Modeling Travel Mode Choice: Application of Discrete-Continuous Model (A Case of Traveler in Yogyakarta, Indonesia)

Muhammad Zudhy IRAWAN [a], Tomonori SUMI [b]
[a,] *Department of Civil and Environmental Engineering, Gadjah Mada University, Yogyakarta, 55281, Indonesia*
[b] *Department of Urban and Environmental Engineering, Kyushu University, Fukuoka,819-0395, Japan*
[a] *E-mail: zudhyirawan@ugm.ac.id*
[b] *E-mail: sumi@doc.kyushu-u.ac.jp*

Abstract: We propose a joint of discrete and continuous model to analyze several factors influencing travel mode choice. This is essential since mode choice is not only affected by spatial constraints, but also temporal constraints, such as departure/arrival time and available time windows for an activity. To do this, a multinomial logit model is examined in terms of discrete choice model. Meanwhile, a disutility model considering the earliness of home departure time and probability of being late is examined in terms of continuous choice model. To evaluate whether or not our proposed model can be suitably implemented, we take high school students on morning commute travel taking public transports (bus and paratransit) and motorcycles in Yogyakarta. The result shows that more than 54 percent of travelers can be precisely estimated in deciding their travel mode choice.

Keywords: Departure Time, Travel Mode Choice, Commuting Travel, Travel Demand Model

1. INTRODUCTION

In the theory of transport modeling, since there are significant drawbacks of four steps model, such assign each step of this method may be given a behavioral interpretation and having the limitations in evaluating demand management policies, some of transport modelers shift to a principle of utility maximization that underlies established microeconomic theory (see for instance: Ben-Akiva, 1984; Bhat, 1998). However, most of these studies consider travel mode choice merely on spatial constraints regardless to the temporal constraints, such as departure time or arrival time (Irawan and Sumi, 2012). In fact, temporal constraints, particularly in commuting travels are critical consideration for travelers in executing their trip and deciding their travel mode (Recker, 2001).

Due to this, this paper aims to model travel mode choice by combining discrete choice model, i.e. multinomial logit model (Ben-Akiva and Lerman, 1985) and continuous choice model, i.e. disutility model proposed by Sumi et al., 1990.

2. MATHEMATICAL MODEL STRUCTURE

2.1 Discrete Choice Model: Multinomial Logit

A multinomial logit (MNL) model assumes that a person selects a choice having the highest utility value. If each alternative mode *m* has a person-specific utility for traveler *n* (U_{nm}), the utility can be expressed by linear function as follows.

$$U_{nm} = V_{nm} + \varepsilon_{nm} \tag{1}$$

where U_{nm} is the utility of traveler n on mode m and ε_{nm} is its random error term.

According to the theory of random utility maximization, traveler n chooses mode m when the utility of that alternative mode is the maximum of all alternative modes (denoted as mode b)

$$U_{nm} > \max_{b=1,2,3...n,m\neq b} U_{nb} \tag{2}$$

Thus, the probability of choosing mode m is calculated by:

$$\Pr\left(U_{nm} > \max_{b=1,2,3...n,m\neq b} U_{nb}\right) = \Pr\left(V_{nb} \leq V_{nm} + \left(\varepsilon_{nm} - \varepsilon_{nb}\right)\right) \tag{3}$$

According to Ben-Akiva and Lerman (1985), by assuming the error term is identically and independently distributed with Gumbel Distribution, the cumulative distribution of the random error term of the chosen alternative (F(ε_{nm})) or the probability of choosing mode m can be written as:

$$\Pr\left(V_{nn} > V_{nb}\right) = F\left(\varepsilon_{nm}\right) = \frac{\exp\left(V_{nm}\right)}{\exp\left(V_{nm}\right) + \sum_{b\neq m} \exp\left(V_{nb}\right)} \tag{4}$$

