Study on activity-based time choice model

J. Zhicai[1], Z. Fang[2] & L. Zhiyao[2]
[1]*Institute of transportation, Shanghai Jiaotong University, China*
[2]*Department of Transportation & Traffic, Jilin University, China*

Abstract

In this paper, the framework of an activity-based travel demand forecasting model is presented. By analysing the factors which affect the residents' travel time decisions, two time choice models which can predict the start and end times of a trip respectively are developed. The models are estimated and the sensitivity of the model is analysed from the 2003 Household Travel Survey in Changchun. With the designed models, we predict how people adjust their start time when the travel cost by auto has been added. Finally the weaknesses of the current system are summarized, and the way to improve the approach of the time forecast is put forward.
Keywords: travel demand, trip, tour, disaggregate model, alternative, intermediate stop, forecast, choice, utility, choice set.

1 Introduction

In order to alleviate the serious traffic congestion in urban areas, in recent years, researchers have paid more and more attention to Travel Demand Management (TDM). Since 1970's, many policies, such as increased peak-hour commuter tolls in the CBD area, flexible work hours, reduced transit fares for off-peak transit users, have been put forward by Europe, US, Japan and other countries. We can learn that many policies are adapted to decrease traffic pressure during morning and evening peak time by decentralizing travel demand on time. But its premise is to master the law of the residents' travel time, forecast the characteristics of the travel time distribution, and provide basis for the establishment and implement of TDM. Thus, in order to avoid misadvising the residents' travel, it is necessary for planners to predict the time demand effectively.

www.witpress.com, ISSN 1743-3509 (on-line)

2 Travel demand forecasting

Travel demand forecasting is the premise of Travel Demand Management. Based on survey and analysis of the travel survey data, it can predict the residents' travel schedules, and evaluate the feasibility of TDM policies. The widely applied forecasting method is the disaggregated choice model based on the theory of the utility maximizing consumer, developed in 1970's. It can be estimated with individual and household data, and describe an individual's travel decision explicitly. In 1990's some researchers develop the daily activity schedule model, which is a kind of activity-based model system, to forecast the daily activity arrangement, the start and end times, and the travel mode of each trip as well [1]. In this paper we present a three-level nested Logit model, as shown in fig. 1, to represent the overall structure of activity-based model system. As a sub-model of the system, the structure of the activity pattern model is shown in fig. 2. Being a Nested Logit (NL) model itself, the activity pattern model consists of four sub-models: primary activity model is designed to forecast the primary tour in the day; secondary activity model is defined as the other tours in the day; intermediate stop model predicts all stops in a tour other than the primary destination; and work-based subtour model predicts a sequence of trip segments that both start and end at work.

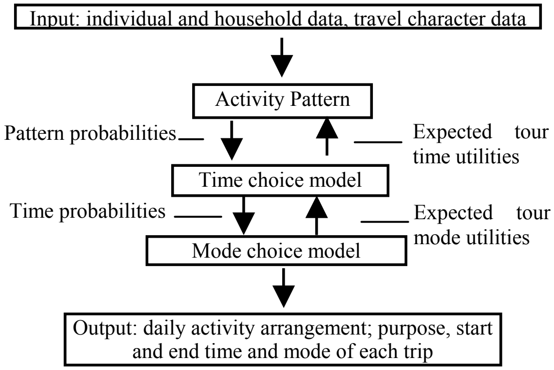

Figure 1: The structure of activity-based travel demand forecasting model.

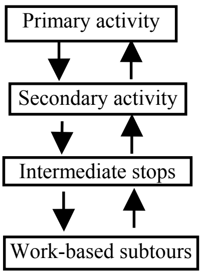

Figure 2: Activity pattern model.

3 Establishment of the time choice model

The behavioural unit for the time choice model is a person-trip, and the model can forecast the start and end time of all the trips. As a result, we can master the law of the time demand, and provide basis for the planners to constitute TDM policies.

3.1 Modelling analysis

For an individual's time decision, the time a trip leaves or arrives at an "attraction" is generally more "scheduled" and thus more stable than the time for the production end of the trip [2]. For example, work is the "attraction" of a work tour, which has regular times, and these times generally affect the time people leave and arrive at home. Besides work tour, many other tours, such as school, day care (exercises in the morning, etc.) also have schedules. Therefore, compared with the production end of the trip, the time choice is based principally on the time at the attraction end. In other words, the destination of a trip is more important to the time choice. So in order to simplify the time choice model, we don't consider a trip is home-based or not temporally, and that's why we don't divide the time choice model into home-based and work-based or other production-based as former models, but forecast the start and end times of all sorts of trips with one model respectively.

