ANALYSIS OF RESIDENTS' TRAVEL CHARACTERISTICS ALONG BEIJING RAIL TRANSIT LINE BASED ON BINARY CHOICE MODEL

Zhaoyang CAI¹, Jianwei YAN²
¹,² Architecture School, Tianjin University, Tianjin, China

Abstract:
From ancient to modern times, in the historical process of urban development, urban transportation has been developing along with the development of the city's political, economic and cultural industries, and the relationship between each other has always been a topic constantly discussed by planning scholars. The development of urban transportation promotes the urban population gathering and industrial development, and promotes the development of urban geographical space. At the same time, it also brings endless convenience to urban residents, so that they can complete the displacement from the beginning to the destination with relatively low cost. However, with the rapid development of urban scale and the rapid growth of urban population, the problems of traffic congestion and land resource shortage in big cities of China seriously restrict the improvement of the quality of life of residents and the further development of the city. In this context, compact city is the inevitable choice for future urban development, while the transportation system supporting compact city form can only be public transportation. As a high-volume, efficient and rapid public transport mode, rail transit can not only solve the traffic congestion problem in high-density areas of cities, but also optimize the development and utilization of urban land and adjust the urban spatial layout, which is of great significance for the sustainable development of cities. The network of rail transit in Beijing becomes the backbone of public passenger transport system, and play an irreplaceable role in guiding the urban space layout adjustment, population migration and the transformation of traffic structure. The study of travel characteristics of the residents who live along the rail transit, in-depth analysis the relationship of the rail transit, the population migration and commuter travel, it is of great significance providing decision support for urban planning. Based on binary choice model, establishing the residents' travel choice model, rail transit impact model of different crowds. Study the relationship between rail transit and residents' travel characteristics, predict the rail transit to guide the trend of population migration.

Key words: rail transit, travel characteristics, site attraction area, binary choice model

To cite this article:

Article is available in open access and licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0)
1. Introduction
With the development of urbanization, the urban space and the urban population are expanding rapidly. For a long time, the city uses the car as the center as the transportation development strategy, the residents' travel brings great pressure to the urban traffic, and the changes of residents' travel characteristics are getting more and more attention. Since the middle of the 20th century, especially after the sustainable development of ecological view widely recognised, western countries began to realize that the car is at the root of many diseases, and paid more attention to public transport. In the works of Garling et al. (2007), COM (2007, 2011) especially the development of rail transit, successively put forward the urban development concepts such as the compact city, TOD, and so on, thus changed the urban space development pattern, and affected the urban space layout and structure. In urban areas, commuting has become a feature of urban traffic. When this kind of daily fixed time, fixed direction of people move more and more and become too concentrated, it will form travel peak, bringing great pressure to urban traffic as Yanfang et al. (2010) mentioned. Networked development of rail transit is the common choice of world's big cities to improve the traffic environment of cities, the rail traffic network in the construction of high density is fundamentally one of the most effective way to solve the problem of urban traffic congestion. In the study of Loukopoulos et al. (Loukopoulos, 2007; Meyer, 1999; Stradling, Meadows and Beatty, 2000; Taylor and Ampt, 2003), rail transit as the backbone of our public passenger transport system, shaping demand for alternative transport modes to private cars, in the new round of urban space development, industrial layout adjustment will play an irreplaceable role, and become the important influencing factors that cannot be ignored when residents choose their traveling way.

