

Modeling Household Car Ownership using Ordered Response Model

Kaijun Cui^{1, a, *}, Chunfu Shao^{1, b} and Guangzheng Yao^{2, c}

¹MOE Key Laboratory for Urban Transportation Complex Systems and Technology, Beijing Jiaotong University, Beijing 100044, China.

²Nanjing Institute of City & Transport Planning Co. Beijing Branch, Beijing 100071, China.

^{a, *} 16120789@bjtu.edu.cn, ^b cfshao@bjtu.edu.cn, ^c 6903880@qq.com

Abstract. Considering the discrete and order attributes of household car ownership, the unordered response model (MNL model) and ordered response model (OL model) are used to predict the household car ownership using the data of fifth Beijing Person Trip Survey in 2015. The results show that the number of driver's licenses has the most significant impact on car ownership, followed by owning house, family size, household income and household registration in Beijing. Different from previous studies, it is found that the age of households between 35-44 and 55-64 will significantly affect the number of car ownership. A brief analysis is carried out on the impact of the travel limit policy on car ownership. The model passes statistical tests and calculates the marginal utility of all significant explanatory variables.

Keywords: Person Trip Survey; Car Ownership; Unordered Response Model; Ordered Response Model; Marginal Utility.

1. Introduction

In recent years, with the continuous improvement of the income level of Chinese residents, more and more families can afford cars. By the end of 2017, the number of motor vehicles in China had reached 310 million. There are 7 cities in the country with more than 3 million cars, including Beijing with 5.64 million, ranking first. There are many factors that affect the household car ownership, which include family factors, economic factors, competitive factors, environmental factors, and policy factors. Generally, the more the household size or the more adults, the more likely the family will buy a car [1]. The number of children in the family or the increase in the number of employees [2] will affect the travel demand, and the increase in travel demand will also lead to changes in the number of family cars. Economic factors have a great impact on the household car ownership. From a macro perspective, the higher the GDP and the more disposable income of urban residents, the higher the possibility of residents buying cars [3], and from the micro level, the higher the family income level [4], the more possibility the family will buy the car. If the car has high purchase cost and high cost of use, it will reduce the probability of family buying a car. Competition factors can be seen as the impact of competitive transportation, which include public transportation, non-motor vehicles, walking, etc. The public transportation services are more convenience, and the fewer families are likely to purchase cars [5]. The environmental factors are mainly the influence of density, including population density and work density, where the population density and the work density is high, the public transportation is relatively developed, so the car ownership in the area is lower [6]. The impact of policies on the household car ownership is mainly to curb the demand for motor vehicles.

Early studies on household car ownership were more dependent on the aggregate model. Many of the aggregate models use time series data for analysis, such as gross domestic product over the years, to predict the total car ownership in a country or region. Button [7] used the logit function, Dargay [8] used the Gompertz function and the S curve. Feng [9] used statistical yearbook data to establish a regression model for analysis. However, because the aggregate model cannot consider individual factors, the disaggregate model seems more suitable. The disaggregated model is based on the family level, including demographic characteristics of the family, income and expenditure, household surrounding environment.

Both of the above models are static models, and it is a static process to predict the total number of cars in a certain area. The dynamic model is a process of describing the number of family cars in a given time period. For example, the purchase, use and replacement of a family car is a dynamic process [10]. The degree of use of cars affects the depreciation, which in turn affects the price of used cars. The price of used cars determines the replacement of cars. If the price is high, the family will sell the existing cars. At this time, the number of family cars decreases, while the price is low, the family will hold the current car, and the number of cars will remain the same.

For the domestic, statistical data are needed for the aggregate model. As a result of the urban and rural economic development is unequal, the use of the aggregate model does not reflect the urban or rural car ownership perfectly. Dynamic models place high demands on data collection. If the disaggregate model is used, analysis by family can reflect the household car ownership. So, this article uses a disaggregate model.

2. Methodology

The disaggregate model can be divided into two categories: unordered response model and ordered response model. The unordered response model conforms to the economic utility maximization model, and the common representative is the multinomial logit model. If we suppose household car ownership are disordered, the use of the multinomial logit model assumes a random utility [11]. The utility of different car ownership to different families are different, and the model goal is to maximize the random utility. The premise of using this model is that each variable needs to meet the IIA assumptions.

