Probability model of driver's choice behavior in toll plaza

Hanmian Liang 1*, Shaowei Yang 1,2 and Binghong Pan 1,2

1 Highway College Chang’an University, Shaanxi 710064, China
2 Key Laboratory for Special Area Highway Engineering of Ministry of Education, Chang’an University, Shaanxi 710064, China
*Corresponding author’s e-mail: gsoshou@qq.com

Abstract. By analyzing the driver's choice behavior at toll plaza, the complexity of driver's toll lane selection behavior is discussed, and the utility function and probability calculation formula of the model are determined. The queue length, the number of lanes changed, the number of large vehicles and the location of the toll lane are selected as the characteristic variables of the model. The binary Logit model of the driver's choice behavior for the toll lane is established. Based on the actual observation data, the model parameters were estimated for each characteristic variable. The research results show that the driver's toll lane selection behavior is affected by many factors, and the fitting accuracy is high. The establishment of the model provides technical support for the traffic organization and toll lane layout.

1. Introduction

With the development of highway in China, the number of toll stations on expressways is increasing, and their safety status has an important impact on expressways. Existing studies[1] show that traffic accidents are relatively clustered at toll stations. The behavior of vehicles arriving at toll stations includes the process of entering toll plaza and looking for appropriate toll lanes until the deceleration stops. Therefore, the behavior characteristics at toll station entrance are complex. Understanding the factors affecting the choice of toll lanes and using these factors to guide the choice will be conducive to the safety management and traffic organization of toll stations. Studying the driver's toll lane selection model will also help to develop a more realistic queuing model for toll squares, and explain the traffic flow process of toll stations, providing a theoretical basis for optimizing toll station design and operation strategies.

Based on evolutionary game theory, Zhang[2] et al. determined the equilibrium point of lane selection which can balance the traffic volume of toll lane and reduce the congestion of toll plaza. K. Komada and T. Nagatani[3] studied the traffic flow and queuing of toll plazas, proposed the relationship between the queuing of toll plazas and the structure of toll lanes, and simulated the situation of reducing the queuing length of toll plazas by expanding toll lanes. Gulewicz and Danko[6] expounded the behavior habits of drivers in toll plazas according to the observed data, and proposed that drivers would prefer to choose toll lanes with shorter queue length and fewer lanes diversion, but they did not give the model of driver's toll lane selection. Ozbay[7] and others suggested that drivers would choose toll lanes with shorter queue length, but the location of vehicles before entering toll plaza was also the influencing factor of toll lane selection. The research on the behavior characteristics of drivers in front of toll plaza is imperfect, this paper intends to adopt the method of Logit model and proceed from the actual traffic condition of toll plazas. By selecting characteristic variables and determining utility functions, a probability model of driver’s toll lane selection behavior at toll plaza is
established, and the probability distribution of different toll lane selection and toll lane selection behavior is studied.

2. Construction of Toll Lane Selection Behavior Model

As the most important theoretical tool for traffic behavior analysis, Disaggregated Model[13] has led to a great leap in the field of traffic behavior research. The "stochastic utility model" proposed from the perspective of disaggregation, referred to as Logit model, is widely used in the field of road traffic because of its explicit and simple algorithm. Based on Binary Logit model, this paper constructs a probabilistic model of driver's toll lane selection behavior.

2.1. Binary Logit model

According to the theory of stochastic utility maximization, drivers will choose the most effective toll lanes. The utility value ($U_{in}$) that driver $n$ ($n=1,2,3,...N$) chooses scheme $J$ ($i=1,2,3,...I$) includes two parts: the observable utility part, which is determined by the characteristics of the toll lane itself and the attributes of the decision making unit, and the unobservable random part, which is determined by the influence of the unobservable factors on the decision making unit. Such a function becomes a stochastic utility function, which can be expressed as:

$$U_{in} = V_{in} + \epsilon_{in}$$  \hspace{1cm} (1)

$V_{in}$ is an observable utility part, i.e. a determinant; $\epsilon_{in}$ is an unobservable utility part, i.e. a random term. In order to solve $V_{in}$, we can assume that it is a linear utility function. Because of its convenience and superposition utility, this assumption is widely used. The expression of $V_{in}$ is as follows:

$$V_{in} = \beta^T x_{in} = \sum_{k=1}^{K} \beta_k x_{ik}$$  \hspace{1cm} (2)

In the formula, $x_{ik}$ is the NO. $k$ eigenvalue of driver $n$ when choosing scheme $i$; $\beta_k$ is the eigenvalue parameter of NO. $k$ characteristic variable to be calibrated, $\beta_k'$ is the equivalent coefficient of $\beta_k$; and $K$ is the sum of eigenvalues. According to the theory of stochastic utility maximization, assuming that the set of all options for driver $n$ to choose toll lane is $A_n$, where the utility of scheme $J$ is $U_{jn}$, the condition for driver $n$ to choose scheme $i$ in $A_n$ as follows:

$$U_{in} > U_{jn}, i \neq j, j \in A_n$$  \hspace{1cm} (3)

