Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Multi-Modal Transportation Choice Modelling of Metropolitan Region of Amsterdam

Varun K Janardhan[a,b] *, Elenna R Dugundji[a,b]

[a]*Vrije Universiteit, Department of Mathematics, De Boelelaan 1018, Amsterdam 1018 HV, Netherlands*
[b]*Centrum Wiskunde & Informatica, Science Park 123, Amsterdam 1098 XG, Netherlands*

Abstract

Transportation is a significant sector that requires much research and innovation in the current state and pace of urbanization, as it connects people's day-to-day lives from reasons varying from work to leisure. It is vital to understand behavioural models for planning any transportation service to meet the demands of the citizens in a cost- effective manner. A Multinomial Logit model has been used to predict the probability that an individual chooses an option from a set of alternatives available using open source Biogeme software. The analysis provides interesting findings which are quite logical and complement the transport policies adopted in The Netherlands. It suggests that it is important to investigate the significance of the variable considered for the model estimation. This study shows that distance, ethnicity and population density are parameters of high significance in the model estimate.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: Transportation; Multinomial Logit Model, Biogeme, Model Estimates

1. Introduction

Transportation is a significant sector that requires much research and innovation in the current state and pace of urbanization, as it connects people's day-to-day lives from reasons varying from work to leisure. Building a strong and efficient transportation system thus becomes essential for any municipal/ government body. Hence, it becomes vital to understand behavioural models for planning any transportation service to meet the demands of the citizens in a cost- effective manner. However, information regarding commuters daily travel i.e origin – destination alone does not suffice the purpose. Other components such as social and land use characteristics of the individual shall also be considered while planning such infrastructure.

* Corresponding author. Tel.: +91-903-379-9858;
E-mail address: varun.kj58@gmail.com

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

This study aims to predict the travel mode choice of individuals in the Metropolitan Region of Amsterdam (MRA) among the various modes of transport available based on their socio-economic, demographic and land-use characteristics.

1.1. Need for the Study

The major purpose of the study is with regard to the upcoming North-South Metro-line in this Region and plan the infrastructure based on the model results. Modelling plays a crucial role in any large scale decision making process. The change in choices of the individuals upon change in demographics or features of the alternate can be predicted using such models. Such models help predict the probability that an individual chooses an option from the set of alternatives available.

1.2. Aim

To develop Multinomial Logit Model for modeling the transportation mode choice of commuters based on socio-economic and built environment factors.

1.3. Objectives

- To study the spatial distribution of demographic and socio-economic factors
- Understanding the use of discrete choice model in transportation
- Identifying the parameters required to develop the model
- Estimation of travel mode choice
- To help formulate policies and planning of transportation infrastructure

1.4. Discrete Choice Modelling

Discrete Choice Models (DCM) help statistically relate the choices made by individuals to their characteristics and the characteristics of the alternatives available to the individual. It is a type of analysis used to analyze and predict travel decisions of individuals.

1.5. Transportation Mode Choices

An estimation is made about which transportation mode does an individual choose from the 4 alternatives available:

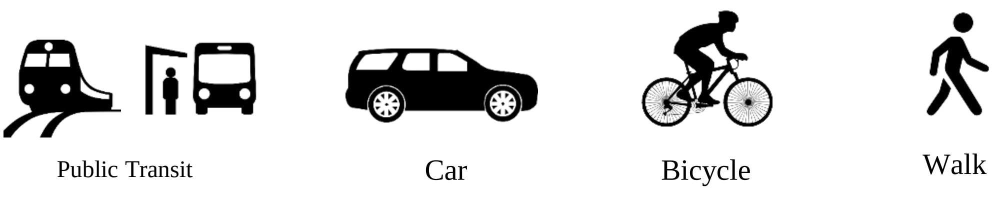


Figure 1: Modes of Transportation

2. Literature Review

The study on Transportation modelling has emerged as it becomes essential for purposes of policy making, planning, capacity calculation etc. The most significant factor for planning a public transport facility is the knowledge of individual behaviour. Modelling helps understand the performance of transport facilities, future developments and influences of the public choices in the network. Many researches has purposed various methods that can be used to model such behaviour and influences which affects the decision making.

Martin Loidl, Gudrun Wallentin, Rita Cyganski, Anita Graser, Johannes Scholz and Eva Haslauer, primarily focused their study on the integration of GIS in transportation modelling. They state that GIS and Transport research are always inter-related wherein through models, simulations and analysis with an explicit consideration of spatial nature, new information can always be generated. The role of geospatial technologies in transport modelling was restricted to cartographic presentation of final results which has started to change in recent years, through the advent of interactive geo-visualization environments, the rise of Visual Analytics and the availability of massive data sets (Loidl, et al., 2016).

Secondly, the growing relevance of geo-visualization in transport modelling can be attributed to the paradigm shift from highly aggregated to disaggregated modelling approaches, which puts a stronger focus on the heterogeneity and inter-dependences of mobility behaviour and result in detailed information on individual activity patterns.

Elenna R. Dugundji, Joan L. Walker, have based their study on Discrete Choice Modelling by categorizing commuters district-wise based on Bicycle, Transit, Cars. Some of the variables that were considered for modelling were Public Transit availability, Car ownership, Age, gender, income, education, Travel time, travel cost (car, bicycle). Additional conditions where, if the travel time by bicycle is > 75min, its availability is considered 0 (Dugundji & Walker, 2005).