At the same time, by analysing the daily travel data of Changchun city in 2003, we get the time distribution of all travel modes, shown in fig. 3. We can learn that the trips are more from 7 to half past 9 o'clock in the morning peak period, and in the evening peak period it is from 16 to 19 o'clock.

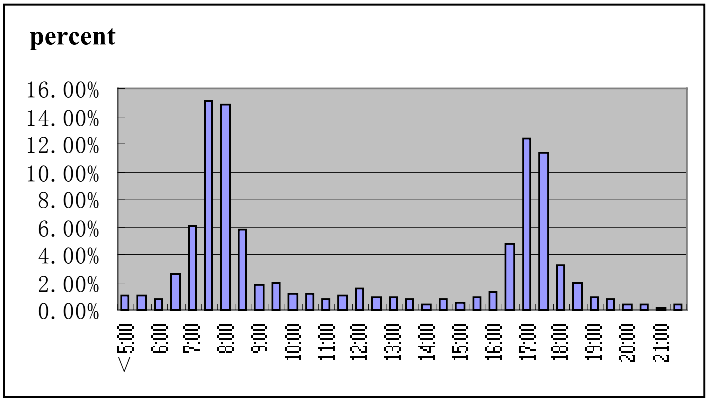

Figure 3: Time distribution of all travel modes.

So we divide a whole day into five time periods each has different travel characteristics:

1) Early (EA) 00:00 to 06:59
2) AM Peak (AM) 07:00 to 09:29

3) Midday (MD) 09:30 to 15:59
4) PM Peak (PM) 16:00 to 18:59
5) Late (LA) 19:00 to 24:00

In the time of day model for Portland in 1994[1], the combination of start and end times was viewed as the choice set, and therefore gets fifteen alternatives shown in table 1. From the statistic of the survey data, we get the probability of each alternative. From table 1, we can learn that not only this method (makes the process of dealing with data very complexly, but also the volume of sample of some combinations (such as EA- LA and AM- PM etc.) is too small for the parameter estimation. But if we design two different models for the start time and end time respectively, then the choice set of each is consisted of five alternatives. Table 2 lists the occupancy distribution of each time period in the two models. By this way, not only the volume of sample is satisfied, but also we can analyse and control residents' start and end times respectively. So in this paper, the time choice model is divided into a start time choice model and an end time choice model.

Table 1: The statistic of the combination of the start and end times.

Alternatives	EA-EA	EA-AM	EA-MD	EA-PM	EA- LA	AM-AM	AM-MD	AM- PM
sample	2923	2671	17	21	4	8291	961	3
percent	7.87%	7.19%	0.05%	0.06%	0.011%	22.31%	2.59%	0.008%
Alternatives	AM- LA	MD-MD	MD-PM	MD-LA	PM-PM	PM-LA	LA-LA	Total
sample	3	11254	1881	4	7482	564	1094	37163
percent	0.008%	30.28%	5.06%	0.011%	20.13%	1.52%	2.94%	100.0%

Table 2: The statistic of the start and end times.

Alternatives	EA	AM	MD	PM	LA	Total
Start time	5609(15.1%)	9249(24.9%)	13120(35.3%)	8065(21.7%)	1108(3.0%)	37151
End time	2971(8.0%)	10945(29.5%)	12212(32.8%)	9360(25.2%)	1665(4.5%)	37153

3.2 Model structure

According to the utility maximization theory, we assume that people will choose a time period whose utility is maximal to him, that is, the probability of an alternative is decided by its utility. Since all the alternatives are independent to each other, the logistic model is used [3]. Its basic theory is the same as the Multinomial Logit (MNL) model, but a ratio form based on the last alternative LA time period is introduced to design model. Its basic equation is shown below:

$$\ln(\frac{p_i}{p_5}) = V_i = \sum_{k=1}^{K} \theta_k x_{ik} \tag{1}$$

Where i is an integer from 1 to 4, p_5 is the probability of LA, and p_i represents the probabilities of the other four time periods except for LA. Compared with LA, V_i is the utility of time period i. And this utility can be expressed with a linear form, where θ_k is the coefficient of factor k, and the variable x_{ik} is the eigenvalue of factor k of alternative i.

