Choice Set Formation Behaviour: A Joint Mode and Route Choice Selection Model

Kiran Shakeel, corresponding author
School of Civil and Environmental Engineering
UNSW Australia
Sydney NSW 2052, Australia, kiran.shakeel@unsw.edu.au

Taha Hossein Rashidi
School of Civil and Environmental Engineering
UNSW Australia
Sydney NSW 2052, Australia, rashidi@unsw.edu.au

Travis S. Waller
School of Civil and Environmental Engineering
UNSW Australia
Sydney NSW 2052, Australia, s.waller@unsw.edu.au

Word count: 5,500 words text + 2 figures + 4 tables + 35 (1000 words) references

Submission Date: 15/11/2015
Abstract
In the perspective of improving level of service of traffic and to predict travel demand, it is essential to analyse the behavioural factors affecting the transportation mode choice and route choice at the individual level. This requires detailed data on the behaviour of people in selecting different modes and routes. This paper presents a unique data collection endeavour to observe how the choice set of alternatives are formed from which the final alternative is selected. The revealed preference information about the route people considered and used for their last work/study trip purpose is targeted in this survey. Google maps API (Application programming interface) which has the capacity to calculate and return both car routes and public transportation routes is employed to program the survey and to adaptively show the respondents the routes according to their reported origin and destination.

A pilot survey has been conducted with this survey tool within the University of New South Wales with a sample of 200 respondents. A preliminary analysis is carried out to analyse the effectiveness of the survey tool and specification of the choice set. Parameters of the preliminary analysis are estimated using three modelling structures of multinomial logit, nested logit and mixed MNL. The results were pretty intuitive as far as signs of the parameters are concerned having travel time to significantly influence the route choice.


**Introduction**

In the advanced and growing economies like Sydney, people have various purposes for travelling to fulfil their needs. In the perspective of improving the level of service of traffic and for forecasting travel demand, it is essential to study factors behind the individual’s decision making for mode and route choice. Quantitative analysis of factors that influence the traveller’s mode choice and route choice play a major role evaluating policies which can result in the more sustainable and environmental friendly transportation system.

This paper presents a novel web based survey tool designed to collect data regarding the travel mode and route choice of individuals. This paper also takes into account various challenges associated with studying behavioural aspects of route choice within the discrete choice modelling framework as compared to the mode choice. The major challenge associated with modelling route choice behaviour is related to the fact that there are infinite number of alternatives available for each individual that satisfy their trip purpose. Further, the alternatives are over-lapping, that is, they are not mutually exclusive or collectively exhaustive as compared to the mode choice case, so route choice modelling follows a specific model specification within the discrete choice framework. This paper presents a unique method for collecting revealed preference information reflecting the observed and subjective choice set for each individual which is later used in the random utility maximization structure where the choice set is provided by the decision maker.

Another unique feature of the survey lies in utilizing google maps to provide people with close to real alternatives as these alternatives are supported by advanced data mining techniques developed by Google. Google maps is by far the most comprehensive route suggestion system designed to guide the road users and inform them of their particular route they want to travel on. With the growing rate of smart phone users, it is not unrealistic to assume that most people use this application to decide about their route selection. Google also provides the Google maps API (Application Programming Interface) which is a set of methods and tools that provides the opportunity to build a software application. This instrument is one of the building stones of the motivation behind the route choice modelling of this paper as it provides us the opportunity to devise the subjective and realistic choice set for revealed preference studies of the route choice modelling within the discrete choice framework that can provide the valuable insights of the routes which the drivers actually considers.

A pilot survey of 200 individuals was conducted in the University of New South Wales, Sydney, Australia. The survey designed will be described later in this paper along with the specification of choice set formation routine for the route choice modelling. A detailed descriptive analysis is then carried out on the data and for more deep understanding of it, results of the developed discrete choice model will be presented for both mode and route choice models. The parameters of the explanatory variables are estimated with three different structures of discrete choice models namely MNL, nested logit and mixed MNL for route choice. The mode choice parameters are estimated through traditional MNL model. Although due to the presence of overlapping route alternatives, these models are not appropriate as
some correction factors should be employed to approximate the overlapping, but only to test the survey tool effectiveness and the preliminary analysis of data, these models utilised. This survey method will then be employed for larger data set of Sydney residents and more advanced models will be developed and compared with these results in the future.

