An Analysis of Residents' Commuting Behavior Considering Household Heterogeneity

Shuhong Ma 1, Xingxing Yin 1*, Dachuan Tang 1 and Chuanqi Liu 1
1 School of Highway, Chang’an University, Xi’an, Shaanxi, 710064, China
*Corresponding author’s e-mail: 1584109777@qq.com

Abstract. In order to analyze the commuting behavior of residents more accurately, the quantitative family heterogeneity is integrated into commuting travel choice behavior analysis. Based on the data of residents’ travel survey in Xi’an, MNL models with neglecting household heterogeneity and Multi-Level MNL models with household heterogeneity are established respectively, and the parameters are calibrated. The model calibration results show that 19% of the residents’ commuting behavior selection is caused by household heterogeneity. Household heterogeneity cannot be ignored. Compared with the multi-layer MNL model, the MNL model overestimates the extent to which some explanatory variables affect the explanatory variables. At the same time, the analysis proves that having a car in the family is one of the key factors in inducing car travel. The transfer of bus and bicycle modes of transportation is highly accepted by commuters. We can increase the share of buses and reduce the use of cars by encouraging passengers to use bus and bicycle transfers.

1. Introduction

Residents' commuting travel choice behavior is the most basic travel of residents, and it is necessary to accurately analyze the commuter travel behavior of residents. Relevant research at home and abroad mainly focuses on two aspects of single travel and travel chain, and builds discrete selection models based on utility maximization theory. Heldt and Johansson developed a comprehensive discrete mode selection model to study commute mode selection by considering potential variables including environmental preferences, safety, comfort and convenience [1]. K Luan constructing a Nested Logit model in which commuters work and travel chain choices interact with each other, the results show that commuters tend to first consider how to organize various activities to participate in the day, and then consider the appropriate choice of travel under the constraints of the travel chain [2]. However, residents' commuting behavior is not independent. It is not only driven by the needs of their own social activities, but also affected by family structure and family resource allocation. F Yue pointed out that with the increase of family monthly income, the probability of choosing a car as a mode of transportation will show an increasing trend [3]; J CH Xianyu pointed out that family income, car and family living position also affect the outing activities and travel choices of the two main members[4]. Personal factors also have a long-term impact on residents' commuting options. J CH Xianyu pointed out that the choice of commuting mode shows a certain gender difference[5]; Zh Chen pointed out that the proportion of residents in the 30-39 age group using private cars as commuter vehicles is much higher than other age groups[6]. It can be seen that personal attributes and family attributes are important factors influencing residents' commuting.

In the analysis of residents' commuting travel choices, many research questions present a multi-level data structure. For example, a certain travel of a resident is also a family or group at the
macro level, and there is a certain nesting relationship. Chikaraishi M’s research shows that the more microscopically the analysis unit is, the more complex the macro level is affected. The research needs to deeply consider the influence of different levels of macro factors on travel behavior [7]. Based on this, this paper intends to define the individual unit and household level factors to define the research unit, establish a MNL model that ignores household heterogeneity and a multi-layer MNL model that considers household heterogeneity. Investigate whether residents’ commuting behaviors are heterogeneous and the impact of household heterogeneity on residents’ commuting behaviors, in order to more accurately analyze residents’ commuting travel choice behaviors, and provide basis for urban traffic demand analysis and traffic management.

2. Multi-layer MNL model

The multi-layered MNL model is a combination of a generalized mixed linear model with a linear predictor and MNL model. In the residents’ travel surveys, household surveys are generally conducted. Therefore, the survey data has a multi-layered structure, with residents representing the first layer of the data structure and households representing the second layer of the data structure. When the data has a multi-layered structure, the two basic assumptions of the inter-individual variance homogeneity and the random error terms of the MNL model are no longer satisfied.

The zero model is a model with no explanatory variables at both the individual and family levels, and is the basis for multi-level modeling. Assuming that there are M classifications for the dependent variables, for each of the M unordered response categories \( m = 1 \ldots M-1 \), and \( M \) is the reference class.

Zero model:

\[
\eta_{mij} = \beta_{0j(m)} + \epsilon_{ij(m)} \\
\beta_{0j(m)} = \theta_{0j(m)} + \delta_{0j} \\
\epsilon_{ij(m)} \\
\]

In the formula, \( \eta_{mij} \) is the dependent variable of the model, indicating the logarithmic probability ratio of the traffic mode \( m \) selected from the family \( j \) (\( j=1 \ldots J \)) traveler \( i \) (\( i=1 \ldots I \)) relative to the reference mode \( M \); \( \beta_{0j(m)} \) is the intercept of the first layer equation, its mean is \( \theta_{0j(m)} \); \( \delta_{0j} \) represents unobserved household-level random error terms; \( \epsilon_{ij(m)} \) represents the random error term of the individual level that is not observed. Random error terms from different layers, \( \epsilon_{ij(m)} \) and \( \delta_{0j} \), are independent of each other and both are subject to the Gumbel distribution.

