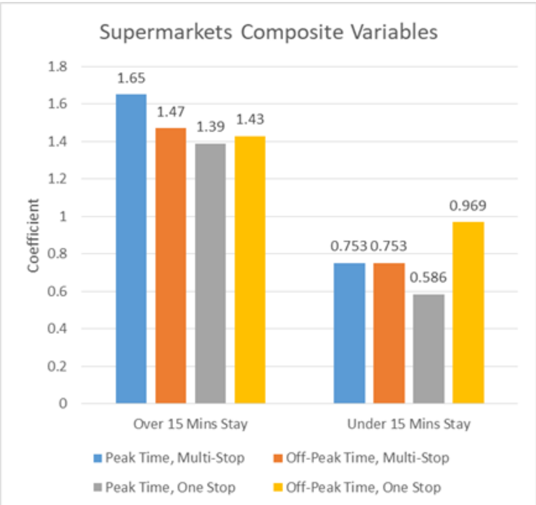


Figure 4 shows that the coefficients for number of shops has the least variability, that is, interacting attributes do not have a very significant effect on how this attribute performs. Although these coefficients also deal with numbers in much larger ranges (I.e. there are many more shops than supermarkets). This means that the interacting trip-specific variables are more likely to influence the coefficients for distance and number of supermarkets.

From the analysis, the composite variables for over 15 minutes and under 15 minutes each have a distinguished range which the composite supermarket coefficient falls within. For example, for over 15 minutes the coefficient is much higher between 1.39 and 1.65, whereas for under 15 minutes the coefficient is much lower between 0.586 and 0.969. However, for the composite shops variables this range is much wider, with over 15 minutes between 0.0385 and 0.0657, and under 15 minutes between 0.0188 and 0.0512. This means that the 15-minute threshold has a more distinct impact on the coefficient of composite supermarket variables than composite shops variable. The nature of grocery shopping that generally happens in supermarkets, perhaps due to the range of items available is favoured by those with the time budget for longer stays in contrast with shops which typically have a number of specific known goods.

The MRS of shops in peak time when staying over 15 minutes at CBD locations (coeff. 0.0608) for shops not in CBD locations (coeff. 0.0385) is 1.58. This indicates the strength of CBD retail and the difficulty for smaller shopping precincts in suburban/inner Melbourne to draw in consumers in the daytime. Yet for short stays under 15 minutes, perhaps where items to be bought are planned and visits are transactional, time of day is not considered a significant factor. For these trips, individuals are not likely to venture into the CBD with shops not in the CBD more than twice as attractive all else, including distance to the destination, equal. MRS of shop not in CBD for shop in CBD is 2.72 for under 15 minute stays.

Another point to note is the lack of personal or household data from the model. The study did assess a range of potentially relevant demographic characteristics and data from the h-table and p-table, although none of them passed the significance test for the goodness of fit. It was concluded that despite various predictions and hypotheses, demographical data has very little to no influence on an individual's choice of destination. This means that the destination characteristics or trip characteristics are the key governing features of shopping trips in Metropolitan Melbourne.

5.1.2. Out of sample validation

The model generated from the training set performed strongly against the validation set, with the base model returning 64% and the composite model returning 67% correct predictions. From various other literature that explores destination choice models, multinomial logit models and combinations of both, that validation using a 'holdout' dataset is common. Validation for these other models range from 11% to a 70.5%, these are assessed from data collected by Parady et al. (2020) for models which use the Percentage of correct predictions or First Preference Recovery (FPR); or Predicted vs observed market outcomes (PVO) validation methods. A summary of these is shown below in Table 4.

In comparison, both the base and composite models from this study perform better than most of the other models listed below. This could be due to the simplicity of the model's aim, the quality of data available and a large sample size of trips to select from (as the model was fit against 8418 trips and validated against 2207 trips). The strength and volume of the data used as part of the model can also be accredited to the accuracy of the validation sets.

Table 4: Summary of multinomial logit model validation

Model/Author	Validation (%)	Model Specification
Composite Model	67.40%	Destination variables combined with personal and trip attributes
Base Model	63.82%	Destination variables (distance, shops and supermarkets at a destination)
Glerum et al. (2014) as cited in Parady et al. (2020)	70.50%	Hybrid choice model for mode choice based on qualitative questions
Gokasar & Gunay (2017) as cited in Parady et al. (2020)	55.90%	Multinomial logit mode choice for airports in Istanbul, Turkey
Chikaraishi & Nakayama (2016) as cited in Parady et al. (2020)	53.00%	Multinomial logit for mode and route choice in the Tokyo Metropolitan Area
Faghih-Imani & Eluru (2015) as cited in Parady et al. (2020)	11.72%	Multinomial logit destination choice for bike sharing in Chicago

5.2. Limitations & recommendation

5.2.1. Origin - destination categorisation

The zones considered in this study as origins and destinations are under a population based geographical zoning categorization. Zones with vastly different sizes of land are compared with one another this reduces the accuracy in using the geographic centroids coordinates representing any trips generated from a zone. It follows that the travel distance or time calculated based on these coordinates can lack in accuracy to varying degrees. Observed trip datasets with more discrete origins and destinations would improve this and allow for greater flexibility in defining destination alternatives and use of further datasets providing more potential explanatory variables.

