

## Research Article

# Daily Commute Time Prediction Based on Genetic Algorithm

**Fang Zong,<sup>1</sup> Haiyun Lin,<sup>2</sup> Bo Yu,<sup>3</sup> and Xiang Pan<sup>4</sup>**

<sup>1</sup> College of Transportation, Jilin University, RenMin Street 5988, Changchun 130022, China

<sup>2</sup> Department of Civil Engineering, City College of New York, 160 Convent Avenue, New York, NY 10031, USA

<sup>3</sup> Transportation College, Dalian Maritime University, Dalian 116026, China

<sup>4</sup> College of Computer Science, Zhejiang University of Technology, 288 Liuhe Road, Hangzhou 310023, China

Correspondence should be addressed to Fang Zong, zongfang@jlu.edu.cn

Received 20 September 2012; Accepted 30 October 2012

Academic Editor: Baozhen Yao

Copyright © 2012 Fang Zong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a joint discrete-continuous model for activity-travel time allocation by employing the ordered probit model for departure time choice and the hazard model for travel time prediction. Genetic algorithm (GA) is employed for optimizing the parameters in the hazard model. The joint model is estimated using data collected in Beijing, 2005. With the developed model, departure and travel times for the daily commute trips are predicted and the influence of socio-demographic variables on activity-travel timing decisions is analyzed. Then the whole time allocation for the typical daily commute activities and trips is derived. The results indicate that the discrete choice model and the continuous model match well in the calculation of activity-travel schedule. The results also show that the genetic algorithm contributes to the optimization and thus the high accuracy of the hazard model. The developed joint discrete-continuous model can be used to predict the agenda of a simple daily activity-travel pattern containing only work, and it provides potential for transportation demand management policy analysis.

## 1. Introduction

The time allocation of individuals for trip making is an important determinant of the temporal pattern of traffic demand on a transportation network. An analysis of individual time allocation decision is, therefore, important for the practical work of transportation planning and management as well as the theoretical work about travel behavior analysis and modeling. Practically, understanding individuals' time allocation decisions is essential for (a) planning the development and construction of new transportation infrastructure by providing

predicted temporal travel demands, (b) examining the potential responses to improved operational measures (such as real-time information), (c) assessing the effectiveness of time-specific transportation demand management policies, such as compressed working week [1], staggered shift [2], road tolling [3], and other similar strategies [4, 5]. Theoretically, understanding time allocation decisions will not only facilitate the efforts toward developing a comprehensive full-scale model of daily activity patterns but also provide useful insights into the nature of the impact of sociodemographic variables and time-space constraints on individual dimensions of activity behavior. Therefore, time allocation has been a focused issue for regional and transportation science since the 1970s [6].

In the time dimension, a daily activity-travel pattern includes all the timing and duration of all the activities and trips in a day. The need to analyze the daily activity pattern makes it essential to consider time of day and activity-travel durations together. However, to date, timing and duration have been largely treated separately in the literature [7], and many existing studies only considered part of daily activity agenda. For example, Vovsha and Bradley [8] focused on the departure-from-home and arrival-back-home time decisions. Fujii and Kitamura [9] and Hamed and Mannering [10] examined the time allocation of postwork activities. Bhat and Singh [11] and Habib et al. [12] modeled the departure time of daily commute trips, without involving the commute travel time. This study is aimed at addressing this issue by developing a joint daily time allocation model to predict a typical daily activity-travel schedule.

One important objective of transportation planning is to relief congestion by improving the level of service during the peak hours on the transportation network. Peak periods' traffic demands, or the source of congestion, are largely contributed by commute trips. For example, based on one survey conducted in Beijing, China, in 2005, over 32% of all the trips in both morning and evening peak periods are commute trips. Therefore, commute trips are at the core of many recent studies, such as Habib et al. [12], Zhang et al. [13], and Bhat and Singh [11].

As stressed above, the study on commute trips and daily time allocation is of great importance for learning the travel behavior during the peak hours as well as obtaining the daily activity pattern, both of which are essential component of for transportation planning and management. However, to the authors' knowledge, there is no study that considered the overall daily time allocation of commute activity-travel pattern as a whole and developed a model system to predict it. Therefore, this paper focuses on predicting timing and duration of daily commute trips, expecting that commuter's typical travel schedule can be obtained based on the developed model. Using data from a 2.5% sample household survey in Beijing, China, a joint discrete-continuous model system for prediction of daily commute time allocation was developed and estimated, including ordered probit models for departure time analyzing and hazard models for travel time forecasting.

The choice of parameters is of great importance for the estimation efficiency and prediction accuracy of the models. As there are a lot of potential factors affecting traveler's decision about time allocation, genetic algorithms (GAs) will be employed in parameters optimization. Being one of the heuristic algorithms, GA has been successfully applied in various optimization problems [14, 15].

The remainder of this paper is organized as follows: Section 2 presents the literature review on activity-travel time allocation in general. In Section 3, the joint discrete-continuous time allocation model is built, in which the ordered probit method, AFT model, and GA are employed. Section 4 predicts the commute activity-travel agenda by using the developed

model. The paper closes with some overall conclusions and a discussion of future research directions.

## 2. Relevant Literature

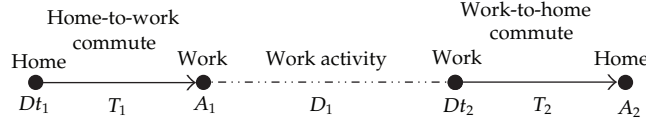
With respect to the two dimensions of time allocation, timing and duration, existing studies can be classified into three categories: those only looked at departure time, those only looked at travel time, and those looked at both.

For the first direction, timing, the major method employed is discrete choice model. For example, Bowman and Ben-Akiva [16] modeled the departure and arrival time of daily trips using the multinomial logit (MNL) model. Small [17] built an MNL model for home-work morning departure time choice. Vovsha and Bradley [8] and Ettema et al. [7] also employed the MNL model but specified a continuous time variable in its utility function. Bhat [18] and Small [19] derived the ordered generalized extreme value (OGEV) model for departure time choice and time of day analysis, respectively. Habib et al. [12] introduced OGEV model into analysis of work start time and work duration because it allows for the accommodation of correlation among time period alternatives. In this way, it resolves the IIA (independence of irrelevant alternatives) problem of MNL model, which assumes zero correlation among different time periods.

For the second aspect (duration analysis), the most widely used model is the hazard model. The hazard model recognizes the dynamics of travel or activity durations by considering the conditional probability of event termination, usually as a function of covariates (explanatory variables) [12]. Bhat [20, 21] applied hazard modeling framework to analyze after-work activity duration. Juan and Xianyu [22] considered daily travel time using hazard-based duration model. Bhat [20, 21] analyzed the duration of shopping activity by employing hazard-based duration model. As a matter of fact, a few studies also used hazard model in timing analysis. Examples include research by Habib et al. [12] on investigation of trip timing and by Bhat and Steed [23] on departure time choice for shopping trip.

