

LOGISTIC REGRESSION APPROACH IN STUDYING TRAVEL DEMAND DURING MORNING PEAK HOUR: A CASE STUDY IN KLANG KOMUTER STATION, MALAYSIA

Nur Fahriza Mohd Ali^{1,*}, Ahmad Farhan Mohd Sadullah², and Adnan Zulkiple³

¹ School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia

² Timbalan Naib Canselor Akademik & Antarabangsa, Pejabat Naib Canselor, Aras 6, Bangunan Canselor, 11800 USM, Universiti Sains Malaysia, Pulau Pinang, Malaysia

³ Faculty of Engineering Technology, Block A2, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia

ABSTRACT – This paper develops a Stated Preference (SP) experiment that provides a way to measure users' travel demand in public transport. This paper introduces an empirical procedure for optimising the SP experiment. This procedure permits the identification of the choice alternatives defining the experiment by simulating the choices of a user sample. By using the data collected from a survey, a Logit model was calibrated. This model is a way of identifying the importance of service quality attributes on global users' satisfaction which provides an operationally appealing measure of current or potential service effectiveness. The result shows that out-of-vehicle time gives a great impact towards users' mode choice

ARTICLE HISTORY

Received: 2nd Nov 2022

Revised: 30th Nov 2022

Accepted: 10th Dec 2022

KEYWORDS

Logit model

Public Transport

Stated Preference

Travel Demand

INTRODUCTION

Transportation plays a significant role in daily life as it associates individuals starting with one place then onto the next. The evolution of urban transportation guarantees that users' activity in that area will be increasingly proficient and their daily movement to reach destination will be quicker. A proficient transport system is a fundamental and capital equipment needed for the functioning of a country or area that is required for an industrial economy to function.

The upcoming infrastructure project by government is the multi-storey Park and Ride at the Klang Komuter Station. The Park and Ride facilities act as a parking lots with public transport connections that allow users headed to city centers by leaving their private vehicles and transfer to the provided Komuter. Park and Ride used to be located in the suburbs or on the outer edges of towns in order to avoid congestion within town centre. However, problem arise when the provision of the Park and Ride at the Klang Town attracts more users to come to the town by their private vehicles causing massive congestion especially during morning peak hour. Therefore, instead of persuading users to leave their cars at home, the service quality of provided public transport should be investigated; thus users will be encouraged to commute by using public transport.

A journey always consists of in-vehicle time and out-of-vehicle time as well as cost for both private vehicles and public transport. However, the users behaviour of public transport is usually affected by out-of-vehicle time. For the reason that during out-of-vehicle time, users need to abide for the public transport disembarkation and walking over few distances to reach the stop or towards end destination.

During morning peak hour, it is important for users to be able to reach their workplace punctually; therefore, lowering the travel time while travelling by public transport is necessary. According to [1], [2], and [3] travel time is the great factor that must be bring out in attracting users towards public transport. Users' travel behaviour is crucial to be explored; therefore, service provider can improve the service quality of public transport in term of in-vehicle and out-of-vehicle aspects that in line with users' travel demand

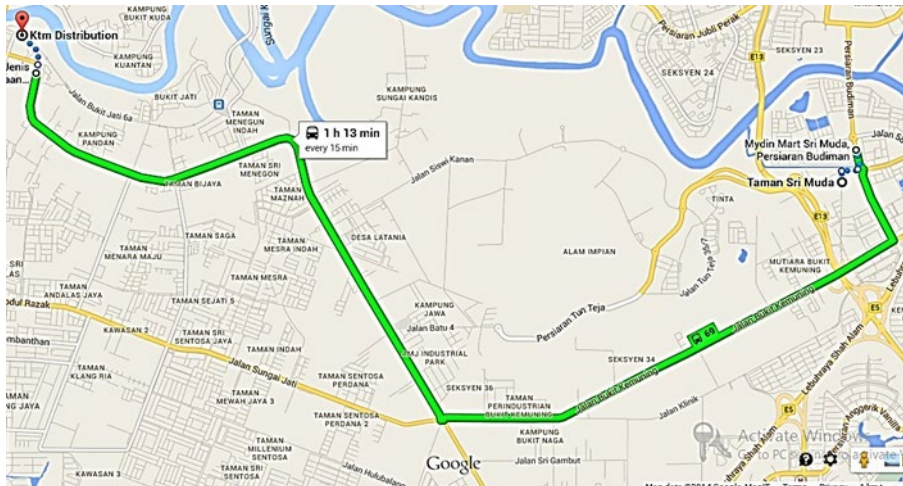


Figure 1. The route map from the origin (Taman Sri Muda) to destination (Klang Komuter Station)

