Study of Clustering Modes based on Choice of Transport across Space: a case study of Tokyo Metropolitan Area

25th – 27th September, 2015 | 14th Behavior Modeling Workshop in Transportation Networks | The University of Tokyo
Presented by IIT Bombay
Objectives of the study

1. To evaluate the spatial variation in modal choice within different cohorts for Tokyo Metropolitan area (age, gender, time of the day).

2. To identify areas which are public transport and non-motorized traffic friendly for the selected cohorts.

3. To analyze the change in travel behavior due to changes in policy attributes (fare, travel time etc., no. of transfers).
The Data

Choice data- Mode choice \{Walk, Bike, Train, Bus, & Car\}
Trip data- Travel time, Purpose, Location etc.

Individual data- Age, Gender etc.
Trip data- Travel time, location etc.
Alternative characteristics- Fare, time, Access, Egress etc.
Data Descriptives

**Modal Share by Age Group**

- Age Group vs Absolute Number
- Bicycle, Bus, Car, Rail
- Uses the Mode, Doesn't Use

**Modal share vs No. of transfers**

- Number of transfers vs Uses the Mode, Doesn't Use
- Bicycle, Bus, Car, Rail, Walk

Pie charts for Bus, Train, and Car modes, showing the proportion of users who use and don't use each mode.
Spatial Variation in Variables

Concentration of car use

Shorter trip times
Spatial Variation in Variables

Concentration of smaller trip distances

Return to home, commute trips
Multi-nominal Logit Model

Walk has been taken as the base case

• X1 or ASC1=Train
• X2 or ASC2=Bus
• X3 or ASC3=Car
• X4 or ASC4=Bike

Variables considered for the analysis

X5= Travel Time
X6= Age

• Log likelihood value=-1273.97
• t-statistics=6.491065, -4.896915, -5.309969, -4.221034, -21.073350, 4.304434
Utility equations

- train <- Data$ModeAvailableTrain*exp(d1*Data$TotalTimeTrain/100 +b1*matrix(1,nrow =hh,ncol=1))
- bus <- Data$ModeAvailableBus *exp(d1*Data$TotalTimeBus/100 +b2*matrix(1,nrow =hh,ncol=1))
- car <- Data$ModeAvailableCar *exp(d1*Data$TimeCar/100 +b3*matrix(1,nrow =hh,ncol=1))
- bike <- Data$ModeAvailableBike *exp(d1*Data$TimeBike/100 +b4*matrix(1,nrow =hh,ncol=1))
- walk <- Data$ModeAvailableWalk *exp(d1*Data$TimeWalk/100 +d2*Data$Age/10)
Predicted Mode Share

- Car: 30%
- Rail: 39%
- Walk: 24%
- Bike: 6%
- BUS: 1%
# Prediction Success Table

<table>
<thead>
<tr>
<th>Predicted</th>
<th>bike</th>
<th>bus</th>
<th>car</th>
<th>Rail</th>
<th>walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>67</td>
<td>0</td>
<td>21</td>
<td>55</td>
<td>68</td>
</tr>
<tr>
<td>Bus</td>
<td>1</td>
<td>0</td>
<td>34</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Car</td>
<td>48</td>
<td>0</td>
<td>326</td>
<td>72</td>
<td>66</td>
</tr>
<tr>
<td>Rail</td>
<td>3</td>
<td>15</td>
<td>62</td>
<td>442</td>
<td>6</td>
</tr>
<tr>
<td>Walk</td>
<td>16</td>
<td>0</td>
<td>4</td>
<td>18</td>
<td>192</td>
</tr>
</tbody>
</table>
Elasticities

• Direct elasticities with respect to time
  • BUS - 1.43
  • Train - 1.73
  • car - -1.29
  • bike - -0.94
  • walk - -1.01

• Cross elasticities with respect to time
  • bus - 0.33
  • train = 0.077
  • car - 0.15
  • bike = 0.47
  • walk = 0.51
Mixed Logit Model 1. (Age group 25-40)

Walk has been taken as the base case

- X1 or ASC1 = Bus
- X2 or ASC2 = Car
- X3 or ASC3 = Bike

Variables considered for the analysis
X4 = mean Travel Time
X5 = variance in TT

- Log likelihood value = -1300.548
- Parameters = -1.93, -1.85, -1.53, -10.62, 0.085
- t-statistics = -10.8363935, -18.7976434, -15.7871061, -21.1553319, 0.2435727