**Literature Review**

In order to mitigate the detrimental effects of congestion, the mode choice as well as route choice modelling within the behaviourally oriented framework is necessary to study the effects of factors influencing them.

In the past, discrete choice models have been in use by transportation research community to study factors influencing the mode choice of individuals as used by Ben Akiva and Lerman (1), Koppleman (4) and Bhat (3, 5, 6). Pinjari et al. (5) showed that urban form and residential location are also said to be influenced by the mode choice. The reliability of travel time has also been included within the framework of the mode choice behaviour in recent studies as shown by Small et al., Bhat and Sardesai (6) and Noland and Polak (7). Johansson et. al(8) analysed personality traits on the mode choice. The household constraints and travel distance were also taken as attributes by Scheiner et.al (9). The willingness to pay for reducing travel time were also analysed by Bhat (10).

As compared to the mode choice modelling where the choice set of different alternatives available to individuals is easy to outline and envisage; the route choice set is not easily definable. Since there are infinite number of route alternatives available to an individual, the choice set has to be formulated based on some principals to form a feasible and relevant set of routes for the origin and destination. For example, in the past, the choice set has been formulated with the k-shortest path algorithms as shown by Kuby et al. (12). Later, some behavioural constraints were introduced in these algorithms either according to the perception of route characteristics by the driver or the preference given to any attribute by the driver (13 and 14). The efficiency and relevance of these procedures require precise data on observed routes to compare the consistency between the actual behaviour of the driver and generated choice set.

In the route choice modelling the data of the route actually used by the individual for his/her origin and destination is an important component and it has been challenging and demanding for researchers. In the past, various researchers have tried to overcome this problem by employing different techniques of data collection. Ben Akiva et al (15) collected data for analysing the factors by stopping the cars and handing over the questionnaires to the owners of car or through mailing to them. Ramming (18) gathered information by asking travellers to describe their chosen path through sequence of route segments. Prato (19) requested respondents to recognize their home-to-work chosen path on a web based map and then fill this information on web based form. Vrtic et al (21) requested the respondents to report their origin and destination and three intermediate locations within their route in their long-distance trip. The interviewing technique has major disputes as the researcher experiences difficulties in filling the missing information and eventually matching the collected data with
the network representation. Many other studies have also made use of GPS and cellular technologies to passively collect the data (22 and 23). This method has several minor issues for obtaining the precise and accurate information of the chosen routes by respondents as information relies on signals and receivers which can be hindered by the atmospheric conditions. Bierlaire and Frejinger (25) proposed the concept of Domain of Data Relevance (DDR) to overcome the problems associated with the missing information of the chosen routes gathered by the surveys which bridges the gap between the model and the data collected for any route but this also depends on the advancement in the technology of the GPS, other cellular technologies and their satellite coverage.

The models are then formulated depending on the specification of the choice set and in addition to it, the overlapping of routes is also considered. These two challenges have opened new paradigms in the discrete choice framework. Cascetta et al (27) introduced the correction factor that measures the overlapping of routes within the deterministic part of the utility of the routes. Ben Akiva and Bierlaire (2) also introduced the Path size logit model that provides the different interpretation of the correction factor. Prashker et. al (32) have developed the paired Combinatorial (PCL) after taking into account the mathematical problem in SUE problem. Prashker and Bekhor (33) also considered the cross nested logit for route choice modelling.

The formation of the route choice based on revealed preference study and then application of the discrete choice framework has been not been given much attention in the past. However, in the past many researchers have used Google maps API (Application programing interface) to facilitate their data collection process related to travelling. For example, Frignani et al (34) made use of Google maps API to collect data for analysing the activity-travel pattern for individuals. They showed with their software tool that with the use of google maps, users could enter the information of their activity travel pattern through visualisation and interaction which reduced the burden on respondents. Guc and May et.al (35) also made use of Google maps API in their study of mobile behaviour of the individuals. They prepared the trajectory annotation model and embedded google maps API in this which provided the visualisation of location of respondent (which was captured by GPS data) and he/she can annotate the data which proved helpful for semantic interpretation of movement information. Jeffrey et. al (36) carried out a bicycle route choice analysis, in which they successfully collected data through the GPS signals. The CycleTracks application was developed for the smart phones in which their trip coordinates were picked up for their trip purpose until the user indicated that the trip was completed. For route choice modelling, so far, the survey was only designed to capture the respondent’s chosen route for the origin and destination. In the current study, the respondents are not only asked about the chosen routes but also the routes they considered in order to define the route choice set for each origin and destination by showing the routes to the respondents. The data of this kind has been collected through a web survey. Google maps as the existing network has been utilised to generate the choice set for a certain origin and destination of an individual for their last/ work study trip. By making use of the Google map API, the respondents will be able to visualise few calculated routes (both car and public transportation) for the OD of their last trip and will be asked which route they
considered or used for their trip purpose. This study is expected to reveal respondent’s route choice as well as the observed set of considered alternatives of route choice which can provide the useful insights towards the actual behaviour of driver.