Comparing with the first two categories of studies, those using the joint analysis of timing and duration are more helpful for the modeling of daily activity schedule, by contributing to an insight into the influence and connection among duration and time of day choice as well as activities and trips. The examples include the study done by Janssens et al. [24] on time allocation of daily activity-travel patterns, by Fujii and Kitamura [9] on timing and duration of commuters' daily activity patterns after work hours, by Habib [25] on work start time and work duration, by Raux et al. [6] on daily time allocation of travel and out-of-home activity, by Schwanen and Dijst [26] on relationship between commuting time and work duration, and by Vovsha and Bradley [8] on departure-time and duration of home-based trips. Similar studies also include Ettema et al. [7], Pendyala and Bhat [27], and Habib et al. [28]. The models involved in these studies include reinforcement machine learning technique [24], structural equations model [9], hazard model [6, 25, 28], and MNL model with a continuous time variable in the utility of its function [7, 8].

There is also a kind of conjunct model that was used in the analysis of timing and duration, which is the discrete-continuous choice model. Pendyala and Bhat [27] examined the relationship between time of day choice and activity episode duration using discrete-continuous simultaneous equations, but it was neither on daily time allocation simulation nor the commute trips. Similar discrete-continuous models were employed by Habib et al. [12], Habib [25], and Hamed and Mannering [10]. Both of the first two studies modeled the joint



**Figure 1:** Basic time allocation of daily commute activity-travel pattern.

mode and commute timing choice with discrete choice model and continuous model, but not focusing on daily activity-travel schedule particularly. Concerning the research by Hamed and Mannering [10], the discrete-continuous models are used to predict the travelers activity-type preference (discrete choice model), the travel time to the activity and back home (continuous model), and activity duration (continuous model). For modeling of time allocation, only continuous model was employed. However, the results of these four studies mentioned above all confirmed that the discrete-continuous choice models work well for the activity-travel behavior related study.

Moreover, the study of Pendyala and Bhat [27] suggests that time of day and activity duration is only loosely related for the commuter sample. Therefore, travel departure time and travel duration will be modeled in this paper, instead of activity duration. In this case, the work duration can be calculated according to the arrival time at work location and departure time of the next activity after work.

### 3. Modeling Time Allocation of Daily Commute Trips

#### 3.1. Analysis of the Commute Activity-Travel Agenda

In this paper, work location refers to the usual work location for a worker and the usual school location for a student. As shown in Figure 1, a typical daily commute activity-travel pattern is home-to-work commute—work activity—work-to-home commute. Key time and duration values in this pattern include home-to-work morning departure time ( $Dt_1$ )—home-to-work travel time ( $T_1$ )—arrival time at work location ( $A_1$ )—duration of work ( $D_1$ )—work-to-home evening departure time ( $Dt_2$ )—work-to-home travel time ( $T_2$ )—arrival time at home ( $A_2$ ). Within this pattern, the departure times  $Dt_1$  and  $Dt_2$  and the travel times  $T_1$  and  $T_2$  are most important. Known their values, one can easily derive the other three times ( $A_1$ ,  $D_1$ , and  $A_2$ ) using the following equations:

$$\begin{aligned} A_1 &= Dt_1 + T_1, \\ D_1 &= Dt_2 - A_1, \\ A_2 &= Dt_2 + T_2. \end{aligned} \tag{3.1}$$

This study employs the ordered probit model (belonging to the discrete choice models) for departure time forecasting and the hazard model (belonging to the continuous models) for travel time analysis. Four models, home-to-work departure time choice model, home-to-work travel time estimation model, work-to-home departure time choice model, and work-to-home travel time estimation model will be developed and the values of  $Dt_1$ ,  $T_1$ ,  $Dt_2$ , and  $T_2$  will be predicted. Values of  $A_1$ ,  $D_1$ , and  $A_2$  are then calculated based on (3.1). Figure 2 is a schematic representation of the entire modeling process.

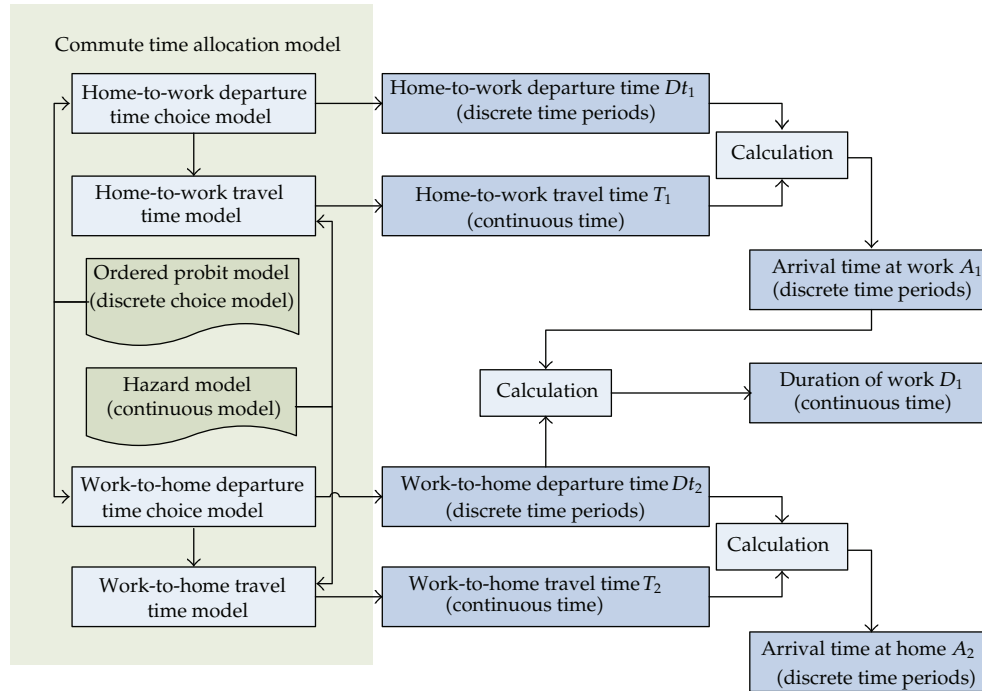


Figure 2: Modeling process of the commute activity-travel time allocation.

### 3.2. Data

This study uses data from a large-scale daily travel survey conducted in Beijing, China, in 2005. A face-to-face interview was given to a sample of 54,398 households, and activity/travel information of all household members on one particular working day is collected. The study area is 1,368 km<sup>2</sup> (covering all 18 districts of Beijing) and had more than 30 millions' population in 2005. In addition to weekday OD information, the survey also collected information regarding household (size, car ownership, home location, monthly income, and mobility), people (age, gender, driving license, and occupation), and trips (departure time, arrival time, purpose, mode, transit path, etc.). With records containing missing values eliminated, our final sample consists of 37,842 commute trips of 37,842 individual workers/students from 28,382.

Based on a preliminary correlation test, 15 sociodemographic and trip characteristics variables were selected from the survey, shown in Table 1.

The statistics of the variables based on the sample data are shown in Table 2.