LITERATURE REVIEW

The correlation between modal split and travel demand

The modal split in transportation planning refers to the estimation of possible volume of travel in a particular mode; strictly speaking, the share of certain transport modes in total travel demands. Travel demand can be explained as a prediction on what travel decisions users would like to make given the generalized travel cost of each mode of transportation. Whilst transportation demand can be described as the amount and type of travel users' would choose under specific conditions, taking account factors such as travel time, walking distance to reach the transport' stop, the quality of transport options available and their prices. In other words, the modal split is the repercussion of users' travel demand.

The modal split is a component of the transport system improvement procedure for the zone under thought. A few different ways to identify the modal split can be utilized, yet depending on research goals, level and character, on information availability, and so on. Models can be aggregate or disaggregate, determined by the basic unit under thought is a spatial unit or an individual user. Binary and multimodal models can be specified distinguished bestow to the number of travel modes.

A transport system user's choice of the mode of transport is influenced by a few distinct components which can be classified into three groups including travel characteristic, passenger characteristic, and transport system characteristic [4]. The primary travel characteristic influencing a transport system user's choice of travel mode are trip purpose, distance, duration, orientation in space, and so forth. For instance, the modal split of work trips is significantly contrary from that of shopping trips. These trips could be classified into three types which are the radial-type trips directed towards a centre, trips from one residential area to another, or trips from residential to work zones. The modal split is likewise influenced by the socio-economic attributes of an individual or household. These attributes including income, the number of cars, family size, the number of employed, age, and education structure.

A standout amongst the most significant variables of the modal split is the motorization degree. This element discovers the size of the population having the decision between a car and public transport as opposed to the individuals who have no vehicle and must utilize public transport. Population density is likewise one of the components utilized in breaking down and gauging the modal split. The trips percentage by public transport reduces with decreasing population density. The low-density areas are barely to be served by public transport at an adequate service level since it is unprofitable.

Travel time and travel cost are the most broadly utilized measures for comparing the productivity or usefulness of transport viz. car or public transport. The consequences of these comparisons are expressed either as travel time or travel cost ratios. The travel time by public transport involves the driving time, walking times at the trip source (origin) and towards destination points, delay time, and transfer time (in-vehicle time). In the meantime, travel time via car is described as driving time with the extensive terminal time at trip destinations which depends on parking conditions [4]. Therefore, while studying travel demand, it is also important to understand user behaviour (such as the travel mode choice) which must consider both

the real aspects (including economic, social, psychological as well as the transport system components) faced by an individual and the characteristics of his lifestyle (household, employment) that influence how user organizes his time and environment.

The importance of aggregate approach to transportation planning

This study applied an aggregate approach to travel demand modelling. An aggregate model is a model estimated with a dependent variable which represents a group of observations. As subsequent derivations to the above definitions, aggregate data are those which are usually assembled on a collective scale, or in the case of travel demand, aggregate data may generally be grouped into traffic zones. Data are normally conveniently collected at disaggregate level, usually at an individual or household stage. In the conventional method of demand analysis, these disaggregated data are then aggregated geographically using a traffic zone system, and travel demand models are then estimated using these aggregated and averaged data. These aggregations would obviously disturb the actual representation of an individual behaviour. Therefore, a model estimated directly from the lowest possible behavioural unit would represent more accurately the behaviour of an individual unit [5].

An aggregate model also imposes inaccuracies as far as zonal aggregation is concerned. To model accurately based on a zonal aggregation system implies that a high degree of homogeneity prevails in the zoning system, which is unlikely to be true. Therefore, a disaggregate model which may also perform some form of aggregation but with more homogeneous groupings would provide a better representation.

A model estimated at the disaggregate level however, requires a transformation to be applied as an aggregate model for prediction exercises. Disaggregate modelling based on discrete choice analysis has been developed to improve on the shortcomings of the conventional methods of travel demand modelling. Disaggregate modelling, which attempts to represent the behaviour of the lowest behavioural unit, has allowed travel demand predictions to be based on the real demand of the community under study, and thus promises more accurate forecasts. The results of disaggregate modelling are of little value if they are not applicable to real life decision making situations, which are usually at an aggregate level. Therefore, having predicted the results at the disaggregate level, a technique of transforming these results to the more applicable aggregate level must be devised.

