ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

Resident Travel Mode Selection based on Logit Model

Yingjie Feng

College of Transport & Communications, Shanghai Maritime University, Shanghai 201306, China.

2831399049@qq.com

Abstract

In order to study the factors that influence travelers' choice of car and public transportation, according to the survey report on the travel mode of Swiss residents, a model of input related to the travel mode selection of Swiss residents was established. It also analyzes the factors that affect residents' travel mode choices, and puts forward relevant suggestions that can attract residents to make corresponding travel choices, and provide reference for attracting more travelers to choose public transportation.

Keywords

Travel Mode Choice; Logit Model; Public Transportation.

1. Introduction

A person's choice of travel mode during a trip is affected by many factors, such as the person's gender, age, travel distance, purpose, time and expense required for different travel modes, etc. According to the theory of maximum utility, during a trip, travelers always choose the most effective way for themselves according to specific conditions among the possible modes of travel, and the travel utility of different plans is determined by each plan, determined by various influencing factors.

2. Literature review

In recent years, with the development of our country's society and economy, private transportation has continued to increase, and the contradiction between the increasing demand for transportation and the shortage of road space resources has become increasingly prominent. It is the focus of scholars at home and abroad to propose transportation development policy recommendations by analyzing the choice behavior of travelers and their influencing factors.

Chen analyzed the classification of urban residents' passenger transportation modes and the factors affecting their choices. Based on the AHP model, he explained the internal mechanism of the current urban residents' choice of passenger transportation modes [1]. Wang analyzed the characteristics of travelers, travel characteristics, and traffic mode characteristics. Based on the results of the correlation analysis of influencing factors, considering the operability of the investigation and analysis, the main influencing factors are screened out [2]. Yan and Meng studied the changes brought about by the opening of rail transit to residents' travel modes and the factors affecting the choice of residents' travel modes [3]. Bo and Teng proposed a prediction method for public transportation passengers' travel choice behavior under the conditions of meteorological disasters, and provided theoretical support for the formulation of emergency warning plans for public transportation management departments [4]. Liang et al. studied the travel choice behavior of residents under the synergistic effect of multiple factors, through the establishment of multiple logistic regression models to analyze the factors that have a significant impact on travelers' choice of transportation mode, and quantitatively display the degree of influence of each influencing factor and the rate of change in the choice intention of various

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

travelers [5]. Liu and Luo used Decision Tree and Support Vector Machine (SVM) models in machine learning methods to compare and analyze multiple Logit models to find the most accurate model for the choice of residents' travel mode [6]. Celikoglu found that the ANNs neural network is very effective for the calibration of the utility function in the travel choice model [7]. Espino et al. used the nested Logit model (NL) to estimate the parameters of the survey data of RP and SP, and obtained the factors that influence the choice of bus travel and private car travel [8]. Joseph F. Wyer established an ECL model based on the travel data of travelers, and verified that the increase in the complexity of the transportation mode will cause travelers to shift to cars [9].

3. Theoretical overview

3.1 Influencing factors of transportation

The transportation modes of the comprehensive transportation plan are divided into: railway transportation, road transportation, water transportation, air transportation, and pipeline transportation. The transportation modes in urban transportation planning are divided into: walking, bus, subway, taxi, bicycle, motorcycle, private car, etc.

Generally speaking, the factors that affect the choice of residents' travel modes can be summarized into three categories: characteristics of travelers, characteristics of travel, and characteristics of transportation modes.

Traveler characteristics include traveler's age, gender, occupation, traveler's family income, vehicle ownership, etc.

Resident travel characteristics mainly include two aspects: travel purpose and travel distance.

The mode of transportation refers to the means of transportation used by travelers from the start point to the end point. With the advancement of science and technology and the improvement of people's living standards, people have more and more choices of transportation.

3.2 Multivariate logit model

When making decisions, people often encounter many selection problems, that is, to choose among multiple options. When faced with different choices, decision makers will choose the most beneficial plan based on their own situation and feelings. The multivariate Logit model is a model that describes the decision maker to choose a certain option from a set of alternatives. Its important assumption is that independent individuals have a clear perception of the choices they are facing, and the decision maker chooses the plan based on the principle of maximizing utility. Since the utility depends on the subjective feelings of the user and is different from the actual situation, the value of the utility includes two parts: the measurable utility and the unmeasured error value.