### 3.3. Departure Time Choice Model

The reported home-to-work morning departure time and the work-to-home evening departure time cover the time period of 4:00 am–12:00 am and 15:00 pm–22:00 pm, respectively, in our sample. Moreover, we observed that the home-to-work morning and the work-to-home evening peak hours in Beijing are 6:00 am–9:00 am and 16:00 pm–19:00 pm using our sample data. In order to reduce the number of alternatives in the models' choice set, we divided

**Table 1:** Variables in the departure time choice models.

Factors	Variables	Values
Gender	Gender	Male: 1, female: 0
Age	Age	Continuous values Below 1500 RMB: 1 1501–2500 RMB: 2 2501–3500 RMB: 3 3501–5500 RMB: 4 5501–10000 RMB: 5 10001–20000 RMB: 6 20001–30000 RMB: 7 Over 30001 RMB: 8
Month income	Income	
Vocation		
Blue-collar worker (manufacture, construction, maintenance, etc.)	Occu-b	Yes: 1, No: 0
Administration	Occu-a	Yes: 1, No: 0
Education	Occu-e	Yes: 1, No: 0
Services	Occu-s	Yes: 1, No: 0
Health care	Occu-h	Yes: 1, No: 0
Travel mode		
Walk	Mode-w	Yes: 1, No: 0
Bike	Mode-bi	Yes: 1, No: 0
Bus	Mode-bu	Yes: 1, No: 0
Auto	Mode-a	Yes: 1, No: 0
Travel distance	Distance	Continuous value (meter)
Dummy variable of going to work	Work	Yes: 1, no: 0 (going to school)

departure times into one-hour segments in peak periods and segments of two or three hours in off-peak periods. Table 3 shows the discrete alternatives for the home-to-work and work-to-home departure time choices.

As shown in Table 2, the alternatives of the departure time choice models are naturally ordered time periods. MNL model, which is commonly used in departure time modeling, would fail to account for the ordinal nature of the dependent variable and have the problem of IIA. This study will employ the ordered multiple choice model for departure time modeling instead.

The ordered multiple choice model assumes the following relationship:

$$\sum_{j=1}^J P_n(j) = F(\alpha_j - \beta_j X_n, \theta), \quad j = 1, \dots, J-1, \quad (3.2)$$

$$P_n(J) = 1 - \sum_{j=1}^J P_n(j),$$

where  $P_n(j)$  is the probability that alternative  $j$  is chosen as departure time of trip  $n$  ( $n = 1, \dots, N$ ),  $\alpha_j$  is an alternative specific constant,  $X_n$  is a vector of attributes of trip  $n$ ,  $\beta_j$  is a vector of

**Table 2:** Statistics index of the variables.

Variables	Percentage of each value	
	1	0
Gender	50.96%	49.04%
Occu-b	42.44%	57.56%
Occu-a	3.48%	96.52%
Occu-e	30.29%	69.71%
Occu-s	9.19%	90.81%
Occu-h	2.49%	97.51%
Mode-w	42.51%	57.49%
Mode-bi	43.89%	56.11%
Mode-bu	0.31%	99.69%
Mode-a	4.67%	95.33%
Work	72.79%	27.21%
Income	1: 13.46%	5: 11.86%
	2: 24.90%	6: 1.41%
	3: 24.18%	7: 0.16%
	4: 23.92%	8: 0.10%
Continuous variables		
Variables	Mean	Standard deviation
Distance	2806.1	3416.91
Age	32.93	15.32

**Table 3:** Alternatives in the departure time choice models.

Home-to-work departure time	Work-to-home departure time	Coded value
(4:00, 6:00]	(15:00, 16:00]	1
(6:00, 7:00]	(16:00, 17:00]	2
(7:00, 8:00]	(17:00, 18:00]	3
(8:00, 9:00]	(18:00, 19:00]	4
(9:00, 12:00]	(19:00, 22:00]	5

estimable coefficients, and  $\theta$  is a parameter that controls the shape of probability distribution  $F$ . Therefore,  $F$  can have various shapes of distribution based on a different value of  $\theta$ .

The ordered probit model, which assumes standard normal distribution for  $F$ , is the most commonly used. The ordered probit model has the following form:

$$\begin{aligned}
 P_n(1) &= \Phi(\alpha_1 - \beta_j X_n), \\
 P_n(j) &= \Phi(\alpha_j - \beta_j X_n) - \Phi(\alpha_{j-1} - \beta_j X_n), \quad j = 2, \dots, j-1, \\
 P_n(J) &= 1 - \sum_{j=1}^{J-1} P_n(j),
 \end{aligned}
 \tag{3.3}$$

where  $P_n(j)$  is the cumulative standard normal distribution function. For all the probabilities to be positive, we must have  $\alpha_1 < \alpha_2 < \dots < \alpha_{J-1}$ . The estimation results of the home-to-work and the work-to-home departure time choice models are shown in Table 4.

The estimation results indicate that high-income travelers are more likely to depart from home or work location later than travelers with low income. Older persons tend to have earlier departure times than younger ones. Commuters whose occupations are administration or health care are more likely to depart from home earlier but back to home (from work location) later than those with other occupations. The probability that students or teachers have early departure times both from home and school is high. Workers and servers tend to depart from workplace late. Concerning gender, men are more likely to leave home earlier but leave work later than women. Regarding travel modes, commuters choosing walk or auto have late home-to-work and work-to-home departure times. Bikers tend to leave home earlier, while commuters taking bus depart from workplace later. Long-distance trips tend to be made later from home while earlier from work location. Workers are more likely to leave home later than students.

### 3.4. Travel Time Estimation Models

#### 3.4.1. AFT Model and KM Estimator

According to the travel survey data in Beijing, the average travel time for the home-to-work commute is 19.36 minutes, with a maximum of 205 minutes and a minimum of 1 minute; the average duration for the work-to-home commute is 18.36 minutes, with a maximum and a minimum of 168 minutes and 1 minute, respectively.

Treating travel times as natural continuous variables, one can use the hazard model to predict both the home-to-work and the work-to-home travel times. Hazard-based duration models are ideally suited to modeling duration data [20, 21], such as travel time and activity duration. The hazard (also called a hazard rate) represents a termination rate of the duration.

Let  $T$  be a nonnegative random variable representing the travel time. The hazard at time  $t$  on the continuous time-scale  $h(t)$  is defined as the instantaneous probability that the travel duration under study will end in an infinitesimal time period  $\Delta t$  after time  $t$ , given that the duration has not elapsed until time  $t$ . A mathematical definition for the hazard function is as follows:

$$h(t) = \lim_{\Delta \rightarrow 0^+} \frac{P(t \leq T < t + \Delta \mid T > t)}{\Delta}. \quad (3.4)$$

Let  $f(\cdot)$  and  $F(\cdot)$  be the density and cumulative distribution function for  $T$ , respectively. Then the probability of ending in an infinitesimal interval of range  $\Delta t$ , after time  $t$ , is  $f(t)\Delta t$ . And the probability that the process lasts for at least  $t$  is given by the survival equation (3.5):

$$S(t) = P(T > t) = 1 - F(t). \quad (3.5)$$

Thus, the hazard function can be further expressed as

$$h(t) = \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{-dS(t)/dt}{S(t)} = \frac{-d \ln S(t)}{dt}. \quad (3.6)$$

The distribution of the hazard can be assumed to be one of many parametric forms or to be nonparametric. Because the distribution of the travel time is unknown, one of