2.2 Continuous Choice Model: Departure Time

Travelers who commute to work or school have a designated starting time for their works/classes. If they arrive late at the destination point, they will be penalized for late arrival. On the other hand, they are also reluctant to depart over-early from home to minimize their lateness. Therefore, they depart by minimizing the disutility related to earliness of home departure time (assumed as D_1) and the disutility related to probability of being late (assumed as D_2). Under the usual assumption of utility-maximizing principle, it can be determined by:

$$\text{Max } U = \text{Min}\left[D_1\left(td_1\right) + D_2\left(td_2\right)\right] \tag{5}$$

In regards to the disutility related to the earliness of home departure time (D_1), since the work/school starting time is designated, travelers are generally unwilling to depart from their home hurriedly. They prefer to depart as late as they can (denoted as td_1). Therefore, the earlier they leave home in the morning, the more disutility will be derived. However, there is an earliest acceptable departure time (td_e) which represents a threshold at which a traveler does not feel disutility if he/she departs later than this time. By assuming D_1 follows a linear function, it can be expressed by:

$$D_1 = \begin{cases} 0 & \text{if } \left(td_e \leq td_1\right) \\ \omega\left(td_e - td_1\right) & \text{if } \left(td_e > td_1\right) \end{cases} \tag{6}$$

$$\text{subject to: } \omega > 0;\ td_e \geq te;\ td_e \leq ts$$

in which, ω = a constant, ts = designated starting time at school/workplace, and te = earliest tolerable time from home.

Regarding the disutility of lateness probability (D_2), since there is a relationship between departure time and probability of being late, it can be modeled as a function which depends on travel time. Therefore, disutility related to lateness probability has already considered the disutility of lateness itself and the disutility of travel time as proposed by Palma and Arnott (1986).

If td_2 is departure time choice regarding lateness disutility derived and tt_{ij} is its travel time, the arrival time (ta) can be specifically calculated by summing td_2 and tt_{ij}. Therefore, travelers generally determine their departure time as td_2 = - tt_{ij}. However, since travel time varies depending on the operational conditions of the road which in turn depends on the presence and number of vehicles on the road, eventually ta cannot be easily determined as described. We thus have divided the time scale into fifteen-minute time intervals. In addition, we applied a probability density function (PDF) describing the variation of travel time within each time interval. By using this approach, PDF of arrival time (ϕta) is determined by:

$$\phi ta\left(t\left|l_{ij}, td_2\right.\right) = \phi tt_{ij}\left(t - td_2\left|l_{ij}\right.\right) \tag{7}$$

where the above equation expresses that either t_a or tt_{ij} is dependent on departure time (td_2) and distance (l_{ij}) from origin-i to destination-j.

If ts is the designated starting time at school/workplace and lateness is defined as a condition where $ta > ts$, the probability of lateness (α) is obtained as a function of td_2 and l_{ij} as follows:

$$\alpha = \int_{ts}^{\infty} \phi ta\left(t\left|l_{ij}, td_2\right.\right) dt \tag{8}$$

Then, substituting Eq. 7 into Eq. 8, it can be obtained:

$$\alpha = \int_{ts}^{\infty} \phi tt_{ij}\left(t - td_2\left|l_{ij}\right.\right) dt = \int_{-td_2}^{\infty} \phi tt_{ij}\left(t\left|l_{ij}\right.\right) dt \tag{9}$$

From the above equation, it is clear that the probability of being late is dependent on the variation of travel time as previously mentioned.

Since the travelers are penalized if they late arrive at the school/workplace, they will minimize its penalty by leaving home early. This behavior can thus be formulated as a function of lateness probability ($D_2 = f(\alpha)$).

Considering the two disutilities above (D_1 and D_2) and referring to Eq. 5, the traveler’s optimal departure time (td_{12}) is determined at the minimum point of the D_{12}. However, since each individual may judge their optimal departure time differently depending on the earliest acceptable departure time and maximum tolerable value of lateness probability, this produces a varying optimal departure time (td_{12}). As a result, td_{12} can be assumed to be normally distributed.

