Analyses of Route Choice and Route Switching Behavior Using Panel ETC Data from Tokyo Metropolitan Expressway

Tawin TIRATANAPAKHOM [a], Takashi OGUCHI [b], Shinji TANAKA [c], Sungjoon HONG [d], Hiroshi WARITA [e]

[a] *Department of Civil Engineering, The University of Tokyo, Tokyo, 153-8505, Japan; E-mail: tawinhkt@gmail.com*
[b,d] *Institute of Industrial Science, The University of Tokyo, Tokyo, 153-8505, Japan;*
[b]*E-mail: takog@iis.u-tokyo.ac.jp*
[d]*E-mail: mrhong@iis.u-tokyo.ac.jp*
[c] *Institute of Urban Innovation, Yokohama National University, Yokohama, 240-8501, Japan; E-mail: stanaka@ynu.ac.jp*
[e] *Maintenance & Traffic Management Department, Metropolitan Expressway Co.,Ltd., Japan; E-mail: h.warita1116@shutoko.jp*

Abstract: The conventional logit-based models with cross-sectional data are not capable to capture the individual's behavior that is correlated over time. It has been known that panel data provides the source of information to overcome this issue. However, collecting panel data, especially from revealed preference (RP) data, is generally expensive and difficult in practice. The present paper shows the use of repeated observations from Electronic Toll Collection (ETC) data with information of attributes derived from detector data to analyze and model route choice and route switching behavior on the selected study area of Tokyo Metropolitan Expressway. In general, the estimated results show that drivers response differently to different levels of congestion information. Moreover, when accounting for the panel data, the factors that capture panel effect are highly significant statistically and the improvement of the models' goodness of fit can be observed from both behavioral models.

Keywords: Route Choice, Switching Behavior, Panel Data, ETC Data

1. INTRODUCTION

Traffic congestion presents a major challenge of the society. The concentration of travel demand during peak period and the reduction of roadway capacity during incidents cause traffic congestion on road networks. Urban expressway is not an exception. Several ITS (Intelligent Transportation Systems) initiatives have been implemented to relieve the severity of traffic congestion. For example, traffic condition information is displayed on the variable message/graphic signboards especially before the major junctions, so that drivers can make decision en route to travel in the most effective way based on traffic information they receive. To maximize the potential of information technology, it is necessary to understand how drivers select the route for their trips and how traffic condition information affects route choice behavior (Bonsall, 1992).

Route choice model is a common tool to analyze and model drivers' route choice behavior. The conventional logit-based route choice models rely heavily on the cross-sectional data. With this data, generally, the standard model can capture the heterogeneity in drivers' taste through the observed individual socioeconomic characteristics and the differences in responsiveness to alternative attributes. In addition to the observed heterogeneity, there is a remaining component that represents random heterogeneity from some unobserved factors in drivers' behavior. This unobserved taste heterogeneity is generally ignored in the standard route choice model. In addition, since the standard model provides the same coefficients of each variable for all drivers, this implies that drivers with the same individual characteristics and under similar choice situation will behave in the same way for making decision. However, in reality, under such condition, a group of the same class of drivers might have different tastes and result in different decisions. This indicated that the results from standard model with fixed coefficients might provide a high rate of error in route choice estimation if there are a large variety of different tastes in each driver group.

To provide more realistic results, the model should be capable to accommodate both observed and unobserved heterogeneity mentioned earlier; particularly, the latter one that is likely to capture the differences in intrinsic route choice preferences. Instead of fixed-coefficient logit model, the taste variation across drivers can be accommodated by random-coefficient logit model. Random-coefficient or Mixed multinomial logit (MMNL) has been popular in various applications since the advances in simulation techniques and computational power (Revelt and Train, 1998; Bhat, 2000; Train, 2009; Hensher and Greene, 2003). In reality, drivers who experience traveling between a particular OD with alternative routes more than once might select a route based on the past experience. In this case, the MMNL with cross-sectional data might not be enough to account for the correlation between repeated choices over time or over choice situation since the cross-sectional data provides only single decision information of each driver. To overcome this issue, it is necessary to obtain the information of how each driver makes a series of decisions over repeated route choice situations. These repeated observations in route choice behavior can be obtained from the panel data.

Depending on the study cases, such information can be obtained from either/both revealed preference (RP) or/and stated preference (SP). RP data allows modeling route choice based on the observed choice decisions while SP data is based on hypothetical choices. Because of the difficulties of collecting actual route choice decisions in RP studies, SP studies become more popular. However, results from the decisions under experiments might be questionable if the analyst cannot set up the experiments for capturing the actual behavior (Ortúzar and Willumsen, 2001). The use of advanced technologies can overcome the difficulties of collecting RP data, specifically, for multi-day or multi-period observations. For example, Li *et al*. (2004) used the multi-day travel information from GPS equipped in vehicles to capture variability in commuting behavior. Tiratanapakhom *et al*. (2012b) examined the intra- and inter-personal variability using the panel ETC (Electronic Toll Collection) data on expressway in Tokyo. The GPS data, however, is based on the information from participants who agree to carry or to install GPS receiver in their vehicles. Sometimes the missing of data can be occurred especially when participants forget to operate or recharge the device. ETC system, in comparison, collects toll automatically when a vehicle passes the toll gate. At the same time, the system records vehicle ID (Identification Number), vehicle type, locations of entrance and exit, entering and exit times. In addition, the same vehicle ID will be recognized when the same vehicle re-enters the system. This feature allows us to observe the multi-day or multi-period travel behavior for the same vehicle on expressways where ETC system is implemented.

