CEE 552 TERM PROJECT: LONDON MODE CHOICE MODEL

Jason Lu
Department of Civil and Environmental Engineering
University of Michigan
Email: lujason@umich.edu

Ruicong Zhang
Department of Civil and Environmental Engineering
University of Michigan
Email: ruicongz@umich.edu

Word Count: 4006 words + 10 table(s) × 250 = 6506 words

Submission Date: December 6, 2023

ABSTRACT

Congestion has become an exponentially growing problem in metropolitan areas worldwide, hindering the effectiveness of mobility. Additionally, the growing recognition of collective and emissions-free mobility solutions has brought upon the potential for greater sustainability and efficiency in mobility. Understanding the travel patterns of individuals is critical to assessing the factors that contribute to certain mode choice alternatives seen in mobility. This paper studies the mode choice patterns of individuals and assesses the contributions of a variety of attributes towards several key mode choice alternatives. It presents a case study for the city of London and evaluates the travel behavior of individuals through four mode choices: walking, cycling, transit, and driving. This study develops a multinomial logit model and utility specifications to effectively analyze the travel behavior of individuals with regards to mode choice. The results reveal several key attributes that contribute towards the utilities of the alternatives with respect to one another, particularly driving vs. non-driving alternatives. A causal inference analysis also expanded the results are gave more insight on the selection of particular mode choices.

Keywords: travel behavior, mode choice, multinomial logit

INTRODUCTION

For years, traffic congestion has been rising significantly, with the majority of cities worldwide experiencing greater number of hours lost in congestion compared to previous years. In 2022, the city of London was the most congested city in the world in terms of hours lost in congestion, at an average of 156 hours per person (*1*). The root of traffic congestion comes from the large number of individuals driving personal vehicles. While other modes of transport exist, such as public transit, cycling, or even walking, driving has been the predominant mode choice for individuals.

While congestion hinders individuals in terms of convenience, the increase in driving personal vehicles also hinders environmental sustainability. On average, personal vehicles emit 1.6 million grams of CO2 yearly (*2*), and in Europe, approximately 72% of all CO2 emissions come from road transportation (*3*). While CO2 trends have generally gone downwards, there is still a need to reduce emissions from transportation. This views collective mobility solutions, such as public transit, or emissions-free mobility solutions, such as cycling or walking, as favorable alternatives towards driving.

In large-scale societies, it is imperative to accurately assess behaviors that affect the mode choice of individuals. Analyzing these behaviors potentially leads to more insight on why individuals favor certain modes of transportation. This analysis could lead to more effective planning of transportation systems and guide individual behaviors towards more desirable systems for all, resulting in greater convenience for travelers and a more sustainable environment.

The mode choice problem is a traditional problem in transportation engineering. The goal is to assess attributes (characteristics pertaining to the mode choices or to the individuals) that will contribute towards the utility of certain alternatives (the mode choices themselves). Several models have been developed to investigate mode choice, including multinomial logit models, nested logit models, demand based models, and activity based models.

Mode choice has been studied extensively in literature and has been applied in different contexts. Hu et. al. utilize a multinomial logit model to assess travel behavior patterns of various mode choices in small cities of China (*4*). Their findings reveal that land use and transportation strategies are the primary factors when individuals determine mode choice, which also vary across the day of the week. Dissanayake and Morikawa utilize a nested logit model to invested household travel behavior while considering travel mode choices in a case study in Bangkok, Thailand (*5*). Their study revealed the key attributes that contribute to mode choices, while also inferring preferences towards travel modes not yet in existence. In a well-known hypothetical mode choice scenario called Swissmetro, Bierlaire et. al. create a study that analyzes factors that would contribute to adoption of Swissmetro if it were implemented (*6*). Several studies have also investigated mode choice towards more sustainable alternatives such as micromobility (e-scooters, e-bikes, etc.) (*7–9*), which have revealed key factors that influence the behavior of people selecting these modes.

This study investigates mode choice travel behavior in the city of London, England from a travel demand survey that was conducted between 2012-2015. Several attributes are recorded in the survey with a total of four alternatives to assess: driving, public transit, cycling, and walking. This study develops a multinomial logit model to analyze attributes that affect the travel mode choices on the survey data, assessing behavior patterns across all individuals. This study also uses causal inference to give additional insight on the choice probabilities between alternatives. A particular focus will be on sustainable modes: 1) emissions-free mobility (cycling and walking) and 2) collective mobility (public transit). Therefore, this study also assesses behaviors that contribute towards the selection of sustainable mode choices and conducts an analysis on these factors.