Intrazonal trips does not take into account when a modelling was made by using aggregate approach. There are numerous transportation studies that treat huge zones yet the model excluded the intrazonal trips although in fact the flows are relatively large. The spatial boundaries established by planners based on their assumptions indeed different from users' actual behaviour. The constraint of the aggregation-based models are they produce great differences of features which included in the model even for the behaviour of homogeneous groups. These models might be incorrect since they do not apply the available data about households that have been collected by surveys as well as they do not clarify individual user behaviour well enough.

METHODOLOGY

The respondents for Revealed Preference/Stated Preference (RP/SP) Survey is comprises of clients from Zone 3 (origin); which the highest number of Klang Komuter users came from this zone. The total respondents that have been interviewed for the RP/SP Survey are 252 clients. The total number of female users (142 clients) at the Klang Komuter Station is higher than male users (110 clients). The survey was conducted in weekdays, during morning peak hour starting from 7 AM until 9 AM.

Research Methodology Flow Chart

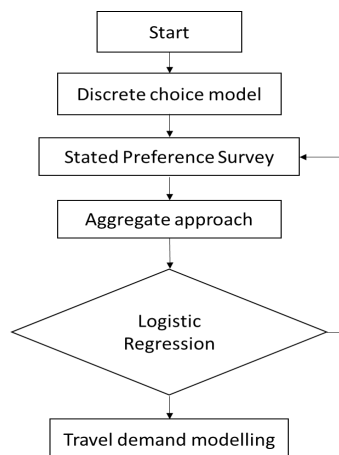


Figure 2. Flow chart of data collection.

Stated Preference Survey

Both revealed preference (RP) and stated preference (SP) applied choice modelling in the studies. RP studies utilize the decisions made by users to evaluate as of now by people to evaluate the qualities (utilities) they attribute to things which they uncover their inclinations by their decisions.

Choice modelling is used in both revealed preference (RP) and stated preference (SP) studies. RP studies use the choices made already by individuals to estimate the values (utilities) they ascribe to items which they reveal their preferences by their choices. SP studies use the choices made by individuals made under experimental conditions to estimate these values which they state their preferences via their choices. The use of SP surveys has become widespread in a range of fields such as marketing, transport, health economics, and agricultural and environmental economics [6]. SP methods have been widely used in the transport field since the 1980s [7], and can be effectively applied to estimate mode choice; as they measure perceptions and attitudes.

RP studies utilize the decisions made as of now by people to evaluate the qualities (utilities) they attribute to things which they uncover their inclinations by their decisions. SP studies utilize the decisions made by people made under trial conditions to assess these qualities which they express their inclinations through their decisions. The utilization of SP studies has turned out to be across the board in a scope of fields, for example, promoting, transport, wellbeing financial matters, and rural and natural financial aspects [6]. SP strategies have been generally utilized in the vehicle field since the 1980s [6] and [7], and can be successfully connected to gauge mode decision; as they measure recognitions and demeanors.

Stated Preference Method also can be defined as family of techniques which use individual respondents' statements about their preferences in a set of transport options to estimate utility functions. The options are typically descriptions of transport situations or contexts constructed by the researcher. Stated Preference Methods were originally developed in marketing research in the early 1970s, and have become widely used since 1978. The use of Stated Preference Methods has become an attractive option in transport research specially to evaluate demand under conditions which do not yet exist. Broadly, these methods are easier to control because the researcher defines the conditions which are being evaluated by the respondents. Besides, they are more flexible since researchers being capable of dealing with a wider variety of variables; and they are cheaper to apply as each respondent provides multiple observations for variations in the explanatory variables which interest the analyst.

The important issue in the design of stated preference experiments is the choice of the context of the experiment and the measurement scale for the dependent variable. "Traditional" stated preference methods provide respondents with a set of descriptions of alternatives and ask them to express their preferences by sorting the alternatives in decreasing order of preferences, or by giving a rating value for each. In the more recently developed choice experiments, respondents are offered combinations of a few alternatives (typically two to five) and are asked to express their choices either by indicating one chosen alternative or by assigning subjective choice probabilities to each of the alternatives.

An important issue in the use of stated preference methods is the quality of the survey and the context in which the survey questions are asked. It is the view of the authors that, useful results are to be obtained from stated preference questions are asked should be as realistic as possible. For this reason, the authors have a strong preference for face-to face interviews, conducted by experienced interviewers and also

structured to ensure that the background to the respondent’s evaluation process (for example, situational constraints, demographic characteristics, planning process) is fully understood by the researcher.