**Methodology**

**Survey Design**

Google maps API (Application Programming Interface) are used in the design of the survey of this paper. JavaScript and PHP programming languages are used for developing the webpages and databases. The survey is an adaptive data collection process. Respondents are initially asked about the origin and destination (OD) and their mode choice for their last work/study trip purpose. The provision was given in the program so that google maps calculate the routes according to the OD provided which were then shown to the respondents. It was programed in a way that Google calculates and returns the car route and public transit routes according to the mechanism Google is working with which is not revealed to public. In our survey, car routes are retrieved from Google for without highways/tolls and with highways/tolls (if there are any for the specific OD). Then public transit routes are calculated with or without transfers whichever route Google considers the best for the given OD. **Google generates six maximum routes for car (three are with highways/tolls and three are without highways and tolls) and maximum four public transit routes.** A checking procedure is implemented in the program to avoid repetition of the similar routes. Respondents are also provided with an opportunity to list keywords of routes they might have considered/selected but not provided by the program. In addition to collecting the information regarding route and mode people selected, they were also asked whether they considered the routes for their trip or not. This leads to the revealed preference route choice set for each of the individual. The respondents were also asked about their socio-demographics. The survey was circulated within the university and a total of 200 samples were collected. Fig 1 shows a snapshot of the survey in which the initial page along with two possible routes are presented.

![Figure 1: A snapshot of survey webpage with the OD question page and possible routes presented to respondents](image)

Figure 1: A snapshot of survey webpage with the OD question page and possible routes presented to respondents
Both for the mode choice and route choice analysis, the individual’s characteristics as well as the attributes of mode and route with the emphasis on travel time and travel cost are taken into account. For public transit routes number of transfers and access time to the bus stop or the train station are also taken into account. As far as the travel cost is concerned, petrol cost, toll cost of roads (if any) and the parking cost are taken into account for car routes. For public transit routes, the concession ticket for students are taken into account as most of respondents reported to use that ticket who chose public transportation route. For the staff members or employees the normal ticket rates for commuters are considered.

**Model formulation for the mode choice and route choice**

**Mode Choice**

For the mode choice analysis, traditional multinomial logit (MNL) model is used. Since maximum of ten routes are available for car and public transit modes (as mentioned above) for individual’s OD, a criterion of minimum travel time was used for the mode choices available for each individual.

The utility for any alternative $i$ by an individual $n$ is given by the equation:

$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

Where, $V_{ni}$ is the deterministic portion of the utility and is given as:

$$V_{ni} = \beta' x_{ni}$$

Where, $x_{ni}$ is the column vector associated with the attribute influencing the choice of mode. $\beta'$ is the related coefficient column vector of parameters to be estimated.

$\varepsilon_{ni}$ is the error component which is assumed to be iid extreme value distributed.

The closed form of this model is given as:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_i e^{V_{ni}}}$$

Where, $P_{ni}$ is the probability of selecting the alternative $i$ by an individual $n$.

**Route Choice**

**Multinomial Logit**

Although due to the challenges described earlier, the traditional MNL models are not appropriate on the route choice and we acknowledge that point of view, but considering the fact that this is the pilot survey only; this data is used for only preliminary analysis to have the comprehensive overview of the survey taken and that of data obtained. The use of the modified models for the route choice will also be tested in the future to estimate the parameters and to carry out the policy sensitivity analysis.