The expression of the maximum utility function can also be written as follows:

$$U_{in} = \max_{j=1,2,...,J} (V_{jn} + \epsilon_{jn})$$  \hspace{1cm} (4)

At this point, the probability $P_{in}$ of driver $n$ choosing toll lane scheme $i$ can be written as follows:

$$P_{in} = \text{prob}(U_{in} > U_{jn}, i \neq j, j \in A_n) = \text{prob}(V_{jn} + \epsilon_{jn} > V_{jn} + \epsilon_{jn}; i \neq j, j \in A_n)$$  \hspace{1cm} (5)

Aiming at the object of this study, it is defined that two values of $I$ (accepting the toll lane) and $\theta$ (rejecting the toll lane) for each toll lane when the driver drives to the toll station are in the selection scheme set $A_n$, and then uses the binary Logit model to analyze the probability. The probability that the driver chooses to accept or reject a toll lane is obtained as follows:

$$P_{in} = \frac{e^{\beta_i V_{in}}}{e^{\beta_i V_{in}} + e^{\beta_0 V_{in}}} = \frac{1}{1 + e^{-(\beta_i - \beta_0)V_{in}}}$$  \hspace{1cm} (6)

$$P_{in} = 1 - P_{in} = \frac{e^{\beta_0 V_{in}}}{e^{\beta_i V_{in}} + e^{\beta_0 V_{in}}} = \frac{1}{1 + e^{-(\beta_i - \beta_0)V_{in}}}$$  \hspace{1cm} (7)

In the formula, $P_{in}$ is the probability of driver $n$ choosing a toll lane scheme $i$ ($i=1,2$) and $V_{in}$ is the utility fixed value of driver $n$ choosing a toll lane scheme $i$ ($i=1,2$).

2.2. Feature variables

The toll plaza is a queuing system with multiple service desks, which belongs to multiple random streams. When the vehicle enters the toll plaza, it will choose the toll lane with idle or relatively short
queue; Gulewicz and Danko[4] explained the driver's behavior habits at the toll booth. It is suggested that the driver will be more inclined to choose the toll lane with shorter queue length and less lane change. So the queue is an important factor affecting the driver's toll lane selection. During the journey of the vehicle into the toll booth, the driver has to change to the target lane. The lane change behavior increases the possibility of collision of the vehicle, and the more the number of lanes changed, the possibility of collision and the driver's operational load increased[7]. Ozbay[5] et al. suggested that the location of the vehicle before entering the toll booth is a factor influencing the choice of toll lanes. Therefore, the number of lane changes is determined as the characteristic variable of the driver's toll lane selection.

When a car is mixed with large vehicles, the psychological pressure of the car driver will increase. At the same time, large vehicles have slow start response and large space, which reduce the capacity of the toll lanes[8]. In view of the impact of large vehicles on the toll booths, the number of large vehicles in the queue is also a characteristic variable.

According to the existing research[2], the driver prefers to choose the toll lane near the central separation zone, because it will reduce the bypass distance. Therefore, the location of the toll lane is considered as a characteristic variable of the driver's toll lane selection.

Based on the above analysis, by substituting these characteristic variables into equation (2), a linear utility function expression can be obtained:

\[
V_{in} = \beta_0 + \beta_1 X_{1}(t) + \beta_2 X_{2}(t) + \beta_3 X_{3}(t) + \beta_4 X_{4}(t) \tag{8}
\]

Where \(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4\) are unknown coefficients; \(X_{1}(t)\) is the queue length of the toll lane; \(X_{2}(t)\) is the number of lane change that the driver needs to drive to the toll lane; \(X_{3}(t)\) is the number of large vehicles in the toll lane queue; \(X_{4}(t)\) is the location of the toll lane from the central divider. By substituting the utility function equation into the driver's acceptance/rejection of a toll lane probability expression, it is obtained that:

\[
P_{in} = \frac{1}{1 + \exp(\beta_0 + \beta_1 X_{1} + \beta_2 X_{2} + \beta_3 X_{3} + \beta_4 X_{4})} \tag{9}
\]

\[
P_{0n} = 1 - P_{in} \tag{10}
\]

3. Instance calculation

3.1. Research section

Through on-the-spot investigation, the survey data of Qinzhou West Toll Station on the Lanhai Expressway was selected for application analysis. The toll booth layout is shown in Figure 1.