On Metropolitan Expressway (MEX) in Tokyo, the ETC system is widely implemented throughout the network. Moreover, since 2010, almost 90% of expressway users have paid their toll using this system (ORSE, 2011). If the route choice information can be observed from this system, it will allow the rich source of RP route choice data in the analyses. This study takes advantage of recognizing the same vehicle's ID in ETC system to observe repeated route choice behavior. Moreover, there are several factors that can influence drivers' route choice behavior. In addition to ETC data, the information of some factors was derived; specifically, traffic condition information such as predicted travel time and length of congestion from detector data. This information is commonly showed on graphic/message signboard.

This paper analyzes route choice behavior on selected study area of MEX using the MMNL model with repeated route choice observations to accommodate both observed and unobserved taste heterogeneity across individuals. In addition to route choice behavior analysis, the paper also examines how drivers diverse from their defined main route (i.e. switching behavior) under the impact of traffic condition information and other factors derived from both ETC and detector data.

The rest of this paper is organized as follows. Section 2 presents a brief review of modeling with repeated choices. Section 3 presents the descriptions of attributes extracted from ETC and detector data. Section 4 presents the model specification and discusses estimation results. Section 5 presents the conclusion.

2. MODELLING WITH REPEATED CHOICES

The logit-based route choice model can be formulated following the principles of discrete choice theory which is based on the utility-maximizing behavior by the decision maker. Random utility model is perhaps the common approach to estimate a behavioral model that satisfies utility maximization. The utility that a driver n associates with route i in choice situation t is formulated in equation (1) as:

$$U_{nit} = \beta_n' x_{nit} + \varepsilon_{nit} \tag{1}$$

where x_{nit} is a vector of explanatory variables that has been observed. An alternative specific constant is included in this component. β_n is a vector of taste coefficients associated with each variable, and ε_{nit} is an unobserved random component that includes all unobserved variables affecting the utility. ε_{nit} is assumed to be independent and identically distributed (iid) extreme value. If the values of β_n are fixed and ε_{nit} is independent over n, i and t, the model then becomes the standard multinomial logit model (MNL).

In MMNL (see Revelt and Train, 1998; Bhat, 2000; Train, 2009; Hensher and Greene, 2003), the β_n is considered to be distributed over individuals with density $f(\beta_n|\theta)$ where θ is a set of parameters of the distribution (e.g. mean and covariance). The logit probability conditional on β_n that driver n chooses route i; where there are $J = \{1, 2,\ldots, j\}$ alternative routes available at choice situation t is presented in equation (2).

$$P_n(i|\beta_n) = \frac{e^{\beta_n' x_{nit}}}{\sum_J e^{\beta_n' x_{njt}}} \tag{2}$$

Under the repeated choice observations during the choice situation T = {1, 2,…, t}, the probability conditional on β_n that individual n makes a sequence of route choice decisions is the product of equation (2) over choice situations.

$$P_{nT}(i|\beta_n) = \prod_{t=1}^{T} \frac{e^{\beta_n' x_{nit}}}{\sum_J e^{\beta_n' x_{njt}}} \tag{3}$$

For cross-sectional data, where the decision is independent over choice situations, the unconditional probability that individual n chooses route i is the integral of equation (2) with respect to all possible values of β_n as expressed in equation (4).

$$P_n(i) = \int P_n(i|\beta_n) f(\beta_n|\theta)\, d\beta_n \tag{4}$$

For repeated choice observations (from panel data), the logit probability portion in equation (4) is replaced by equation (3). Consequently, the unconditional probability that individual n chooses route i with respect to all possible values of β_n is formulated as equation (5).

$$P_{nT}(i) = \int P_{nT}(i|\beta_n) f(\beta_n|\theta)\, d\beta_n \tag{5}$$

The integral in equation (4) or (5) can be calculated approximately using simulation technique described in Train (2009). This simulation process draws $R = \{1, 2,\ldots, r\}$ values of β_n and puts each value into the logit formula (equation (2) or (3)). The average value of all (logit) probabilities results in a simulated probability $SP_n(i)$ in equation (6) which is an unbiased estimator of unconditional probability of equation (4) or (5).

$$SP_n(i) = \frac{1}{R}\sum_{r=1}^{R} P_{n(T)}(\beta_n^r) \tag{6}$$

After the simulated probabilities are obtained, the parameters θ can be estimated by maximizing the simulated log likelihood in equation (7).

$$SLL = \sum_{n=1}^{N}\sum_{j=1}^{J} \delta_{nj} \ln SP_{nj} \tag{7}$$

where $\delta_{nj} = 1$ if individual n choose route j and zero otherwise.