The MNL was developed separately for car and public transportation (R1 to R6 choices for car routes and R7 to R10 routes for public transportation). For the car routes the attributes
that were considered were travel time, travel cost and toll cost was also added as a dummy variable. For public transit in addition to above variables, number of transfers and the access time were also taken into account. As the alternate specific constants were not included in the model for the route choice, the socio-demographics were interacted with the attributes related to alternatives to see their effect. For example, the income scales the cost. In the model, only those routes were made available to the respondents which were considered by the individuals. So the choice set was bound to change for every individual.

**Nested logit**

Since both car routes and public transit routes are available, a nested logit model formulation was also used on the data. Nested logit belongs to the family of GEV (Generalized extreme value models) in which the error term or the unobserved portion of the utility are jointly distributed as a generalized extreme value. It relaxes the IIA substitution patterns property of MNL by allowing to specify the alternatives in different nests.

Here, 
\[ U_{ni} = V_{ni} + \varepsilon_{ni} \]

Where, \( i = 1 \) to 6 for \( n \) individuals.

\( i = \text{R1 to R6} \) (Alternatives in first nest related to car routes)

Its error terms or unobserved parts of the utility are \( \varepsilon_1 \) to \( \varepsilon_6 \) which are correlated.

\( j = \text{R7 to R10} \) (Alternatives in second nest related to public transit routes)

The error terms or unobserved parts of the utility are \( \varepsilon_7 \) to \( \varepsilon_{10} \) are correlated.

Here, \( \text{Cov} (\varepsilon_i, \varepsilon_j) = 0 \)

\( \lambda \) is the parameter to be estimated which tells us the degree of correlation among the alternatives in each nest.

So,

For \( i = 1, \ldots, 6 \) the probability of alternative \( i \) to be selected by the individual \( n \) is given as:

\[
P_{ni} = \frac{e^{V_{ni}/\lambda} \left[ \sum_{j=1}^{6} e^{V_{nj}/\lambda} \right]}{\left[ \sum_{j=1}^{6} e^{V_{nj}/\lambda} \right]^\lambda + \left[ \sum_{j=7}^{10} e^{V_{nj}/\lambda} \right]}
\]

For \( i = 7, \ldots, 10 \) the probability of alternative \( i \) to be selected by the individual \( n \) is given as:

\[
P_{ni} = \frac{e^{V_{ni}/\lambda} \left[ \sum_{j=7}^{10} e^{V_{nj}/\lambda} \right]}{\left[ \sum_{j=1}^{6} e^{V_{nj}/\lambda} \right]^\lambda + \left[ \sum_{j=7}^{10} e^{V_{nj}/\lambda} \right]}
\]

**Mixed Logit**
For the route choice, mixed logit is expected to provide a better model fit as it averts the assumptions of the traditional MNL model by allowing for correlation between the unobserved factors, unobstructed substitution patterns and the random taste variation.

The utility obtained by an individual $n$ for an alternative $i$ is given as:

$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni}$$

Where $\varepsilon_{ni}$ is iid extreme value distributed.

$\beta_n$ is the random parameter to be estimated and it depends upon the distribution over the population.

$$\beta_n \sim f(\beta|\emptyset)$$

Here, $\emptyset$ includes parameters of the distribution such as mean and variance of $\beta_n$ over the population.

The probability of the choosing any alternative $i$ by an individual $n$ is given by the following integral form. Since $\beta_n$ is random in this case, the probability is integrated over the density of $\beta_n$

$$P_{ni} = \int \frac{e^{\beta_n x_{ni}}}{\sum_j e^{\beta_n x_{nj}}} f(\beta|\emptyset) d\beta$$

A mixed logit model was also developed separately on car and public transit routes. In the model, only those routes were made available to the respondents which were considered by the individuals. So the choice set was bound to change for every individual. The cost coefficient and the time coefficient were specified as lognormal distributed over the individuals as they are expected to have the negative sign theoretically.

**Descriptive Analysis**

This paper presents the results of the pilot survey taken among university students and staff to analyse the route choice and mode choice behaviour of the individuals. A total data of 200 individuals were interviewed. Due to the missing values, and some other reasons, 38 responses were not taken into account. Around 5% of all people who received the email opened the survey among which 50% completed the survey.

The survey was programmed to provide a maximum of ten routes for every origin and destination, of which six represent the routes with the car mode and four represent the public transportation mode. A total of 1194 routes were shown to the respondents for a total of 162 origin and destinations. Almost 47% were the public transportation routes and 53% were car routes among them.

About 70% of the individuals who used public transportation considered car routes also and about 60% of the car route users considered public transportation.