Table 1. Presentation of option.

Option set	Travel Time	Cost	Walking distance from home to the 1 st stop (WD1)	Waiting time at the stop (WT)	Walking distance from last stop to the end destination (WD2)
1	30 minutes	2	200 meters	10 minutes	400 meters
2	30 minutes	4	200 meters	10 minutes	200 meters
3	30 minutes	6	400 meters	20 minutes	200 meters

In this research, users were given few options and they need to choose their preferred options on transportation services; as shown in Table 1. For a choice based approach, the appraisal relates to a choice decision between alternative available choices presented within each designed option. The respondents would be asked to select her/his preferred choice (cost, WD1, WT, and WD2) for a particular journey from option set 1, and then from option set 2, and so on.

Though face-to-face interviews are preferable, this does not mean that other survey methods cannot be used. Nowadays, internet survey has been used widely used for stated preference methods which provided the task for the respondents is fairly easy and well explained. For straightforward choice tasks, for instance, they appear to give good results. An important disadvantage of internet surveys; however, is that they do not allow the stated preference alternatives to be specifically adapted to each respondent’s situation. Face-to-face interviews allow for that flexibility.

RESULTS

Technique of analysis

Logistic regression, also called a logit model, is used to model dichotomous outcome variables. In the logit model the log odds of the outcome is modeled as a linear combination of the predictor variables. Below we use the logistic regression command to run a model predicting the outcome variable mode choice, using cost, walking distance, and waiting time. The output is shown in sections, each of which is discussed below.

A binomial logistic regression (often referred to simply as logistic regression), predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical. If, on the other hand, your dependent variable is a count, see our Poisson regression guide. Alternatively, if you have more than two categories of the dependent variable, see our multinomial logistic regression guide.

Table 2. Case Processing Summary

		N	Percent
Unweighted Cases ^a			
Selected Cases	Included in Analysis	4826	100.0
	Missing Cases	0	.0
	Total	4826	100.0
Unselected Cases		0	.0
Total		4826	100.0

a. If weight is in effect, see classification table for the total number of cases.

Table 3. Dependent Variable Encoding

Original Value	Internal Value
N	0
P	1

The Table 2 shows a breakdown of the number of cases used and not used in the analysis. The Table 3 gives the coding for the outcome variable, mode choice.

Table 4. Classification Table^{a,b}

Observed		Predicted		Percentage Correct
		RES		
RES	N	4343	0	100.0
	P	483	0	.0
Overall Percentage				90.0

- a. Constant is included in the model.
- b. The cut value is .500

Table 5. Variable in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Constant	-2.196	.048	2096.692	1	.000	.111

Table 6. Variables not in equation

Variables		Score	df	Sig.
	Cost	7.683	1	.006
	WD1	188.215	1	.000
	WT	504.426	1	.000
	WD2	93.285	1	.000
Overall Statistic		799.253	4	.000

- The first model in the output is a null model, that is, a model with no predictors.
- The constant in the Table 5 gives the unconditional log odds of mode choice (i.e., P=1).
- The Table 6 gives the results of a score test, also known as a Lagrange multiplier test. The column labeled Score gives the estimated change in model fit if the term is added to the model, the other two columns give the degrees of freedom, and p-value (labeled Sig.) for the estimated change. Based on the Table 6 above, all four of the predictors, cost, walking distance from origin, waiting time, and walking distance from end stop, are expected to improve the fit of the model.

Table 7. Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step	955.849	4	.000
Block	955.849	4	.000
Model	955.849	4	.000

Table 8. Model Summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
2183.609 ^a	.180	.376

- a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.
- Table 7 gives the overall test for the model that includes the predictors. The chi-square value of 955.849 with a p-value of less than 0.0005 tells us that our model fits significantly better than an empty model (i.e., a model with no predictors).
- The -2*log likelihood (2183.609) in Table 8 can be used in comparisons of nested models. This table also gives two measures of pseudo-R-square.
- The Cox & Snell R Square and the Nagelkerke R Square values provide an indication of the amount of variation in the dependent variable explained by the model (from a minimum value of 0 to a maximum of approximately 1). These are described as pseudo-R square statistics, rather than the true R square values that will be seen provided in the multiple regression output.
- In this example, the two values are .180 and .376, suggesting that between 18 % and 37.6 % of the variability is explained by this set of variables.