The utility U_{ij} obtained by the decision maker i choosing option j can be expressed as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{1}$$

Where, V_{ij} is the measurable part of the utility; ε_{ij} is the unmeasured error term. To simplify the calculation, the utility function is usually set as a linear function:

$$U_{ij} = \beta_i X_{ij} + \varepsilon_{ij} \tag{2}$$

Where, X_{ij} is the explanatory variable vector of the decision-maker i's choice of option j; β_i is the parameter vector to be estimated. If the alternatives are independent of each other, that is, the error terms are independent and obey the same Gumbel distribution. Then the probability p_{ij} that the decision maker i chooses option j is:

$$p_{ij} = \frac{e^{V_{ij}}}{\sum_{j=1}^{J} e^{V_{ij}}} \tag{3}$$

For the unknown parameters of the multivariate Logit model, the maximum likelihood method can be used to estimate. Define variable f_{ij} as the situation of the decision-maker's choice of plan:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

$$f_{ij} = \begin{cases} 1, & decision\ maker\ i\ chooses\ option\ j \\ 0, & else \end{cases} \tag{4}$$

Since the decision maker can only choose one of the alternatives, the likelihood function of the multivariate Logit model can be derived:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} p_{ij} \tag{5}$$

Convert formula (5) into logarithmic function form:

$$ln\, L = \sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij}\, lnp_{ij} \tag{6}$$

The value of the parameter vector can be obtained by maximizing $ln\, L$.

4. Model construction

4.1 Case introduction

This case uses revealed preference data to evaluate the model choice behavior of Swiss residents. The survey was carried out by CarPostal, the public transport branch that serves the Swiss Post. The main purpose of this survey is to collect data to analyze the travel behavior of people in low-density areas. The next study proposed new alternatives to public transportation based on respondents' willingness to pay for these potential services to increase the market share of public transportation.

4.2 Data

The survey covers the French and German-speaking regions of Switzerland. Questions and answers are sent by mail to people living in rural areas. Respondents were asked to register all trips during the day. The information collected includes origin, destination, cost, travel time, choice of methods and destination activities. In addition, socio-economic information of the respondents and their families was also collected.

Collected 1124 completed surveys. For each respondent, a cyclic stroke sequence (start and end at the same location) is detected, and its main mode of transportation is determined. The resulting database includes 1,906 travel sequences related to the interviewee's psychological indicators and socio-economic attributes. It should be noted that each observation is a series of trips that begin and end at home. A respondent may have several consecutive trips a day.

4.3 Variables

TimePT: The duration of the loop performed in public transport (in minutes).

TimeCar: The total duration of a loop made using the car (in min-utes).

MarginalCostPT: The total cost of a loop performed in public transports, taking into account the ownership of a seasonal ticket by the respondent. If the respondent has a "GA" (full Swiss season ticket), a seasonal ticket for the line or the area, this variable takes value zero. If the respondent has a half-fare travelcard, this variable corresponds to half the cost of the trip by public transport.

CostCar: The total gas cost of a loop performed with the car in euros.

TripPurpose: The main purpose of the loop: 1 =Work-related trips; 2 =Work- and leisure-related trips; 3 =Leisure related trips. -1 represents missing values

distance_km: Total distance performed for the loop.

NbChild: Number of kids (< 15) in the household. -1 for missing value.

NbCar: Number of cars in the household.-1 for missing value.

NbBicy: Number of bikes in the household. -1 for missing value.

ModeToSchool: Most often mode used by the respondent to go to school as a kid (> 10), 1 is car (passenger), 2 is train, 3 is public transport, 4 is walking, 5 is biking, 6 is motorbike, 7 is other, 8 is multiple modes, -1 is for missing data and -2 if respondent didn't answer to any opinion questions.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

FreqCarPar: Frequency of the usage of car by the respondent’s parents (or adults in charge) during childhood (< 18), 1 is never, 2 is occasionally, 3 is regularly, 4 is exclusively, -1 is for miss-ing data and -2 if respondent didn’t answer to any opinion questions.

FreqTrainPar: Frequency of the usage of train by the respondent’s parents (or adults in charge) during childhood (< 18), 1 is never, 2 is occasionally, 3 is regularly, 4 is exclusively, -1 is for miss-ing data and -2 if respondent didn’t answer to any opinion questions.

FreqOthPar: Frequency of the usage of tram, bus and other public trans-port (not train) by the respondent’s parents (or adults in charge) during childhood (< 18), 1 is never, 2 is occasion-ally, 3 is regularly, 4 is exclusively, -1 is for missing data and -2 if respondent didn’t answer to any opinion questions.