**Table 4:** Estimation results of the departure time choice models.

Variables	Home-to-work departure time choice model		Work-to-home departure time choice model	
	Coef.	<i>t</i> -stat.	Coef.	<i>t</i> -stat.
Income	0.09	19.14*	0.06	16.18*
Age	-0.01	-10.49*	-0.01	-29.41*
Occu-w	—	—	0.43	32.00*
Occu-a	-0.14	-4.60*	0.35	11.44*
Occu-e	-0.72	-27.09*	-0.25	-12.96*
Occu-s	—	—	0.27	11.68*
Occu-h	-0.48	-12.96*	0.33	9.24*
Gender	-0.11	-9.34*	0.08	8.36*
Mode-w	0.18	7.89*	0.42	35.49*
Mode-bi	-0.41	-19.30*	—	—
Mode-bu	—	—	0.59	4.76*
Mode-a	0.10	3.12*	0.41	15.29*
Distance	0.000067	-32.89*	-6.26e - 06	-3.11*
Work	0.39	12.92*	—	—
$\alpha_1$	-2.42	—	-1.30	—
$\alpha_2$	-0.98	—	0.01	—
$\alpha_3$	0.58	—	0.97	—
$\alpha_4$	1.60	—	2.07	—
Hit ratio		64.36%		61.00%
<i>P</i> value		0.0000		0.0000
<i>N</i>			37842	

\*Significant at 1% level.

the nonparametric methods, the Kaplan-Meier (KM) product limit estimator, is conducted to explore the covariate effects and the potential distribution.

As a nonparametric method, the KM estimator produces an empirical approximation of survival and hazard but hardly takes any covariate effects into consideration. It is similar to an exploratory data analysis. Denoting the distinct failure times of individuals  $n$  as  $t_1 < t_2 < \dots < t_m$ , the KM estimator of survival at time  $t_i$  is computed as the product of the conditional survival proportions:

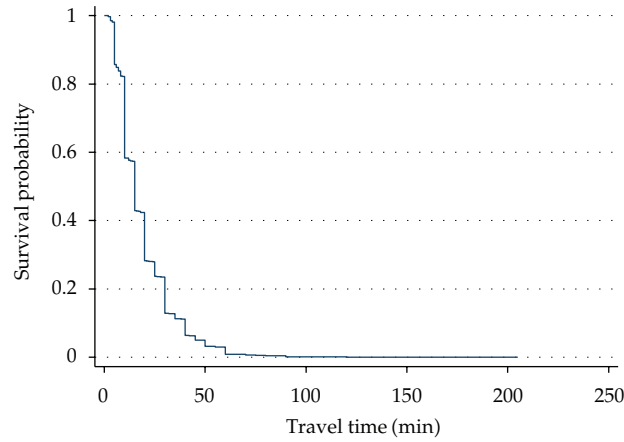
$$S_{KM}(t_i) = \prod_{k=1}^i \frac{r(t_k) - d(t_k)}{r(t_k)}, \quad (3.7)$$

where  $r(t_k)$  is the total trips at risk for ending at  $t_k$  and  $d(t_k)$  is the number of trips stopping at  $t_k$ .

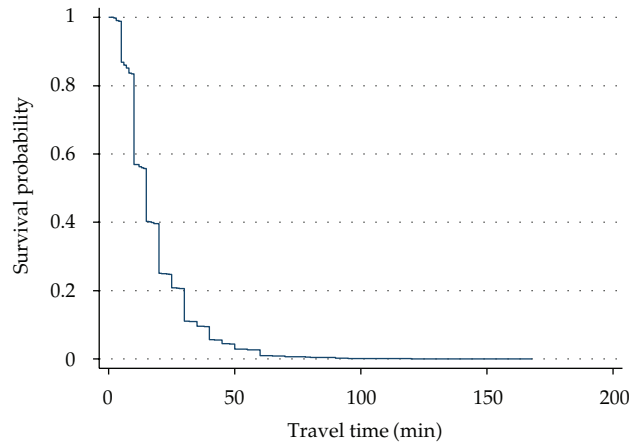
By using the KM estimator, the survival function curves of the home-to-work and the work-to-home travel time are estimated, which are shown in Figures 3 and 4, respectively. The results indicate that the survival probability decreases with travel time, which implies an accelerated failure time (AFT) model with Weibull or Exponential distribution should be employed. Therefore, the AFT model is developed to examine the linkages between travel time and covariates relative to individual and household.

The AFT model is one of the popular parametric forms of hazard model. It permits the covariates to affect the duration dependence. The survival function of AFT model is given as

$$S(t) = S_0 [t \cdot \exp(-\beta'X)], \quad (3.8)$$



**Figure 3:** Survival curve of the home-to-work travel time.



**Figure 4:** Survival curve of work-to-home travel time.

where  $S_0(\cdot)$  is the baseline survival function. The corresponding hazard function is

$$h(t) = \frac{-\partial S(t)/\partial t}{S(t)} = h_0 [t \cdot \exp(-\beta'X)] \exp(-\beta'X). \quad (3.9)$$

The AFT model can be expressed as a log-linear model:

$$\ln t = \beta'X + \varepsilon. \quad (3.10)$$

Assuming the random error  $\varepsilon$  follows either a Weibull distribution or an Exponential distribution, one can get two kinds of AFT models, and both of them are often used in duration analysis.

### 3.4.2. GA for Parameter Optimization

The parameters in the AFT models will influence the estimation efficiency and prediction accuracy of the models greatly, especially for large-scale or real-time feature practice application. Therefore, this paper attempts to find the appropriate parameters in AFT models by using GA. GA is a part of evolutionary computing, which is a rapidly growing area of artificial intelligence. The process of GA is as follows.

#### *Encoding of Chromosome*

GA is started with a set of solutions (represented by chromosomes) called population. The individuals comprising the population are known as chromosomes. In most GA applications, the chromosomes are coded as a series of zeroes and ones, or a binary bit string. For the travel time forecasting models, some parameters are continuous valued (like *distance* and *age*) while some are discrete valued (such as the variables about mode and occupation). Therefore, the real encodings were adopted for continuous-valued parameters, and the binary bit string was adopted for discrete-valued parameters. Thus, each chromosome consists of  $n$  "genes",  $gen_1^t, gen_2^t, \dots, gen_n^t$ , which represents  $n$  parameters, respectively.

#### *Crossover*

Crossover is a reproduction technique that takes two parent chromosomes and produces two child chromosomes. A commonly used method for crossover called one-point crossover [29, 30] will be employed in this study. In this method, both parent chromosomes are split into left and right subchromosomes, where the left subchromosomes of each parent are the same length, and the right subchromosomes of each parent are the same length. Then each child gets the left subchromosome of one parent and the right subchromosome of the other parent, as shown in Figure 5. The split position (between two successive genes) is called the crossover point. For example, if the parent chromosomes are 011 10010 and 100 11110 and the crossover point is between bits 3 and 4 (where bits are numbered from left to right starting at 1), then the children are 01111110 and 100 10010. We will call crossover applied at the bit level to bit strings binary crossover, and crossover applied at the real parameter level real crossover.

#### *Mutation*

Mutation is a common reproduction operator used for finding new points in the search space to evaluate. A genetic mutation operation [31] is used in this paper.