G.-M. Miletić, S. Gašparović, T. Carić, conducted an analysis of socio-spatial differentiation on transport mode choice preferences considering transport mode choice is one of those aspects of travel behaviour that, to a great extent, affects the efficiency of the transport system. This paper analyses the factors that contribute to the use of public and car transport to obtain insight into the preferences for using these two modes of transport in Croatia and also to find the extent of demographic and socioeconomic characteristics on it. The use of binary logistic regression analysis has determined that the preferences towards the frequent use of car or public transport are significantly influenced by the age of the respondents, size of the settlement, accessibility of the destinations by public transport, the number of vehicles in the household and whether the respondent is the main car user in the household (Miletić, Gašparović, & Carić, 2017).

M. Alqhatani, S. Bajwa, S. Setunge, in their research modelled the influence of socio-economic and land-use factors on mode choice in cities of Riyadh and Melbourne using multinomial (MNL) and nested logit (NL) model. It shows that the urban structure which is a framework of housing, development and employment combined with the socio-economic parameters, influence and shape the travel patterns. Some of the socio-economic factors considered was age, income, household size, car ownership. MNL model allows for service improvements to already functioning modes by decreasing the likelihood of other existing modes in terms of changes. In this model, there is an equivalence assumption in all alternatives. To reduce the effect of equal competition between all alternatives, a more comprehensive and fluid model can be created using NL model, which accommodates a variety of differences in the extent of similarities across different structures of nests and access model alternatives. The model results conclude that urban sprawl is prominent factor that affects the mode choice, income and house type (Alqhatani, Bajwa, & Setunge, 2013).

3. Study Area

The study area for this project is Metropolitan Region of Amsterdam (MRA) which consists of 36 municipalities which are a part of Noord-Holland and Flevoland provinces in The Netherlands. The regions has more than 2.4 million citizens and accounts for 19% of the GDP of the Netherlands (Board, n.d.). The MRA is a growing region where development mostly happens in 'new towns' in the outskirts of the city, because of the limitation in built-up space in the main city.

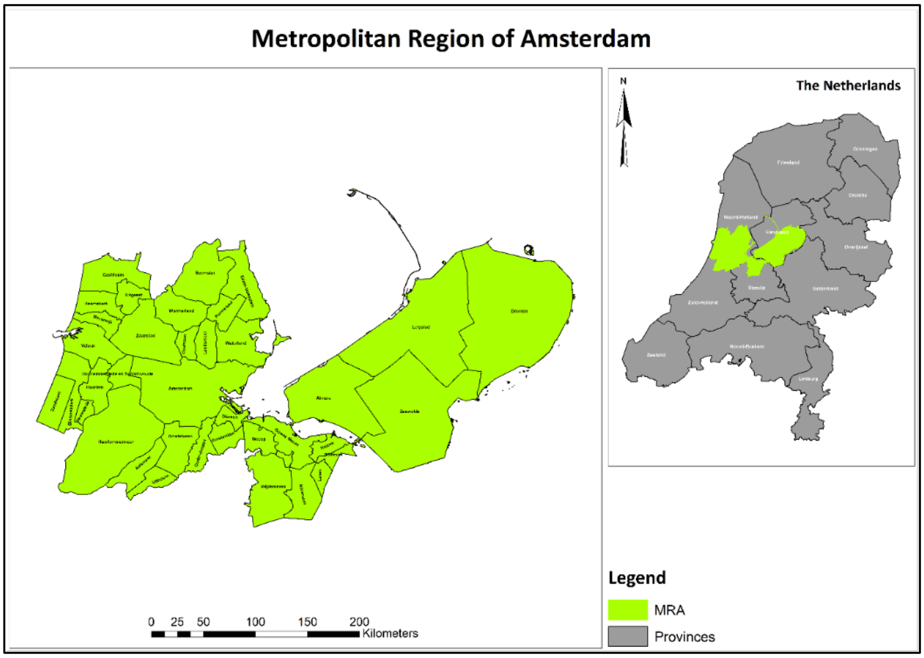


Figure 2: Metropolitan Region of Amsterdam

4. Methodology

The flowchart Figure 3: Methodology for the Study gives an overview about the various steps that have been followed for estimating the mode choice of individuals. The various steps involve collection of socio-economic, demographic and administrative boundaries for analysis about the spatial distribution of the explanatory variables and determining the factors that could be influential for mode choice. The data modelling phase involves defining variables, baselines, alternatives, utility functions and defining the type of model for estimating the mode choice. A detailed description of the same has been provided below:

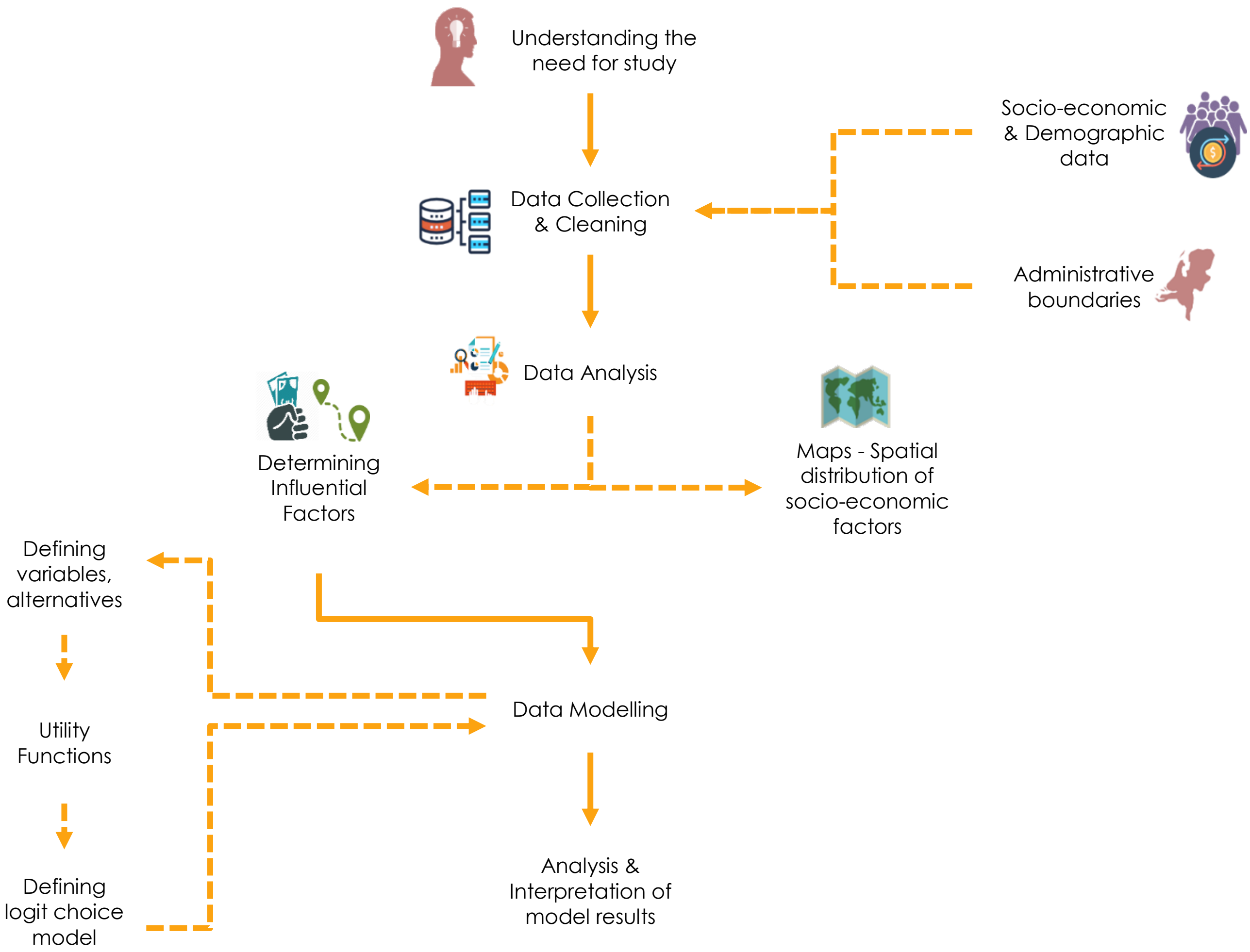
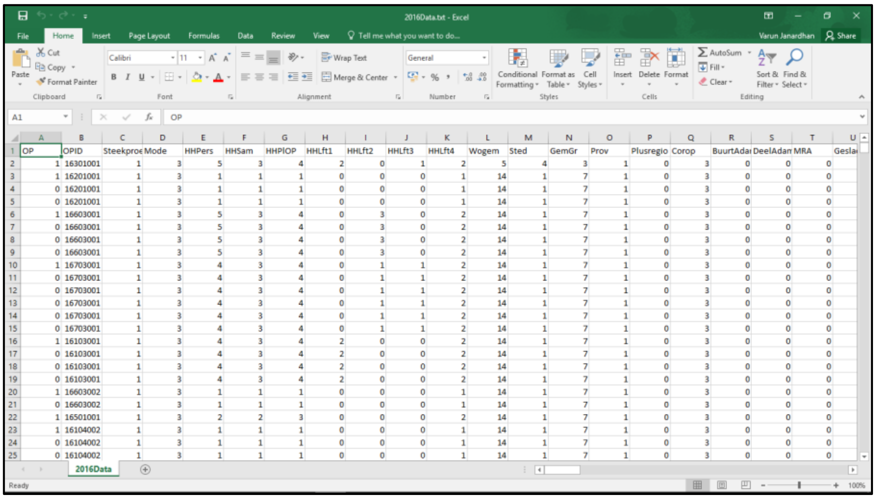


Figure 3: Methodology for the Study

4.1. Data Used

The data collection process involves the procurement of the datasets related to socio-economic and demographics of the study area and the respective GIS shapefiles. The socio-economic and travel information datasets are in the form of excel sheets which are the OViN (ONDERZOEK VERPLAATSINGEN IN NEDERLAND – Research movements in The Netherlands) Dataset as shown in Figure 4: Datasheet of 2015, 2016 which is a continuous daily investigation conducted by the Central Bureau of Statistics (CBS) into the travel behavior of the Dutch. The dataset consists information about the respondent's:

- Travel Mode
- Travel Time
- Purpose of Travel
- Household Composition
- Income etc.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	OP	OPID	Steekproe	Mode	HHPers	HHSam	HHPlOP	HHLft1	HHLft2	HHLft3	HHLft4	Wogem	Sted	GemGr	Prov	Plusregio	Corop	BuurtAdai	DeelAdam	MRA	Geslai
2	1	16301001	1	3	5	3	4	2	0	1	2	5	4	3	1	0	3	0	0	0	
3	1	16201001	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
4	0	16201001	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
5	0	16201001	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
6	1	16603001	1	3	5	3	4	0	3	0	2	14	1	7	1	0	3	0	0	0	
7	0	16603001	1	3	5	3	4	0	3	0	2	14	1	7	1	0	3	0	0	0	
8	0	16603001	1	3	5	3	4	0	3	0	2	14	1	7	1	0	3	0	0	0	
9	0	16603001	1	3	5	3	4	0	3	0	2	14	1	7	1	0	3	0	0	0	
10	1	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
11	0	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
12	0	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
13	0	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
14	0	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
15	0	16703001	1	3	4	3	4	0	1	1	2	14	1	7	1	0	3	0	0	0	
16	1	16103001	1	3	4	3	4	2	0	0	2	14	1	7	1	0	3	0	0	0	
17	0	16103001	1	3	4	3	4	2	0	0	2	14	1	7	1	0	3	0	0	0	
18	0	16103001	1	3	4	3	4	2	0	0	2	14	1	7	1	0	3	0	0	0	
19	0	16103001	1	3	4	3	4	2	0	0	2	14	1	7	1	0	3	0	0	0	
20	1	16603002	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
21	0	16603002	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
22	1	16501001	1	3	2	2	3	0	0	0	2	14	1	7	1	0	3	0	0	0	
23	1	16104002	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
24	0	16104002	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	
25	0	16104002	1	3	1	1	1	0	0	0	1	14	1	7	1	0	3	0	0	0	