Multinomial Logit model takes the form:

$$U_i = V_i + \varepsilon_i \quad (1)$$

Where, U is the total utility, V is the fixed utility and ε is the random utility. Generally, this paper chooses the greatest utility of all options, so the probability of the maxim utility is:

$$P_i = \text{Prob}(U_i > \max U_j; i \neq j) = \text{Prob}(V_i + \varepsilon_i > \max(V_j + \varepsilon_j); i \neq j) \quad (2)$$

Where, $0 \leq P_i \leq 1, \sum_i P_i = 1$

If the household car ownership is modeled following an ordered response mechanism, the ordered logistic regression is needed, and the common representative is the ordered logistic regression model. Different to the multinomial logit model, the explanatory variables of the car ownerships are different, while the ordered logistic regression model explains the same variable parameters and uses different thresholds to distinguish the number of cars.

Ordered logistic regression models take the form:

$$Y^* = x' \beta + \varepsilon \quad (3)$$

Where, Y^* is the unobserved dependent variable, x' is the vector of independent variables, β is the vector of regression coefficient to be estimated, and ε is a random disturbance term. Y^* is the estimated value of Y , which has various category thresholds. Different Y this paper chooses due to the rule that:

$$Y = \begin{cases} 1 & Y^* \leq \mu_1 \\ 2 & \mu_1 < Y^* \leq \mu_2 \\ 3 & Y^* > \mu_2 \end{cases} \quad (4)$$

Where, μ_1 and μ_2 are threshold. If ε assumed to be independently Gumbel distributed across observations, then an ordered logistic regression model could be seen as follows:

$$\left. \begin{aligned} P(Y = 1) &= \phi(-\beta x) \\ P(Y = 2) &= \phi(\mu_1 - \beta x) - \phi(-\beta x) \\ P(Y = 3) &= 1 - \phi(\mu_2 - \beta x) \end{aligned} \right\} \quad (5)$$

Where, ϕ is the cumulative distribution function of logistic distribution.

3. Data Collection

The data used for car ownership model is from the fifth Beijing Person Trip Survey. The scope of the fifth Person Trip survey in Beijing included about 50000 permanent residents in 265 street towns and towns in 16 districts and counties throughout the city. The data includes the number of family cars and household income, and the personal characteristics of the residents include age, gender, occupation, and the daily travel log of the residents.

The raw survey data includes 227125 travel data of 101815 family members in 40003 households, excluding household income and missing population data. Since the cars are counted on a household basis, family member data is processed and integrated into family units and finally there are a total of 16723 data for modeling use.

3.1 Dependent Variable

Table 1 shows the statistics of the number of family cars. About 43% of households do not buy a car, and 48% of households who purchase a car. Only a small number of families (0.3%) purchase three cars or more. So, this paper puts the family that buys two cars and three cars into the same category considering the sample size is very small.

Table 1. Observed frequency of household car ownership

car_number	Freq.	Percent	Cum.	Corresponding categories used in model
0	7,209	43.11	43.11	Category 0:zero car
1	8,011	47.9	91.01	Category 1:one car
2	1,454	8.69	99.71	Category2:two-plus car
3	49	0.29	100	Category2:two-plus car
Total	16,723	100		

3.2 Explanatory Variables

Several statistical tests were conducted on survey data, and persuasive explanatory variables were screened out. The economic characteristics of household and family travel characteristics were chosen as explanatory variables. The household size is an important factor in determining the number of family cars. The more members there are, the more travel needs will be generated, and the car will be purchased to satisfy the growing travel needs. However, different types of family members have different travel needs, and working family members have the most travel needs, and the staff members are also potential drivers in the family. The characteristics of the head of the household are also an important factor affecting the number of cars. According to the actual situation, the household head age is divided into five categories (25-34, 35-44, 45-54, 55-64, >65), which is used as a dummy variable because age has a nonlinear effect on the number of cars. The household car ownership is also affected by the gender of the head of the household. The type of work and the highest degree of education of the head of the household are category variables. The location of the household registration also has an impact on the number of cars.

Whether or not the households buy a car is determined by family income. If the income is not high enough, it will not be able to afford the car. The income will be divided into 8 categories (0-5w, 5w-10w, 10w-15w, 15w-20w, 20w-25w, 25w- 30w, 30w-50w, >50w). Generally speaking, if renting a house, the family will usually not buy a car. The number of driver's licenses determines whether

family members could drive. Non-motor vehicles and small cars are in a competitive relationship. The more trips mean higher travel needs. Related explanatory variables are shown in Table 2.