The integral portion might be multi-dimensional integral depending on the number of variables in β_n. Moreover, the simulation process is computationally burdensome. To assist the calculation process, this paper used the open source software package BIOGEME for model estimations (Bierlaire, 2003).

3. EMPIRICAL DATA AND INFORMATION EXTRACTION

This paper reports the route choice and route switching behavior analyses on expressway in Tokyo using the individual travel information from ETC data and traffic condition information from detector data. ETC data can provide the information of each vehicle at the toll gate location only. In general, it does not give individual route choice information directly. However, as presented in Tiratanapakhom *et al*. (2012a), it is possible to derive route choice information from ETC and other data such as detector data. Identifying route choice with this method might include some errors from fault identification. Since the present paper intends to show the use of ETC data for examining route choice behavior, the study area was selected in such a way that the actual route choice can be observed directly. For example, from a particular OD pair where there are two alternative routes and there is a pair of off-ramp locations (i.e. two off-ramps ETC toll gates are located on the opposite sides of the road).

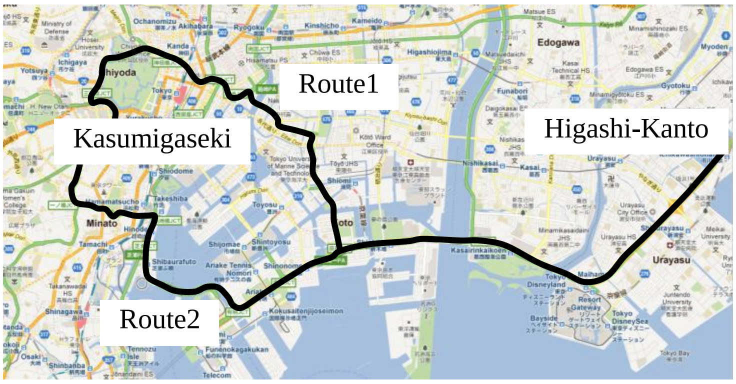

Figure 1. Selected study area

The present paper used the ETC data and the corresponding detector data between Higashi-Kanto expressway transition gate (i.e. toll gate connecting between MEX and Higashi-Kanto expressway) and Kasumigaseki toll gates inbound direction (Figure 1) where the condition of pair of off-ramp is satisfied. The multi-day ETC data which allows the repeated route choice observations was collected between August 2010 and February 2011 with the entry time during 06:00-24:00 hours. To focus more on working trips, the analyses mainly consider the travel information on common working days (141 days) of small (excluding motorcycle) and medium vehicles.

Route choice and route switching behavior analyses require a utility function which includes several attributes that influence drivers' route choice decisions. Without further field survey for RP data, the relevant attributes were derived from the available data sources (i.e. ETC and detector data). The attributes considered in the models can be categorized into 1) drivers' traveling characteristics, and 2) level-of-service characteristics.

3.1 Drivers' Traveling Characteristics

In general, the socioeconomic variables are used to reflect the differences in preferences of individual when choosing alternatives. Without such information, for expressway users, it is possible to assume the homogeneous drivers. In this paper, the information used in the model is derived from the ETC and detector data only. These data sources do not provide the socioeconomic information. However, it is possible to derive other information that can reflect the heterogeneity in drivers' preference from the available data sources. The following sub-sections present the definitions of derived information.

3.1.1 Frequency of travel

This variable is defined based on the number of trips made by each driver in the study area. Decision making by drivers with high frequency of travel might be different from those with low frequency. The frequency of travel represents the demand for travel; hence, might reflect through the purpose of travel, income, and so on. Using clustering method, the analyses classify drivers into two main groups: frequent and infrequent drivers. Along the study period (seven months), the total number of trips made by each vehicle in each month constitutes various patterns (Table 1) that can be recognized by the clustering method. This classification technique is similar to that in Tiratanapakhom *et al*. (2012b).

Table 1. Image of the patterns of monthly trip number made by each individual

ID	Aug	Sep	Oct	Nov	Dec	Jan	Feb
1	5	5	2	6	3	4	6
2	20	16	18	20	17	17	19
3	3	5	7	5	1	0	0
4	17	20	20	20	19	18	18
…	…	…	…	…	…	…	…

The dendrogram in Figure 2 provides the image of clustering results. To choose the number of clusters, using the simple approach, the dendrogram is cut at the large changes in distances as shown by the dashed line in Figure 2. As a result, the sample can be partitioned into two clusters; frequent and infrequent drivers, defined based on the average number of trips. The analyses adopted the k-mean clustering method with the pre-defined number of groups (i.e. two groups) for this classification.