Codebook OViN 2016

Variable name OViN 2016	Variable label OViN 2016	Level: P = Person; V = displacement; R = drive; W= weighting factor;	Code OViN 2016	Code label
HHPers	Number of people in household	P		
			1..9	
			10	10 or more
HHSam	household Composition	P		
			1	single person household
			2	Couple
			3	Couple + child (ren)
			4	Couple + child (ren) + other (s)
			5	Pair + other (s)
			6	One parent family + child (ren)
			7	One parent family + child (ren) + other (s)
			8	Another composition
HHPlOP	Place OP household compared to domestic core	P		
			1	Single (single household)
			2	Single Household Kern (multi-person household)
			3	Spouse (o) t (e) / partner (core member household)
			4	Child
			5	Father / mother or father / mother
			6	Brother / sister or brother / sister
			7	In-law / daughter-in
			8	Little child
			9	Miscellaneous: family or relatives
			10	Other: no family or relatives

Figure 4: Datasheet

4.2. Software Used:

- Microsoft Excel
- Notepad++
- ArcGIS
- QGIS
- BIOGEME

4.2.1. About Biogeme

Biogeme is an open source freeware designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models.

Bison Biogeme is designed to estimate the parameters of a list of predetermined discrete choice models such as logit, binary, nested logit, etc. It is based on a formal and simple language for model specification.

Input:

Model File: .mod

Data File: .dat, .txt, .csv

4.3. About Modelling

Discrete Choice Modelling technique has certain general assumptions

- Decision-maker
- Alternatives
- Attributes
- Decision rule

Discrete choice models are also referred to as disaggregate models, meaning that the decision-maker is assumed to be an individual. The "individual" decision-making entity depends on the particular application (Ben-Akiva & Bierlaire). A discrete choice set contains a finite number of alternatives that can be explicitly listed. The choice of a travel mode is a typical example of a choice from a discrete choice set (Ben-Akiva & Bierlaire). Every option available for an individual from the set of alternatives has some characteristics attributed with it, which is generally considered as a function of the data available. Some attributes are generic to all alternatives and some are alternative-specific. The decision rule is the process used by the decision-maker to evaluate the alternatives in the choice set and determine a choice. Most models used for travel behaviour applications are based on utility theory, which assumes that the decision-maker's preference for an alternative is captured by a value, called utility, and the decision-maker selects the alternative in the choice set with the highest utility (Ben-Akiva & Bierlaire).

4.3.1. Types of Modelling

- Binary Logit Model
- Multinomial Logit Model
- Nested Logit Model
- Cross-nested Logit Models

This study uses multinomial logit modelling technique for analyse the travel mode choice of individual among the set of alternatives available.

Multinomial Logit Model:

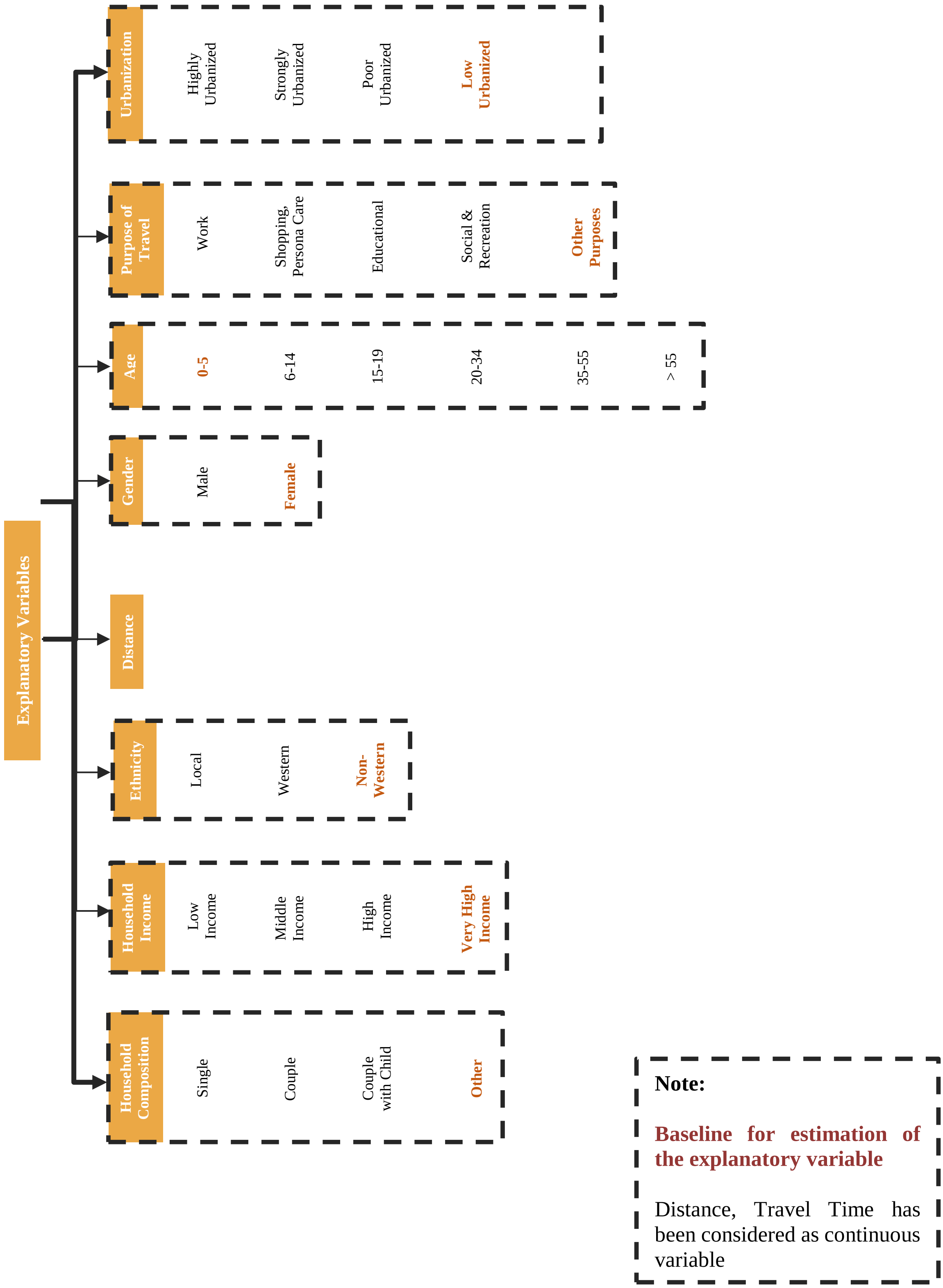
Multinomial Logit model or Multinomial logit regression is a classification method that generalizes logistic regression to multiclass problems with more than two possible discrete outcomes. It is used to predict the probabilities of the different possible outcomes of a dependent variable given a set of independent variables.