Table 2. Descriptive statistic of explanatory variables

Variable	Min	Max	Mean	Std. Dev.	Variable	Min	Max	Mean	Std. Dev.
age_25_34	0	1	0.28	0.45	own_house	0	1	0.79	0.40
age_35_44	0	1	0.30	0.46	rent_house	0	1	0.20	0.40
age_45_54	0	1	0.27	0.44	household_~a	7	948	95.70	69.20
age_55_64	0	1	0.09	0.28	non_motor_~r	0	8	1.31	1.13
age_65	0	1	0.00	0.09	hukou_beij~g	0	1	0.79	0.40
sex	0	1	0.58	0.49	edu	0	8	5.46	1.70
income	1	8	2.28	1.19	work_categ~y	0	3	1.53	0.68
driver_lic~e	0	1	0.66	0.47	trip_num	0	35	6.48	3.50
fam_num	1	5	2.64	0.95	trip_dis	0.41	604.16	32.80	26.30
work_num	0	4	1.82	0.63					

4. Modeling and Analysis

4.1 Model Parameter Estimation

The maximum likelihood estimation is used to estimate the model. The multinomial logit model and ordered logistic regression model are used in car ownership model. Table 3 shows the results of the two models.

Table 3. The parameters and t statistic of the car ownership model

Variables	MNL				OLOGIT	
	Segment-1		Segment-2		Parameters	t ratios
	Parameters	t ratios	Parameters	t ratios		
age_25_34	0.104	0.409	0.415	0.153	0.133	0.256
age_35_44	0.461***	0	0.855***	0.003	0.417***	0
age_45_54	0.328**	0.011	0.893***	0.002	0.388***	0.001
age_55_64	0.427***	0.003	1.208***	0	0.550***	0
age_65	-0.149	0.59	-0.258	0.747	-0.205	0.438
sex	-0.484***	0	-0.689***	0	-0.439***	0
income	0.268***	0	0.520***	0	0.290***	0
driver_license	2.371***	0	3.671***	0	2.477***	0
fam_num	0.354***	0	0.717***	0	0.406***	0
work_num	0.333***	0	0.703***	0	0.400***	0
own_house	0.507***	0	0.929***	0	0.541***	0
o.rent_house	-		-		-	
household_area	0.00247***	0	0.00400***	0	0.00244***	0
non_motor_number	-0.242***	0	-0.411***	0	-0.253***	0
hukou_beijing	0.414***	0	0.608***	0	0.382***	0
edu	0.0305**	0.035	-0.0596**	0.013	-0.00991	0.431
work_category	0.287***	0	0.577***	0	0.324***	0
trip_num	-0.0589***	0	-0.105***	0	-0.0625***	0
trip_dis	0.00845***	0	0.0148***	0	0.00869***	0
Constant	-4.700***	0	-10.76***	0	-	
Constant cut1	-				4.822***	
Constant cut2	-				8.264***	

Notes: *** p<0.01, ** p<0.05, * p<0.1

As a result of the multinomial logit model has IIA assumptions, this paper uses Hausmann statistical test of the model, the results are as follows: $\text{Prob} > \chi^2 = 0.0000$. It can be seen that the P value is small and the IIA hypothesis can be rejected, so it is not suitable to use the multinomial logit model, so this paper uses the ordered logistic regression model (ologit model).

The parameter for the people aged 25-64 are positive and people aged more than 65 are negative, people who aged 35-44 and 55-64 are more positive. It seems that people older than 65 are less likely to buy a car, and people aged 35-44 and 55-64 seem more willing to buy a car. The possible reason is that the income of people aged 35-44 is gradually stable, and the child's travel demand increases due to schooling and other reasons, so the possibility of buying a car increase. The reason why people aged 55-64 are more willing to buy a car is because they have more free time after retirement and have more opportunities to travel. The coefficient of gender is negative, indicating that families with female heads are more reluctant to buy a car than families with males. The income coefficient is positive, indicating that the higher the income, the more probabilities for families to buy a car, also households with a large number of cars tend to have high incomes. Whether or not people can drive on the road is determined by driver's license, so it has the biggest impact on car ownership. The increase in the number of households and the number of workers will also increase the likelihood of buying a car. Whether or not to own a house will significantly affect the number of car ownership. Generally speaking, households in Beijing are more likely to own a car. Household car ownership is little affected by the space of the house. The number of non-motor vehicles and the highest educational level of the head of household are negative. For job categories, self-employed people are more willing to buy a car, perhaps a car is a symbol of identity. The parameter for travel distance is also positive.

It is worth noting that the number of trips has a negative but significant impact on the number of family cars. For the family who do not have a car that can only choose to travel by public transport. The accessibility of public transport is not as good as that of private car, so they have to through more transfers to reach their destination. In addition, the travel limit policy has resulted in residents only taking public transportation or walking on the restricted day, which undoubtedly increased the number of residents traveling.

4.2 Model Validation

To verify the validity of the model, Table 4 shows the accuracy of the predictions and correctly predicts the number of household cars at 68.7%. It can be seen that the prediction results of 0 vehicles and 1 vehicle are more accurate, especially the prediction accuracy of 1 vehicle is 81.9%, but the prediction probability of 2 or more vehicles and above is low. The possible reason is that the sample of 2 or more vehicles are rather small compared to the others.