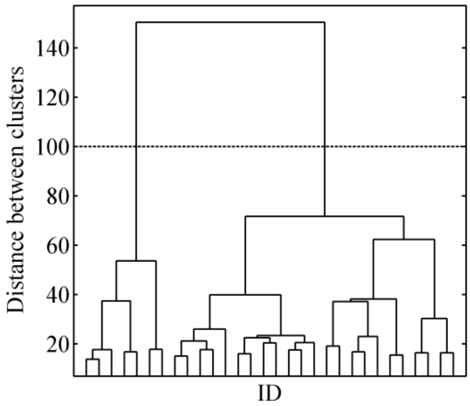

Figure 2. Image of clustering results

3.1.2 Time of use

In daily travel, drivers, especially commuters, are more likely to enter the expressway system at the similar period on the common working days. The time-of-use variable represents the point of time the vehicle enters the system. Since the study area observed the inbound traffic, this entry time is the time point when a vehicle passes the Higashi-Kanto toll gate. In general, it is expected that the demand of commuting trip is more likely to be observed during the morning peak rather than the evening peak. This variable is classified into three durations: morning peak, evening peak, and off-peak. The entry time 07:00-09:00 hours is defined for morning peak, and 17:00-19:00 hours is defined for evening peak. The trips enters the system other than these two periods are defined as off-peak trips. Using the effects coding system, the three qualitative levels are transformed into two effects coded variables (Bech and Gyrd-Hansen, 2005) as illustrated in Table 2.

Table 2. Effects coding variables for attribute time-of-use

Levels	Effects code1	Effects code2
Morning peak	+1	0
Evening peak	0	+1
Off-peak	-1	-1

3.1.3 Habitual route choice behavior

Drivers who are always observed traveling on the same route will be identified as habitual drivers. The habitual drivers are not aware of changing route decision, for example, even when it is faster to travel on the alternative route or there is an incident on their preferred route. Drivers other than habitual drivers will be considered as the non-habitual drivers. To classify the habitual drivers, it is necessary to observe repeated route choice decisions of each driver from multi-day and multi-period ETC data.

3.1.4 Driver's main route

This variable will be used in the route switching behavior analysis. A driver's main route may refer to the most frequently used route observed for that driver. The total number of trips made on one route for each driver is observed by multi-day and multi-period ETC data. Comparing the total number of trips that one particular driver makes on route 1 and route 2: for example, if the number of trips on route 1 is greater than on route 2, in this case, route 1 will be defined as the driver's main route.

3.1.5 Driver's preference for driving speed

The driver's preference for driving speed may be defined as the observed average speed that driver used for traveling on her/his chosen route under a particular traffic condition. We may assume that drivers try to achieve their preferred speed (as they recognize by experience) for traveling between this OD. They may evaluate the opportunity of using driving speed on each route according to the current traffic condition or other information. Such behavior might influence the route choice decision. The paper does not attempt to observe or to create a model to obtain the information of experience in driving speed for each individual. Instead, the paper assumes that drivers satisfy the driving speed that they actually used. This speed can be calculated based on the observed ETC travel time of each individual on each trip.

3.2 Level-of-Service (LOS) Characteristics

Drivers may evaluate the attractiveness of selecting a route for travel from the information about traffic condition or LOS on the available route choices. On expressway system, drivers can receive such information from several media such as message signboard, usually installed at the on-ramp locations, or graphic signboard, usually installed before junctions. Particularly on these signboards, travel time and the length of congestion on the routes are commonly displayed. The paper derived these two sources of traffic condition information from both ETC and detector data. The detector data used in this study provides traffic information aggregated every five-minute interval. In practice, usually the expressway operator evaluates the traffic condition by using the traffic information at the time before the current time point. To be more realistic, the traffic condition information is derived from traffic data at one time step (5 min.) before the time that vehicle enters the system.

3.2.1 Predicted travel time

Travel time prediction from speed data obtained by detector system is perhaps the most common way to obtain information of travel time based on the available traffic condition information. Li *et al*. (2006) evaluated the performance and provided some general formulations of the common speed-based travel time estimation models. Among the available travel time estimation models, the instantaneous model is generally adopted by expressway operators to provide travel time based on current traffic information measured by detector system since this model can perform on-line with real time traffic information. The instantaneous model calculates travel time on the route by using the speed information, $v(i,k)$, from each detector section along this route collected at the same time interval k. The travel time on each detector section is calculated using equation (8).

$$t(i,k) = \frac{l_i}{v(i,k)} \tag{8}$$

where $t(i,k)$ denotes the travel time on detector section i at time k. l_i is the length of detector section i. The route travel time under traffic condition at time k can be obtained by the summation of travel times from all detector sections along this route.

3.2.2 Length of congestion

On graphic road map signboards, the congestion information on routes may be displayed by illuminating the colors representing the levels of congestion based on the current traffic condition corresponding to the detector location. On MEX, at the locations under the free flow condition, the graphic map is shown as black color or not illuminating. The free flow condition is defined when the travel speed is over 40 km/h. The section that highlights with orange color represents slight congestion when travel speed is over 20 km/h and below 40 km/h. For the section of congested traffic condition, where the travel speed is less than or equal to 20 km/h, the red color will be displayed. Drivers can evaluate the severity of congestion from both the different types of colors and the length of these colors. Following the way to display this information, two variables are introduced: slight congestion length and heavy congestion length. The values of these variables on each route were calculated by the total detector section lengths in which the travel speeds meet the criteria of congestion level used for graphic signboards.