It uses a linear predictor function

$$f(k,i) = \beta_{0,k} + \beta_{1,k}x_{1,i} + \beta_{2,k}x_{2,i} + \cdots + \beta_{M,k}x_{M,i},$$

where, ßk is the set of regression coefficients with outcome k, and xi is the set of explanatory variables associated with observation i.

4.4. Determining Variables of Influence



4.5. Spatial Distribution of Transportation Mode Choice

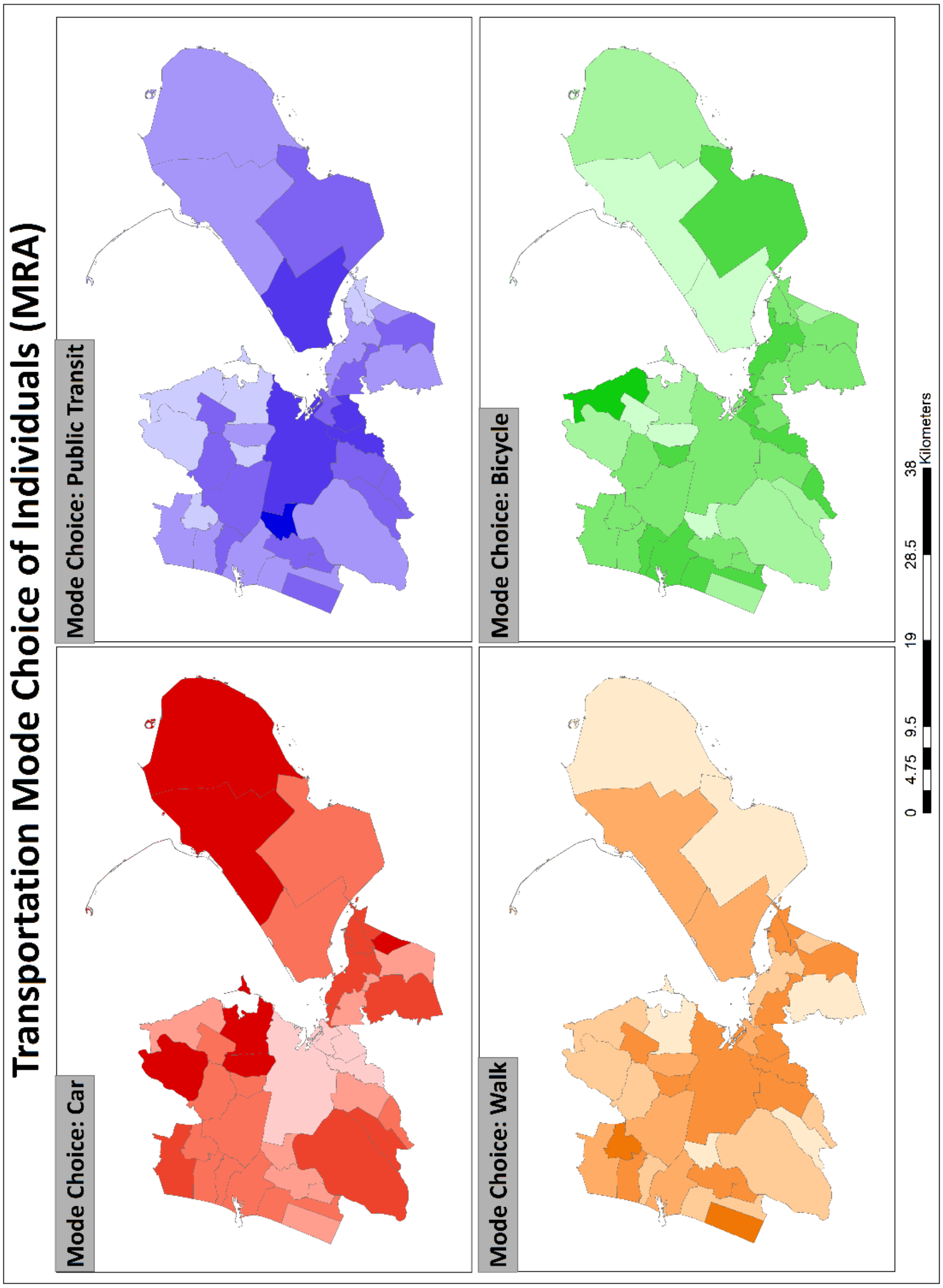


Figure 5: Map showing Transportation Mode Choice Distribution

5. Data Modelling

The model file consists of the model specifications of the discrete choice model to be estimated. The sections of the file are:

- [Model Description]

This section is required for providing description about the model. Each line must be in double-quotes as mentioned in the figure below which is an example of the model code.