Table 4. Confusion matrix of the prediction and observation

prediction	observation			prediction total
	zero car	one car	two plus car	
zero car	4843	1362	63	6268
one car	2355	6566	1352	10273
two plus car	11	83	88	182
observation total	7209	8011	1503	

4.3 Marginal Effects Analysis

To assess the overall effect of the explanatory variables on the three car ownership levels, marginal effects are calculated for each observation and then averaged across all observations. Table 5 shows the marginal utility of all significant explanatory variables. The results show that the number of driver licenses is the variable with the highest impact on car ownership with a marginal effect of 0.26. Regarding driver licenses, it would say that for a one-unit increase in driver licenses, there is a 26% chance for a 2-driver license household of "own zero car" to "own one car" or "own two-plus car" combined with a 1 driver license household. Such as the income factor, for every unit of income

added, the probability of a family buying a car increases by 2%. If work types change, such as from a government institution to a self-employed business, the probability of buying a car increases by 2%. Non-motor vehicles and private cars are substitutes for each other. The more non-motor vehicles, the less people need to buy a car. It is worth noting that the marginal utility of the number of trips is negative, but the value is only -0.005. The reason is that the high accessibility of private cars, which can reach the destination by only one trip. However, due to the travel limit policy, private cars cannot be used on the restricted day, and only public transportation can be used. The accessibility of the bus is not as good as that of the private car. People need more transfers to reach the destination.

Table 5. Marginal effects of the explanatory variables

variable	marginal effect	variable	marginal effect
age_35_44	0.043619	own_house	0.04867
age_45_54	0.015242	household_area	0.000267
age_55_64	0.017253	non_motor_number	-0.02495
sex	-0.05746	hukou_beijing	0.048058
income	0.024133	edu	0.009348
driver_license	0.264623	work_category	0.024715
fam_num	0.030223	trip_num	-0.00583
work_num	0.02681	trip_dis	0.000848

5. Summary

This paper uses the multinomial logit model and ordered logistic regression model to analyze the car ownership. However, due to the existence of the IIA characteristics of the multinomial logit model and the order relationship of the car ownership, the sequential response model is adopted, and finally the ordered logistic regression model is selected. All significant explanatory variables are marked for each significant level. At the same time, the marginal utility is calculated, that is, quantify the probability change in the household car ownership following a one-unit change in the explanatory variables. The Nagelkerke R² of the model is 0.235. In the actual situation, the Nagelkerke R² range is between 0.2 and 0.4, and the model can be considered to be highly fitted. Therefore, compared with the traditional regression model and the multinomial logit model, it is more ideal to predict the car ownership by using the ordered logistic regression model.

Acknowledgments

This work was supported by the Hebei Natural Science Foundation under Grant E2016513016.

References

- [1]. Shay E, Khattak A J. Household travel decision chains: Residential environment, automobile ownership, trips and mode choice[J]. *International Journal of Sustainable Transportation*, 2012, 6(2): 88-110.
- [2]. Giuliano G, Dargay J. Car ownership, travel and land use: a comparison of the US and Great Britain[J]. *Transportation Research Part A: Policy and Practice*, 2006, 40(2): 106-124.
- [3]. Zegras C. The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile[J]. *Urban Studies*, 2010, 47(8): 1793-1817.
- [4]. Guo Z. Does residential parking supply affect household car ownership? The case of New York City[J]. *Journal of Transport Geography*, 2013, 26: 18-28.

- [5]. Aditjandra P T, Cao X J, Mulley C. Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data[J]. *Transportation research part A: policy and practice*, 2012, 46(1): 22-32.
- [6]. Paleti R, Bhat C, Pendyala R. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2013 (2382): 162-172.
- [7]. Button K J, Pearman A D, Fowkes A S. *Car ownership modelling and forecasting*[M]. 1982.
- [8]. Dargay J, Gately D. Income's effect on car and vehicle ownership, worldwide: 1960–2015[J]. *Transportation Research Part A: Policy and Practice*, 1999, 33(2): 101-138.
- [9]. Feng Zekun. Analysis of the Factors Affecting the Private Car Ownership in China[J]. *China Market*, 2014, (31): 135-139.
- [10]. Page Mattewandwhelan G AA J. Modelling the Factors which Influence New Car Purchasing[C]// *TRANSPORT MODELLING: PROCEEDINGS OF SEMINAR K HELD AT THE EUROPEAN TRANSPORT CONFERENCE*. 2000.
- [11]. Wong K I. An analysis of car and motorcycle ownership in Macao[J]. *International Journal of Sustainable Transportation*, 2013, 7(3): 204-225.