Assume a chromosome is  $G = (gen_1^t, gen_2^t, \dots, gen_n^t)$  if the  $gen_i^t$  ( $i = 1, \dots, n$ ) is selected for the mutation, the mutation can be shown in

$$G' = (gen_1^{t-1}, gen_2^{t-1}, \dots, gen_n^{t-1}),$$

$$gen_i^t = \begin{cases} gen_i^{t-1} + \Delta(t, gen_{i,max}^t - gen_i^{t-1}) & \text{if random}(0, 1) = 0, \\ gen_i^{t-1} + \Delta(t, gen_i^{t-1} - gen_{i,min}^t) & \text{if random}(0, 1) = 1, \end{cases} \quad (3.11)$$

where  $n$  is the total number of the parameters.

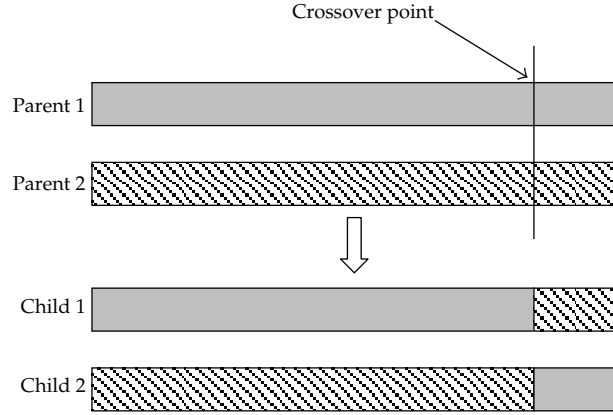


Figure 5: Crossover diagram.

The function  $\Delta(t, y)$  returns a value between  $[0, y]$  given in

$$\Delta(t, y) = y \times \left(1 - r^{(1-t/T_{\max})^n}\right), \quad (3.12)$$

where  $r$  is a random number between  $[0, 1]$ ;  $T_{\max}$  is a maximum number of generations. This property causes this operation to make a uniform search in the initial space when  $t$  is small and a very local one in later stages.

To deal with the problem that the mutation may violate the parameters constraints, we will assign a relatively high weight to reduce their probability of being selected in the following search [31].

### Termination

There are four GA parameters, namely,  $p_c$ ,  $p_m$ ,  $p_{size}$ , and  $T_{\max}$ , that need to be predetermined. Considering the features of this problem and our experiences in GA, the values of four GA parameters are set to be 0.6, 0.05, 80, and 5000, respectively.

### 3.4.3. Estimation Results

#### Home-To-Work Travel Time Estimation Model

Figure 6 illustrates the survival curves for the home-to-work travel time influenced by several major variables. It shows that factors of departure time, travel mode, income, gender, and going to work have influences on home-to-work travel time.

Two AFT models are estimated, each of which assumes the random error in (3.10) follows a Weibull distribution, and an Exponential distribution, respectively. The parameters are optimized using GA and then estimated using maximum likelihood estimation (MLE). The estimation results are shown in Table 5.

The mean absolute percentage error (MAPE), which looks at the average percentage difference between predicted values and observed ones, is adopted to examine the accuracy

**Table 5:** Estimation results of the home-to-work travel time model.

Variables	Weibull distribution		Exponential distribution	
	Coef.	z-stat.	Coef.	z-stat.
Constant	3.11	163.26	2.84	73.09
Income	-0.01	-4.82	-0.01	-2.65
Age	0.01	22.49	0.00	7.00
Occu-e	-0.20	-19.86	-0.10	-4.49
Gender	-0.04	-7.38	-0.04	-3.72
Mode-w	-0.18	-17.28	-0.14	-7.65
Mode-bi	0.07	7.04	0.15	8.93
Mode-bu	0.20	4.28	—	—
Mode-a	-0.04	-2.97	—	—
Distance	0.00013	100.58	0.00012	57.51
Departure time	-0.09	-29.84	-0.09	-13.06
Work	-0.32	-29.82	-0.19	-7.36
$\gamma$	0.50	—	1	—
Prob > chi <sup>2</sup>	0.0000		0.0000	
N	37842			

**Table 6:** Goodness of fit index and estimated distribution statistics of the home-to-work travel time model.

Model statistics	Weibull distribution	Exponential distribution
MAPE value	0.4449	0.3882
Mean (min)	18.46	13.66
Maximum (min)	6.02	5.09
Minimum (min)	158.64	114.08

of the developed home-to-work travel time model. MAPE is calculated as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - P_i}{A_i} \right|, \tag{3.13}$$

where  $A_i$  is the observed value and  $P_i$  is the predicted value for observation  $i$ . The MAPE values of the two AFT models are shown in Table 6.

According to the results shown in Table 5, the MAPE value of the Exponential distribution is less than that of the Weibull distribution, indicating that the values predicted by the AFT model with the Exponential distribution is more close to the actual travel time. Therefore, the Exponential distribution function is chosen. The hazard function and survival function are shown as follows:

$$\begin{aligned}
 h(t) = & \exp (2.84 - 0.01 \text{ Income} + 0.0098 \text{ Age} - 0.1 \text{ Occue} - 0.04 \text{ Gender} - 0.14 \text{ Modew} \\
 & + 0.15 \text{ Modebi} + 0.00012 \text{ Distance} - 0.09 \text{ Departure time} - 0.19 \text{ Work}), \\
 s(t) = & \exp ( - \exp(2.84 - 0.01 \text{ Income} + 0.0098 \text{ Age} - 0.1 \text{ Occue} - 0.04 \text{ Gender} - 0.14 \text{ Modew} \\
 & + 0.15 \text{ Modebi} + 0.00012 \text{ Distance} - 0.09 \text{ Departure time} - 0.19 \text{ Work})).
 \end{aligned}
 \tag{3.14}$$