As far as the mode share is concerned, about 75% of respondents used public transportation for their study/ work trip and about 20% respondents used car as shown in figure 2. About
73% of the respondents were male respondents. Most of the age group among the respondents was of 18 to 25 years (about 70%). The income group varied among the respondents with the highest share being of medium income. The income was categorised as 20K to 60K AUD Dollars as low income, 60K to 150K AUD dollars as medium income and >150K AUD dollars as high income. The gender wise route share depicts that about 70% of both the genders are using public transportation routes. About 80% of people in the 18-25 years age group which was the most reported age-group in the survey reported to choose the public transportation. The car mode or routes are more popular in age group of 45-60 years. The route share for income group is also shown in figure 2 below. It can be seen that about 75% of all the income groups are prone to use public transportation; hence income is not playing significant role in decision making of mode choice. About 89% among the respondents are students and the rest of them were staff members. About 80% of the students used public transportation and 60% of the staff members used public transportation.

The presentation of the descriptive analysis shows the distinct inclination towards the public transportation mode for university students and staff. This finding is consistent with the results of the UNSW travel survey which is carried out on a larger scale by the facilities management to plan the public transportation use better for the students. The UNSW survey reported that about 60% of the respondents were inclined towards public transport mode. The unavailability of free parking inside the university, concession tickets available for students and better organization of most public transportation routes for the university might be the reason of continuing trend towards the use of more sustainable public transport over the seven years.
Figure 2: Mode share among individuals, Gender-wise, age-wise and income-wise mode share for individuals (top to bottom)

Results and Analysis of Estimated Parameters

Multinomial logit for mode choice

The choice set for this model was taken as car, public transport and cycle. The attributes related to alternatives that were taken into account were travel time, travel cost, access time, number of transfers. The socio-demographic variables that were included in the model were household income, age, gender, number of cars household has access to. The socio demographic variables came out to be quite insignificant. The less variability due to limited data set might be the reason of insignificance of these variables. Travel time and travel cost coefficients had intuitive negative signs which is consistent with the hypothesis that the utility of the mode increases with the decrease in the travel cost, travel time, access time and the number of transfers. Travel cost coefficient was coming as significant, which is consistent with the hypothesis that the travel cost affects the decision making process quite significantly. The rho-square of the model was 0.618 through which can be concluded that all the attributes related to alternatives were adequately and sufficiently describing the behaviour of the mode choice. However, the significance of the individual variables will be tested in the future for larger data set. (see table 1).
Table 1: Estimated parameters for mode choice with MNL

<table>
<thead>
<tr>
<th>Name</th>
<th>Coefficient</th>
<th>Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC_Car</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ASC_Cycle</td>
<td>-5.31</td>
<td>-5.83</td>
</tr>
<tr>
<td>ASC_PT</td>
<td>0.874</td>
<td>1.62</td>
</tr>
<tr>
<td>Travel time (min)</td>
<td>-0.00153</td>
<td>-0.14</td>
</tr>
<tr>
<td>Travel cost (AUD)</td>
<td>-0.215</td>
<td>-3.03</td>
</tr>
<tr>
<td>Access time (min)</td>
<td>-0.04</td>
<td>-1.19</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-0.168</td>
<td>-0.77</td>
</tr>
<tr>
<td>Toll</td>
<td>-0.408</td>
<td>-0.66</td>
</tr>
<tr>
<td>Initial LL</td>
<td>-164.373</td>
<td></td>
</tr>
<tr>
<td>Final LL</td>
<td>-62.751</td>
<td></td>
</tr>
<tr>
<td>Rho-square</td>
<td>0.618</td>
<td></td>
</tr>
</tbody>
</table>

Multinomial Logit for route choice

The multinomial logit model nested logit and mixed logit models are developed for route choice using Biogeme.

The MNL model are developed on the car routes and public transportation routes separately to analyse the weight of respective attributes given by the individuals towards choosing public transportation and car routes. For car routes, the cost coefficient and time coefficient signs came out to be negative, that is the utility of the route decreased as the cost and the travel time increased which matches the intuition about these variables. Considering the significance of the variables, after carrying out the correction of heteroscedasticity in the data, i.e. according to the robust t-test, the cost coefficient came out to be significant as compared to the travel time. Before this correction, the t-test values were quite different, so it can be concluded that there is a presence of heteroscedasticity among the observations in data for cost. From the final significance of the variables, it can be extracted that travel cost influences most significantly in the route choice decision when using the car mode rather than travel time. The Toll variable was also included as a dummy variable in the model, but it was not quite significant. The rho-square of the model was 0.121, which implies that more alternate related attributes should be included in the model, but the current variables are also valid to study their impact on the car route choice.