(If the values are between -1.96 to 1.96, it is considered to be significant)
Mixed Logit Mode Error

```r
+ 
+ # print the calculation process 
+ print(x) 
+ print(SimLL/R) 
+ 
+ # divide by the number of the repetition 
+ SimLL <- SimLL / R 
+ 
+ > 
+ > ################ maximize Log-likelihood function, fr ################## 
+ > 
+ > ## Parameter optimization 
+ > res <- optim(b0, fr, method = "BFGS", hessian = TRUE, control=list(fnscale=-1)) 
+ > [1] 0 0 0 0 0 0 
+ > [1] NA 
+ Error in optim(b0, fr, method = "BFGS", hessian = TRUE, control = list(fnscale = -1)) : 
+ initial value in 'vmmin' is not finite 
+ In addition: There were 50 or more warnings (use warnings() to see the first 50) 
+ > 
+ > ## Parameter estimation Hessian matrix calculation 
+ > b <- res$par 
+ > hhh <- res$hessian 
+ > 
+ > ## Calculate the t-statistic 
+ > tval <- b/sqrt(-diag(solve(hhh))) 
+ > 
+ > ## L(0), Log-Likelihood when all parameters are 0 
```
Mixed logit Model (without Age segmentation)

• X1 or ASC 1=Train
• X2 or ASC2=Bus
• X3 or ASC3=Car
• X4 or ASC4=Bike

Variables considered for the analysis
X5= mean Travel Time
X6= variance in TT
X7= mean age
X8= variance in age

• Log likelihood value= -1276
• Parameters= 1.16, -1.11, -0.97, -0.75, 11.09, 0.13, 0.19, -0.01
• t-statistics= 6.48, -4.83, -5.33, -4.24, -21.06, 0.37, 4.25, -0.29
Mixed Logit Model

```
> ## L(0)
> print(L0)
> [1] -2135.675
> ## LL
> print(LL)
> [1] -1276.333
> ## rho-square
> print((L0-LL)/L0)
> [1] 0.402375
> ## adjusted rho-square
> print((L0-(LL-length(b)))/L0)
> [1] 0.3986291
> ## estimated parameter values
> print(b)
> [5] -11.089914999  0.132997882  0.191448906  -0.007237038
> ## t-statistic
> print(tval)
```
The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}} \tag{1}$$

where $x_j$ is the attribute value for feature $j$, $w_{i,j}$ is the spatial weight between feature $i$ and $j$, $n$ is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{3}$$

The $G_i^*$ statistic is a $z$-score so no further calculations are required.
Spatial Distribution of probability of modal choice: Train
Spatial Distribution of probability of modal choice: Car
Spatial Distribution of probability of modal choice: Bike
Spatial Distribution of probability of modal choice: Walk
<table>
<thead>
<tr>
<th>Mode</th>
<th>Bike</th>
<th>Bus</th>
<th>Car</th>
<th>Train</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>67</td>
<td>0</td>
<td>21</td>
<td>55</td>
<td>68</td>
</tr>
<tr>
<td>Bus</td>
<td>1</td>
<td>0</td>
<td>34</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Car</td>
<td>49</td>
<td>0</td>
<td>326</td>
<td>72</td>
<td>65</td>
</tr>
<tr>
<td>Rail</td>
<td>3</td>
<td>15</td>
<td>62</td>
<td>442</td>
<td>6</td>
</tr>
<tr>
<td>Walk</td>
<td>16</td>
<td>0</td>
<td>4</td>
<td>18</td>
<td>192</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mode</th>
<th>Direct Elasticities</th>
<th>Cross Elasticities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>-1.42</td>
<td>0.33</td>
</tr>
<tr>
<td>Train</td>
<td>-1.73</td>
<td>0.029</td>
</tr>
<tr>
<td>Car</td>
<td>1.29</td>
<td>3</td>
</tr>
<tr>
<td>Bike</td>
<td>0.32</td>
<td>0.94</td>
</tr>
<tr>
<td>Walk</td>
<td>1.01</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Conclusion

• Mode choice is significantly affected by the age factor

• Spatial variation in P values across the Tokyo metropolitan region gives an insight into travel behavior across space
Thank You

Rohan Joshi
Sayantani Sarkar
Naveen Krishna
Varun Varghese

Under the guidance of Prof. Arnab Jana
Indian Institute of Technology Bombay