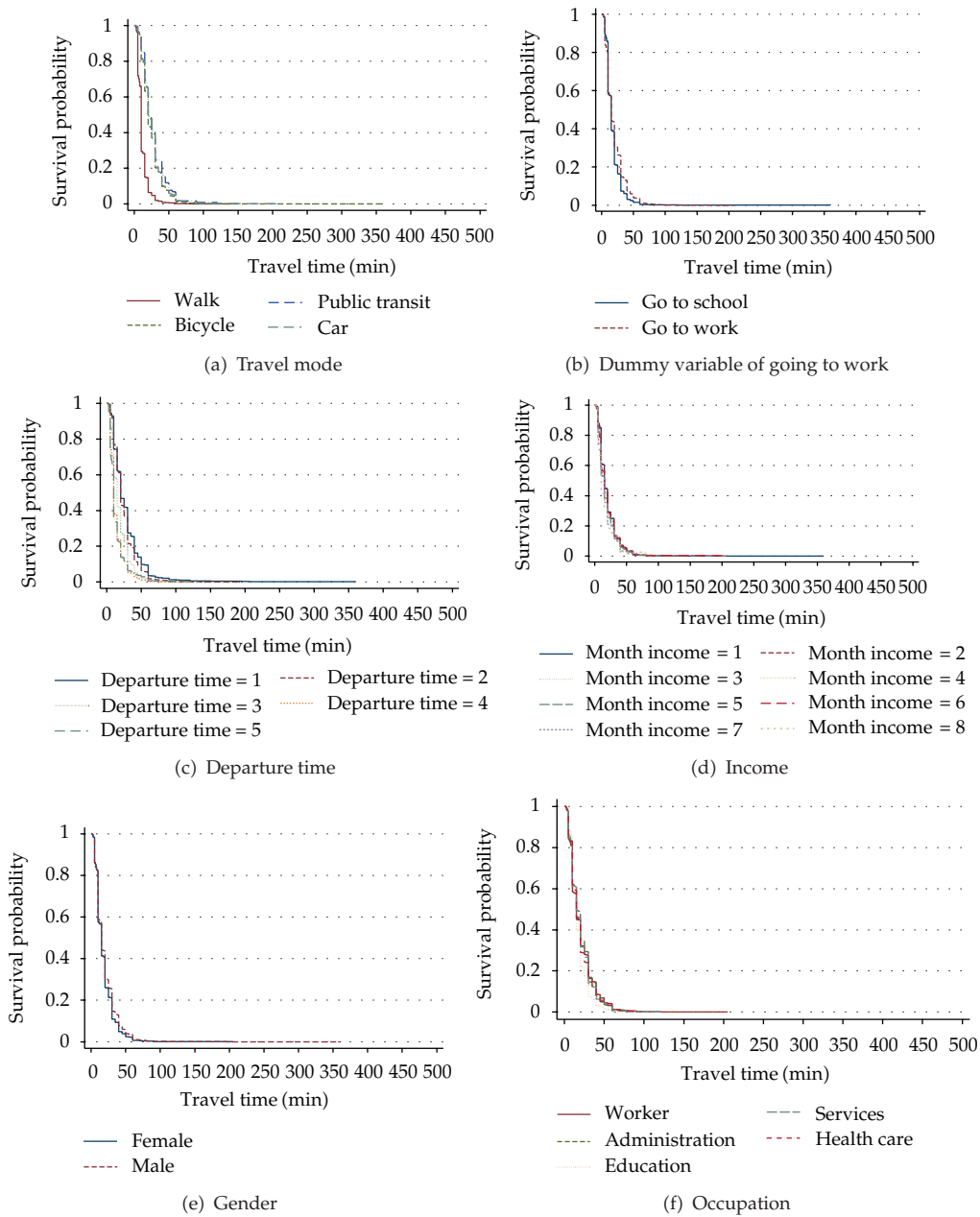


Figure 6: Survival curves of the home-to-work travel time impacted by several factors.

The estimation results indicate that the most essential factor of travel time is distance. The longer the distance from home to work is, the more time it will take. Comparing with other travel modes, the travel times of walk are 14% lower while those of bike are 15% higher. The reason is that, at first short distance encourages short travel time according to the estimation results, and then distance has influence on mode decisions, that is, walking usually belongs to short-distance travel comparing to biking. As for the factor of departure time,

the results show that the later the departure times are, the longer the travel times will be. In the above three parameters, the factors of distance and mode are related to the transportation network, while the real traffic condition is considered by using departure time as a factor, because the traffic condition depends regularly on departure time. For instance, if a traveler departs from home in the morning peak time, the probability that he/she encounters traffic congestion is much larger than that in the nonpeak time. The results also show that the higher travelers' income is, the less their travel time will be. The travel times of students or teachers are about 10% less than those of other travelers. The older the travelers are, the longer the travel time will be. Women have longer travel time than men. The travel times of travelers whose occupation is administration are 19% longer than that of education.

#### *Work-To-Home Travel Time Estimation Model*

Figure 7 illustrates the survival curves for the work-to-home travel time influenced by several variables of the interest. The AFT models with both Weibull distribution and Exponential distribution are employed for the work-to-home travel time modeling. The estimation results and the MAPE values of the two AFT models are shown in Tables 7 and 8, respectively.

Same as the home-to-work model, the AFT model of work-to-home travel times with Exponential distribution is better than that with Weibull distribution. Therefore, the former is selected. The hazard function and survival function are as follows:

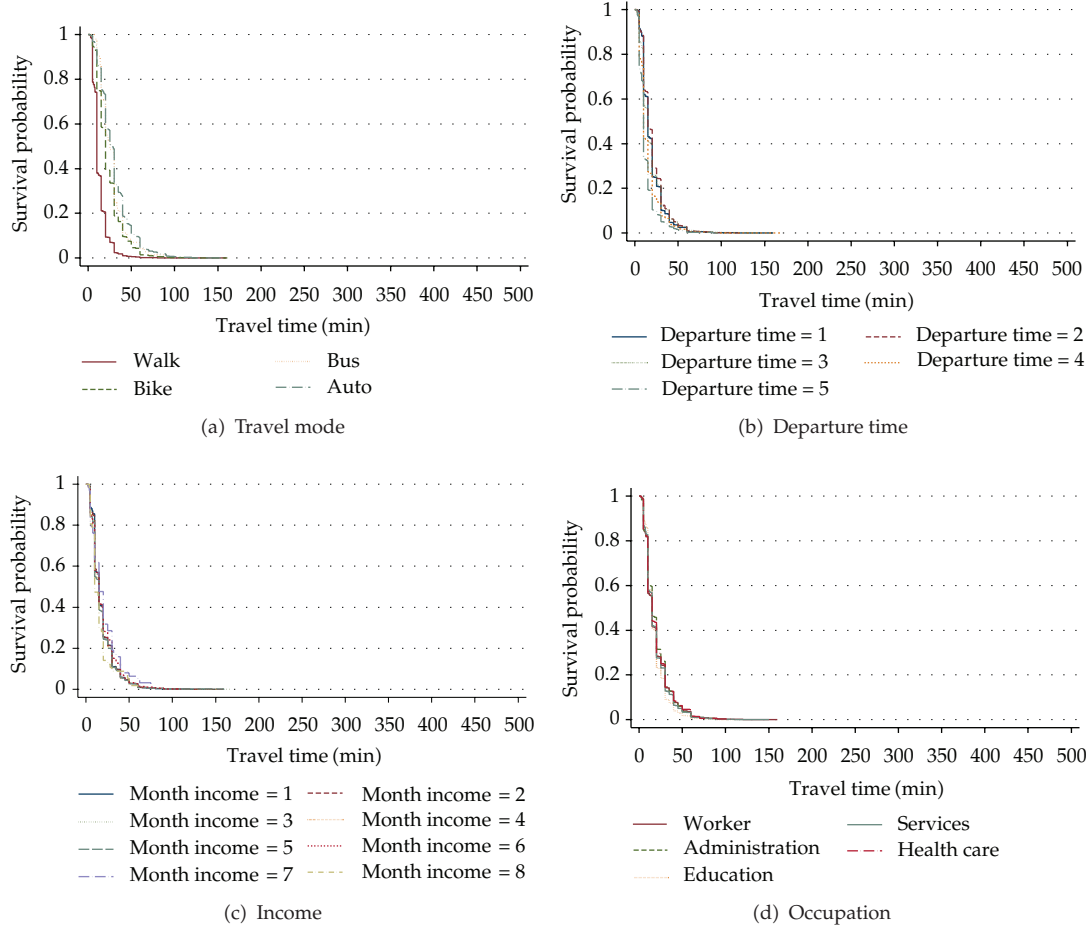
$$\begin{aligned}
 h(t) &= \exp(2.53 - 0.01 \text{ Income} + 0.0033 \text{ Age} - 0.03 \text{ Occuw} + 0.07 \text{ Occue} - 0.06 \text{ Modew} \\
 &\quad + 0.08 \text{ Modebi} + 0.00015 * \text{Distance} - 0.06 \text{ Departure time}), \\
 s(t) &= \exp(-\exp(2.53 - 0.01 \text{ Income} + 0.0033 \text{ Age} - 0.03 \text{ Occuw} + 0.07 \text{ Occue} \\
 &\quad - 0.06 \text{ Modew} + 0.08 \text{ Modebi} + 0.00015 * \text{Distance} - 0.06 \text{ Departure time})).
 \end{aligned}
 \tag{3.15}$$