![Figure 1. Floor plan of toll plaza.](image)

The toll booth has eight toll lanes, of which the toll lanes 1 and 2 are non-stop toll lanes, and the toll lanes 7 and 8 are closed during data collection. Therefore, in this paper, four charging lanes of 1, 2, 7, and 8 are not considered in the data collection, and only four artificial toll lanes of 3, 4, 5, and 6 are collected.
3.2. Data collection
The driver's toll lane selection behavior is observed at the observation site. The vehicle movement process of each toll lane is recorded at the position shown in Figure 1. According to the actual observation, the driver will generally select the toll lane before the toll island 100m to 30m, so the following data should be recorded in the actual observation: The location of the toll lane; The queue length of each toll lane when the vehicle is 100m away from the toll island; The lateral position of the vehicle when it is 100m away from the toll island; The number of large vehicles in each toll lane when the vehicle is 100m away from the toll island; The toll lane that the driver finally chooses to enter.

In the regression analysis model, variables are affected by qualitative variables and quantitative variables. For some independent variables that cannot be quantitatively measured, it is called a dummy variable[9]. When recording the location of each toll lane, the dummy variable is used to represent the position of each toll lane, and the calculated regression result has a reasonable and clear meaning.

3.3. Model parameter estimation
To solve the parameter estimation value of each characteristic variable, that is, to calculate the maximum likelihood estimation value \( \hat{\beta} \) of the model likelihood function \( L \). For the \( N \) samples of this field observation, the probability that the driver \( n \) selects the scheme \( i \) is \( P_{in} \), and assuming that the selection result is \( \delta_{in} \), then:

\[
P_{in} = P_{in}^{\delta_{in}} P_{0n}^{\delta_{0in}}
\]  

(11)

In the formula, \( \delta_{in} = i_0 \), when the selection result is the same as \( i \), \( 1 \) is taken, and when the selection result is different from \( i \), \( 0 \) is taken. From this, a likelihood function expression can be obtained:

\[
L = \prod_{n=1}^{N} P_{in}^{\delta_{in}} P_{0n}^{\delta_{0in}}
\]  

(12)

The maximum likelihood estimation method is used to simultaneously calculate the parameter \( \beta \) at which the likelihood function \( L \) is maximum. Since the logarithm \( L' = -lnL \) is the largest when \( L \) is maximum. Therefore, the parameter \( \beta \) at which the likelihood function \( L \) is maximum can be obtained by estimating the parameter \( \hat{\beta} \) at the maximum \( L' \). The logarithm expression of the likelihood function is:

\[
L' = \sum_{n=1}^{N} (\delta_{in} \ln P_{in} + \delta_{0in} \ln P_{0in})
\]  

(13)

The \( K \times K \) order Hessian matrix is established according to the logarithm expression of the likelihood function, and solve the \( K \) order simultaneous nonlinear equations when the gradient vector is zero. Using the Newton-Raphson (NR) method and the DGP method (Davidon-Fetcher-Powell method) for numerical calculation, the parameter estimates of the maximum likelihood function can be obtained.

3.4. Data sorting
A charging lane selection process for 487 passenger cars was collected at the observation site. According to the original data collected at the site, the percentage of the driver's choice of the toll lane under different characteristic variables is obtained, as shown in the following figure.
Figure 2. Driver’s choice percentage for different queue lengths

Figure 3. Driver’s choice percentage for different times of lane change

Figure 4. Driver’s choice percentage for different large vehicle numbers

Figure 5. Driver’s choice percentage for different toll lane locations

The statistical analysis calculation is performed by the maximum likelihood estimation method described above, the calculation and analysis results of the respective statistics are shown in Table 1.

| Influence Variable | $\beta$  | S.E.  | Wald  | df   | Sig.  | Exp($\beta$) |
|--------------------|---------|-------|-------|------|-------|--------------|
| $x_1$              | -1.077  | 0.112 | 92.007| 1    | 0.000 | 0.341        |
| $x_2$              | -0.708  | 0.109 | 41.839| 1    | 0.000 | 0.493        |
| $x_3$              | -0.526  | 0.211 | 6.233 | 1    | 0.013 | 0.591        |
| $x_4$              | 16.498  |       | 1     |      |       |              |
| $x_4=3$            | 0.399   | 0.315 | 16.071| 1    | 0.021 | 1.491        |
| $x_4=4$            | 0.226   | 0.355 | 4.040 | 1    | 0.042 | 1.253        |
| $x_4=5$            | -0.654  | 0.268 | 5.955 | 1    | 0.015 | 0.520        |
|                    | 2.273   | 0.378 | 36.193| 1    | 0.000 | 9.711        |

Note: $\beta$ is the regression coefficient of the characteristic variable; S.E. is the regression standard deviation; Wald is the chi-square value; df is the degree of freedom; Sig. is significant, when Sig. < 0.05, the variable influence is significant, when Sig. > 0.05, the effect of the variable is not significant; Exp($\beta$) is the index of the coefficient $\beta$.