The text mentioned in the model description will be specified in the output files.

- [Choice]

The choice section in the model file is required for specifying the dependent variable in the model being developed. Usually, the field name of the dependent variable as per the Data sheet is mentioned.

- [Beta]

Each line of this section corresponds to a parameter in the utility functions. Five entries must be provided for each parameter:

1. Name: Name of parameter to be estimated
2. Value: Default value that will be used as a starting point for the estimation, or used directly for the simulation in BIOSIM.
3. Lower bound: Minimum search range for estimation; can be increased based on the model results
4. Upper bound: Maximum search range for estimation; can be increased based on the model results
5. Status, which is 0 if the parameter must be estimated, or 1 if the parameter has to be maintained at the given default value.

- [Utilities]

Each row corresponds to an alternative. Four entries are specified:

1. This column is the identifier of the alternative, with numbering as per the choice definition;
2. Name of the Alternative
3. Availability condition variable of the alternative; the conditions are specified in the [Expressions] section.
4. The linear parameter utility function which consists of the list of terms separated by +; i.e. the name of the parameter and the name of the attribute separated by *.

- [Expressions]

This section is for defining all the expression for the utility functions or availability conditions in [Utilities] sections. The expressions for all the dummy variables, grouping of data from the datasheet can be directly mentioned in this section.

- [Exclude]

This section is used to define all the exclusion i.e. the data rows that need to be omitted during the estimation process which is done by specifying certain criteria.

- [Model]

The model section specifies the type of model that is to be used for the model estimation from the built-in library of the software.

Ex:
$BP: Binary Probit
$MNL: Multinomial Logit model
$ NL: Nested logit model

```
Mode_Choice2.mod - Notepad
File  Edit  Format  View  Help
    3 A4_BC     BC_AV        ASC_BC * one
                             + B_BC_HHSAM1 * hhsam1 + B_BC_HHSAM2 * hhsam2 + B_BC_HHSAM3 * hhsam3
                             + B_BC_HERKOMST1 * herkomst1 + B_BC_HERKOMST2 * herkomst2 + B_BC_HERKOMST3 * herkomst3
                             + B_BC_HHBESTINK1 * hhbestink1 + B_BC_HHBESTINK2 * hhbestink2 + B_BC_HHBESTINK3 * hhbestink3 + B_BC_HHBESTINK4 * hhbestink4
                             + B_BC_KAFSTV1 * kafstv1 + B_BC_KAFSTV2 * kafstv2 + B_BC_KAFSTV3 * kafstv3 + B_BC_KAFSTV4 * kafstv4 + B_BC_KAFSTV5 * kafstv5

 [Expressions] // This can be viewed as the preprocessing steps for the model to function.
CAR_AV = 1 // For this example we assume that the choice of using a car is available to everyone in the sample data set
PB_AV = 1 // For this example we also assume that the choice of using a public transport is available to everyone in the sample data set
BC_AV = 1 // For this example we also assume that the choice of using a bicycle is available to everyone in the sample data set
WALK_AV = 1 // For this example we also assume that the choice of walking is available to everyone in the sample data set
one = 1
hhsam1 = ( HHSam = 1 )
hhsam2 = ( HHSam = 2 )
hhsam3 = ( HHSam = 3 )
herkomst1 = ( Herkomst = 1 )
herkomst2 = ( Herkomst = 2 )
herkomst3 = ( Herkomst = 3 )
hhbestink1 = ( HHBestInk = 1 || HHBestInk = 2 )
hhbestink2 = ( HHBestInk = 3 )
hhbestink3 = ( HHBestInk = 4 || HHBestInk = 5 )
hhbestink4 = ( HHBestInk = 6 )
kafstv1 = ( KAfstV = 1 || KAfstV = 2 || KAfstV = 3 )
kafstv2 = ( KAfstV = 4 || KAfstV = 5 )
kafstv3 = ( KAfstV = 6 || KAfstV = 7 )
kafstv4 = ( KAfstV = 8 || KAfstV = 9 )
kafstv5 = ( KAfstV = 10 || KAfstV = 11 || KAfstV = 12 || KAfstV = 13 )

// as the model gets more complicated, expressions can be added to improve the estimation results

[Exclude]
(( KHvm == -1 || KHvm == 5 || KHvm == 8 ) + ( HHBestInk == 7 ) + ( KAfstV == 0 ) + ( Weggeweest != 1 ))

[Model] // Model type that we are using
$MNL // Model specification that you are using. In this case we use a multinomial logit
```

Figure 6: Sample Code File

6. Analysis

The model results are provided in a html file, which gives information about the log likelihood, rho-square, standard error, t-test, covariance, correlation values etc.
The following explanatory variables are considered for the prediction of this model:

- Distance (continuous)
- Ethnicity
- Household Income
- Household Composition
- Purpose of Travel
- Urbanization
- Population Density
- Gender
- Age

Also note that in this model a special condition has been set where the availability of bicycle has been provided only when the travel time is less than 75 min;

Sample Size: 19322
No. of Parameters estimated: 78
Final log likelihood: -18236.381
Rho-square: 0.309
Adjusted Rho-square: 0.306

Note: It is important to consider that all these analyses are relative to public transit.