The development of urban rail transit in China is not yet mature, and the rail transit network is also moving forward. The construction of Beijing rail transit network caters to the development trend of Beijing's "two-axis—two-belt—multi-center" urban pattern planning. There are some problems in the actual development and operation of Beijing rail transit:

(1) The synchronous development of urbanization and suburbanization in Beijing, the current development of the old city center is relatively intensive, and the population of the central city is out-migration, and the new urban population is growing rapidly in the outskirts of the city. The dense construction of the old city and the disordered development of the suburbs have caused the poor quality of the community environment, which has had a negative effect on the transportation system and the ecological environment as Jacyna et al. (Jacyna and Merkisz, 2014; Jacyna, Wasiak, Lewczuk and Klodawski, 2014) mentioned. In addition, due to the lack of urban planning theory of some planners, lead to rail way transportation construction and urban development planning failed to take effective development strategy from the macroscopic aspect, and caused in the process of urban construction too much focus on solving the local problems of urban construction, ignoring the coordinated development of land use and urban traffic, Lee et al. (Lee, Kim, Rho, et al., 2016; Paz, Suarez, Gil, et al., 2015) thought the construction of land development and rail transit is relatively isolated and cannot be implemented synchronously. Therefore, rail transit cannot maximize the service of urban development, and the land along the rail transit line has not reached the maximum development and utilization.

(2) The commutes traffic pressure is huge in Beijing, and the traffic congestion is serious in the morning and evening. The total amount of travel and the distance of travel are increasing year by year, and the phenomenon of "tidal" of residents' commuting trips is obvious. Due to land development along the rail transportation is slow, and the supporting facilities in the new town construction area are not perfect, the real travel demand of residents cannot be met. As a result, people do not choose rail transit as the main mode of transportation. Therefore, the resources of the facilities along the rail transit cannot be fully utilized, the pressure on urban traffic caused by work travel has not been alleviated effectively as well, these will seriously affect the sustainable development of Beijing. Therefore, it is of great significance that master the law of the selection and distribution trend of urban residents to promote the formation of new pattern in Beijing.

(3) Urban rail transit has obvious agglomeration effect on urban development. Because of the improper development of surrounding land, the
passenger flow volume in some remote sites is scarce. The utilization of the site is not high, and there is no new attraction. Therefore, it is necessary to fully explore the location and characteristic advantages of different sites, make reasonable use of its surrounding resources, and realize the benign interaction between the site and surrounding resources. For different station domains, to strengthen the mixed use of land around the site, and gradually achieve urban public transport priority development in the process of urban development, improve the traffic problem, make coordinated development of urban land development and traffic system construction.

As the capital and cosmopolitan metropolis of China, the development of Beijing is faced with the above situation. With the rapid development of suburbanization in Beijing, the problem of spatial distribution imbalance is prominent. The urban traffic pressure is huge, which seriously disturbs the sustainable development of Beijing (Jian and Yixing, 2004). Fanyu and Wenbin (2007) proposed the "tidal" phenomenon of residents’ commuting trips is obvious, and the "centripetal" traffic in the early peak accounts for more than the total amount of transportation. Chmielewski et al. (Chmielewski and Szczuraszek, 2004; Jacyna, 1999) proposed the rapid development of rail transit which can solve the passengers streams in public transportation systems and cargo streams in transportation systems will gradually become the main force of Beijing transportation. In this paper, qualitative and quantitative methods are both used to study. Tuansheng (2007) put forward according to the characteristics of the urban area of Beijing rail transit along the representative sample area for related data collection, on the basis of statistical analysis, analysis of urban rail transit network development impact on urban residents' travel choice, based on the software to describe residents' travel characteristics.

2. Binary choice model

The distribution of large numbers of people in rail transit station provides a source of development vitality for the region, urban areas centered on rail transit stations has become an important development area of the city, no matter from capacity, layout, or public space organization have the particularity as Jacyna et al. (Jacyna, 2009; Yuan, Wu, Feng, et al., 2014; Zochowska and Karon, 2016) mentioned. Through the quantitative analysis method and the model construction, expect the quantitative analysis method with big data opening as prospect can be support tools for planning prediction and rational decision making, make the affinity of people's daily travel and activities become important principles and the basis for building sustainable cities. This model mainly studies the influence of objective conditions and subjective conditions on residents’ choice of rail transit. Model based on the results of the survey and combine with the basic theory, comprehensive consideration about the factors that affect the choice of rail transit travel, prevent the problems that due to the lack of explanatory variables caused by the main factors influencing the residents' travel choice regression results appear too high or too low estimation. Because the model is explained in the variable is a binary choice, namely whether people choose rail transit travel, so set the binary choice model, because people often choose the way to travel is for the sake of utility maximization angle, so use the binary Logit choice model of binary choice model.