3.5. Goodness of fit

After the model parameters are calculated, it is necessary to evaluate the effectiveness of the model. In this paper, the Hosmer-Lemeshow test is used for analysis. The null hypothesis is that 10 decile groups are applied to the fitting probability $P$, the difference between the fitted value and the observed value in each group should be small. The statistic is approximately a chi-square statistic of $df=8$. The test result is shown in Table 2.

| Step | Chi-square | df | Sig  |
|------|------------|----|------|
| 1    | 8.443      | 8  | 0.391|

The Sig value evaluation standard of the Hosmer-Lemeshow test is different from the Sig value evaluation standard of the binary Logit model. In the test, the fitting effect is better when Sig > 0.05. It can be seen from Table 2 that the Sig>0.1, indicating that the fitting result of the model is great.
4. Conclusion

According to the parameter estimation results of the binary Logit model, the relationship between the driver's behavior selection probability and the characteristic variable is established, namely:

\[ V_{in} - V_{on} = 2.273 - 1.077x_1 - 0.708x_2 - 0.526x_3 + 0.399(x_4 = 3) + 0.226(x_4 = 4) - 0.654(x_4 = 5) \]  

(13)

\[ P_n = \frac{1}{1 + e^{[2.273 - 1.077x_1 - 0.708x_2 - 0.526x_3 + 0.399(x_4 = 3) + 0.226(x_4 = 4) - 0.654(x_4 = 5)]}} \]  

(14)

The following analysis results can be obtained by studying the calculation results:

1. From the results of parameter estimation, it can be seen that the estimated values of queue length, number of lanes changed and number of large vehicles are all negative, which is consistent with the theoretical analysis results. The utility values of the two lanes closest to the central line of the road are both positive, while the estimated values of parameters of toll lane 5 are negative. This indicates that drivers are more inclined to choose toll lanes directly connected to the straight lanes.

2. From the parameter estimation of each characteristic variable, it can be seen that the absolute value of queue length is greater than the absolute value of other parameters, and the influence is significant, which shows that queue length is the primary factor affecting drivers' choice of toll lane.

3. The number of changing lanes is the second important factor. If the number of changing lanes increases by 1, the utility value of the toll lane will be reduced by 0.708, and the probability of drivers choosing the toll lane will be 0.493 times that of the original one.

4. Although the number of vehicles choosing this toll lane significantly decreases when there are large vehicles in the queue, the influence of the number of large vehicles on the toll lane is smaller than the queue length and the number of lanes changing.

References

[1] Chen, H., Wu, X. (2005) Traffic Safety Evaluation of Expressway Toll-Gate. In: Proceedings of the 5th International Conference on Transportation. Xian. pp. 594-599.

[2] Zhang, G., Zhang, C., Cao, M., et al. (2015) Selection of Highway Toll Channels Based on Evolutionary Game. Journal of Transportation Systems Engineering and Information Technology, 15: 29-35.

[3] Komada, K., Nagatani T. (2010) Traffic Flow Through Multilane Tollbooths on a Toll Highway. Physica A: Statistical Mechanics and Its Applications, 389: 2268-2279.

[4] Gulewicz, V., Danko, J. (1995) Simulation-based Approach to Evaluating Optimal Lane-staffing Requirements for Toll Plazas. Transportation Research Board, 33-39.

[5] Ozbay, K., Mudigonda, S., Bartin B. (2005) Development and Calibration of an Integrated Freeway and Toll Plaza Model for New Jersey Turnpike Using Paramics Microscopic Simulation Toll. Piscataway, 8: 13-16.

[6] Guan, H. (2004) Disaggregate Model: A Toll of Traffic Behavior Analysis. China Communications Press, Beijing.

[7] Zhang, Z. (2014) Research of Traffic Conflict of Freeway Diverging and Merging Sections and Toll Booth. Harbin Institute of Technology, Harbin.

[8] Wang, H. (2015) Study on Management and Control Strategy of Toll-Gate with High Occupancy of Freight Vehicle. Hebei University of Technology, Tianjin.

[9] Li, H., Wang, K., Hu, M., et al. (2017) Relative Importance Index of Dummy Variables in Regression Model. Journal of Computer Applications, 37: 3048-3052.

[10] Sadoun, B. (2005) Optimizing the Operation of a Toll Plaza System Using Simulation: A Methodology. Simulation, 81: 657-667.

[11] Ji, Y., Zhou, J. (2018) Lane Allocation of Highway Toll Gate Based on Cost Analysis. Journal of Chongqing Jiaotong University(Natural Science), 37: 85-91.