Further, to validate our estimated departure times and its disutility values, we compare between simulated and observed arrival times at the destination place. First, it is obvious that the effect of departure time decision at origin-i on arrival time at destination-j is highly dependent on traffic condition from i to j.

Given an l_{ij} distance section and a travel speed of v_{ij}, it is clear that travel time can be determined as $tt_{ij} = l_{ij}/ v_{ij}$. However, since travel speed varies during every time interval, we can use the probabilistic distribution to depict these travel speed fluctuations. Considering v_{ij} is always positive, the distribution of travel speed v_{ij} (denoted as $\Phi[v_{ij}]$) is then calculated by following a log normal distribution as shown by:

$$\Phi\left[v_{ij}\right] = \left(1 / v_{ij}.\sigma\left(v_{ij}\right).\sqrt{2\pi}\right).\exp\left[-\left(\ln v_{ij} - \mu\left(v_{ij}\right)\right)^2 / 2.\sigma\left(v_{ij}\right)^2\right] \tag{10}$$

where μ and σ are the average and standard deviation of ln (v_{ij}) at a given time interval respectively.

After the distribution of travel speed is known, PDF of travel time can be easily determined as follows:

$$\phi tt_{ij}\left(tt_{ij}\left|l_{ij}\right.\right) = \phi\left(l_{ij}/tt_{ij}\right)\left|dv_{ij}/dtt_{ij}\right| = l_{ij}/tt_{ij}^2\,\phi\left(l_{ij}/tt_{ij}\right) \tag{11}$$

Thus, by substituting Eq. 11 into Eq. 9, the value of α can be calculated at every departure time.

Considering tt_{ij} in Eq. 11 represents the total travel time from door to door, for a traveler who uses public transportation, total travel time can be expressed by (see Figure.1):

$$tt_{ij} = tt_{ac} + tt_w + \sum_{x=1}^{X} tt_{r(x)} + tt_{eg} \tag{12}$$

where tt_{ac} is access travel time, tt_w is public transport waiting time, tt_r is riding travel time, tt_{eg} is egress travel time, and x is the number of sections dividing bus route.

Assuming that the traveler departs from home at $t = td$ along l_{ac} , the PDF of arrival time at the place where he/she boards the public transport (denoted as ϕta_{ac}) is given by:

$$\phi ta_{ac}\left(t\left|td, l_{ac}\right.\right) = \phi tt_{ac}\left(t - td\left|l_{ac}\right.\right) \tag{13}$$

where tt_{ac} is the access travel time obtained by observation and is log normally distributed due to consideration that travel speed is always greater than zero.

Since there is no schedule of public transports arrival time, causes difficulties in determining an exact arrival time such that the waiting time is also difficult to determine. It is important to note that several studies indicate that waiting time, in reality, is perceived more inconvenient than travel time (Hensher, 2001). Therefore, waiting time cannot be assumed to be negligible. To simplify this problem, we assume that passengers are familiar with the time when the bus/paratransit is usually passing at the boarding location. By this, we thus calculate the probability of a passenger to ride on b^{th} bus at u^{th} boarding location (P_b^u).

Since ϕta_{ac} is known from Eq. 13 and arrival/departure time of b^{th} bus at u^{th} boarding location also known from field data (denoted as td_b^u), denoting the number of available buses as B, the probability that a passenger can board the b^{th} bus at the u^{th} bus stop (P_b^u) is derived by:

$$P_b^u\left(td,l_{ac}\right)=\prod_{b=1}^{B}\left(1-P_{b-1}^u\right)\int_{-\infty}^{\infty}\phi ta_{ac}\left(t\left|td,l_{ac}\right.\right)\int_{t}^{\infty}\phi td_b^u\left(\tau\right)d\tau dt \tag{14}$$

A very important factor in deciding home departure time is the variation of riding time along the route from origin to destination. To obtain a good estimate of riding time, a route along l_x distance will be divided into x road sections (see Figure.1). Total riding time is calculated by summing the riding time along all x road sections during the specific time interval when the vehicle passes on that road section.