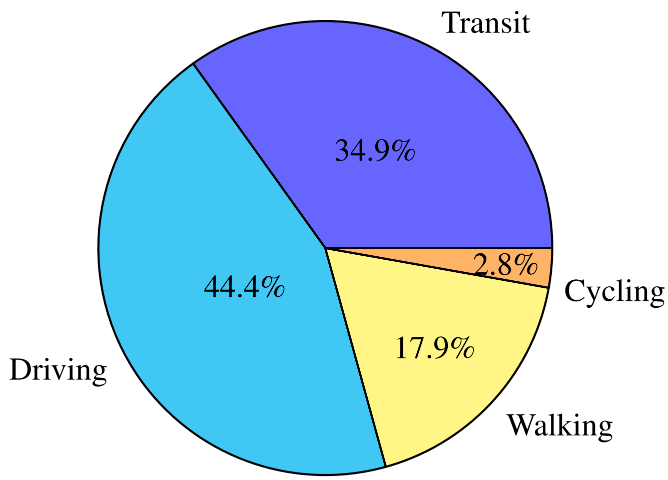


FIGURE 1: Distribution of Mode Choice Selections Across Trips

The rest of this paper is as follows. Section 3 discusses the background of the study and data descriptions. Section 4 details the model specifications, multinomial logit choice model, and causal inference analysis designed for this study. Section 5 assess the results from the methodology applied to the data. Section 6 discusses broad perspectives of the findings and potential future work.

DATA DESCRIPTION

This study utilizes travel demand survey data that was conducted between April 2012 and March 2015 on riders travelling in London (*10*). The historical revealed-preference data was obtained from the London Travel Demand Survey, a household survey conducted by Transport for London (TfL). To include details of choice set within the data, an online journey planner was used to determining level-of-service attributes for the 4 mode choices used in this study and appending them to the survey data. The choice set in this study is defined as $C_n = \{1,2,3,4\}$, with 1 = walking, 2 = cycling, 3 = transit, and 4 = driving. Due to computational constraints, this study does not consider all trips in the survey. In total, this study considers a portion of the survey conducted across 7703 households, resulting in 36080 total collected trips made by the individuals. A distribution of the proportion of mode choices selected in these trips is shown in Figure 1. Evidently, the majority of trips are driving trips while transit also has a significant number of trips. While walking and especially cycling account for a minority of the trips, it is still imperative to include these as part of the choice set as these modes are distinguishable from the other modes in the choice set.

A detailed description of the initial variables from the dataset is presented in Table 1, which lists all variables, descriptions of each variables (includes the units when applicable), and variable types. Note that Public Transit (denoted as PT in the table) requires multiple attributes to determine the total travel time. Additionally, the cost for driving is split between two attributes as well.

From the way the data was collected, we propose two improvements to improve the instrument and methodology of acquiring the data:

1. **Accountability of Availabilities**: When acquiring the data, considering constraints for availabilities of alternatives for each individual would improve the data when assessing

mode choice options. This could be done when determining level-of-service attributes to consider choice set in the historical trips.

2. **Expand Survey Attributes**: While the data does consider a wide range of attributes that are adequate for analysis, expanding the range of attributes to include other individual-level characteristics (e.g. ethnicity, disabilities) would give more information to inform the models. Additionally, modifying certain attributes such as distance (which is the same across alternatives) to vary across alternative (e.g. accounting for areas where only walking/cycling is permitted and cuts the distance) would also be beneficial.

The initial set of variables were processed according to the needs of the study. The variables of *survey_year*, $travel\_year$, $travel\_month$, $travel\_day$, day_of_week, and $start\_time$ were eliminated as we wanted to focus on the assessment of other factors in the model.

Additionally, the distance was scaled from meters to kilometers in order to be in line with the time attributes of the study, which are given in hours. A public transit duration time variable, which is denoted *dur_transit*, is created as the summation of *dur_pt_access*, *dur_pt_rail*, *dur_pt_bus*, and *dur_pt_int*, displacing the latter four variables. A cost for driving variable, which is denoted *cost_driving*, is created as the summation of $cost\_driving\_fuel$ and *cost_driving_ccharge*, displacing the latter two variables. The creation of *dur_transit* and *cost_driving* allows us to consider duration and cost fairly and holistically with respect to the other alternatives.

The final set of variables used in the model is given in Table 3, which also gives more detailed descriptions on the meaning of each variable. A summary of the descriptive statistics for variables used in the model is given in Table 2, which considers all variable changes described previously. This table considers variables that are not categorical, as descriptive statistics for categorical variables would not be meaningful. An exception was made for the *car_ownership* variable as the car ownership categories follow an increasing magnitude with how it's represented. Descriptive statistics calculated for the variables are minimum, 25th percentile (25%), 50th percentile (50%), 75th percentile (75%), maximum, mean, and standard deviation across all collected trips. From the descriptive statistics, we can see that the duration variables for each alternative are very sensible and fall in line with what is expected compared to one another (with walking being the slowest by a large margin and driving being the fastest). Additionally, costs for driving skew higher, with costs being relatively lower than transit at 50th and 75th quartiles but the overall mean being higher. This shows very high driving costs from fuel and congestion for a good portion of trips.