From this equation we get the ratio of the former four time periods to the LA respectively.

$$\frac{P_i}{P_5} = e^{V_i} \tag{2}$$

The sum of all the alternatives is 1:

$$\sum_{i=1}^{5} P_i = 1 \tag{3}$$

From eqns (2) and (3), we can get the probabilities of all the time periods.

It not only makes the calculating more simple to adopt the model of logarithm ratio form, but more important Logistic model is based on the comparison between the alternatives, and can recognize the merits and shortcomings of all the alternatives. And it is more complicated to do this kind of analysis with the MNL model.

Table 3: Variables in the time choice model.

Type	Factors		Variables
Individual attributes	Gender		Gender
	age	Age under 20	AGE A
		Age from 20 to 40	AGE B
		Age from 40 to 60	AGE C
		Age 60+	AGE D
household attributes	household income every month（Yuan）	0-500	INC A
		501-1000	INC B
		1001-2000	INC C
		2001-3000	INC D
		3001-5000	INC E
		5000+	INC F
	Children in hhld	Children under age 6 in hhld	Child6
	Vehicles in hhld	Car in hhld	Car
		Motor in hhld	Motor
		2+ bicycles in hhld	Bike
Travel attributes	Primary activity（work、maintenance、discretionary、at home）		Pri
	secondary activity（work, maintenance, discretionary, no secondary tours, combination of maintenance and discretionary）		Sec
	Intermediate stops（on way from home, back home all, no stop）		Inter
	Work-based subtour		Sub

3.3 Choice of the variable x_{ik}

Generally, the factors that affect time choice are mostly derived from the features of individual and household. Besides, we think an individual's activity pattern such as tour purpose, intermediate stops etc. are also important factors, and so we introduce a series of travel characteristic variables to the model. Table 3 reports the variables utilized for this model.

There are two reasons for introduce travel characteristic variables: First, time decision is restricted by the schedule of the "attraction" or the purpose of a tour; Second, the intermediate stops during a tour (such as shopping), which affect the travel time, can greatly influence the start and end time of the tour. But in order to make the model simple, we don't account for the influence of network volume to the travel time. If we can add travel assignment model to activity-based model system in the future, this portion of influence can be considered.

Table 4: Estimation results of the start and end time choice.

Variable	Start time choice model								End time choice model							
	EA		AM		MD		PM		EA		AM		MD		PM	
	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V	T
Constant	1.34	7.11	2.84	10.62	5.48	21.41	2.68	10.42	.70	3.12	2.64	10.13	5.12	23.15	2.51	11.89
Logsum	-.01	-0.50	.02	1.80	-.003	-0.30	.01	1.40	-.14	-5.38	.03	1.62	-.01	-1.50	.001	0.13
Gender	.21	3.13	.18	2.78	.33	5.19	.13	1.97	.41	6.06	.13	2.25	.32	5.83	.17	3.17
AGE A											-.96	-7.60	-.34	-4.23	-.35	-4.42
AGE B	-1.07	-12.64	.12	1.42	-.41	-6.06	-.12	-1.74	-1.60	-19.06	-.50	-4.37	-.49	-7.73	-.23	-3.66
AGE C	-.71	-8.35	.37	4.39					-.84	-10.82	-.20	-1.81				
INC B									.20	2.35						
INC C									.16	2.21						
Child6	.24	1.85	.46	3.80	.32	2.65	.40	3.28	.33	2.67	.42	4.11	.35	3.39	.46	4.49
Car									2.56	5.13	-.64	-1.51				
Motor	-.19	-1.50	-.21	-1.76	-.16	-1.38	-.23	-1.89	.15	1.16	-.140	-1.28				
Bike									.24	2.93						
Pri	.13	2.05	.55	8.70	.71	11.50	-.49	-7.48	.42	6.97	.09	1.51	.84	14.95	-.47	-8.16
Sec	.18	6.59	.10	3.65	-.19	-7.12	.10	3.81	.04	1.42	.07	3.00	-.26	-11.77	.03	1.14
Inter			-.22	-5.29	-.28	-7.05	-.07	-1.59			-.13	-3.71	-.29	-8.79	-.04	-1.12
Sub	-.09	-2.56	-.83	-13.45	-1.48	-24.18	-.20	-3.63	-.06	-1.56	-.32	-5.75	-1.44	-25.71	-.13	-3.47

3.4 Estimation results and analysis

3.4.1 Estimation results

Logistic model is also estimated with Maximum likelihood method as MNL model. And the two time choice models are estimated with the statistic software SPSS [3]. The data we use are collected from a household travel survey carried out in Changchun in 2003, including a total of 37,276 trips of 5,101 households and 14,655 people. Because the time choice model is affected by the mode choice model which is in the lower level, according to the law of NL model, we view logsum which is the logarithm of the sum of the exponentiated utility among the available lower dimension alternatives as a variable in the time choice models. Due to the limit of the content, we'll introduce not the course of how the logsum is calculated, but instead the results of parameters estimation the start and end time choice models, as shown in table 4.