Table 9. Classification Table^a

		Predicted		
		RES		
Observed		N	P	Percentage Correct
RES	N	4278	65	98.5
	P	364	119	24.6
Overall Percentage				91.1

a. The cut value is .500

Table 10. Variables in the equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Cost	-.054	.016	10.872	1	.001	.948
WD1	-.003	.000	201.594	1	.000	.997
WT	-.149	.008	381.568	1	.000	.862
WD2	-.002	.000	111.811	1	.000	.998
Constant	3.865	.255	229.537	1	.000	47.689

a. Variable(s) entered on step 1: Cost, WD1, WT, WD2.

- In the Table 10 we see the coefficients, their standard errors, the Wald test statistic with associated degrees of freedom and p-values, and the exponentiated coefficient (also known as an odds ratio). All variables are statistically significant. The logistic regression coefficients give the change in the log odds of the outcome for a one unit increase in the predictor variable.
- For every one unit change in cost, the log odds of mode choice for public transport (versus non-public transport) reduces by 0.054.
- For a one-unit increase in walking distance from origin (WD1), the log odds of public transport being chosen as users' mode choice reduces by 0.003.
- For a one-unit increase in waiting time (WT), the log odds of public transport being chosen as users' mode choice reduces by 0.149.
- Meanwhile, for a one-unit increase in walking distance from end stop (WD2), the log odds of public transport being chosen as users' mode choice reduces by 0.002.

The Wald test in Table 10 is similar to the Logistic Regression test but here it is used to test the hypothesis that each $\beta = 0$. In the sig column, the p-values are all below 0.05. When interpreting the differences for each variable, look at the exp (β) column which represents the odds ratio for the individual variable. For example, WD2 were 0.998 times more likely to influence user's preferences towards mode choice. Then, the odds ratio is followed by WD1 (0.997), COST (0.948), and WT (0.862). Many literatures show that WT is the most influential towards user's mode choice, however, this study shows that walking distance from end stop to Klang Komuter (egress) and walking distance to reach to nearest stop (access) highly potential in influencing users' mode choice.

The full mode being tested is:

$$\ln \frac{p}{1-p} = 3.865 - 0.054x_{\text{COST}} - 0.003x_{\text{WD1}} - 0.149x_{\text{WT}} - 0.002x_{\text{WD2}} \tag{1}$$

DISCUSSION

Logistic Regression is a function approximation algorithm that uses training data to directly estimate $P(Y|X)$. In this sense, Logistic Regression is often referred to as a discriminative classifier because we can view the distribution $P(Y|X)$ as directly discriminating the value of the target value Y for any given instance X. Logistic Regression is a linear classifier over X. The result from the Logistic Regression gives us the probability in cumulative perspective for the respective research area. Based on the odds ratio, WD2 were 0.998 times more likely to influence user's preferences towards mode choice. Then, the odds ratio is followed by WD1 (0.997), COST (0.948), and WT (0.862).

The result proves that out-of-vehicle time affected the most towards users' mode choice. The location of the multi-storey park and ride which is very near to the Klang Komuter encourage users to bring along their car instead of riding bus. Besides, the hot weather condition in Malaysia causes most users refuse to walk under the sun. Since users need to walk for few distances from the end stop towards the Klang Komuter, it is advisable to provide pedestrian walkway with shade in order to encourage users to walk from the end stop to the Klang Komuter.

In order to improve the quality of service in public transport, effective road-based also necessary. Effective road-based public transport is central to economic growth of developing cities. For the majority of residents, road-based public transport (bus and paratransit) is the only means to access employment, education, and public services. In medium and large developing cities, such destinations are beyond viable walking and cycling distances while vast numbers of

individuals have limited access to automobiles. Unfortunately, the current state of road-based public transport services in many developing cities does not serve the mobility needs of the population adequately. Formal bus services are often unreliable, inconvenient, uncomfortable, or even dangerous. Informal paratransit services, while providing benefits including on-demand mobility for the transit-dependent, jobs for low-skilled workers, and service coverage in areas devoid of formal transit supply, carry major costs, such as increased traffic congestion, air and noise pollution, traffic accidents, and even violence among route cartels [8].

In recent decades, the creation of bus lanes on existing roads (painting of a lane in a different color from the rest of the asphalt) has been a common low-cost strategy for improving the quality of bus systems throughout the world. New technologies allow vehicles in bus lanes to gain priority at intersections, with lights automatically turning red for cars and green for buses whenever the later approach shared intersections. Here, a more effective intervention in favor of public transport is the construction of busways that are physically segregated from other traffic by means of barriers, cones, or other well defined physical features. Located on the curb or in the median of a roadway, they are permanently and exclusively for the use of public transport vehicles although emergency vehicles are often allowed to use the lane.