4.4 Model Construction

Select the travel purpose related to work (TripPurpose=1) and work-entertainment (TripPurpose=2) data, and eliminate invalid data in the survey:

TripPurpose=1 or 2

ModeToSchool, NbChild, NbCar, NbBicy = -1 missing data

FreqCarPar, FreqTrainPar, FreqOthPar = -1 or -2

missing data or respondent didn't answer to any opinion questions.

Table 1. Related parameters of different modes

Parameter	V_{PMM}	V_{PT}	V_{SM}
ASC_{CAR}	1	-	-
ASC_{PT}	-	-	-
ASC_{SM}	-	-	1
β_{cost}	CostCar	MarginalCostPT	-
β_{time}	TimeCar	TimePT	-
$\beta_{distance}$	-	-	distance_km
$\beta_{nbchild}$	NbChild	-	NbChild
β_{nbcar}	NbCar	-	-
β_{nbbicy}	-	-	NbBicy
$\beta_{freqcarpar}$	FreqCarPar = 3 or 4	-	-
$\beta_{freqtrainpa}$	-	FreqTrainPar = 3 or 4	-
$\beta_{freqothpar}$	-	-	FreqOthPar = 3 or 4
$\beta_{modecar}$	ModeToSchool=1	-	-
β_{modept}	-	ModeToSchool=2 or 3	-
β_{modesm}	-	-	ModeToSchool=4 or 5

There are three options for this model:

(1) Private Motorized Modes (PMM), including cars, motorcycles and taxis,

(2) Public Transport (PT), including buses, trains and car postal, and

(3) Slow Modes (SM), including walking and bike.

According to the relevant factors listed in Table 1, write the corresponding effect function, establish the basic base model, and name it final test base.py.

After that, based on the final test base.py, consider the piecewise linear model of distance, divide distance into three segments, A, B, and C, and build a model distance_piecewise.py:

A: 0-30km

B: 30-90km

C: >90km

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

In addition, on the basis of final test base.py, considering the influence of the proportion of walking and waiting time in travel time on behavior selection, consider the following time:

LowTime: WalkingTimePT+WaitingTimePT≤40min

HighTime: WalkingTimePT+WaitingTimePT>40min

Finally, based on the final test base.py, do non-generic processing for NbChild to see if the result is significant.

5. Result analysis

5.1 The Base Model

The effect function of the basic model:

$$V_{CAR,n} = ASC_{CAR} + \beta_{cost} \cdot CostCar + \beta_{time} \cdot TimeCar + \beta_{nbchild} \cdot NbChild + \beta_{nbcar} \cdot NbCar$$
$$+\beta_{freqcarpar} \cdot FreqCarPar + \beta_{modecar} \cdot ModeToSchoolCar$$

$$V_{PT,n} = 0 + \beta_{cost} \cdot MarginalCostPT + \beta_{time} \cdot TimePT + \beta_{freqtrainpar} \cdot FreqTrainPar$$
$$+\beta_{modePT} \cdot ModeToSchoolPT$$

$$V_{SM,n} = ASC_{SM} + \beta_{distance} \cdot distance\ km + \beta_{nbchild} \cdot NbChild + \beta_{nbbicy} \cdot NbBicy$$
$$+\beta_{freqothpar} \cdot FreqOthPar + \beta_{modesm} \cdot ModeToSchoolSM$$

The results of the Base model in Biogeme are as follows:

Estimation report

Number of estimated parameters:	14
Sample size:	1906
Excluded observations:	359
Init log likelihood:	-2093.955
Final log likelihood:	-1247.516
Likelihood ratio test for the init. model:	1692.878
Rho-square for the init. model:	0.404
Rho-square-bar for the init. model:	0.398
Akaike Information Criterion:	2523.032
Bayesian Information Criterion:	2600.770
Final gradient norm:	+5.859e-003
Diagnostic:	Trust region algorithm with simple bounds (CGT2000): C
Iterations:	13
Data processing time:	00:00
Run time:	00:02
Nbr of threads:	2

Figure 1. The results

It can be seen from the above data that the effect of the model is relatively ideal and can be used as a reference. The relevant results such as the p-value of the estimated parameter are as follows: The relevant results such as the p-value of the estimated parameters are shown in Table 2.

It can be seen that the frequency of the respondents' parents using trams and other public transportation (non-trains) in childhood, and the frequency of using cars or slow transportation by their parents in childhood are not significant for their choice of travel mode.