Figure 2 displays the distribution of the proportion of trips for the categorical and binary variables considered in the model. Specifically, Figure 2a displays the distribution for the $female$ variable, Figure 2b displays the distribution for the *purpose* variable, Figure 2c displays the distribution for the *car_ownership* variable, and Figure 2d displays the distribution for the *driving_license* variable. From the charts, we see that the distribution of gender across trips is very even. The majority of trips are home based trips (specifically for the "Other" category), meaning that the overall distance for these trips are likely at shorter distances. The majority of trips are taken by individuals with < 1 car per adult in household, with a relatively even split between the other two categories. Finally, the majority of individuals hold a drivers license.

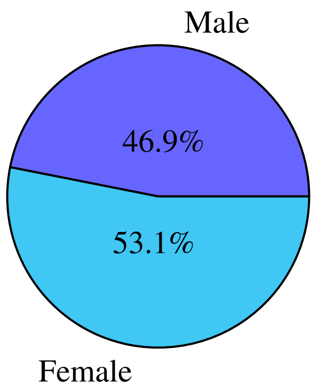
To account for *purpose* and *car_ownership* categorical variables in the study, we transform these into dummy variables. For *purpose*, the "non-home purpose" category serves as the base category and dummy variables are created for the other categories. For *car_ownership*, the category of owning less than one car per adult in the household is used as the base category. Dummy variables are created for the other categories.

Variable	**Description**	**Variable Type**
travel_mode	Mode choice to predict	Categorical
purpose	home vs. work based	Categorical
survey_year	Year survey was conducted	Categorical
travel_year	Year of trip	Discrete
travel_month	Month of trip	Discrete
travel_day	Day of month of trip	Discrete
day_of_week	Day of week of trip	Categorical
start_time	Start time of trip in the day (linearized (0-24))	Discrete
age	Age of person making trip (years)	Discrete
female	If trip person is female or not	Binary
driving_license	If trip person has license or not	Binary
car_ownership	Number of cars person owns	Categorical
distance	Distance of trip (meters)	Continuous
dur_walking	Walk time for trip (hours)	Continuous
dur_cycling	Cycle time for trip (hours)	Continuous
dur_pt_access	Access/egress time for PT (hours)	Continuous
dur_pt_rail	PT Rail time for trip (hours)	Continuous
dur_pt_bus	PT Bus time for trip (hours)	Continuous
dur_pt_int	Transfer time in PT for trip (hours)	Continuous
pt_interchanges	Number of transfers in PT for trip	Discrete
dur_driving	Driving time for trip (hours)	Continuous
cost_transit	Cost of PT (GBP)	Continuous
cost_driving_fuel	Cost of fuel driving (GBP)	Continuous
cost_driving_ccharge	Congestion cost for driving (GBP)	Continuous
driving_traffic_percent	Traffic variability	Continuous

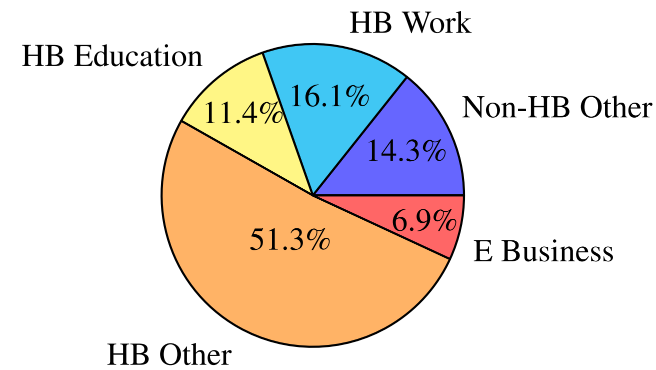
TABLE 1: Attributes of Initial Dataset

METHODOLOGY

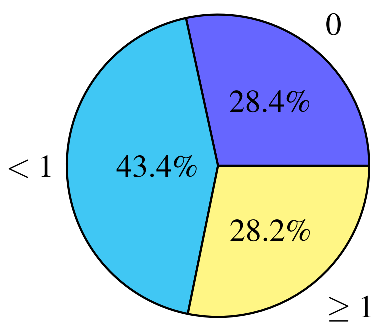
This section details the choice model used to assess travel behavior for the study. Section 4.1 discusses the multinomial logit model, which is the focal choice model in the study. Section 4.2 details the specifications of alternatives. Section 4.3 describes the framework used for causal inference.