First, assuming that a passenger can always board the b^{th} bus at the u^{th} bus stop ($P_b^u = 1$). Given bus travel time and its PDF along l_{x1} distance as $tt_{r(x1)}$ and $\phi tt_{r(x1)}$ respectively, the PDF of arrival time at point $x1$ (denoted as ϕta_{x1}) conditional on l_{x1} distance is given by:

$$\phi ta_{x1}\left(t\left|l_{x1}\right.\right)=\int_{-\infty}^{\infty}\phi td_b^u\left(td_b^u\left|l_{x1}\right.\right)\phi tt_{r(x1)}\left(t-td_b^u\left|l_{x1}\right.\right)dtd_b^u \tag{15}$$

Then, by using above equation, the calculation process is repeated until x^{th} road section. Using the same equation for access travel time in Eq. 13, arrival time at destination point (ta) conditional on bus arrival time at destination bus stop (ta_x) and egress distance to destination point (l_{eg}) can be easily determined. However, since there is distribution of travelers' arrival time at each arrival points (ta_{ac}, ta_{x1}, …, ta), we employ the convolution integration to find the PDF of arrival time at the destination point.

By assuming a traveler arrives at destination point by using the b^{th} bus, the probability of being late if traveler uses the b^{th} bus according to Eq. 8 is given by:

$$\alpha_b=\int_{ts}^{\infty}\phi ta\left(t\left|b,l_{ij},td\right.\right)dt \tag{16}$$

Then, if we remove the assumed certainty that a passenger can always board the b^{th} bus at u^{th} boarding location ($P_b^u = 1$) and instead use the Eq. 14 to determine P_b^u, the probability of being late is given as a function of departure time as follows:

$$\alpha=\sum_{b=1}^{B}P_b^u\left(td,l_{ac}\right)\alpha_b \tag{17}$$

Meanwhile, the calculation model of motorcycle is similar to the calculation method for public transportation with the exception that we will consider a non-existent public transport waiting time. We will regard the length time to find parking space as negligible since a person only needs a few seconds to find a parking space for his/her motorcycle. Also, it should be noted that schools in Yogyakarta always provide a huge parking space for its students, especially for motorcycle so that the students never have the experience to lack of parking lot.

Further, assuming traveler-n chooses transport mode-m as his/her mode, Since there is a distribution of disutility values, the probability of choosing travel mode-m (P_m) for the entire range of disutility can be calculated by:

$$P_m(D_m) = \int_0^{\infty} \phi[D_m] \left[\prod_{b \neq m}^{B} \int_{D_m}^{\infty} \phi[D_b] dD_b \right] dD_m \tag{18}$$

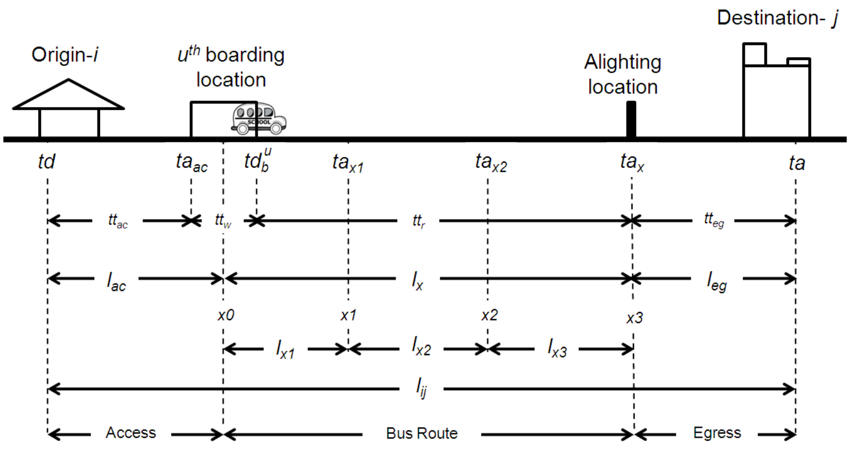

Figure 1. Trip by public transportation from origin to destination

2.3 Joint Model

After the equations of choice probability is obtained either for discrete or continuous model and since the error random terms are different between the two (Gumbel distribution for discrete choice model and normal distribution for continuous choice model), we can transform them into an equivalent standard normal variable as shown by (Lee, 1983):