The time, distance, fuel consumption cost of cars and slow traffic, respondents' most frequent use of cars or slow transportation during childhood, and the frequency of respondents' parents using trams and other public transportation (not trains) during childhood all have a negative impact on their choice of travel mode.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202106_7(6).0055

Table 2. Results in Biogeme

Name	Value	Std err	t-test	p-value	
ASC_CAR	-0.408	0.128	-3.20	0.00	
ASC_SM	-0.566	0.282	-2.01	0.04	
B_COST	-0.0602	0.00721	-8.35	0.00	
B_DISTANCE	-0.184	0.0194	-9.50	0.00	
B_FREQCARPAR	0.251	0.114	2.20	0.03	
B_FREQOTHPAR	-0.248	0.283	-0.87	0.38	*
B_FREQTRAINPAR	0.324	0.151	2.15	0.03	
B_MODECAR	-0.971	0.554	-1.75	0.08	*
B_MODEPT	0.801	0.131	6.10	0.00	
B_MODESM	-0.00860	0.245	-0.04	0.97	*
B_NBBICY	0.291	0.0537	5.42	0.00	
B_NBCAR	0.660	0.0677	9.75	0.00	
B_NBCHILD	0.00587	0.0571	0.10	0.92	*
B_TIME	-0.00387	0.00124	-3.13	0.00	

Respondents used public transportation most often during childhood; the frequency of using cars and trains by their parents during childhood; the number of bicycles, cars, and children in the family all had a positive influence on their choice of travel mode.

5.2 The Distance Segmented Model

The effect function of the Distance segmented model is:

$$V_{CAR,n} = ASC_{CAR} + \beta_{cost} \cdot CostCar + \beta_{time} \cdot TimeCar + \beta_{nbchild} \cdot NbChild + \beta_{nbcar} \cdot NbCar$$
$$+\beta_{freqcarpar} \cdot FreqCarPar + \beta_{modecar} \cdot ModeToSchoolCar$$
$$V_{PT,n} = 0 + \beta_{cost} \cdot MarginalCostPT + \beta_{time} \cdot TimePT + \beta_{freqtrainpar} \cdot FreqTrainPar$$
$$+\beta_{modePT} \cdot ModeToSchoolPT$$
$$V_{SM,n} = ASC_{SM} + \beta_{distanceA} \cdot distanceA + \beta_{distanceB} \cdot distanceB$$
$$+\beta_{distanceC} \cdot distanceC + \beta_{nbchild} \cdot NbChild + \beta_{nbbicy} \cdot NbBicy$$
$$+\beta_{freqothpar} \cdot FreqOthPar + \beta_{modesm} \cdot ModeToSchoolSM$$

The running results of the Distance segmentation model verify that different travel distances will have different effects on the choice of slow traffic travel mode: a distance of 30-90km has a positive effect on the SM mode, while a distance of less than 30km and greater than 90km is negative Effect, when the distance is relatively long, people are very unwilling to choose slow traffic travel.

5.3 The Time Model

Considering the influence of the proportion of walking and waiting time in travel time on behavior choices, the model established is time_interaction.py, and its effect function is:

$$V_{CAR,n} = ASC_{CAR} + \beta_{cost} \cdot CostCar + \beta_{time} \cdot TimeCar + \beta_{nbchild} \cdot NbChild + \beta_{nbcar} \cdot NbCar$$
$$+\beta_{freqcarpar} \cdot FreqCarPar + \beta_{modecar} \cdot ModeToSchoolCar$$
$$V_{PT,n} = 0 + \beta_{cost} \cdot MarginalCostPT + \beta_{hightime} \cdot \frac{HighTime}{TimePT} + \beta_{lowtime} \cdot \frac{LowTime}{TimePT}$$
$$+\beta_{freqtrainpar} \cdot FreqTrainPar + \beta_{modePT} \cdot ModeToSchoolPT$$
$$V_{SM,n} = ASC_{SM} + \beta_{distance} \cdot distance\_km + \beta_{nbchild} \cdot NbChild + \beta_{nbbicy} \cdot NbBicy$$
$$+\beta_{freqothpar} \cdot FreqOthPar + \beta_{modesm} \cdot ModeToSchoolSM$$

The running result of Time_interaction verifies how the proportion of walking time and waiting time in the total travel time will affect behavior choices during the journey.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202106_7(6).0055

According to the running results of the model in Biogeme, it can be seen that the proportion of walking time and waiting time in travel time has a positive effect on PT. When the proportion is relatively small, PT is more willing to choose PT as the mode of travel.