2.1. Utility model of binary choice

In classical models, explanatory variables are usually assumed to be continuous variables. However, it is often faced with a number of decision issues, or choice issues, that people must choose among the options available. With such decision results as the model established by interpreted variables, it is called discrete variable data model, or discrete selection model. If there are only two alternatives to an explanatory variable, it is called a binary choice model. Therefore, for modelling of the residents' travel choice that takes into account in urban transportation systems the following description is required (Ambroziak and Jacyna, 2002; Jacyna, 2009; COM, 2007)

In order to make the study of binary choice problem possible, firstly we must establish a stochastic utility model. Set \( U_i^1 \), \( U_i^0 \) as the probability of the first choice result (\( y = 1, y = 0 \)) of the first individual \( i \), that is, the effect of various property variables of the individual on the selection result. The utility is unobservable, and the observed value we can get is still the selection result, which is 1 and 0. But obviously, when the utility value is \( U_i^1 > U_i^0 \),
that corresponds to the observed value \( y = 1 \); In contrast, when the utility value is \( U_i^1 \leq U_i^0 \), it corresponds to the observed value \( y = 0 \). \( y_i^* \), \( X_i \), \( B \), \( \mu_i^* \), respectively explained variables and explanatory variables, estimated parameters and random error terms for the model. The probability of individual selection is:

\[
P(y_i = 1) = P(y_i^* > 0) = P(\mu_i^* > -X_iB) \tag{1}
\]

2.2. Maximum likelihood estimation

If the utility model of binary choice can be estimated, a particular probability distribution must be selected for \( \mu_i^* \). The two most commonly used distributions are standard normal distribution and logistic distribution, so two of the most commonly used binary choice models - Probit model and Logit model - are formed. Whether it's a standard normal distribution or a Logistic distribution, they're symmetric. The likelihood function of the utility model of binary choice is:

\[
L = \prod_{i=1}^{n} (F(X_iB))^y_i (1 - F(X_iB))^{1-y_i} \tag{2}
\]

The first order condition of logarithmic likelihood function is:

\[
\frac{\partial \ln L}{\partial B} = \sum_{i=1}^{n} \frac{y_i f_i}{F_i} + \frac{(1 - y_i)}{(1 - F_i)} X_i = 0 \tag{3}
\]

\( f_i \) is the probability density function. Obviously, with the support of the sample data, if the logical probability distribution function \( F(t) = \frac{1}{1 + e^{-t}} \) and the probability density function \( f(t) = \frac{e^{-t}}{(1 + e^{-t})^2} \) are known, the model parameter estimation can be obtained by solving the system.

2.3. Binary Logit selection model and parameter estimation

The Logit model is derived by using the logistic distribution as the probability distribution of \( u_i^* \) in the utility model of binary choice. If the choice is made according to the utility maximization, the logical distribution with the limit value is a better choice. In this case, the binary choice model should adopt the Logit model. Logit model based on Logistic probability distribution function:

\[
p_i = F(y_i^*) = F(X_iB + \mu_i) = \frac{1}{1 + e^{-(X_iB+\mu_i)}} \tag{4}
\]

For a given \( X_i \), \( p_i \) is the probability that the corresponding individual makes a certain choice. \( 0 \leq p_i \leq 1 \), and when \( p_i \) in the condition of \( x \to \pm\infty \), the change is slower and slower.