5.2.2. Choice set

In this study, choice sets are randomly generated, with alternative destinations limited only by a 10km radius from the origin. Further work on choice set specification would improve understanding of the destination alternatives which a trip maker considers including trip purpose, travel time. This determination is made after the failed testing on setting the boundary to 5km, in which case the coefficient for travel distance calculated from the model becomes positive. It is then realised as a significant finding that most chosen destination has a travel distance greater than 5km. For this reason, it is recommended to test other ranges of travel distance, generate choice sets based on travel time (eg. Below 30 mins, below 15 mins), time budget or other factors to reflect travellers constraints (Scott and He 2012) or change the size of the choice set (Shobeirinejad et al. 2013) to avoid mis-specified choice set and therefore unintuitive explanations of destination choice (Thill 1992).

5.2.3. Sequence of related decisions

The model is built based on a few assumptions where some factors are considered as exogenous to the destination choice. the fundamental one is that the mode choice of travellers is formed before the decision of destinations is made. This does not capture the dependency between mode and destination choice in reality. This can be addressed by creating a joint destination-mode, cross nested logit model to incorporate the decision for both choices simultaneously. Alternatively, a simplified solution can be to gain empirical data to verify the sequence of the decision-making process. Simlar to the mode choice in travelling, other attributes such as the time of the day or the intended stay duration at destination does not necessarily stand independently from the choice of destination in reality. Future study could look into the possible, more likely sequences of decision making around destination choice to enable a more holistic, accurate representation of the trip behaviours.

5.2.4. Trip inclusion relating to trip chaining issues

The model only considered trips in VISTA data with "Destination Purpose" being "Buy Something". This means other trips with the component of shopping in a section of it, such as a stop-by shopping trip on the way home from work, are not captured in the model. Resolving this could enhance the understanding of shopping trip behaviour and the interconnection between different purposes of trips.

5.2.5. Trip purpose breakdown

To extend on the sub-section 5.2.4, as the observed trips in the model are filtered by the broad definition, "Buy Something", the model does not further distinguish the incentives affecting different kinds of shopping trips. For example, travellers' main destination purpose is grocery shopping would factor the number of supermarkets more than shops as opposed to those who would need clothes shopping. To specify these sub-categories of shopping trips could improve the accuracy of the model.

6. Conclusions

The study has developed a discrete destination choice model using random utility theory to estimate people's choice of shopping destination in Metropolitan Melbourne. The multinomial logit models (MNL) produce utility functions to quantify the attractiveness of a destination to individuals from three main aspects, trip, destination and personal attributes. It is also crucial to ensure the choice set generation methodology was appropriate for shopping type of trips.

There are two versions of the model produced as part of this study. The base model has a utility function consist of three main variables, linear aerial distance, number of shops and number of supermarkets at destinations as they appear to provide the most impact on people's decisions on choosing a shopping destination; whereas the composite model incorporates more trip-specific attributes such as the duration of stay at a destination, whether or not the destination is in the CBD, into the utility function for enhancing precision. The composite model is validated with 67% successful prediction rate. Though highly specific and mostly applicable to Melbourne, this compared favourably to existing destination choice models in the literature.

The purpose of the base model is to improve the understanding of destination and trip related attributes in affecting destination attractiveness in Melbourne for industry application, such as an integration for the Victorian Integrated Transport Model (VITM). The composite model captures the characteristics of individuals travelling in terms of their trip-specific attributes. Among the three attributes, travellers are similarly sensitive to the distance and the number of supermarkets at the destination. However, travellers are shown to be notably less elastic to the number shops in a destination zone. In comparison, personal attributes do not have as great an influence on people's destination choice for shopping as trip or destination attributes including travel distance and number of shops and supermarkets.

The main challenge in this study is the choice set generation process as it has a crucial impact on the outcome of the model. It can be prioritised to be further tested and improved in future study. It is recommended for further investigation to relax assumptions made in this research including mode choice before destination choice and population-based categorisation of zones by considering additional sources of data or incorporate other theories such as trip chaining, nested logit model.

7. Acknowledgements

The report authors would like to express their gratitude to Joseph Leong and Patrick Jankowski who volunteered his time to assist with technical and conceptual issues in the research. Their contributions are appreciated profusely. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org

8. References

Arentze TA and Timmermans, HJP (2001) 'Deriving Performance Indicators from Models of Multipurpose Shopping Behaviour', *Journal of Retailing and Consumer Services*, 8(6): 325-334.