The estimation results indicate that the travel time of high-income travelers is 1% lower than that of low-income travelers. Moreover, old persons are likely to spend longer time in work-home trip. Regarding commuter's occupation, blue-collar workers are likely to spend shorter time for evening commute trip, while the travel times for teachers or students are longer. Comparing with other modes, walking trips have shorter time, while cycling trips have longer time. Long-distance trip takes longer travel time. The later the commuters depart from work, the longer the travel times will be.

#### **4. Prediction of the Commute Activity-Travel Agenda**

As explained in Section 3.1, the key timings and durations  $A_1$ ,  $D_1$ , and  $A_2$  can be calculated once the values of  $Dt_1$ ,  $T_1$ ,  $Dt_2$ , and  $T_2$  are predicted. Here is an example: the first member in the family with ID number 010104065 in our sample, Mr. Chen, is 45 years old, has an occupation of services, and earns 0–1500 RMB every month. His commute mode is walk, and the distance of one-way commute trip is 1500 meters. He had a typical commute activity-travel pattern on the survey day, which is shown in Figure 8.

Based on the developed models, his daily commute time allocation is predicted and shown in Figure 9.



**Figure 7:** Survival curves of the work-to-home travel time impacted by several factors.

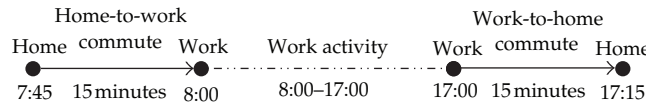
**Table 7:** Estimation results of the work-to-home travel time model.

Variables	Weibull distribution		Exponential distribution	
	Coef.	z-stat.	Coef.	z-stat.
Constant	2.65	164.15	2.53	79.98
Income	-0.01	-8.13	-0.01	-3.57
Age	0.0034	19.23	0.0033	8.80
Occu-w	-0.03	-6.20	-0.03	-2.22
Occu-e	0.05	6.56	0.07	4.25
Mode-w	-0.06	-5.53	-0.06	-3.51
Mode-bi	0.05	4.38	0.08	-3.51
Mode-a	0.05	3.33	—	—
Distance	0.00015	119.81	0.00015	66.12
Departure time	-0.06	-23.50	-0.06	-11.96
$\gamma$	0.4835	—	1	—
Prob > $\chi^2$	0.0000		0.0000	
N	37842			

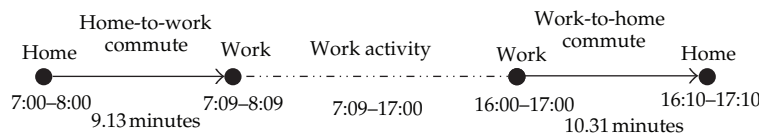


**Table 8:** Goodness of fit index and estimated distribution statistics of the work-to-home travel time model.

Model statistics	Weibull distribution	Exponential distribution
MAPE value	0.4541	0.3905
Mean (min)	18.50	13.36
Maximum (min)	8.44	6.17
Minimum (min)	269.11	172.80



**Figure 8:** Observed commute activity-travel pattern and time allocation.



**Figure 9:** Predicted commute time allocation.

Comparing with the observed values, the errors of the predicted results can be calculated as follows.

- (i) The errors of home-to-work departure time: the maximum error is 45 minutes, the minimum error is 15 minutes.
- (ii) The error of home-to-work travel time: 5.47 minutes.
- (iii) The errors of arrival time at work location: the maximum error is 51 minutes, and the minimum error is 0 minute.
- (iv) The errors of work-to-home departure time: the maximum error is 60 minutes, and the minimum error is 0 minute.
- (v) The error of work-to-home travel time: 4.29 minutes.
- (vi) The errors of arrival time at home: the maximum error is 65 minutes, and the minimum error is 5 minutes.

By comparing the predicted values of the developed daily time allocation model with the observed values of our sample, the maximum errors for all the departure times and the activity-travel durations are as follows:  $error(Dt_1) = 21.38$  minutes;  $error(T_1) = 7.40$  minutes;  $error(A_1) = 28.78$  minutes;  $error(D_1) = 28.78$  minutes;  $error(Dt_2) = 23.40$  minutes;  $error(T_2) = 7.17$  minutes;  $error(A_2) = 30.57$  minutes. These statistics indicate that the overall goodness of fit of the model is rather satisfying.

Results also show that the errors of departure and arrival time are much larger than that of the travel time. The main reason is that we divided the natural continuous departure time into discrete time interval artificially, which reduces the predictive accuracy of the model. It has been tested that the smaller the interval is, the higher the predictive accuracy will be. However, as there are already five alternatives for both of the departure time choice models, at least half-an-hour interval will make the number of the alternatives double. Then the model will be more complex and the efficiency of the model will be lower.

## 5. Conclusions

In this paper, we have formulated and estimated a joint model of departure time choice and travel duration for commuters' daily activity-travel time allocation. Two ordered probit models have been employed to forecast the home-to-work and work-to-home commute departure time. By doing so, we were able to recognize the natural temporal ordering among the departure time alternatives and address the IIA limitation of the standard MNL model. Furthermore, two AFT models were built and estimated to predict home-to-work and work-to-home travel times by using GA as parameters optimization. Then timing choice of a simple daily activity-travel pattern has been calculated.

Comparing with the previous studies, this paper developed a joint discrete-continuous model system to predict all the departure times and the activity-travel durations of a typical daily commute activity pattern. Results of this study not only contribute to developing a full-scale daily activity pattern forecasting model but also provide useful insights in the influence of sociodemographic variables on activity-trip timing decisions as well as the time constraint between daily activities and trips. Moreover, GA contributes to the optimization and thus the high accuracy of the travel time prediction model. In addition, this analysis of daily commute time allocation can be applied to a wide range of TDM policies, especially the measures aimed at adjusting the commute times, such as flexible work and compressed working week. For examining the effects of the traffic demand strategies, the developed model cannot only describe the overall change of the daily activity schedule caused by the strategies but also explore the time tradeoff between the connected trips as well as trips and activities. Besides evaluating the effects of the transportation demand management strategies, this study is also essential for planning the development and construction of new transportation infrastructure as well as examining the potential responses to improved traffic operational measures.