Transportation Mode Choice **Alternatives**: Car, Public Transit, Bicycle, Walk		**Estimated Value**	**Std. Error**	**t-statistic**
Alternative Specific Constants	Bicycle	0.446	0.257	1.73
	Car	1.30	0.221	5.87
	Walk	1.18	0.290	4.06

The t-test values for car are higher than the other modes of travel i.e Walk, Bicycle and Public Transit.

Distance	Bicycle	-0.0260	0.000630	-41.27
	Car	0.00416	0.000792	-8.96
	Walk	-0.0701	0.00159	-44.06
	Public Transit	0.00	Fixed	Fixed

The t-test values indicate that as the distance increases public transit is most preferred. Whereas, the t-test values for Bicycle, Walk is significant but negative which means that with increase in distance people tend to walk/ bicycle less. This result is quite logical.

Ethnicity	**Bicycle**	Local	1.49	0.0745	20.03
		Western	1.08	0.0963	11.17
		Non-Western	0.00	Fixed	Fixed
	Car	Local	0.725	0.0621	11.68
		Western	0.274	0.0830	3.30
		Non-Western	0.00	Fixed	Fixed
	Walk	Local	0.702	0.0817	8.59
		Western	0.336	0.109	3.08
		Non-Western	0.00	Fixed	Fixed

The t-test values for ethnicity as an explanatory variable shows that local residents and western migrants use the mode bicycle, car and walk mode of commute more than the non-western migrants. Comparing the t-test values of each mode shows that people use bicycle more than car or walking.

Household Income	**Bicycle**	Low	-0.199	0.0967	-2.06
		Mid	-0.139	0.0872	-1.59
		High	-0.0979	0.0656	-1.49
		Very High	0.00	Fixed	Fixed
	Car	Low	-0.166	0.0867	-1.91
		Mid	0.0828	0.0749	1.10
		High	0.0583	0.0558	1.04
		Very High	0.00	Fixed	Fixed
	Walk	Low	-0.188	0.109	-1.72
		Mid	0.0962	0.0972	0.99
		High	-0.0184	0.0753	-0.24
		Very High	0.00	Fixed	Fixed

The t-test values for household income as an explanatory variable indicates that as the household income increase people use the bicycle/ walking mode of transport. This is a special case which can be noticed in countries like The Netherlands or Denmark. Comparing the overall t-test values suggests that car mode of transport is much preferred. We should also take into consideration here that this analysis is being done on the Metropolitan Region of Amsterdam as a whole, therefore, the results could be quite different when compared to Amsterdam city alone.

Household Composition	**Bicycle**	Single	-0.291	0.100	-2.89
		Couple	0.178	0.102	1.74
		Couple with Child	0.141	0.0888	1.59
	Car	Single	-0.343	0.0878	-3.90
		Couple	0.317	0.0877	3.61
		Couple with Child	0.538	0.0783	6.87
	Walk	Single	0.0124	0.116	0.11
		Couple	0.398	0.118	3.36
		Couple with Child	0.168	0.102	1.65

Higher household composition than couple with child is considered fixed i.e. the estimation is relative to this.

The t-test values for household composition shows that as the family size increase people tend to car mode of transport, whereas in the case of walking it can be seen that couple tend to walk much more than couple with a child, which is also logical. Comparing the overall t-test values indicates that car mode of transport is use more often when the family size increases which is sensible.

Purpose of Travel	**Bicycle**	Work	-0.281	0.125	-2.25
		Shopping, Personal Care	-0.124	0.130	-0.95
		Education	-0.466	0.143	-3.25
		Social & recreational activities	0.192	0.131	1.47
	Car	Work	-1.20	0.109	-11.04
		Shopping, Personal Care	-0.384	0.118	-3.27
		Education	-2.28	0.132	-17.23
		Social & recreational activities	-0.718	0.116	-6.21
	Walk	Work	-0.716	0.142	-5.04
		Shopping, Personal Care	-0.160	0.153	-1.05
		Education	-0.646	0.160	-4.05
		Social & recreational activities	0.968	0.141	6.87

All other purposes are considered fixed i.e. the estimation is relative to this.

The t-test values for purpose of travel as an explanatory variable implies that people do not use car mode of transport much and prefer bicycle or walking to their destination. It also suggests that use of bicycle or walking is most in the case of Social and recreational activities.

Urbanization	**Bicycle**	Highly Urbanized	0.322	0.164	1.97
		Strongly	0.0668	0.149	0.45
		Poor	0.192	0.154	1.24
		Few	0.00	Fixed	Fixed
	Car	Highly Urbanized	-0.663	0.138	-4.80
		Strongly	-0.541	0.126	-4.29
		Poor	-0.171	0.131	-1.30
		Few	0.00	Fixed	Fixed
	Walk	Highly Urbanized	0.703	0.194	3.62
		Strongly	0.505	0.179	2.82
		Poor	0.484	0.184	2.63
		Few	0.00	Fixed	Fixed

The t-test values for urbanization as an explanatory variable indicates that as there is increase in urbanization the use of bicycle or walking for commuting is highly preferred over the use of car. Such results are true in the case of The Netherlands or the MRA.

Population Density	**Bicycle**	Upto 50000	0.872	0.120	7.28
		50,000 – 150,000	0.665	0.0970	6.85
		150,000 – 250,000	0.251	0.0937	2.68
	Car	Upto 50000	0.981	0.103	9.50
		50,000 – 150,000	1.24	0.0831	14.96
		150,000 – 250,000	0.678	0.0795	8.52
	Walk	Upto 50000	0.686	0.136	5.06
		50,000 – 150,000	0.676	0.108	6.24
		150,000 – 250,000	0.410	0.104	3.95

Population Density above 250,000 is considered fixed i.e. the estimation is relative to this.