Other factors that influence users' mode choice also should be discovered. Users are more likely to stop using public transit when they experience delays they can blame on the transit agency, according to researchers at the University of California Berkeley [9]. They are more likely to forgive delays caused by traffic, emergencies or mechanical failures. The most significant negative experiences that drove a reduction in transit use were delays perceived to be the fault of the transit agency, long waits at transfer points, and being prevented from boarding due to crowding.

Researchers found that passengers care about much more than just when the bus arrives; a factor traditionally considered to influence perceptions of reliability. Passengers care about the types of delays they endure and when in the trip they occur. For example, passengers are more likely to be angered by delays at a transfer stop than an origin stop, where they first board the bus or train. At an origin stop passengers may be able to wait at home, and if they have access to real-time arrival information, they may be able to consider alternative ways to travel. The worst experience of the bus riders is when people get stuck inside the bus.. The top reasons people give up on public transit, according to the researchers:

1. Delayed on board due to transit vehicles backed up or problems on the transit route downstream.
2. Experienced long wait at a transfer stop.
3. Missed departure due to wrong real-time information.
4. Unable to board or denied boarding due to crowding.
5. Delayed on board due to emergency or mechanical failure.
6. Experienced long wait at origin stop.
7. Ran to stop but the bus or train pulled away.
8. Delayed on board due to traffic.

The researchers found that comfort is the least important factor influencing decisions to stop using public transit. Riders don't mind standing in crowded buses or trains as long as the vehicles move without delay and run frequently. Commuters are willing to wait 10.2 minutes, on average, before they consider a wait too long, the study found. When holding up buses for schedule reasons, hold up the empty buses, not the full ones. Small vehicles running more often may be better than large vehicles running less often. Pay particular attention to the wait times for transfers. When buses and trains are delayed, let passengers know what's happening, especially if the delay is not the transit agency's fault.

CONCLUSION

It is therefore the aim of this study to gain a better understanding of the value that users place on different aspects of public transport service quality; for both in-vehicle and out-of-vehicle aspects to estimate the percentage value of each parameters that affect users' preferences, so that alternative public transport investment scenarios can be more accurately appraised. Revealed Preference (RP) refers to observations of actual behaviour, for example the mode choices that decision-makers currently make or made in the past, and Stated Preference (SP) refers to observations of hypothetical behaviour under controlled experimental conditions were applied which successfully provide the information on users' travel demand that is important in transportation planning.

REFERENCES

- [1] L. Eboli and G. Mazzulla, "A Stated Preference Experiment for Measuring Service Quality in Public Transport," <http://dx.doi.org/10.1080/03081060802364471>, vol. 31, no. 5, pp. 509–523, 2008.
- [2] J. Walker, "How Flexible Should Bus Service Be? Frequent Networks As a Tool For Permanent Change," Jan. 2009.
- [3] J. Thøgersen, "Promoting public transport as a subscription service: Effects of a free month travel card," *Transp. Policy*, vol. 16, no. 6, pp. 335–343, Nov. 2009.
- [4] J. J. Jovi] and M. M. Popovi, "Modal Split Modelling Using Multicriteria Analysis and Discrete Fuzzy Sets," *Yugosl. J. Oper. Res.*, vol. 11, pp. 221–233, 2001.
- [5] E. Cherchi and D. A. Hensher, "Workshop Synthesis: Stated Preference Surveys and Experimental Design, an Audit of the Journey so far and Future Research Perspectives," *Transp. Res. Procedia*, vol. 11, pp. 154–164, Jan. 2015.
- [6] J. Bates, "Econometric Issues In Stated Preference Analysis," *J. Transp. Econ. Policy*, vol. 22, no. 1, Jan. 1988.
- [7] D. A. Hensher, "Stated preference analysis of travel choices: the state of practice," *Transp. 1994 212*, vol. 21, no. 2, pp. 107–

133, May 1994.

- [8] R. Cervero and A. Golub, "Informal transport: A global perspective," *Transp. Policy*, vol. 14, no. 6, pp. 445–457, Nov. 2007.
- [9] J. McMahon, "Top Eight Reasons People Give Up On Public Transit," *www.forbes.com*, 2013. [Online]. Available: <https://www.forbes.com/sites/jeffmcmahon/2013/03/06/top-eight-reasons-people-give-up-on-public-transit/?sh=36f6c7cbfd84>. [Accessed: 17-Dec-2022].