$$\begin{aligned} \varepsilon_m &= J_1(\varepsilon_m) = \Phi^{-1}\left[F(\varepsilon_m)\right] \\ \alpha_m &= J_2(\alpha_m) = \Phi^{-1}\left[F(\alpha_{mtd})\right] \end{aligned} \tag{19}$$

where Φ^{-1} indicates an inverse of the cumulative standard normal variable.

Finally, referring to Habib et al. (2009) and taken into our case, the joint probability of choosing mode-*m* and a corresponding departure time *td* (*P* [*departure time* = *td* $\cap$ *mode* = *m*]) can be expressed as:

$$P(td \cap \varepsilon \le J_1(\varepsilon_m)) = \frac{1}{\sigma_{td}} \phi\left(\frac{td - \mu_{td}}{\sigma_{td}}\right) \Phi\left(\frac{J_1(\varepsilon_m) - \rho_{mtd} . J_2(\alpha_{mtd})}{\sqrt{1 - \rho_{mtd}^2}}\right) \tag{20}$$

in which ρ_{mtd} is the correlation of two distributions (bivariate normal distribution).

3. DATA COLLECTION

3.1 Arrival Time and Questionnaire Survey

Students from Senior High School 1 in Yogyakarta were sampled as respondents. The questionnaire survey was conducted on March 17, 2009. Out of 365 randomly distributed forms, 312 forms (85.5%) were recollected effectively. From 312 respondents, 74.36% were motorcycle users, 11.22% were bus users, 3.53% were bicycle users, 8.97% walked to school, and 1.92% were escorted by their parents/family members using a car. However, due to limits of this research scope which focuses on travel behavior for both public transport and motorcycle users, only 267 (85.56%) sets of data will be used in this research.

The questionnaire items were divided into of three components: (1) individual information such as gender, age, driving license and motorcycle owning, and home address, (2) household characteristic information such as the interaction of family members relating to the school morning commute, their travel diary on morning commute, their workplace/school location, and their arrival and starting time at workplace/school, and (3) detailed information on the school morning commute such as trip chain and its route, travel mode and its weekly frequency, the location of departure and alighting of the public transport as well as its route number, and transport cost.

Further, the student arrival time survey was conducted in conjunction with the questionnaire survey. Surveyors note the respondent arrival time at school in the questionnaire form before the surveyors hand over the questionnaire form to respondent. Then the respondent fulfill the questionnaire form and submit to the surveyor.

3.2 Travel Speed Survey

Based on the questionnaire survey, there are six routes and three routes were identified for respondents using city buses and Trans-Jogja buses respectively, and 27 routes for motorcycle user respondents. It will be impossible to survey all motorcycle possible routes. Therefore, we minimized the number of routes by lumping together the road sections with similar characteristics and then surveying a limited number of road sections as the representative of all road sections with the same characteristic.

A vehicle plate number survey was used to obtain this data. By checking the passing time of every vehicle with same number at the different points of a section and measuring the distance, a distribution of vehicle travel speeds in different time duration will be obtained (mean and standard deviation).

Due to the many observation points, this survey had to be conducted in a span of three days at times with similar morning traffic patterns for those days (Tuesday – Thursday) between June 23th - 25th, 2009. The observed mean and standard deviation of vehicle travel speed in per 15 minutes in every road section of different routes from 06.00 a.m. to 08.00 a.m. are shown in Figure 2 for motorcycle mode, Figure 3 for public transport mode.