For public transit routes, the explanatory variables that are included are travel time, travel cost, number of transfers and access time. All these variables had the negative signs which are deemed to be consistent with the theoretical aspects. The coefficients which are significant are travel time and number of transfers which supports the fact that the travel time and number of transfers play a significant role in the decision-making of the route using public transportation as a mode. The cost coefficient was almost negligible, as most of them were the students and they have a concession ticket which makes the ticket cost almost the
same for everyone travelling to the university. The rho-square value with these parameters was 0.149 which articulates that more attributes related to alternatives should be included in the model. The household income was interacted with travel cost and was added as the alternate specific variable which did not improve the rho square and it came out to be insignificant. (see table 2)

Table 2: Estimated parameters of route choice with MNL

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>CAR</th>
<th>TRANSIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Robust t-test</td>
</tr>
<tr>
<td>Toll</td>
<td>2.19</td>
<td>1.2</td>
</tr>
<tr>
<td>Travel time (min)</td>
<td>-0.143</td>
<td>-1.47</td>
</tr>
<tr>
<td>Travel cost (AUD)</td>
<td>-0.553</td>
<td>-2.16</td>
</tr>
<tr>
<td>Access time (min)</td>
<td>-0.0588</td>
<td>-1.01</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-0.733</td>
<td>-2.38</td>
</tr>
<tr>
<td>Initial LL</td>
<td>-33.156</td>
<td></td>
</tr>
<tr>
<td>Final LL</td>
<td>-29.146</td>
<td></td>
</tr>
<tr>
<td>Rho-square</td>
<td>0.121</td>
<td></td>
</tr>
</tbody>
</table>

A nested logit model was also developed on the data. In this nest structure the routes from 1 to 6 were kept in the car mode nest and routes 7 to 10 were kept in the public transit nest. The socio demographic variables were also included on the upper level of the nest that is car and transit choice set and car mode was taken as reference. Household income, age, gender and vehicle ownership were socio-demographic variables. These variables did not come out to be significant. As shown above in the descriptive analysis, there is not much variability in the respondent’s socio demographics in the limited data set as this was only the pilot survey carried out to test the survey tool. This might be the reason of the socio-demographics not coming significant in the model. So the socio-demographics were excluded from the model. The attributes of the routes such as travel time, travel cost, number of transfers, access time to the public transport and toll variable as dummy were included in the model. The signs of these coefficients were negative and consistent with the hypothesis and theoretical aspects, however these variables were not significant. The log-sum coefficient for the car nest was 0.16 (1/ 5.98) and degree of correlation was 83% which describes quite a reasonable correlation among the unobserved factors of the alternatives in the nest. The log-sum coefficient for the public transportation nest was 0.454 (1/2.2) and degree of correlation was 54% which also describes the reasonable correlation among the unobserved factors of the alternatives of this nest. The rho-square of the model was 0.243 which shows that the current alternate specific attributes are adequately describing the route choice behaviour (see table 3).
Table 3: Estimated parameters with nested logit

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Coefficient</th>
<th>Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC_CAR</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ASC_PT</td>
<td>2.56</td>
<td>3.65</td>
</tr>
<tr>
<td>Travel time (min)</td>
<td>-0.0341</td>
<td>-1.52</td>
</tr>
<tr>
<td>Travel cost (AUD)</td>
<td>-0.163</td>
<td>-1.13</td>
</tr>
<tr>
<td>Access time (min)</td>
<td>-0.0755</td>
<td>-1.85</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-0.241</td>
<td>-1.06</td>
</tr>
<tr>
<td>λ_car</td>
<td>5.98</td>
<td>1.44</td>
</tr>
<tr>
<td>λ_PT</td>
<td>2.2</td>
<td>1.02</td>
</tr>
<tr>
<td>Initial LL</td>
<td>-102.427</td>
<td></td>
</tr>
<tr>
<td>Final LL</td>
<td>-77.512</td>
<td></td>
</tr>
<tr>
<td>Rho-square</td>
<td>0.243</td>
<td></td>
</tr>
</tbody>
</table>