The t-test values for population density as an explanatory variable shows that where the density is low people tend to use car more often than in highly dense areas.

Gender (Male)	Bicycle	0.131	0.0533	2.45
	Car	0.245	0.0458	5.35
	Walk	0.136	0.0608	2.24

The t-test values for gender as an explanatory variable with female set as baseline suggests that the male population use car more compared to bicycle or walk.

Age	**Bicycle**	6-14 Years	1.19	0.153	7.82
		15-19 Years	-0.122	0.152	-0.80
		20-34 Years	-0.339	0.141	-2.42
		35-55 Years	0.154	0.136	1.13
		Above 55 Years	-0.00302	0.146	-0.02
	Car	6-14 Years	0.445	0.145	3.06
		15-19 Years	-2.00	0.147	-13.62
		20-34 Years	-1.05	0.124	-8.47
		35-55 Years	-0.334	0.122	-2.74
		Above 55 Years	-0.361	0.129	-2.79
	Walk	6-14 Years	0.674	0.161	4.19
		15-19 Years	-1.06	0.180	-5.88
		20-34 Years	-0.491	0.152	-3.23
		35-55 Years	-0.114	0.145	-0.79
		Above 55 Years	-0.0543	0.158	-0.34

The t-test values for age as an explanatory variable (Categorical) suggests that 6-14 Years age group use bicycle which is quite logical, the significance of the same in the car alternative maybe because they are accompanied by an adult. 15-35 Years age group prefer to use public transit more based on the t-test values. 35 Years & above age group use bicycle more than car, walk although the t-test values do not show much significance.

7. Results

	Sample Size	***Parameters Estimated***	***R-Squared***	***Adjusted R-Squared***
Model-I	19538	42	0.254	0.252
Model-II (Gender + Age)	19455	48	0.261	0.259
Model-III (Purpose of Travel)	19455	60	0.284	0.282
Model-IV (Urbanization)	19455	69	0.295	0.292
Model-V (Bicycle Availability)	19322	78	0.309	0.306

Comparing the model results after including more explanatory variables with each model run shows a significant increase in the Adjusted R-squared value which was initially 0.252 improved to 0.306 resulting in a better fit model for the study conducted. The parameters that were used in the final model are mentioned below:

S. No	**Variable**	**Type of Variable**
1	Household Composition	**Categorical**
2	Ethnicity	
3	Household Income	
4	Gender	
5	Purpose of Travel	
6	Urbanization	
7	Population Density	
8	Age	
9	Travel Distance	**Continuous**

The final model also included an additional condition wherein the alternative of bicycle was made available only when the travel time is under 75 minutes.

8. Conclusion

This study presents various choice models estimated considering the socio-economic characteristics of individuals with the help of multinomial logit model using open source software Biogeme. The analysis provides interesting results for e.g. the income variable indicates that the higher income group respondents do not prefer to choose car as a mode of transport which is true in the case of The Netherlands. The age group variable also suggests that people prefer to commute by bicycle, walking or public transit which includes bus, tram, metro and train rather than opting for car. The travel time variable suggests that people prefer to use public transit more when the travel time is more which is quite logical as well. The findings of the study complement the transport policies adopted in The Netherlands.

It is also important to investigate the significance of the variables considered for the model estimation. This study shows that distance, ethnicity and population density are parameters of high significance in the model estimate.

9. Future Research

- Modelling the departure time choice of individuals and analysis of people travelling by which mode avoid the rush hour.
- Studying the influence neighborhood (residential location) has on the transportation mode choice of individuals
- Understanding, modelling and analysis of combined mode transportation (e.g. Walk + Train) of individuals and looking into the last mile problem
- Implementing the model in an Indian Scenario

Acknowledgements

We would like to express sincere thanks to Dr. Anjana Vyas, Ms. Shaily Gandhi, Dr. Bindi Dave, CEPT University, Ahmedabad for their guidance and inputs during the work. We would also like to thank Jullian van Kampen, Thomas Koch, Centrum Wiskunde & Informatica, Amsterdam for their valuable inputs. Finally, VKJ would like to thank his brother, parents and friends for their help and support throughout.

References

Alqhatani, M., Bajwa, S., & Setunge, S. (2013). Modeling the Influence of Socioeconomic and Land-Use Factors on Mode Choice: A Comparison of Riyadh, Saudi Arabia and Melbourne, Australia. *International Journal of Environmental and Ecological Engineering, Vol: 7, No:2*.

Ben-Akiva, M., & Bierlaire, M. (n.d.). Chapter 2: Discrete choice models with applications to Departure time and Route Choice. In *Handbook of Transportation Science.*

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. *In: Proceedings of the 3rd Swiss Transportation Research.* Ascona, Switzerland.

Board, A. E. (n.d.). *Metropolitan Region of the Future*. Retrieved from https://www.amsterdameconomicboard.com/metropolitan-region-the-future

Dugundji, E. R., & Walker, J. L. (2005). Discrete Choice with Social and Spatial Network Interdependencies. *Transportation Research Record*.

Loidl, M., Wallentin, G., Cyganski, R., Graser, A., Scholz, J., & Haslauer, E. (2016). GIS and Transport Modeling—Strengthening the Spatial Perspective. *International Journal of Geo-Information*.

Miletić, G.-M., Gašparović, S., & Carić, T. (2017). Analysis of Socio-spatial Differentiation in Transport Mode Choice Preferences. *Promet - Traffic & Transportation, Vol. 29, No. 2*, 233-242.