4. MODELS ESTIMATION RESULTS AND DISCUSSION

Three types of models were estimated for route choice and route switching behavior: 1) MNL, 2) MMNL with pooled data and 3) MMNL with repeated observations from panel data. The two MMNL models estimated the random parameters based on the assumed theoretical distribution. Since this paper aims mainly to show the use of repeated observations from ETC data to accommodate the effect of heterogeneity in drivers' behavior, it is adequate to assume the normal distribution for each random variable. In addition, the paper examines only generic variables: travel time and length of congestion to be randomly distributed across the population. In the simulation estimation process for MMNL, 1,000 draws were used. To avoid the issues of misinterpretation especially from the constant terms (Bech and Gyrd-Hansen, 2005), the categorical variables were coded using the effects coding system. The general information of empirical data used in the analyses is presented in Table 3.

Table 3. General information of empirical data

Category	Statistics: Pooled data	Statistics: Panel data
A. Individual characteristics		
Total observations	8363 veh. (100.0)	1718 veh.ID (100.0)
Frequent drivers	1705 veh. (20.4)	32 veh.ID (1.9)
Infrequent drivers	6658 veh. (79.6)	1686 veh.ID (98.1)
Habitual drivers	6213 veh. (74.3)	1391 veh.ID (81.0)
Non-habitual drivers	2150 veh. (25.7)	327 veh.ID (19.0)
Main route		
Route 1	3140 veh. (37.5)	664 veh.ID (38.6)
Route 2	5223 veh. (62.5)	1054 veh.ID (61.4)
Morning peak usage	2040 obs. (24.4)	
Evening peak usage	807 obs. (9.6)	
Off-peak usage	5516 obs. (66.0)	
Speed preference (km/h)		
Route 1: mean	54.33	
: SD	18.33	
Route 2: mean	56.26	
: SD	19.54	
B. Level of service characteristics		
Travel distance (m.)		
Route 1	24510	
Route 2	24790	
Predicted route travel time (min.)		
Route 1: mean	30.26	
: SD	12.39	
Route 2: mean	29.72	
: SD	12.22	
Heavy congestion length (m.)		
Route 1: mean	2051.2	
: SD	1764.0	
Route 2: mean	1356.6	
: SD	1121.8	
Slight congestion length (m.)		
Route 1: mean	2050.5	
: SD	1087.6	
Route 2: mean	2363.3	
: SD	1705.3	
C. Choice observations		
Route 1	3193 obs. (38.2)	
Route 2	5170 obs. (61.8)	
Switch route (using alternative route)	463 obs. (5.5)	
No switch route (using main route)	7900 obs. (94.5)	

Note: veh.ID: number of vehicle ID, where each ID is generated by ETC system.
obs.: number of observations.
SD : standard deviation.
Percentage unit is given in parentheses.

4.1 Route Choice Behavior

The general forms of utility functions of MNL, MMNL with pooled and panel data of route choice behavior analyses are demonstrated in equations (9), (10) and (11), respectively.

MNL route choice behavior:

$$\begin{aligned}U_{ni} = ASC_i &+\beta_{traveltime}TT_{ni} + \beta_{heavyCongestion}HVY_{ni} \\ &+\beta_{slightCongestion}SLT_{ni} \\ &+ \beta_{speedPreference}SPEED\ PREFERENCE + \beta_{freqUser}FREQ\ USER \\ &+ \beta_{morning}MORNING + \beta_{evening}EVENING \\ &+ \beta_{habitual}HABITUAL + \varepsilon_{ni}\end{aligned} \tag{9}$$

MMNL route choice behavior with pooled data:

$$\begin{aligned}U_{ni} = ASC_i &+N(\beta_{traveltime},\sigma_{traveltime})TT_{ni} \\ &+ N(\beta_{heavyCongestion},\sigma_{heavyCongestion})HVY_{ni} \\ &+N(\beta_{slightCongestion},\sigma_{slightCongestion})SLT_{ni} \\ &+ \beta_{speedPreference}SPEED\ PREFERENCE + \beta_{freqUser}FREQ\ USER \\ &+ \beta_{morning}MORNING + \beta_{evening}EVENING \\ &+ \beta_{habitual}HABITUAL + \varepsilon_{ni}\end{aligned} \tag{10}$$

MMNL route choice behavior with panel data:

$$\begin{aligned}U_{ni} = ASC_i &+N(\beta_{traveltime},\sigma_{traveltime})TT_{ni} \\ &+ N(\beta_{heavyCongestion},\sigma_{heavyCongestion})HVY_{ni} \\ &+N(\beta_{slightCongestion},\sigma_{slightCongestion})SLT_{ni} \\ &+ \beta_{speedPreference}SPEED\ PREFERENCE + \beta_{freqUser}FREQ\ USER \\ &+ \beta_{morning}MORNING + \beta_{evening}EVENING \\ &+ \beta_{habitual}HABITUAL + \varepsilon_{ni} +N(0, \xi_{ni})\end{aligned} \tag{11}$$

where,

U_{ni}: The utility of the individual *n* for alternative *i*.

ASC_i: Alternative specific constant.

$\beta_{traveltime}$: Coefficient for estimated travel time (TT_{ni}), where the unit of travel time is in minutes. In MMNL models, this coefficient is assumed to distribute normally, $N(\beta_{traveltime}, \sigma_{traveltime})$, with $\beta_{traveltime}$ mean and $\sigma_{traveltime}$ standard deviation.