To measure the extent to which the overall difference in the dependent variable is caused by family differences, Grilli L indicates that intra-class correlation coefficient can be used to characterize the degree of family heterogeneity [8], and derive the ICC formula of the multi-layer MNL model, as shown in formula (3), where \( Var(\delta_{0j}) \) represents the variance of \( \delta_{0j} \) and \( Var(\epsilon_{ij(m)}) \) represents the variance of \( \epsilon_{ij(m)} \).

\[
ICC = \frac{Var(\delta_{0j})}{Var(\delta_{0j}) + Var(\epsilon_{ij(m)})} = \frac{Var(\delta_{0j})^2}{Var(\delta_{0j})^2 + \frac{\pi^2}{3}} \\
\]

The larger the ICC, the greater the degree of family heterogeneity, and the greater the impact on the dependent variable due to differences in family conditions. The ICC is between 0 and 1. According to the classification of heterogeneity degree proposed by Snijder, when ICC>0.138, the degree of high heterogeneity is indicated, and the variation of the dependent variable between families is not negligible; when 0.059 <ICC<0.138, it means moderate heterogeneity; ICC <0.059 indicates the degree of low heterogeneity [9]. In order to gain a deeper understanding of the impact of family and individual factors on residents' commuting behavior, a complete model is needed for analysis.

Complete model:
\[
\eta_{mj} = \log\left(\frac{P_m}{P_M}\right) = \beta_{0j(m)} + \sum_{q=1}^{Q} \beta_{qj(m)} X_{qj} + \epsilon_{ij(m)} \tag{4}
\]

\[
\beta_{0j(m)} = r_{00(m)} + \sum_{i=1}^{S} r_{0i(m)} + \delta_{0j} \tag{5}
\]

\[
\beta_{qj(m)} = r_{q0(m)} + \delta_{qj} \tag{6}
\]

In the formula, \(P_m, P_M\) respectively choose the probability of traffic mode \(m, M\) for this traveler, \(X_{qj}\) is an explanatory variable containing the individual attributes of the traveler, \(\beta_{0j(m)}\) is the coefficient of the explanatory variable, \(r_{0i(m)}\) is the explanatory variable of the traveler's household attribute. If the coefficient of an explanatory variable in (4) is a random coefficient, it is expressed by (6). \(\beta_{qj(m)}\) is a random coefficient, its mean is \(r_{q0(m)}\), \(\delta_{qj}\) is a random item. In the study, the adaptive maximum Gaussian orthogonal method is used to obtain the maximum likelihood value [10]. The solution of the model is used by Stata.

3. Data sources and variable interpretation

This paper uses Xi'an 2011 resident travel survey data as a data source to study residents' commuting travel choice behavior, eliminating data with incomplete information and incorrect information. The final sample contains 13,704 commuters' travel data for 13,704 residents. The main statistical variables of the model are shown in Table 1.

In addition, commuting modes include walking, bicycle, electric vehicle, bus and car. The proportion of each mode of transportation is 26.3%, 13.6%, 15.8%, 15.5%, 28.7%. The selected variables are shown in Table 1, in which the age and travel time are continuous variables, and the remaining variables are dummy variables. In the final sample, the average age of the respondents was 38.73 years, and the average commuting time of residents was 32.24 minutes. Among the family characteristics, 84.2% of the households do not have a population of less than 6 years old, and more than 70% of the households have an average monthly income of 2000-6000, indicating that middle-income families are mostly. In general, the basic situation of the sample is consistent with the overall distribution of Xi'an families.

| Level          | Variable Name | Variable Description | Proportion |
|----------------|---------------|----------------------|------------|
| Individual level | Age           | age (38.73 mean, 115 variance) |
| Gender          | gender        | 0: female, 1: male   | 49.9%, 50.1% |
| Travel time     | time          | 32.24 (mean, 113 variance) | 0.5%, 22.7% |
| Household level | The total family population | nfamily | 61.8%, 10.7%, 4.2% |
| The population of less than six years old | less 6 | 0, 1, 2 | 84.2%, 13.8%, 2.0% |
| The average monthly income of the family | income | 1: <¥2000, 2: ¥2000-¥4000, 3: ¥4000-¥6000, 4: ¥6000-¥10000, 5: >¥10000 | 16.4%, 42.5%, 28.7%, 10.5%, 1.8% |

Table 1. Statistical model of the main variables.
4. Model analysis

4.1 Zero model.
First, establish a zero model without any explanatory variables. The zero model calibration results are shown in Table 2.