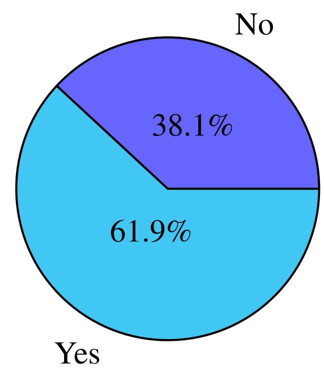
(a) Distribution of Gender Across Trips



(b) Distribution of Purpose Across Trips (HB = Home-based, E=Employer)



(c) Distribution of Car Ownership per Adult in Household Across Trips



(d) Distribution of Driving License Ownership Across Trips

FIGURE 2: Distribution of Trips Across Categorical and Binary Variables Used in Model

Variable	**Minimum**	**25%**	**50%**	**75%**	**Maximum**	**Mean**	**Standard Deviation**
age	5	25	38	52	99	39.46	19.23
female	0	0	1	1	1	.53	.5
driving_license	0	0	1	1	1	.62	.49
car_ownership	0	0	1	2	2	.98	.75
distance	.08	1.31	2.81	6.18	40.941	4.605	4.782
dur_walking	.03	.35	.72	1.51	9.28	1.13	1.12
dur_cycling	.01	.12	.23	.48	3.05	.36	.35
dur_transit	.01	.23	.39	.64	2.73	.47	.31
dur_driving	.0003	.11	.19	.37	2.06	.28	.25
cost_transit	0	0	1.5	2.4	13.49	1.56	1.54
cost_driving	0	.29	.57	1.29	17.16	1.90	3.49
driving_traffic_percent	0	.17	.31	.48	1.25	.33	.2

TABLE 2: Descriptive Statistics for Variables

Variable	Description	Variable Type
travel_mode	Mode choice selection of individual (1: walking, 2: cycling, 3: transit, 4: driving)	The Alternative
purpose	Purpose of trip (1: home-based work, 2: home-based education, 3: home-based other, 4: employers' business, 5: non-home based other)	Categorical
age	Age of person making trip (years)	Discrete
female	If trip person is female or not	Binary
driving_license	If trip person has license or not	Binary
car_ownership	Number of cars owned in household (0: 0 cars in household, 1: <1 car per adult, 2: ≥ 1 cars per adult	Categorical
distance	Distance of trip (kilometers)	Continuous
dur_walking	Walk time for trip (hours)	Continuous
dur_cycling	Cycle time for trip (hours)	Continuous
dur_transit	Transit time for trip (hours)	Continuous
dur_driving	Driving time for trip (hours)	Continuous
cost_transit	Cost of PT (GBP)	Continuous
cost_driving	Cost of driving (GBP)	Continuous
driving_traffic_percent	Traffic variability (%)	Continuous

TABLE 3: Attributes Considered in Model

Multinomial Logit Model

This study designs a multinomial logit model to effectively assess travel behavior mode choices on the data detailed in Section 3. The four alternatives of walking, cycling, transit, and driving make up the choice set $C_n = \{1,2,3,4\}$.

The multinomial logit model follows three assumptions on the specification:

- The error terms e_{jn} of our data (for alternative j in the choice set and an individual n) are independently and identically distributed.
- The error terms $e_{jn} \sim EV(0,\mu)$ (extreme value distirbuted with location 0 ans scale μ).
- Variance of the specification is given by $\frac{\pi^2}{6\mu^2}$

The derivation of the probability of individual n for selecting alternative $i \in C_n$ is given by the initial form in Equation 1:

$$P_n(i) = P(V_{in} + \varepsilon_{in} \geq max_{j \in C_n/i}(V_{jn} + \varepsilon_{jn})) \quad (1)$$

V_{in} and ε_{in} is the systematic and random utility, respectively, of alternative i for individual n. Additionally, $max_{j \in C_n/i}(V_{jn} + \varepsilon_{jn}) = V_n^* + \varepsilon_n^*$. $V_n^* = \frac{1}{\mu} ln(\sum_{j \in C_n/i} e^{\mu V_{jn}})$ and $e_n^* \sim EV(0, \mu)$, where μ is a scale parameter on the error terms. This sets the probability as:

$$P_n(i) = P(V_{in} + \varepsilon_{in} \geq V_n^* + \varepsilon_n^*) \tag{2}$$

$$P_n(i) = P((\varepsilon_n^* - \varepsilon_{in}) \leq (V_{in} - V_n^*)) \tag{3}$$

Which shows that the probability is given by the difference in the random utilities between other alternatives and alternative i being less than or equal to the difference in the systematic utilities between alternative i and other alternatives. Essentially, the systematic, and thus deterministic, utility must be greater than or equal to the random, unexplained utility. This leads to the final derivation of the probability of individual n for selecting an alternative $i \in C_n$, given by Equation 6:

$$P_n(i) = \frac{1}{1 + e^{\mu(V_n^* - V_{in})}} \tag{4}$$

$$P_n(i) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{ln(\sum_{j \in C_n/i} e^{\mu V_{jn}})}} \tag{5}$$

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{6}$$

With the multinomial logit model formulation, this study assumes the property of Independence from Irrelevant Alternatives. This holds that for any particular individual, the ratio of the choice probabilities is unaffected by the existence of any other alternatives in the whole set of alternatives.

Utility Specifications

The multinomial logit model was tuned significantly according to the variables from Table 3, with parts of the specification being adjusted if there were variables that were statistically insignificant and weren't logically sensible in the final model specifications. Due to this, additional variables were removed or certain alternatives were not considered with certain parameters.