$\beta_{heavyCongestion}$: Coefficient for length of heavy congestion (HVY_{ni}), where the length of congestion is divided by the distance between junction and destination. In MMNL models, this coefficient is assumed to distribute normally, $N(\beta_{heavyCongestion}, \sigma_{heavyCongestion})$, with $\beta_{heavyCongestion}$ mean and $\sigma_{heavyCongestion}$ standard deviation.

$\beta_{slightCongestion}$: Coefficient for length of slight congestion (SLT_{ni}), where the length of congestion is divided by the distance between junction and destination. In MMNL models, this coefficient is assumed to distribute normally, $N(\beta_{slightCongestion}, \sigma_{slightCongestion})$, with mean $\beta_{slightCongestion}$ and standard deviation $\sigma_{slightCongestion}$.

$\beta_{speedPreference}$: Coefficient for drivers' preference for driving speed (SPEED PREFERENCE), where the speed is divided by the free speed condition defined for message sign bard (i.e. 40 km/h).

$\beta_{freqUser}$: Coefficient for frequent/infrequent users (FREQ USER), where frequent/infrequent users are coded by effect coding system: frequent user (+1), infrequent user (-1).

$\beta_{morning}$: Coefficient for morning peak usage (MORNING) coded by effects coding system (Table 1).

$\beta_{evening}$: Coefficient for evening peak usage (EVENING) coded by effects coding system (Table 1).

$\beta_{habitual}$: Coefficient for habitual users (HABITUAL), where habitual users are coded by effects coding system: habitual user (+1), non-habitual user (-1).

ε_{ni}: The unobserved portion of utility that is distributed iid (independent and identically distributed) extreme value with zero mean.

ξ_{ni}: The unobserved portion of utility which is assumed to be normally distributed with zero mean and is correlated over choice situations for each individual.

The estimation results for route choice behavior are illustrated in Table 4. In general, the mean values of parameters estimated in MMNL with panel data are greater than those in MNL and MMNL with pooled data. To make the comparison between different models, the parameters estimated in both MMNL models have been scaled with respect to the estimated travel time parameter in MNL model. This scaling technique is referred to Frejinger and Bierlaire (2007).

Table 4. Route choice model estimation results

Independent Variables	MNL		MMNL (pooled data)		MMNL (panel data)	
	Parm.	t-stat	Parm.	t-stat	Parm.	t-stat
ASC_{R1}	0.338	3.54	0.270	3.12	0.974	7.44
Estimated travel time						
$\beta_{traveltime}$	-0.028	-4.22	-0.028	-3.22	-0.028	-4.42
$\sigma_{traveltime}$	-	-	0.017*	0.41	0.036	3.40
Heavy congestion length						
$\beta_{heavyCongestion}$	-4.780	-12.83	-4.935	-9.67	-3.210	-7.71
$\sigma_{heavyCongestion}$	-	-	3.375	3.38	1.341	2.00
Slight congestion length						
$\beta_{slightCongestion}$	-1.440	-6.02	-1.238	-5.42	-1.382	-5.85
$\sigma_{slightCongestion}$	-	-	2.790	2.12	1.058	3.95
$\beta_{speedPreference}$	-0.391	-6.47	-0.322	-5.96	-0.283	-4.04
$\beta_{freqUser}$	0.145	5.03	0.136	5.07	0.925	13.09
$\beta_{morning}$	0.047*	0.92	0.050*	1.06	0.064*	1.19
$\beta_{evening}$	-0.088*	-1.48	-0.088*	-1.58	-0.039*	-0.64
$\beta_{habitual}$	-0.158	-6.06	-0.138	-5.92	-0.039	-6.66
ξ-panel	-	-	-	-	2.111	14.11
Scale factor**	-	-	0.848	-	0.272	-
Parameters	9		12		13	
Final log-likelihood	-5272.505		-5267.955		-2506.228	
Adjusted $\bar{\rho}^2$	0.089		0.089		0.565	

- * rejected at 5% level, **with respect to the parameter "estimated travel time" in MNL.

The final log-likelihood for the MMNL with pooled data is slightly greater than that for MNL model. This indicates that there is an improvement of model estimation when accounting for the unobserved heterogeneity across population in drivers' behavior through the randomness of the selected variables. For MMNL with panel data, in addition to the standard deviations that capture the unobserved taste variation in the defined random coefficients, the error term (ξ-panel) is introduced to capture the remaining panel effects of other factors that are not included in the model. When the panel effect is taken into account, obviously, the significant improvement of log-likelihood value can be observed. The adjusted rho-squared ($\bar{\rho}^2$) values from each model show the similar indication with the final log-likelihood values.

For the estimated parameters, after scaling, the values of all parameters are consistent across the three models. The variables relating to estimated travel time and congestion length show the negative sign. This implies that drivers are less likely to travel on route 1 when they receive the increasing of congestion information through travel time or length of congestion. In addition, the heavy congestion length influences drivers' preference around more than twice as much as that under slight congestion. Moreover, it is found that drivers with higher average speed and more habitual in route choice are prone to travel on route 2 rather than route 1. Conversely, the results show that drivers who travel between this OD more often are more attracted to travel on route 1. Two variables relating to the time of use: morning peak and evening peak (with respect to off-peak hours) were found statistically insignificant. However, comparing the t-stat values across three models, it is observed that the absolute value of t-stat of morning-peak variable is greater than the value of evening-peak variable after accounting for the panel effect. Since the data is based on inbound traffic, the t-stat values of these variables in MMNL with panel data would be more realistic than other two models.