5.4 The NbChild Model

In the final test base, the result of generic processing for NbChild is not very significant, so non-generic processing for NbChild is performed again to check whether the result is more significant, and the model NbChild_.py is established, and its effect function is:

$$V_{CAR,n} = ASC_{CAR} + \beta_{cost} \cdot CostCar + \beta_{time} \cdot TimeCar + \beta_{nbchildPMM} \cdot NbChild + \beta_{nbcar} \cdot NbCar + \beta_{freqcarpar} \cdot FreqCarPar + \beta_{mode car} \cdot ModeToSchoolCar$$

$$V_{PT,n} = 0 + \beta_{cost} \cdot MarginalCostPT + \beta_{time} \cdot TimePT + \beta_{freqtrainpar} \cdot FreqTrainPar + \beta_{mod ePT} \cdot ModeToSchoolPT$$

$$V_{SM,n} = ASC_{SM} + \beta_{distance} \cdot distance\ \ km + \beta_{nbchildSM} \cdot NbChild + \beta_{nbbicy} \cdot NbBicy + \beta_{freqothpar} \cdot FreqOthPar + \beta_{mod esm} \cdot ModeToSchoolSM$$

Run the program in biogeme and get the results. By comparison, it can be seen that the number of children in the family has a greater impact on people who choose cars, while for people who choose chronic traffic, the impact is relatively small. After doing the generic treatment, the result is not significant, but changing to the non-generic treatment method, the result is relatively significant.

6. Conclusion

In the above result table, value represents the corresponding parameter value, Std err represents the standard error, and t-test is the t-test value of the parameter value. The larger the t-test value, it indicates that the influence of this factor is not significant. And in the process of statistics, a part of invalid data is discarded. Finally, it can be known from the data that the probability of choosing a car as a mode of travel is far greater than that of choosing public transportation as a mode of travel.

The tighter the parking spaces near the work place, the higher the parking fee near the work place. And the farther the place of work is from the place of residence, the more likely travelers are to choose a car. This may be because the farther the distance, the more obvious the advantages of choosing a car travel mode in terms of travel expenses and travel time costs; the more children in the family, the more inclined to choose a car to travel. In addition, travel expenses have a significant negative impact on travellers' choice of car sharing, car sharing transfer to subway, and bus interchange to subway. Travel cost is also a factor that people often consider.

This article analyzes by establishing a number of Logit models for travellers' choice of travel modes, discussing the factors that affect travellers' choice of cars and public transport modes, and provides relevant measures that can be referred to in order to attract more travellers to choose public transport, such as: When choosing car-sharing to transfer to the subway, since travelers are more sensitive to travel expenses, the joint fare policy of car-sharing and subway can be considered to reduce travel expenses.

References

[1] Chen Xiao. Research on the Inner Mechanism of Urban Residents' Choice of Passenger Transport Modes (Adhesion, China 2021), 45(02):184-186+191. (In Chinese)

[2] Wang Lei. Analysis of factors affecting the choice of rail transit mode[J]. Science and Technology Vision, 2014(30): 307+323. (In Chinese)

[3] Yan Liying, Meng Xiangpei. Research on the influence of rail transit on the choice of residents' transportation mode (Railway Transport and Economy, China 2018), 40(01): 100-105. (In Chinese)

[4] Bo Kun, Teng Jing. Travel mode choice of public transportation passengers under meteorological disasters (Comprehensive Transportation, China 2017), 39(08): 40-43+67. (In Chinese)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0055

[5] Liang Wenbo, Liu Chen, Wang Yanling. Analysis of influencing factors of residents' travel mode based on Logistic regression model[J]. Productivity Research, 2020(01): 125-127+136. (In Chinese)

[6] Liu Chunyu, Luo Xia. Choice of travel mode: A comparative study of machine learning methods and multiple Logit models (Comprehensive Transportation, China 2018), 40(08): 57-63. (In Chinese)

[7] Celikoglu H B. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling[J]. Mathematical and Computer Modelling, 2006, 44(7-8): 640-658.

[8] Espino R, de Dios Ortúzar J, Román C. Understanding suburban travel demand: Flexible modelling with revealed and stated choice data[J]. Transportation Research Part A: Policy and Practice, 2007, 41(10): 899-912.

[9] Wyer J F. Urban Transportation Mode Choice and Trip Complexity: Bicyclists Stick to Their Gears[J]. Economic Inquiry, 2018, 56(3): 1777-1787.