A list of the final utility specifications for the MNL model is given in Table 4. These make up the set of alternative specific constants, individual specific variables, and alternative specific variables that are part of the model specifications. It also includes how each variable will be represented in the utility specifications, which are discussed below.

The final model specifications hold linear-in-parameters properties. Variables that are discrete or continuous in nature hold linearity. Categorical variables were converted to linear representation as binary dummy variables. Equations 7, 8, 9, and 10 give the final utility specifications for the alternatives of walking, cycling, transit, and driving, respectively, for a particular individual n. These are in line with Table 4. Each equation signifies the total utility with both systematic and random components included.

$$U_{1n} = \beta_1 + \beta_{1,TT}TT_1 + \beta_{1\&2,CO0}CO0 + \beta_F F + \beta_{1,PHW}PHW + \beta_{1,PHE}PHE + \varepsilon_{1n} \tag{7}$$

$$\begin{split} U_{2n} = \beta_2 + \beta_{2,TT}TT_2 + \beta_{2,DIST}DIST + \beta_{1\&2,CO0}CO0 + \beta_F F + \\ \beta_{2,PHW}PHW + \beta_{2,PHE}PHE + \beta_{2,PHO}PHO + \beta_{2,PB}PB + \varepsilon_{2n} \end{split} \tag{8}$$

$$\begin{split} U_{3n} = \beta_3 + \beta_{3,TT}TT_3 + \beta_{3,DIST}DIST + \beta_{3,CO0}CO0 + \beta_{AGE}AGE + \\ \beta_{3,PHW}PHW + \beta_{3,PHE}PHE + \beta_{3,PHO}PHO + \beta_{3,PB}PB + \beta_{3,COST}COST_3 + \varepsilon_{3n} \end{split} \tag{9}$$

$$\begin{split} U_{4n} = \beta_{4,TT}TT_4 + \beta_{4,DIST}DIST + \beta_{DL}DL + \beta_{CO2}CO2 + \beta_{AGE}AGE + \\ \beta_{4,COST}COST_4 + \beta_{TRAF}TRAF + \varepsilon_{4n} \end{split} \tag{10}$$

With the specifications set, the model is estimated using maximum likelihood estimation method to obtain parameter estimates that best fit the data. The model was solved using the PyLogit module.

Causal Inference

Causal inference is a science that conducts counterfactual estimation based on observational data and analyzes the causal relationship between intervention and outcome. This study uses it to analyze the relationship between each variable change and the mode selection probability, which will be beneficial to policy formulation. This study uses the DoWhy framework to conduct causal inference modeling, which provides a principled method to transform a given problem into a causal diagram that ensures the clarity of all hypotheses. The DoWhy framework also provides a unified interface for a variety of commonly used causal inference methods, and combines two main causal inference frameworks (graph models and potential causal models), while automatically testing the accuracy of hypotheses and robustness of the estimate. The entire causal inference process of DoWhy can be divided into four major steps:

- **Model**: using assumptions (prior knowledge) to model causal inference problems
- **Identify**: an expression (causal estimator) that identifies a causal effect under a hypothesis (model)
- **Estimate**: using statistical methods to estimate expressions
- **Refute**: using various robustness checks to verify the correctness of estimates

We conducted causal inference analysis on the four modes of walking, cycling, public transit, and driving.

RESULTS

This section details the results of the multinomial logit model, including interpreting parameter estimates for the utility specifications and statistical tests on the model. The causal inference analysis is also detailed.

Table 5 gives the full list of parameter estimates associated with the model, as well as the t-statistics and p-values associated with each parameter. The final model specifications were ad-

Variable	**Representation**	**Associated Alternatives**
ASC		1,2,3
Travel Time	TT	1,2,3,4
Driving License	DL	4
Car Ownership: 0	$CO0$	[1,2], 3
Car Ownership: 2	$CO2$	4
Female	F	[1,2]
Age	AGE	[3,4]
Distance	$DIST$	2,3,4
Purpose: Home-based Work	PHW	1,2,3
Purpose: Home-based Education	PHE	1,2,3
Purpose: Home-based Other	PHO	2,3
Purpose: Business	PB	2,3
Travel Cost	$COST$	3,4
Traffic Percent	$TRAF$	4

TABLE 4: Final Utility Specifications for Multinomial Logit Model

justed through multiple iterations to account for insignificant parameters with regards to a 95% confidence level. Therefore, as seen in Table 5, all parameters in the final specification are statistically significant according to the t-statistics and p-values.

To begin, the signs of all parameter estimates are sensible and line up with what is expected. Namely, all cost and time magnitudes are negative. Additionally, individual specific variables like "Car Ownership: 0" correlate positively with alternatives when compared to the base of driving. Of note is the "Distance" variable, which shows that the estimates are positive. But recall that distance is the same regardless of the mode choice, and as the base is set to the walking alternative, other alternatives are expected to handle longer distance more effectively.