4.2 Route Switching Behavior

In this study, route switching or diversion occurs when drivers opt to divert from their defined main route to alternative route. The major attributes for investigating the impact to route switching behavior are mostly similar to the route choice behavior analysis. The variable related to habitual drivers is excluded in this analysis since all habitual drivers will not choose to use the alternative route. Instead, the variable main route is considered to investigate the behavior of drivers whose main route is route 1 with respect to route 2.

The general forms of utility functions of MNL, MMNL with pooled and panel data of route switching behavior analysis are demonstrated in equations (12), (13) and (14), respectively.

MNL route switching behavior:

$$\begin{aligned} U_{ni} = {} & ASC_i + \beta_{traveltime}TT_{ni} + \beta_{heavyCongestion}HVY_{ni} \\ & + \beta_{slightCongestion}SLT_{ni} \\ & + \beta_{speedPreference}\text{SPEED PREFERENCE} + \beta_{freqUser}\text{FREQ USER} \\ & + \beta_{morning}\text{MORNING} + \beta_{evening}\text{EVENING} \\ & + \beta_{mainRoute}\text{MAIN ROUTE} + \varepsilon_{ni} \end{aligned} \quad (12)$$

MMNL route switching behavior with pooled data:

$$\begin{aligned} U_{ni} = {} & ASC_i + N(\beta_{traveltime},\sigma_{traveltime})TT_{ni} \\ & + N(\beta_{heavyCongestion},\sigma_{heavyCongestion})HVY_{ni} \\ & + N(\beta_{slightCongestion},\sigma_{slightCongestion})SLT_{ni} \\ & + \beta_{speedPreference}\text{SPEED PREFERENCE} + \beta_{freqUser}\text{FREQ USER} \\ & + \beta_{morning}\text{MORNING} + \beta_{evening}\text{EVENING} \\ & + \beta_{mainRoute}\text{MAIN ROUTE} + \varepsilon_{ni} \end{aligned} \quad (13)$$

MMNL route switching behavior with panel data:

$$
\begin{aligned}
U_{ni} = {} & ASC_i + N(\beta_{traveltime}, \sigma_{traveltime})TT_{ni} \\
& + N(\beta_{heavyCongestion}, \sigma_{heavyCongestion})HVY_{ni} \\
& + N(\beta_{slightCongestion}, \sigma_{slightCongestion})SLT_{ni} \\
& + \beta_{speedPreference} \text{SPEED PREFERENCE} + \beta_{freqUser} \text{FREQ USER} \\
& + \beta_{morning} \text{MORNING} + \beta_{evening} \text{EVENING} \\
& + \beta_{mainRoute} \text{MAIN ROUTE} + \varepsilon_{ni} + N(0, \xi_{ni})
\end{aligned}
\tag{14}
$$

where,

$\beta_{mainRoute}$: Coefficient for driver's main route (MAIN ROUTE), where driver's main route are coded by effects coding system: main route is route1 (+1), main route is route2 (-1).

Table 5 displays the model estimation results for route switching behavior analysis. In general, the estimated mean values for each variable before scaling are quite close across three models. This might be due to the very low number of observations in the case of switching route (using alternative route) compared to the choice of non-switching (Table 3).

Table 5. Route switching model estimation results

Independent Variables	MNL		MMNL (pooled data)		MMNL (panel data)	
	Parm.	t-stat	Parm.	t-stat	Parm.	t-stat
ASC_{main}	3.890	18.38	4.067	16.25	4.338	14.29
Estimated travel time						
$\beta_{traveltime}$	-0.070	-4.86	-0.070	-5.00	-0.070	-4.88
$\sigma_{traveltime}$	-	-	0.070	2.64	0.089	4.76
Heavy congestion length						
$\beta_{heavyCongestion}$	-6.610	-8.59	-6.493	-8.40	-6.528	-7.94
$\sigma_{heavyCongestion}$	-	-	3.545*	1.68	3.809	2.76
Slight congestion length						
$\beta_{slightCongestion}$	-2.330	-4.68	-2.127	-4.54	-2.507	-4.71
$\sigma_{slightCongestion}$	-	-	3.368	3.00	3.117	4.86
$\beta_{speedPreference}$	-0.445	-3.28	-0.600	-4.18	-0.419	-2.34
$\beta_{freqUser}$	0.454	6.04	0.451	6.16	0.615	4.26
$\beta_{morning}$	-0.262	-2.62	-0.341	-3.34	-0.175*	-1.45
$\beta_{evening}$	0.177*	1.42	0.242*	1.86	0.167*	1.17
$\beta_{mainRoute}$	0.021*	0.39	0.017*	0.34	0.030*	0.45
ξ-panel	-	-	-	-	1.155	13.74
Scale factor**	-	-	0.933	-	0.814	-
Parameters	9		12		13	
Final log-likelihood	-1626.08		-1619.62		-1506.49	
Adjusted $\bar{\rho}^2$	0.718		0.719		0.738	

- * rejected at 5% level, **with respect to the parameter "estimated travel time" in MNL.