3.4.2 Analysis

In table 4, V is the coefficient of every variable and T is the *T*-statistic of it. In this paper we evaluate whether a variable influenced choice with the *T*-statistic that are levels of uncertainty of each coefficient. It is the ratio of the estimated coefficient to its standard error. And a "one-tailed" *T*-test for a coefficient represents that if the absolute value of *T* is greater than 1.65 and the coefficient has the correct sign, there is a 95 percent confidence that the variable influenced choice [4]. Failure of the *T*-test is not necessarily proof of insignificance, nor is it a warrant for disqualifying a variable. In many estimation of MNL model, although the *T*-statistic of a variable doesn't meet the requirement, but if we are certain that it makes influence to the result by experience, we can still save it by other statistic [4].

We can comprehend the results of model parameter estimation in this way: Take the EA time period of the start model for example, the coefficient of the variable Child6 is 0.235 represents that when there are children under six years old in household, the ratio of leaving home during EA period to LA period is $\boldsymbol{e^{0.235} = 1.265}$. It is probably because that for a person who has child in household, in order to sent the child to the kindergarten, he is more inclined to leave home earlier. We can also learn from the AM time period of the end time model, that with the increasing of age, people are prefer to arrive at the destination during the AM period. Because there're more adults who are busy in this time period, and the "attraction" usually have regular times. With this method, we can analyse other Estimation results as well.

4 Sensitivity analysis

This sensitivity analysis means the change of the ratio of time probability in response to the change of a certain factor. And in this way the influence of the variable on the time choice can be explained. In this paper, we mainly analyse the change of EA / LA($p_1/\boldsymbol{p_5}$) and AM /LA($p_2/\boldsymbol{p_5}$) in response to changes in the age of individual.

Let's assume that someone will begin daily activities, and the attributes of him, his family and the activity pattern are described as: a man who have a kid under six years old, and has a motor in household. Today he will go to work (school) by motor in the morning, at noon he will walk out for lunch and personal business, returning to work for the afternoon., and go shopping on the way home in the afternoon. The distance from home to work (school) is ten kilometres, and the travel speed of auto is influenced in urban areas during peak hours, so the speed of motor is set to 25km/h, and the walk time and wait time is zero, thus the final travel is 24 minutes.

Since the logsum variable which come from the motor mode of the lower level will be used in the model, we list all variables that are used in the EA / LA, AM /LA and motor mode choice model, including: AGE A(x_1), AGE B(x_2), AGE C(x_3), Gender(x_4), Child 6(x_5), Motor(x_6), Logsum(x_7), Pri(x_8), Sec(x_9), Sub(x_{10}), Inter(x_{11}), Time B(travel time from 20 to 40 minutes, x_{12}) and Time C(travel time from 40 to 60 minutes, x_{13}). From the description of the attributes, we get the values of these variables: $x_4=1$, $x_5=1$, $x_6=1$, $x_8=1$, $x_9=0$, $x_{10}=1$, $x_{11}=2$, $x_{12}=1$, $x_{13}=0$.

By simplifying these three models, we get the formulas of logsum, EA / LA and AM /LA:

$$\log sum = \ln(-0.376 - 1.08\,x_1 - 0.447\,x_5 + 4.842\,x_6 - 1.192\,x_8 + 0.285\,x_{11} + 0.451\,x_{12} + 0.388\,x_{13}), \tag{4}$$

$$\frac{p_1}{p_5} = \exp(1.343 - 1.074x_2 - 0.710x_3 + 0.21x_4 + 0.235x_5 - 0.186x_6 - 0.005x_7 + 0.133x_8 + 0.178x_9 - 0.087x_{10}), \tag{5}$$

$$\frac{p_2}{p_5} = \exp(2.835 + 0.119x_2 + 0.373x_3 + 0.181x_4 + 0.456x_5 - 0.21x_6 + 0.018x_7 + 0.548x_8 + 0.095x_9 - 0.843x_{10} - 0.217x_{11}). \tag{6}$$

We only examine the influence of the change of age (x_1, x_2, x_3) on the time choice. For comparison, the age of individual is assumed to 19, 35 and 50 respectively, and the changes of the time choice are shown in table 5.