From the estimates in the table, we observe that all attributes played a role in determining the utilities of each alternative But, there are several key attributes that heavily contribute to the utilities of particular alternatives with respect to others. In travel time, we observe that both walking and driving modes heavily deter their respective utilities with longer times compared to cycling and transit. When focusing on driving, this shows how switching to other alternatives for time efficiency would potentially yield more efficient travel. Additionally of particular note is the "Car Ownership: 0" variable, which indicates that individuals having no car will be much more inclined to seek non-driving alternatives for travel. This is critical in deciding other mobility solutions to utilize outside of driving. The categorical dummy variables associated with "Purpose" also show important contributions to utilities for certain alternatives. With the base alternative for these variables set as the driving alternative or both driving/walking alternatives, the estimates are all positive. This shows that biking and transit, and additionally walking when it is not set as a base, all have greater utility gains compared to driving. Furthermore, it indicates greater liking towards modes other than driving and indicates potential for other alternatives to displace driving. For example, the "Home-based Work" category shows particularly high estimates for the alternatives, indicating a strong potential for non-driving mode choices when utilizing this mode. Lastly, the "Traffic percent" variable shows negative heavy contribution to the utility of driving when there's

Variable	Associated Alternatives	β Estimate	t-stat	p-value
ASC	1	2.1625	31.777	0
	2	-3.2651	-24.888	0
	3	-1.4320	-23.788	0
Travel Time	1	-5.0489	-11.680	0
	2	-1.6100	-2.196	0.028
	3	-2.6270	-24.514	0
	4	-4.4680	-19.974	0
Driving License	4	0.8325	25.238	0
Car Ownership: 0	1,2	2.0142	39.179	0
	3	2.6574	57.388	0
Car Ownership: 2	4	0.5830	18.550	0
Female	1,2	-0.3315	-9.962	0
Age	3,4	0.0044	4.910	0
Distance	2	0.6379	5.373	0
	3	0.9475	8.919	0
	4	0.9432	8.891	0
Purpose: Home-based Work	1	0.7355	9.558	0
	2	1.8989	14.313	0
	3	1.0320	17.656	0
Purpose: Home-based Education	1	0.4271	7.395	0
	2	0.4987	3.082	0.002
	3	0.8700	13.789	0
Purpose: Home-based Other	2	0.4458	3.577	0
	3	0.2420	5.244	0
Purpose: Business	2	1.2836	8.343	0
	3	0.4925	7.025	0
Travel Cost	3	-0.1887	-14.885	0
	4	-0.1192	-17.279	0
Traffic Percent	4	-2.5356	-25.118	0

TABLE 5: Parameter Estimates and Statistical Tests From Final Model Specifications

a large amount of congestion. The usage of non-driving alternatives would be a huge alleviation to the burdens and inefficiencies of congestion.

Additional Specification Tests

We undertook a series of informal and statistical tests on the multinomial logit model specifications. For an informal test, the value of time was calculated for both the transit and driving alternatives, which have time and cost attributes associated with them. The results for both alternatives are presented in Equations 11 and 12, respectively. The value of time results for both alternatives are reasonable which authenticates the estimates for both alternatives.

$$\frac{\beta_{3,TT}}{\beta_{3,COST}} = 13.92 \text{ GBP/hour} \tag{11}$$

$$\frac{\beta_{4,TT}}{\beta_{4,COST}} = 37.48 \text{ GBP/hour} \tag{12}$$

Summary Statistic	Value
No. Observations	36080
Degrees of Freedom	29
$\mathscr{L}(0)$	-50017.501
$\mathscr{L}(\beta)$	-24982.436
ρ^2	.501
$\bar{\rho}^2$	.500