| Model Component                  | Parameter Estimation | Standard Error |
|----------------------------------|----------------------|----------------|
| Intercept (walking)              | 0.74                 | 0.024          |
| Intercept (bicycle)              | 0.08                 | 0.027          |
| Intercept (electric vehicle)     | 0.23                 | 0.027          |
| Intercept (bus)                  | 0.83                 | 0.026          |

\( \delta_{0j} \) = 0.77(variance) 0.05(covariance)

-2LL 63405.45

\[ ICC = \frac{Var(\delta_{0j})^2}{2\pi^2} = \frac{0.77^2}{2\pi^2} = 0.19 \]

ICC=0.19 indicates that among the reasons for residents’ differences in commuting choice behaviors, 19% of differences are caused by differences in families (inter-group variation), 81% of the variability is caused by differences in individual factors (intra-group variation). It also illustrates the necessity and correctness of multi-layer model analysis.

4.2 Complete model.
In order to further explain the behavior of residents' commuting modes of travel, a complete model is established. The parameter calibration results of this model are shown in Table 3.

From the analysis results of the MNL model in Table 3, the multi-layer MNL model estimation results show that the direction and coefficient values of most variable parameter estimates are consistent, but the standard error of most variables of multi-layer MNL is larger than that of MNL. In addition, there are some differences between the two models in characterizing the significance of variables. In the multi-layer MNL model, the significance of the explanatory variable of whether the family owns the bicycle in the electric vehicle mode is degraded, and the significance of the total number of households in the bus mode is also reduced.

After the household level and the individual level explanatory variables were included in the zero model, the inter-group variation value of residents' commuting modes was 0.31, which was significantly reduced. The random error term at the family level in the two models is compared, and the function expression is as follows:
Therefore, the multi-layer MNL model $R^2$ is 0.60, while the $R^2$ of the MNL is only 0.47. In contrast, the multi-layer MNL model has a better fitting effect.

After considering the family heterogeneity, some family factors also have a significant impact on the commuter behavior of commuters. Compared to cars, the increase in household monthly income will reduce the effectiveness of the other four modes of transportation. Having a car in the family also has a negative effect on the other four modes of transportation, indicating that having a car is one of the key factors in inducing a car to travel. The presence of a population of less than 6 years in the family reduces the probability of commuters using walking and cycling. Family-owned bicycles have a significant positive impact on the use of bicycles, indicating that commuters use bicycles significantly when they have bicycle use conditions. The increase in the monthly average income of the family will reduce the residents' choice of buses. The family-owned bicycles will increase the effect of the bus, indicating that the transfer of the two modes of transportation between the bus and the bicycle is highly accepted by the commuters.