The final log-likelihood values improve slightly when using the MMNL with pooled data instead of MNL. As expected, more improvement in term of final log-likelihood value can be obtained when applying the MMNL with panel. The similar conclusions can also be drawn from the adjusted rho-squared ($\bar{\rho}^2$) values.

The negative sign of the values of estimated parameters for estimated travel time, length of congestion and speed recognition indicates that drivers are more likely to switch route when the congestion condition on their main route increase. The effect of congestion under heavy situation is found to be as much as around three times greater than the slight congestion. In addition, those who prefer to drive at higher speeds may have higher preference to switch to the alternative route. Drivers who travel to the downtown area during the morning peak are more likely to switch route than those who travel during the evening peak. It is also found that drivers who travel between this OD more often and those who usually travel on route 1 are more reluctant to divert their route.

For the estimated parameter values, the independent variables of evening peak and main route are not statistically significant across the three models. When accounting for panel data, in addition to those variables, the variable morning peak in MMNL with panel data also exhibits to be statistically insignificant. In MMNL with pooled data, the standard deviation of variable heavy congestion length is not significant from zero. This implies a more homogeneity in drivers’ behavior under this variable. However, the standard deviation becomes statistically significant when using panel data. This indicates that accounting for panel data, not only can improve the goodness-of-fit of the model estimation, but also might provide different information from the basic models.

5. CONCLUSION

In this paper, RP route choice data derived from multi-day ETC data between August 2010 and February 2011 on the selected study area of MEX have been used. From this data, the repeated observations of route choice decision of each individual can be obtained. In addition to the route choice data, the information of several attributes that influence route choice behavior were derived from the corresponding detector data. These long term RP data sources allow the models to capture more actual heterogeneity in route choice behavior. The paper analyzes route choice and route switching behavior from those data sources. Three types of model have been estimated and compared in each analysis: 1) MNL, 2) MMNL with pooled data, and 3) MMNL with panel data or repeated observations. In general, the results show that MMNL models provide the improvement in estimation results in terms of final log-likelihood and goodness-of-fit statistics.

In route choice and route switching behavior analysis, drivers select their route in response to the informed congested situations in terms of travel time and congestion length. Moreover, it is found that the heavy congestion length influences drivers’ decision more than the slight congestion length. In route switching analysis, particularly, the estimated results of three models show the different information of statistical significance in parameter estimations. Moreover, as can be observed from both analyses, the panel effect (ξ-panel) is statistically significant. When using repeated observations to accommodate panel effect, the model estimation is superior to the MMNL with pooled data. This confirms that the panel data should be taken into account in modeling route choice.

In summary, collecting panel data is generally expensive and difficult in practice. This study shows the possibility to use ETC data and available data sources to analyze and model route choice behavior especially with repeated observations on urban expressway.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Metropolitan Expressway Company Limited in Tokyo for providing the data used in this research.

REFERENCES

Bech, M., Gyrd-Hansen, D. (2005) Effects coding in discrete choice experiments. *Health Economics*, 14, 1079-1083.

Bhat, C.P. (2000) Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science,* 34(2), 228-238.

Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bonsall, P. (1992) The influence of route guidance advice on route choice in urban networks. *Transportation,* 19(1), 1-23.

Frejinger, E., Bierlaire, M. (2007) Capturing correlation with subnetworks in route choice models. *Transportation Research Part B,* 41, 363-378.

Hensher, D.A.,Greene, W.H. (2003) The mixed logit model: the state of practice. *Transportation*, 30(2), 133-176.

Li, H., Guensler, R., Ogle, J., Wang, J. (2004) Using GPS data to understand day-to-day dynamics of morning commute behavior. *Transportation Research Record*, 1985, 78-84.

Li, R., Rose, G., Sarvi, M. (2006) Evaluation of speed-based travel time estimation models. *Journal of Transportation Engineering*, 132(7), 540-547.

Organization for Road System Enhancement (ORSE) (2011) Electronic toll collection system. (in Japanese)

Ortúzar, J.D, Willumsen, L.G. (2001) *Modelling Transport*. John Wiley & Sons, Chichester.

Revelt, D., Train, K. (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level. *The review of economics and statistics*, 80(4), 647-657.

Tiratanapakhom, T., Oguchi, T., Tanaka, S., Hong, S., Warita, H. (2012a) Using ETC data and detector data to derive information of route choice on Tokyo Metropolitan Expressway (MEX). Paper presented at the 19th ITS World Congress, Vienna, Austria, October 22-26.

Tiratanapakhom, T., Oguchi, T., Tanaka, S., Hong, S., Warita, H. (2012b) Analysis of daily variability characteristics of travel behavior on Tokyo Metropolitan Expressway using ETC data. Paper presented at the 11th ITS Symposium, Aichi, Japan, December 13-14.

Train, K. (2009) *Discrete choice methods with simulation*. Cambridge, New York.