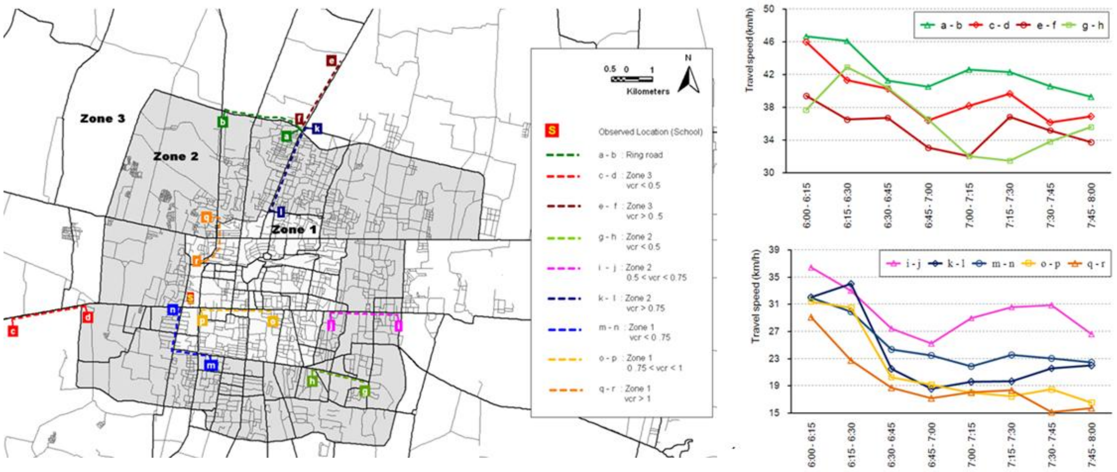

Figure 2. Motorcycle selected routes, observing points, and average travel speeds

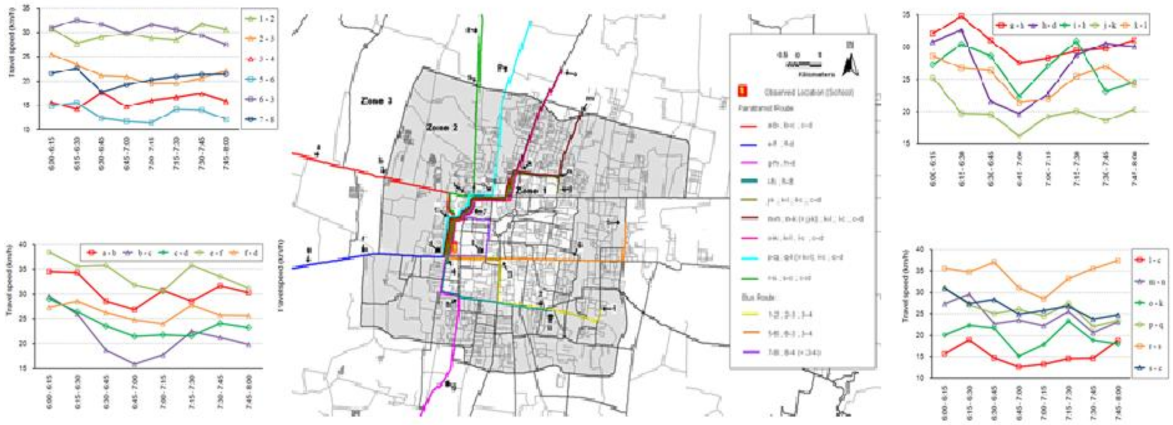

Figure 3. Public transport selected routes, observing points, and average travel speeds

4. MODEL APPLICATION AND RESULTS

Applying the proposed model above, the estimated parameters are shown in Table 1. In obtaining those parameters, the calculation process is as follows. First, to express the individual difference in recognition of earliest home departure time (td_e), the threshold of td_e regarding D_1 was assumed as a random variable following the normal distribution. It is obvious that if the mean (μ) and standard deviation (σ) are known, by using a Monte Carlo simulation and follows normal distribution, the distribution of td_e can be obtained (Fishman, 1995) such that by giving arbitrarily initial values to the parameters of ω, D_1 can be determined for each departure time.