For the binary logit selection model, the model parameters can be estimated using maximum likelihood method. Likelihood function:

\[
L(B) = P(y_1, y_2, \ldots y_N) = P(y_1)P(y_2)\cdots P(y_N) = p_1 \cdots p_n(1-p_{n+1}) \cdots (1-p_N) \tag{5}
\]

\[
= \prod_{i=1}^{n} p_i \prod_{i=n+1}^{N} (1-p_i)
\]

The logarithmic likelihood function is:

\[
\ln L(B) = \sum_{i=1}^{n} \ln p_i + \sum_{i=n+1}^{N} \ln(1-p_i) \tag{6}
\]

The partial derivative \( b_i \) of the above equation is obtained, and it is 0, and the maximum likelihood estimator \( b_i \) can be obtained by solving the system.

In the actual calculation, we can use professional econometric software (such as stata) to return the binary Logit selection model to obtain the parameter estimation.

2.4. Significance test of regression parameters

In order to better explain the influence of subjective factors and objective factors on the choice of rail transit travel, multiple linear regression models can be established to find out the relationship between travel conditions and various factors. Due to the limited number of samples in the survey, the conclusion is unstable, so in multiple linear regression analysis, if we can prove that an independent variables \( x_i \) to \( y \) the linear relationship was not significant, we can remove the arguments \( x_i \) from the regression equation, to establish the regression equation, can improve the
whole model of optimal benign. In multiple linear regression, the regression equation of each independent variable on the dependent variable does not mean that significantly impact is significant, so need to each factor, including constant term significance test (t test), at the same time, the variance expansion factor VIF is used to judge whether there is multiple collinear in the model. If the variance expansion factor VIF>10, the independent variables in the model have a high degree of co-linearity, that is, the self-correlation between the independent variables is strong, and the independent variables need to be examined and eliminated. t test shows that the influence of constants and their variables on the dependent variable is significant, the common linear test shows that there is no multiple collinear problem between each variable, then the equation is established.

(1) Put forward the original hypothesis \( H_0 : \beta_i = 0 \);
The alternative hypothesis \( H_1 : \beta_i \neq 0 \).

(2) Structural statistic \( t = \frac{\hat{\beta}_i - \beta_i}{S(\hat{\beta}_i)} \). When \( \beta_i = 0 \) is formed, the statistic \( t = \frac{\hat{\beta}_i}{S(\hat{\beta}_i)} \sim t(n-k-1) \).

\( S(\hat{\beta}_i) = \sigma(\hat{\beta}_i) = \sqrt{C_oS^2} \) is the standard deviation of \( \hat{\beta}_i \); \( S^2 = \frac{e'e}{n-k-1} \); \( k \) is explanatory variable.

(3) Given significant level to \( \alpha \), check \( t \) distribution table that the degree of freedom of \( n-k-1 \), get the critical value \( t_{\alpha}(n-k-1) \).

(4) If \( |t| \geq t_{\alpha}(n-k-1) \), then refused to \( H_0 : \beta_i = 0 \), accept \( H_1 : \beta_i \neq 0 \), the idea that \( \beta_i \) significantly to zero.

3. Site attraction area calculation method
Residents use shuttle transportation rail transportation can be divided into two categories, one kind is walking directly into the rail transit system, the other is a mode of transport other than by walking to rail transit system. Whether or not the traveler chooses rail transit is largely determined by the ease with which the traveler arrives at the site. Site attraction area is the acceptable area for travelers to reach the rail transit site by some means of transportation, in the study of Jingshuai and Dianye (2008) which can be divided into walking attraction area and traffic attraction area. The short distance travel often chooses the dense block, to walk primarily. Middle distance travel generally needs to go through a few blocks, mainly by bicycle, so the road connecting with the block is preferred. Long-distance travel is mainly for motor vehicles, and urban expressway is the first choice. Therefore, it can be used to express the accessibility value of the analysis radius of 500m, 2000m and 5000m respectively by means of pedestrian network, bicycle network and motor vehicle network respectively. Through the analysis of the relationship between traffic network and traffic flow distribution in rail transit station, found that more than 70% of the rail in passenger traffic is determined by the accessibility of rail system, more than 60% of the motor vehicle flow influenced by road accessibility, while 50% of road traffic by rail transit accessibility and decide together walking accessibility.