## Table 3. Multi-layer MNL model and MNL model calibration results.

| Mode       | Variable | MNL      | Multi-layered MNL |
|------------|----------|----------|--------------------|
|            |          | B        | Std. Err. | Sig. | B        | Std. Err. | Sig. |
| Walking    | Intercept| 4.578    | 0.137     | 0.000 | 4.879    | 0.150     | 0.000 |
|            | time     | -0.079   | 0.001     | 0.001 | -0.080   | 0.001     | 0.001 |
|            | age      | 0.027    | 0.002     | 0.000 | 0.027    | 0.002     | 0.000 |
|            | gender   | -1.519   | 0.044     | 0.000 | -1.607   | 0.047     | 0.000 |
|            | Ev       | -0.342   | 0.049     | 0.000 | -0.303   | 0.052     | 0.003 |
|            | car      | -2.892   | 0.049     | 0.000 | -3.077   | 0.058     | 0.000 |
|            | income   | -0.292   | 0.023     | 0.002 | -0.310   | 0.024     | 0.002 |
|            | less6    | -0.031   | 0.047     | 0.004 | -0.032   | 0.050     | 0.008 |
| Bicycle    | Intercept| 1.721    | 0.157     | 0.000 | 2.017    | 0.168     | 0.000 |
|            | time     | -0.031   | 0.001     | 0.000 | -0.032   | 0.001     | 0.000 |
|            | age      | 0.021    | 0.002     | 0.000 | 0.022    | 0.002     | 0.005 |
|            | gender   | -1.4     | 0.049     | 0.003 | -1.487   | 0.052     | 0.003 |
|            | bike     | 2.592    | 0.056     | 0.000 | 2.620    | 0.058     | 0.000 |
|            | car      | -3.241   | 0.065     | 0.000 | -3.426   | 0.072     | 0.001 |
|            | income   | -0.396   | 0.026     | 0.000 | -0.415   | 0.028     | 0.000 |
|            | less6    | -0.106   | 0.055     | 0.003 | -0.107   | 0.057     | 0.007 |
| Electric Vehicle | Intercept | 1.144 | 0.155 | 0.000 | 1.434 | 0.166 | 0.000 |
|            | time     | -0.014   | 0.001 | 0.003 | -0.015 | 0.001 | 0.005 |
|            | gender   | -0.892   | 0.049 | 0.009 | -0.978 | 0.052 | 0.009 |
|            | Ev       | 3.243    | 0.059 | 0.000 | 3.284 | 0.061 | 0.001 |
|            | bike     | -0.191   | 0.048 | 0.007 | -0.219 | 0.051 | 0.009 |
|            | Car      | -3.029   | 0.062 | 0.000 | -3.209 | 0.069 | 0.000 |
|            | income   | -0.324   | 0.026 | 0.000 | -0.344 | 0.027 | 0.000 |
| Bus        | Intercept| 3.259    | 0.13   | 0.000 | 3.549   | 0.143   | 0.000 |
|            | time     | 0.014    | 0.001 | 0.000 | 0.013 | 0.001 | 0.001 |
|            | Age      | -0.017   | 0.002 | 0.000 | -0.016 | 0.002 | 0.000 |
|            | gender   | -1.507   | 0.043 | 0.000 | -1.595 | 0.046 | 0.000 |
|            | Bike     | 0.178    | 0.042 | 0.000 | 0.206   | 0.045   | 0.002 |
|            | Car      | -3.303   | 0.048 | 0.000 | -3.501 | 0.058   | 0.000 |
|            | income   | -0.167   | 0.022 | 0.000 | -0.185 | 0.023   | 0.000 |
|            | Nfamily  | 0.052    | 0.028 | 0.005 | 0.045 | 0.030 | 0.008 |
5. Conclusions
The application of the multi-layered MNL model in commute selection behavior suggests that if family heterogeneity is ignored, data fit will be reduced and misleading assessments of traffic control measures will be developed. The results of two different methods of ignoring the family heterogeneity model and the model considering family heterogeneity are analyzed and compared. In the estimation of model parameters, the direction and coefficient values of the two model parameters are basically the same, but there are large differences in the significance of some variables. In particular, model results that ignore family heterogeneity tend to overestimate the extent to which some explanatory variables affect the explanatory variables. In terms of model calibration results, the multi-layer MNL model considering family heterogeneity shows better advantages. Therefore, when analyzing commute travel selection behavior, family heterogeneity factors cannot be ignored. In addition, the results show that having a car in the family is one of the key factors in inducing car travel. Restricting the car's possession can effectively reduce car travel; when there are bicycle use conditions, it will significantly improve the use of bicycles by commuters, and the family having a bicycle increases the effect of the bus, so ensuring that commuters have the conditions for using the bicycle will increase the share of the bus to a certain extent.

Acknowledgments
We are grateful to Professor Ma for her help during our research, and we are very grateful to the Chang'an University Transportation Research Center for providing data.

References
[1] Johansson, M.V., Heldt, T., Johansson, P. (2006) The effects of attitudes and personality traits on mode choice [J]. Transportation Research Part A Policy & Practice, 40(6):507-525.
[2] Luan, K, Juan Zh.C., Zong, F. (2010) Research on Commuter's Choice Behavior between Travel Mode and Trip Chain [J]. Journal of Highway and Transportation Research and Development, 27 (6): 107-111.
[3] Yue, F. (2008) Study on the characteristics of commuting behavior [D]. Beijing: Beijing Jiaotong University.
[4] Xianyu, J.CH., JUAN, Zh.C. (2012) Research on Activity-Travel Choice Behavior Interactions between Household Heads [J]. Journal of Systems & Management, 21(2):252-257.
[5] Xianyu, J.CH., JUAN, Zh.C., ZHU, T.Y. (2013) Selection of Commute Trip Timing and Mode [J], Journal of Shanghai Jiaotong University, 47 (10): 1601-1605.
[6] Chen, Zh., Zhou, H., Liu, Y.S. (2006) Study on Selection of Traffic Modes for commuter Trips in Suzhou [J]. Road Traffic and Safety, (8): 34-37.
[7] Chikaraishi, M., Fujiwara, A., Zhang, J., et al. Decomposing variations and co-variations in discrete travel choice behavior with multilevel cross-classified logit model [J]. 2009.
[8] Grilli, L., Rampichini, C. (2007) A multilevel multinomial logit model for the analysis of graduates' skills [J]. Statistical Methods & Applications, 16(3):381-393.
[9] Snijders, T. (1999) Multilevel analysis [J]. An Introduction to Basic & Advanced Multilevel Modeling, 1999, 64(2):785–793.
[10] Skrondal, A., Pickles, A. (2004) GLLAMM Manual [J]. Sophia Rabe-Hesketh, 19(1):467–472.