The results of this paper confirm that the discrete choice models and the continuous models can match well in the calculation of a whole-day activity-travel schedule, although comparing with the continuous models, the predictive accuracies of the discrete choice models are a little lower, as they divide the naturally continuous time into artificially defined time periods. Similar studies were also found to employ the discrete-continuous methods to model coupled mode and commute timing choice [12, 25]; joint activity-type preference, travel time, and activity duration [10]; as well as other activity-travel behaviors [27]. Therefore, it can be expected that we can further employ the combination of the discrete and continuous models to predict all the dimensions of the entire-day activities and trips. This future study can provide more useful insights into the nature of travelers' daily activity-travel decision making.

It should be pointed out that only the typical commute activity-travel pattern, comprising two commute trips and one work activity, has been considered in this paper. In reality, it is also common to observe other commute activity-travel patterns such as those including work-based subtour or home-based nonwork trips, or those having stops during commute travel. Further study may be done to model such daily activity-travel patterns. It will also be very important to examine one's activity-travel patterns over multiple days if the multiday travel survey data can be obtained.

## Acknowledgment

The research is funded by the National Natural Science Foundation of China 50908099 and 51108053 and China Postdoctoral Science special Foundation 201104526.

## References

- [1] M. B. Sundo and S. Fujii, "The effects of a compressed working week on commuters' daily activity patterns," *Transportation Research Part A*, vol. 39, no. 10, pp. 835–848, 2005.
- [2] F. Zong, Z. Juan, and H. Jia, "Examination of staggered shifts impacts on travel behavior: a case study of Beijing," *Transport*, STRA-2011-0129.R1, Forthcoming, 2011.
- [3] K. Ozbay and O. Yanmaz-Tuzel, "Valuation of travel time and departure time choice in the presence of time-of-day pricing," *Transportation Research Part A*, vol. 42, no. 4, pp. 577–590, 2008.
- [4] R. Hung, "Using compressed workweeks to reduce work commuting," *Transportation Research Part A*, vol. 30, no. 1, pp. 11–19, 1996.
- [5] G. Giuliano and T. F. Golob, "Staggered work hours for traffic management: a case study," *Transport Research Record*, vol. 1280, pp. 46–58, 1990.
- [6] C. Raux, T. Y. Ma, I. Joly, V. Kaufmann, E. Cornelis, and N. Ovtracht, "Travel and activity time allocation: an empirical comparison between eight cities in Europe," *Transport Policy*, vol. 18, no. 2, pp. 401–412, 2011.
- [7] D. Ettema, F. Bastin, J. Polak, and O. Ashiru, "Modelling the joint choice of activity timing and duration," *Transportation Research Part A*, vol. 41, no. 9, pp. 827–841, 2007.
- [8] P. Vovsha and M. Bradley, "Hybrid discrete choice departure-time and duration model for scheduling travel tours," *Transportation Research Record*, no. 1894, pp. 46–56, 2004.
- [9] S. Fujii and R. Kitamura, "Evaluation of trip-inducing effects of new freeways using a structural equations model system of commuters' time use and travel," *Transportation Research Part B*, vol. 34, no. 5, pp. 339–354, 2000.
- [10] M. M. Hamed and F. L. Mannering, "Modeling travelers' postwork activity involvement: toward a new methodology," *Transportation Science*, vol. 27, no. 4, pp. 381–394, 1993.
- [11] C. R. Bhat and S. K. Singh, "A comprehensive daily activity-travel generation model system for workers," *Transportation Research Part A*, vol. 34, no. 1, pp. 1–22, 2000.
- [12] K. M. N. Habib, N. Day, and E. J. Miller, "An investigation of commuting trip timing and mode choice in the Greater Toronto Area: application of a joint discrete-continuous model," *Transportation Research Part A*, vol. 43, no. 7, pp. 639–653, 2009.
- [13] X. Zhang, H. J. Huang, and H. M. Zhang, "Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city," *Transportation Research Part B*, vol. 42, no. 1, pp. 38–56, 2008.
- [14] B. Dong, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005.
- [15] P. F. Pai, "System reliability forecasting by support vector machines with genetic algorithms," *Mathematical and Computer Modelling*, vol. 43, no. 3–4, pp. 262–274, 2006.
- [16] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transportation Research Part A*, vol. 35, no. 1, pp. 1–28, 2000.
- [17] K. A. Small, "The scheduling of consumer activities: work trips," *American Economic Review*, vol. 72, no. 3, pp. 467–479, 1982.
- [18] C. R. Bhat, "Analysis of travel mode and departure time choice for urban shopping trips," *Transportation Research Part B*, vol. 32B, no. 6, pp. 361–371, 1998.
- [19] K. A. Small, "A discrete choice model for ordered alternatives," *Econometrica*, vol. 55, no. 2, pp. 409–424, 1987.
- [20] C. R. Bhat, "A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity," *Transportation Research Part B*, vol. 30, no. 3, pp. 189–207, 1996.
- [21] C. R. Bhat, "A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute," *Transportation Research Part B*, vol. 30, no. 6, pp. 465–480, 1996.
- [22] Z. C. Juan and J. C. Xianyu, "Daily travel time analysis with duration model," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 4, pp. 62–67, 2010.
- [23] C. R. Bhat and J. L. Steed, "A continuous-time model of departure time choice for urban shopping trips," *Transportation Research Part B*, vol. 36, no. 3, pp. 207–224, 2002.
- [24] D. Janssens, Y. Lan, G. Wets, and G. Chen, "Allocating time and location information to activity-travel patterns through reinforcement learning," *Knowledge-Based Systems*, vol. 20, no. 5, pp. 466–477, 2007.
- [25] K. M. N. Habib, "Modeling commuting mode choice jointly with work start time and work duration," *Transportation Research Part A*, vol. 46, pp. 33–47, 2012.

- [26] T. Schwanen and M. Dijst, "Travel-time ratios for visits to the workplace: the relationship between commuting time and work duration," *Transportation Research Part A*, vol. 36, no. 7, pp. 573–592, 2002.
- [27] R. M. Pendyala and C. R. Bhat, "An exploration of the relationship between timing and duration of maintenance activities," *Transportation*, vol. 31, no. 4, pp. 429–456, 2004.
- [28] K. M. N. Habib, J. A. Carrasco, and E. J. Miller, "Social context of activity scheduling: discrete-continuous model of relationship between "with whom" and episode start time and duration," *Transportation Research Record*, no. 2076, pp. 81–87, 2008.
- [29] A. H. Wright, "Genetic algorithms for real parameter optimization," in *Foundations of Genetic Algorithms*, J. E. Rawlins, Ed., pp. 205–218, Morgan Kaufmann, San Mateo, Calif, USA, 1991.
- [30] N. Ahmad and A. F. M. A. Haque, "Optimization of process planning parameters for rotational components by genetic algorithms," in *Proceedings of the 4th International Conference on Mechanical Engineering*, vol. 7, pp. 227–233, Dhaka, Bangladesh, 2001.
- [31] J. B. . Yao, B. Z. Yao, L. Li, and Y. L. Jiang, "Hybrid model for displacement prediction of tunnel surrounding," *Neural Network World*, vol. 22, no. 3, pp. 263–275, 2012.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