The index that measures the attraction area of the site is the weighted attraction distance under the common action of various modes of transportation. Due to the difference of road network structure, the distance of the transport is generally greater than the straight distance from the point of departure to the station. In general, the rail transit station surrounding road network structure for the grid layout or a diagonal channel grid layout, under the assumption that the direction of travel is under the same proportion of two kinds of grid layout of the conversion coefficient calculation as follows.

\( L \) means the distance from the site, \( Z \) is the conversion coefficient. The weighted average transport distance of the station is:

\[
L_{\text{transport}} = \sum_{i=1}^{n} p_i \times l_i
\]  

(7)

In the above equation, \( L_{\text{transport}} \) is the weighted average distance of the station, \( p_i \) is the sharing rate of all kinds of transportation connections, and \( l_i \) is the average connection distance of the connection mode.
The station weighted average distance $L_{transport}$ and the commutation coefficient $Z$ can calculate the reasonable attraction area of the station $L_{attraction}$ (the linear distance between the place of residence and the station) is as follows:

$$L_{attraction} = L_{transport} \times Z \quad (8)$$

4. Residents' rail transit travel choice model samples and regression results

The interpreted variable $y$ in the model are the people's choice, $y = 1$ means using the rail transit to travel, $y = 0$ means not the rail transit. As shown in Fig.1, the passenger volume and passenger flow intensity of 18 lines in Beijing. Explanatory variables $x$ mean factors will respectively influence whether the residents choose rail transportation (whether $y$ is equal to 1), each factor influences the utility of rail travel but does not directly determine whether residents choose rail transportation, such as scale factors (population, housing and employment, etc.), density factors (population density, the density of land use, etc.), design factors (block space, business services, etc.), layout factors (urban structure, land use structure, etc.). On the basis of relevant scholars, it is concluded that land use density is the main factor affecting urban traffic. With the increase of land use density, the number of traffic trips will be reduced, but the decrease of travel speed may lead to the increase of travel distance. And the degree of land mixing has a weak effect on the type of transportation. In addition, the change of land use density has a profound impact on traffic flow. Under the influence of decentralization, Degang et al. (2005) investigated the probability of residents choosing public transportation is reduced from 61% to 46%. The feedback loop mechanism of land use and transportation is often used to explain the dynamic relationship between land use, transportation facilities and accessibility changes. Because different land properties have different requirements for accessibility, the transportation system can adapt to the functional layout and development intensity by changing the network layout or increasing or decreasing the traffic facilities. The change of traffic facilities brings about the change of accessibility pattern, thus changing the distribution mode of user activity and the value of land, and thus affecting the further decision of investors on land use. This cycle continuously until the three balance or intervention with external factors, generally speaking, the new traffic facilities can bring immediate accessibility change, but the accessibility of the impact of land use has a relatively long time lag effect, so the feedback loop mechanism of land-traffic-accessibility usually need a long time to reach equilibrium.

When the effect of all kinds of factors is more than zero, residents will decide to choose rail transportation, namely $y = 1$; When the effect of all factors is less than 0, the residents will not choose the rail transit, namely $y = 0$. Therefore, there is no linear relationship between the influence factors and $y$ is 0 or 1, but there is a linear relation between the influence factors and the utility of rail transportation, and the last option is determined by the utility $y^*$ of the positive and negative, can build the following relations:

$$y^* = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_7x_7 + a_8x_8 + a_9x_9 + a_{10}x_{10} + a_{11}x_{11} + a_{12}x_{12} + a_{13}x_{13} + a_{14}x_{14} + a_{15}x_{15} + a_{16}(x_5)^2 + a_{17}(x_4)^2 + a_{18}(x_5)^2 + a_{19}(x_{12})^2 + v_i \quad (9)$$