TABLE 6: Summary Statistic Values from Multinomial Logit Model Specifications

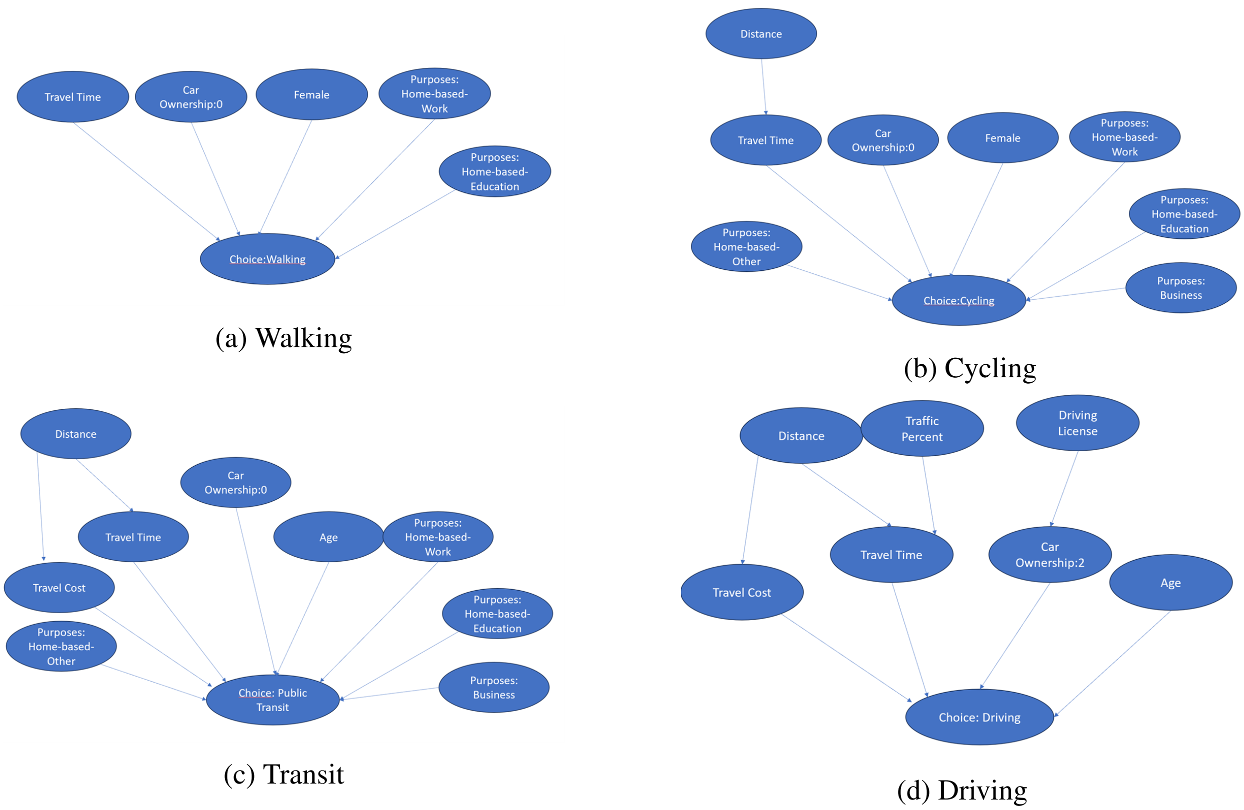
Using the summary statistical values of the model displayed in Table 6, we undertook statistical tests on the model as well. These will be conducted on the model as a whole, as we have already undertook t-tests on the t-statistics and p-values previously for each individual parameter estimate. We conducted a likelihood ratio test based on the log-likelihood of restricted null model $\mathscr{L}(0)$ (when all estimates are zero) and the log-likelihood of the unrestricted full model $\mathscr{L}(\beta)$ (with our model parameter estimates). This leads to the statistic value of $-2(\mathscr{L}(0) - \mathscr{L}(\beta)) = 50070.142 \geq \chi^2_{29}$. Here, the statistic is performed with a χ^2 test of 29 degrees of freedom, which is the difference in degrees of freedom between unrestricted and restricted model. The test result shows that the full model is statistically significant. Additionally, the ρ^2 and $\bar{\rho}^2$ values of .501 and .500, respectively, indicate adequate goodness of fit for the model.

Causal Inference Results

The cause-and-effect diagrams of the four models are shown in Figure 3. Figure 3a, Figure 3b, Figure 3c, and Figure 3d, represent the cause-effect diagrams of walking, cycling, transit, and driving, respectively.

Treatment Variable	Causal Effect	Outcome
Travel Time	Yes	-0.13784
Car Ownership:0	Yes	0.06132
Female	No	-
Purpose:Home-based Work	No	-
Purpose:Home-based Education	No	-

TABLE 7: Causal Inference for Choice Walking



(a) Walking

(b) Cycling

(c) Transit

(d) Driving

FIGURE 3: Cause-Effect Diagrams Across Different Modes

Treatment Variable	Causal Effect	Outcome
Travel Time	Yes	-0.09784
Car Ownership:0	Yes	0.08132
Female	No	-
Distance	Yes	-0.02952
Purpose:Home-based Work	No	-
Purpose:Home-based Education	No	-
Purpose:Home-based Other	No	-
Purpose:Business	No	-

TABLE 8: Causal Inference for Choice Cycling

From the analysis results in Table 7, we can see that only travel time and Car Ownership 0 have a causal effect on choosing to walk. From a practical point of view, the probability of choosing to walk can only be changed by changing travel time.

From the analysis results in Table 8, we can see that only travel time, Car Ownership: 0 and distance have a causal effect on the choice of cycling. From a practical point of view, the probability of choosing cycling can only be changed by changing travel time and distance. Changing distance is preferred.

From the analysis results in Table 9, we can see that only travel time, Car Ownership: 0, distance and travel cost have a causal effect on the choice of public transit. From a practical point

Treatment Variable	Causal Effect	Outcome
Travel Time	Yes	-0.00533
Car Ownership:0	Yes	0.10116
Age	No	-
Distance	Yes	-0.00052
Purpose:Home-based Work	No	-
Purpose:Home-based Education	No	-
Purpose:Home-based Other	No	-
Purpose:Business	No	-
Travel Cost	Yes	-0.01382

TABLE 9: Causal Inference for Choice Transit

Treatment Variable	Causal Effect	Outcome
Travel Time	Yes	0.42103
Driving License	Yes	0.03392
Car Ownership:2	Yes	0.54116
Age	No	-
Distance	Yes	0.00392
Travel Cost	Yes	-0.01382
Traffic Percent	Yes	-0.00182

TABLE 10: Causal Inference for Choice Driving

of view, the probability of choosing public transit can only be changed by changing travel time, distance, and travel cost. Changing distance is preferred.