ABS (Australian Bureau of Statistics) (2010) Victoria Mesh Blocks ASGS Ed 2011 Digital Boundaries in ESRI Shapefile Format.

ABS (Australian Bureau of Statistics) (2016) *1270.0.55.001 – Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2016*, Australian Bureau of Statistics, https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1270.0.55.001Main+Features10018July%202016?OpenDocument

Barnard PO (1987) 'Modelling shopping destination choice behaviour using the basic multinomial logit model and some of its extensions', *Transport Reviews*, 7(1): 17-51.

Ben-Akiva M and Lerman S (1985) *Discrete Choice Analysis: theory and application to travel demand*, MIT Press.

Bernardin VL, Koppelman F and Boyce D (2009) 'Enhanced destination choice models incorporating agglomeration related to trip chaining while controlling for spatial competition', *Transportation Research Record*, 2132: 143-151.

Bierlaire M (2020) A short introduction to PandasBiogeme.

Chatzopoulou E and Tsimonis G (2010) 'Shopping Malls: Just Another Shopping Place? The Case of Physical Environment Quality', *Academy of Marketing Conference*.

ATAP (Australian Transport Assessment and Planning) (2016) T1 Travel Demand Modelling, ATAP, Australian Government, https://www.atap.gov.au/sites/default/files/T1_Travel_Demand_Modelling.pdf

Currie G and Delbosc A (2011) 'Exploring The Trip Chaining Behaviour of Public Transport Users in Melbourne', *Transport Policy*, 18(1):204-210

Dahlberg M and Eklöf M (2003) 'Relaxing the IIA Assumption in Locational Choice Models: A Comparison Between Conditional Logit, Mixed Logit, and Multinomial Probit Models'

DEDJTR (Department of Economic Development, Jobs, Transport and Resources) (2017) Victorian Integrated Survey of Travel and Activity 2012-2016 Data [data set].

Hassan M (2019) 'A two-stage recreational destination choice study incorporating fuzzy logic in discrete choice modelling', *Transportation Research Part F*, 67:123-141.

Koppelman FS and Bhat C (2006) *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*.

Kristoffersson I, Daly A and Algers S (2018) 'Modelling the attraction of travel to shopping destinations in large-scale modelling', *Transport Policy*, 68:52-62.

Leong J (2020) Tripgo Routing API Parser. https://github.com/leongjoseph/tripgo_parser

McFadden D (1974) 'The Measurement of Urban Travel Demand', *Journal of Public Economics*, 3:303-328.

Mishra S. Wang Y, Zhu X and Moeckel R (2012) 'Comparison between Gravity and Destination Choice Models for Trip Distribution in Maryland', Transportation Research Board 92nd Annual Meeting.

OpenStreetMap contributors (2015) Planet dump [Data file from August 2021]. https://planet.openstreetmap.org

Pagliara F and Timmermans HJP (2009) 'Choice set generation in spatial contexts', *Transportation Letters, The International Journal of Transportation Research*, 1(3):181-196.

Parady GT, Ory D and Walker J (2020) 'The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature', *Journal of Choice Modelling*, 28(5).

Recker WW and Kostyniuk LP (1978) 'Factors Influencing Destination Choice For The Urban Grocery Shopping Trip', *Transportation*, 7:19-33.

Richards MG and Ben-Akiva M (1974) 'A Simultaneous Destination and Mode Choice Model for Shopping Trips', *Transportation*, 3: 343-356.

Scott DM and He SY (2012) 'Modeling constrained destination choice for shopping: A GIS-based, time-geographic approach', *Journal of Transport Geography*, 23: 60-71.

Shobeirinejad M, Veitch T, Smart J, Sipe N and Burke M (2013) 'Destination choice decisions of retail travellers: results from discrete choice modelling in Brisbane', *Australasian Transport Research Forum 2013 Proceedings*.

SkedGo Pty Limited, n.d. TripGo API, https://skedgo.com/tripgo-api

Spiridonov GB (2010) Choice set formation and its implication for welfare measures in a random utility model of recreation demand, *ProQuest Dissertations and Theses*, 129.

Strathman JG and Ducker KJ (1995) 'Understanding Trip Chaining', *Special Reports on Trip and Vehicle Attributes: Based on Data from the 1990 Nationwide Personal Transportation Survey (NTPS) :6-28*, Washington, DC: Department of Transportation, Federal Highway Administration.

Thill JC (1992) 'Choice set formation for destination choice modelling', *Progress in Human Geography*, 16(3): 361-382.

Wu L, Zhang J and Fujiwara A (2011) 'Representing tourists' heterogeneous choices of destination and travel party with an integrated latent class and nested logit model', *Tourism Management*, 32(6):1407-1413.