Among them, $v_i$ obey Logistic distribution, when $y^* > 0$, $y = 1$; when $y^* < 0$, $y = 0$. Therefore, the probability of residents choose traveling by rail:

$$P(y = 1 | x) = P(y^* > 0 | x) = P(v_i > -Xa) = 1 - F(-Xa) = F(Xa) \quad (10)$$

Among them, $X = [x_1, x_2, \cdots x_n]$, $a = [a_1, a_2, \cdots a_n]$

The maximum likelihood method is used to obtain the coefficient size, and the regression coefficient can be obtained by the maximum likelihood estimation. The coefficient $a$ of the initial regression is the impact size of $x$ to $y^*$, and the marginal impact $b$ that $x$ to $P(y = 1)$ is also calculated. The model can be written as:
\[ P(y = 1) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_6 + b_7 x_7 + b_8 x_8 + b_9 x_9 + b_{10} x_{10} + b_{11} x_{11} + b_{12} x_{12} + b_{13} x_{13} + b_{14} x_{14} + b_{15} x_{15} + b_{16} (x_6)^2 + b_{17} (x_7)^2 + b_{18} (x_8)^2 + b_{19} (x_{12})^2 \]  

As shown in Fig.2, the coefficient of residents' rail transit travel choice. After obtaining the regression results of the model, the regression degree of the model needs to be judged, and the accuracy rate is tested in this paper. The accuracy rate refers to the result of regression, which brings each sample data into the regression model, calculates whether \( y \) is 1, and then compares it with the actual observation value, so as to judge the accuracy rate of the prediction. Regression analysis was performed on all influencing factors using total sample travel data, and the regression results were obtained. The predictive accuracy rate of this model is 94.05%.

Fig. 1. Passenger volume and passenger flow intensity of Beijing rail transit lines

Fig. 2. Passengers flow in and out rail transit stations peak hour coefficient
5. Discussion and conclusion
Based on the binary choice model, this paper establishes the model of the choice of the residents' rail transit, and analyzes whether the residents choose the rail transit. The overall prediction accuracy is high, which indicates that the model setting and considering factors are in line with the reality. The closer the place of residence or workplace is to the rail transit station, the more likely it is to travel by rail. If the residents have to transfer to get to the destination, the more times, the greater the probability that the residents will choose the rail transit. The commuting consumption and the distance between work and residence, and rail travel is in parabolic relation. This paper mainly from the angle of the micro analysis of various factors on the influence of the residents' travel choice, analyze from the density of land use, social and economic attributes of the individual citizens, commuter travel condition and land use structure, etc. Further research could analyze comprehensively by adding policy guiding, residents factors such as social ideology and the nature of urban land use, with application in regional population forecast, population distribution along the rail traffic prediction, rail transit passenger flow forecast, etc. The establishment and development of the public transport system of rail transit will inevitably have an important influence on the structure of urban traffic. Rail transit has outstanding advantages that other modes of transportation do not have, and can synchronize with the development of the urban population and urban scale. It is a good way to connect with other flow points and guide urban layout adjustment. The construction of high-density rail transit network is one of the most effective ways to solve the problem of urban traffic congestion. The spatial structure of urban layout in Beijing has gone through the process from centralization to sprawl, rail transport as the main body of public passenger transport system will play an irreplaceable role in the new round of urban space development and industrial layout adjustment, and become the residents travel mode choice important influencing factors that cannot be ignored. China's urban traffic should gradually realize the priority development of public transport, the mixed-use of urban land use, the coordinated development of urban transportation and land use. Gradually adjust the spatial structure of the city, establish a comprehensive transportation system with public transportation as the core, and integrate urban spatial layout and resource allocation, clear the development direction of urban transportation based on rail transit, form a new urban space structure and new city construction development mode with rail transit as the axis, it will become an important issue to promote the sustainable development of Beijing.

Acknowledgment
The research is carried out under the project "Study on culture communication of metro station in new media era" (17BXW062) funded by the National social science foundation.

References