Table 5: Time choice in response to changes in age.

Age	Variable vector(x_1, x_2, x_3)	Logsum	EA / LA	AM/ LA
19	(1, 0, 0)	1.02	5.15	12.81
35	(0, 1, 0)	1.35	1.75	14.59
50	(0, 0, 1)	1.35	2.51	18.17

By analysing the results in table 5, we can learn, for the individuals of all stages of age, the probabilities of EA or AM for work (school) are much higher than LA. And between AM and EA, people prefer to leave home during the am peak. In the AM/LA model, with the increase of age, the possibility of AM increases as well. The reason is the possibility people choose evening to go out decreases with the increase of age. In the EA model, for many students who have to do morning exercises, the depart time is very early, so the probabilities of EA is higher than people in the other two stages of age. And people from 40 to 60 are more likely to depart during EA than from 20 to 40. In this way, we have analysed other variables' sensitivities. From above analysis, the models we developed can well model the features of time distribution, and can be applied to the prediction of travel behaviour.

5 Application of the models

We focus on people's time adjustment in response to changes in travel costs. The model application is to provide forecasts for possible two congestion pricing policies: a 10% increase in all car travel costs, and a 100% increase in car fuel and operating costs only during the peak periods. By forecasting and comparing the change of the original survey data, we reach the results shown in table 6.

Table 6: Model application-percent changes in travel time under two policies.

Mode	Time period	Changes in depart time (%)	
		10% increase in car travel cost	100% increase in peak periods car cost
All modes	EA/LA	0.002	0.013
	AM/LA	-0.026	-0.133
	MD/LA	0.004	0.024
	PM/LA	-0.015	-0.091
Car	EA/LA	0.112	1.000
	AM/LA	-1.563	-10.616
	MD/LA	0.267	1.897
	PM/LA	-0.878	-7.265

When the travel costs of car increases by 10%, for all modes or for car only, there're some shifts of tours out of the AM and PM peak periods into the midday and off-peak periods. And for the second policy, the change of time choice is the same, but more obvious. For both policies, the change for the trip by car is much larger than the mean value of all modes. In summary, the policy of increasing car costs not only causes shift from the peak periods to the off-peak periods, but also conducts people to adopt other travel modes, such as bus and light rail an so on, and solve congestion problem by means of both time and space constrain.

6 Conclusions

In this paper, we introduced the basic structure of activity-based travel demand forecasting model, then developed a time choice model by analysing the survey data and concluding the former models. Finally, we completed the parameter estimation, sensitivity analysis and application of the model. Taking time choice model as an example, we presented the process of design and prediction of the activity-based models. And try to study the people's travel behaviour in a new way, in order to provide decision basis for the conduct of residents' trips. Compared with former researches, we divided time choice model into start and end time models, and use Logistic model to replace the MNL model. Furthermore, we use the statistic software SPSS to carry out model estimation, which can make the model more appropriate for calculation with PC.

However there're some weaknesses in this model. First, we only divide a day into five time periods. Second, we didn't consider the influence of network volume on travel time. In order to improve the model, we can divide a day time into more periods, and add variables (such as congesting degree, travel time, etc.) which are the output of the traffic assignment model, to make the estimation and prediction more accurate. Furthermore, we can integrate activity-based model with the "four-stage" model [5], introduce the assignment model and the aggregate idea of the latter, in order to simplify data processing and forecasting, and make the activity-based model better adapt to large scale of prediction and planning.

References

[1] Bowman J., Bradley M, et al., Demonstration Of An Activity-based Model System For Portland. 8th World Conference on Transport Research, pp. 12-17, 1998.

[2] Sacramento Area Council of Governments, Sacramento Regional Travel Demand Model Version 2001, http://www.sacog.org/publications/SACOG02003.pdf, 2002.

[3] Wang, J., Guo, Z., Logistic Regression Models: Methods and Application. Higher Education Press, 2001.

[4] Jin an, On Methodology of Parameter Estimation in LOGIT Model, Journal of Transportation Systems Engineering and Information Technology, 4(1), pp.71-75, 2004.

[5] Lu Hua-pu, The Theory and Method of Transportation Planning, Tsinghua University Press, 1998.