Mixed Logit for route choice

The parameters were estimated on the car routes where the choice set includes R1 to R6 for every individual. The routes reported to be considered were made available in the model. It was observed in the MNL model previously that there was significant difference between the estimated standard errors and robust standard errors of the of travel time and travel cost coefficients as well cost coefficient to have a significant coefficient. In mixed logit model, travel time and travel cost variables are considered having a lognormal distribution as these are expected to have a negative sign for the individuals. The Toll variable which was previously positive in MNL model was given an independent normal distribution to observe its preference heterogeneity. The mean and standard deviation of toll variable were not significant hence this variable was taken out of model. Further, the travel cost standard deviation was also not significant, so it was not taken as random variable in the final model. The travel time coefficient mean and standard deviation were quite significant suggesting that travel time significantly affects the car route choice behaviour with significant heterogeneity among the individuals. For public transportation routes, the travel cost variable with a lognormal distribution, the singular value of hessian came out to be very small suggesting that there is lack of variability in the data regarding travel cost. The travel cost mean and standard deviation were quite insignificant. This can be due to the fact that students have the concession ticket and the cost is more or less same for each route. The travel time coefficient suggested that it significantly affects the route choice and its variance was also significant. The variances of number of transfers and access time (taken as log normal distributed) were also quite insignificant so they were taken as fixed parameters in the final model. The coefficient of number of transfers was significant and had intuitive negative sign. (see table 4).
Table 4: Estimated parameters for route choice with mixed logit.

Conclusions and Future Work

In this analysis, a unique internet-based survey tool was developed to capture the mode choice and route choice behaviour of the individuals. The choice set formation of people regarding their mode and route choice was targeted in this study using the Google maps API (Application programming Interface). The routes were calculated according to the reported origin and destination of the respondents of their last work/study trips. These routes were then shown to the respondents to inquire them which route they considered and finally selected for their last work/study trip. This survey gave the opportunity to analyse the observed and revealed preference choice set for the route choice modelling. Nonetheless, a mixed logit model was also estimated for the route choice selection. Car route alternatives with and without highway and public transport route alternatives are presented to respondent and they are asked to report whether each of the presented alternatives were considered for the trip or not. A maximum of six car routes and four public transit route alternatives were shown to the respondents. A pilot survey was carried out in two schools in the University of New South Wales, Australia. 200 individuals completed the survey whose data was used in the modelling exercise of this paper.

A preliminary analysis was done on the data and the parameters were estimated with three different structures of discrete choice models (MNL, NL and mixed MNL) and results were compared. Two MNL models for route selection were developed on the car routes and public transport routes to study their respective attributes. In both the models the alternate specific attributes had intuitive signs. For car routes, the travel cost coefficient came out to be significant in the model as compared to the other attributes but suffered with the heteroscedasticity errors as there was a marked difference between the standard error of maximum likelihood estimator and robust standard errors. On the contrary, the travel time coefficient was significant when estimated with the mixed logit model where both the cost coefficient and time coefficient were taken as log normally distributed for car routes. As far as public transit routes are concerned, the travel time and number of transfers came out to be significant as estimated by MNL model. Travel time coefficient was considered having a lognormal distribution in the mixed logit model where the signs of estimated coefficients of
travel time and travel cost were consistent with the intuitive expectations. The attributes namely travel time, number of transfers and access time had negative signs as far as public transportation routes were concerned where travel time and number of transfers were statistically significant at the 95% confidence level. In nested logit model, the socio demographics were also included on the upper level of nests but these came out to be not insignificant. In this model the attributes related to alternatives had intuitive signs but were not significant. This will also be tested in the future with the larger dataset which is not limited to university students and staff.

The mixed logit model gave the best model fit among the other model specifications for both car and public transportation route selection behaviour.

In mode choice model, the socio-demographic coefficients were insignificant. The coefficients of travel time and travel cost were negative which is pretty intuitive. The non-significance can be due to the limited set of data.

This survey tool will be used for the future studies of route choice modelling on larger data set. The choice set formation will be modelled and the correlation between the alternatives will also be taken into account for the route choice modelling.

References


Shakeel, Rashidi, Waller

Theory, Lyon, France, 24-26 July 1996.