From the analysis results in Table 10, we can see that except for the variable “Age”, other variables will have a causal effect on the choice of driving. From a practical point of view, the probability of choosing driving can only be changed by changing travel time, car ownership: 2, distance, and travel cost. The priority is to change distance and car ownership: 2.

In conclusion, shortening the travel time and distance of walking and cycling will lead to a significant increase in the number of people choosing non-driving, and thus more sustainable, mode choice alternatives.

CONCLUSION

This study assessed mode choice behavior of individuals in the city of London using travel demand survey data. A multinomial logit choice model was developed to assess factors that contributed to particular mode alternatives. Afterwards, the utility specifications were set for the four modes and the model was estimated. We used the results with a causal inference framework to give some guidance to change the probability of choosing every mode. Our results reasonably assessed various attributes that contributed to certain alternatives, with an emphasis on comparison between driving and non-driving alternative mode choices. Furthermore, we validated our results with a series of tests and expanded the results with causal inference analysis.

A few limitations were present in the study. First, the computational limitations resulted in the inability to utilize all the original data. Furthermore, we did not consider all households in the original data. Having better computational resources and considering a greater proportion of households (e.g. sampling one trip randomly from each household) would lead to more effective results. Second, the assumption on availability of all modes for each individual is rather unreasonable, especially for walking in long-distance trips. Adding in additional constraints to the model to account for availabilities effectively would yield more informative results. Last, there was a clear disproportion of the mode choices assessed from the trips. Processing the data in a way such that a more even distribution across all alternatives is in place would yield better results as well.

Despite the limitations in our study, the overall model proves to be effective in assessing travel behavior across a variety of attributes. In particular, the attributes related to travel times, car ownership, trip purposes, and congestion all impacted the utilities of alternatives significantly with respect to one another. Additionally, the causal inference showed that distance and travel cost could potentially impact the likelihood of selecting a particular mode choice. From these results, we can inform several key policies contributing towards more sustainable and collective mobility solutions. First, incentives towards using collective and emissions-free mobility solutions, such as subsidy programs and credit schemes for purpose-specific travel, would contribute favorably towards these options. Additionally, the increase in accessibility towards micro-mobility options, such as more e-bikes or more cycling lanes, would be a huge factor in bringing greater efficiency for these options. Greater accessibility for transit, such as with point-to-point transportation, would bring greater efficiency to transit as well and make this options more favorable. Lastly, greater awareness on sustainability and the deterrent of congestion towards sustainability initiatives would bring greater adoption towards non-driving mode choice alternatives.

REFERENCES

1. Pishue, B., Global Traffic Scorecard. *INRIX*, 2022.
2. Yurday, E., *Average CO2 emissions per car in the UK*, 2023.
3. Neugebauer, M., A. Żebrowski, and O. Esmer, Cumulative emissions of CO2 for Electric and combustion cars: A case study on specific models. *Energies*, Vol. 15, No. 7, 2022, p. 2703.
4. Hu, H., J. Xu, Q. Shen, F. Shi, and Y. Chen, Travel mode choices in small cities of China: A case study of changting. *Transportation Research Part D: Transport and Environment*, Vol. 59, 2018, p. 361–374.
5. Dissanayake, D. and T. Morikawa, Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference nested logit model: Case study in bangkok metropolitan region. *Journal of Transport Geography*, Vol. 18, No. 3, 2010, p. 402–410.
6. Bierlaire, M., K. Axhausen, and G. Abay, The acceptance of modal innovation: The case of Swissmetro. *1st Swiss Transport Research Conference*, 2001.
7. Fan, A., X. Chen, and T. Wan, How have travelers changed mode choices for first/last mile trips after the introduction of bicycle-sharing systems: An empirical study in Beijing, China. *Journal of Advanced Transportation*, Vol. 2019, 2019, p. 1–16.
8. Tzouras, P. G., L. Mitropoulos, K. Koliou, E. Stavropoulou, C. Karolemeas, E. Antoniou, A. Karaloulis, K. Mitropoulos, E. I. Vlahogianni, and K. Kepaptsoglou, Describing micromobility first/last-mile routing behavior in urban road networks through a novel modeling approach. *Sustainability*, Vol. 15, No. 4, 2023, p. 3095.
9. Eom, J. K., K.-S. Lee, and J. Lee, Exploring micromobility mode preferences for last-mile trips from subway stations. *Journal of Public Transportation*, Vol. 25, 2023, p. 100054.
10. Hillel, T., M. Z. Elshafie, and Y. Jin, Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction*, Vol. 171, No. 1, 2018, p. 29–42.