Looking into D_2, which represents a function of the probability of being late (α), it could be directly represented by α itself when the penalty of being late is assumed to be 1. According to Eq. 9, where tt_{ij} was recognized from survey data and calculated by Eq. 12, the value of α can be computed also for each minute of departure time.

By minimizing the sum of D_1 and D_2 and thus combining with the multinomial logit model by using Eq. 19 and Eq. 20, the estimated parameters can be achieved. To check whether or not our estimated parameters result travel mode choice which is similar to the observation results, Boolean variables y_m and y_b are then created in which $y_m = 1$ and $y_b = 0$ if $P_m^n \geq 0.5$ and a value of $y_m = 0$ and $y_b = 1$ if $P_m^n < 0.5$. The percentage of comparison among theoretical and observed travel mode choice is shown in Table 2.

Table 1. The estimated parameters

Correlation coefficients	Para. (public trans.)	Para. (motorcycle)
ρ	0.08	0.32
Continuous model	Para. (public trans.)	Para. (motorcycle)
ω	0.63	0.56
μtd_e	6:10 a.m.	6:26 a.m.
σtd_e(minutes)	21	23
Discrete model (motorcycle as a base mode)	Para. (public trans.)	SE
Age	- 0.687	0.433
Gender (Male)	0.160	0.624
Driving License	**- 3.743**	**0.801***
Number of vehicles	**- 2.885**	**0.444**
Number of school-going children	0.022	0.498
Characteristics of household members	**1.216**	**0.387***
Distance	- 0.013	0.083

Bold indicates the parameter is statistically significant at the 95% level, whereas * is statistically significant at the 99% level

Table 2. Percent correct of mode choice model

Travel Mode	Observed	Estimated		Percentage
		Hit	Miss	
Public Transport	35	19	16	54.29
Motorcycle	232	187	45	80.60

From Table 1, it can be recognized that by using motorcycle mode, traveler can lately depart around 16 minutes from destination place than using public transport. Public transport users also tend to obtain higher disutility (0.63) than motorcycle users (0.56).

In regards to discrete choice model, driving license becomes the most significant factor for the travelers in choosing their transport mode (-3.743). The other influence factors are number of vehicle ownership (-2.885) and characteristic of household member (1.261). More household members who work/study with designated starting time at workplace/school, less probability for the students to use motorcycle mode. Meanwhile, the factor of age, gender, number of school-going children, and distance have no influence for the travelers to choose between motorcycle mode and public transport mode.

REFERENCES

Ben-Akiva, M., Cyna, M., De Palma, A. (1984) Dynamic model of peak period congestion. *Transportation Research Part B*, 18, 339-355.

Ben-Akiva, M., Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge.

Bhat, C.R. (1998) Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B*, 32, 361-371.

Fishman, G.S. (1995) *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York.

Habib, K.M.N., Day, N., Miller, E.J. (2009) An investigation of commuting trip timing and mode choice in the Greater Toronto Area: application of a discrete-continuous model. *Transportation Research Part A*, 43, 639-653.

Hensher, D.A. (2001) Measurement of the valuation of travel time saving. *Journal of Transport Economics and Policy*, 35, 71-98.

Irawan, M.Z., Sumi, T. (2012). Motorcycle-based adolescents' travel behaviour during the school morning commute and the effect of intra-household interaction on departure time and mode choice. *Transportation Planning and Technology*, 35(3), 263-279.

Lee, L.F. (1983) Notes and comments generalized econometric models with selectivity. *Econometrica*, 51, 507-512

Palma, A. Arnott, R. (1986) Usage-dependent peak-load pricing. *Economics Letter*, 20, 101-105

Recker, W.W. (2001) A bridge between travel demand modeling and activity-based travel analysis. *Transportation Research Part B*, 35, 481-506

Sumi, T., Matsumoto, Y., Miyaki, Y. (1990) Departure time and route choice of commuters on mass transit system. *Transportation Research Part B*, 24 (4), 247- 262.