Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-based Recursive Logit Model Featuring Link Awareness

Noriko KANEKO[*,a], Hideki OKA[a,b], Makoto CHIKARAISHI[c], Henrik BECKER[d], Daisuke FUKUDA[a]

[a]Tokyo Institute of Technology, Ookayama 2-12-1-M1-11, Meguro-ku, Tokyo 152-8552, Japan
[b]The Institute of Behavioral Sciences, Ichigayahonmuracho 2-9, Shinjyuku-ku, Tokyo 162-0845, Japan
[c]Hiroshima University, Kagamiyama 1-5-1, Higashi-Hiroshima 739-8529, Japan
[d]Swiss Federal Institute of Technology, Zürich CH-8093, Switzerland

Abstract

Identification of an appropriate route choice model to understand travel behavior remains challenging. To this end, Fosgerau et al. (2013) have recently developed a link-based route choice model termed the "recursive logit" (RL) model. A decision-maker is assumed to choose the next link recursively that maximizes the sum of instantaneous utility and expected downstream utility at each node. However, in practical application, some computational issues remain, including large (and often ill-defined) matrix inversions. Here, we develop an alternative RL model that considers the probability of awareness of the next link that improves the stability of model estimations. The model was estimated using vehicle trajectory data from the ETC (Electronic Toll Collection) 2.0 dataset of the Tokyo Metropolitan area, and the results were compared to those of a conventional RL model in terms of predictive accuracy and computational efficiency.

Keywords: Recursive logit model, route choice model, choice set, link awareness

1. Introduction

A metropolitan inter-city expressway and an outer ring road are currently under construction in the Tokyo Metropolitan Area (TMA); these are expected to drastically reduce downtown through-traffic. Also, the highway tolling systems have changed dramatically. For example, a seamless highway toll system was introduced in 2016 and

* Corresponding author. Tel.: +81-3-5734-2577; Fax: +81-3-5734-3578.
E-mail address: n.kaneko@plan.cv.titech.ac.jp

a congestion tax on arterial roads is under consideration. Thus, an accurate and detailed route choice model is essential for quantitative evaluation of travel plans and policies. Path-based models have been used to construct route choice models. However, given very large traffic networks such as the TMA, the number of route alternatives is enormous and the construction of feasible sets of choices is computationally demanding. To cope with this issue, Fosgerau et al. (2013) recently proposed a recursive logit (RL) model in which a traveler chooses the next link to maximize the utility at each node, rather than choosing a single route ab initio. Several applications of the RL model later appeared in the literature. Mai et al. (2015) developed a nested RL model that considered link correlations. Mai et al. (2016) used a decomposition method to evaluate multiple destinations efficiently. Recently, Chikaraishi et al. (2016) and Oka et al. (2017) applied an RL model to a large proportion of the TMA using trajectory data from large trucks.

The RL model has a considerable potential for modeling large-scale route choices behavior. However, the model is inherently unstable and computationally burdensome when applied to large-scale networks. To solve these issues, we here integrate the concept of awareness probability into an RL model. Specifically, we introduce "IAP" (the perception/availability of an alternative implicit in the choice model, as proposed by Cascetta et al. (2001)) into the RL model. It then becomes possible to restrict many alternatives to a more feasible set of choices in a stochastic manner, contributing to the avoidance of cyclic routes. This is the first study to use the detailed vehicle trajectory data of personal cars traveling in the TMA (the "ETC 2.0" dataset), combined with detailed road network attributes such as tolls, road widths, and numbers of lanes.

2. Data

We used a digital road map (DRM) [1] as a basic network data, and vehicle trajectories from the ETC 2.0 [2] database as observational data. The DRM was prepared by the Japan Digital Road Map Association and contains the original link IDs, node IDs, link lengths and widths, longitudes, latitudes, and lane numbers. The ETC 2.0 is collected by the Ministry of Land, Infrastructure, Transport, and Tourism of Japan, and contains vehicle trajectory data for private cars fitted with on-board devices of generation 2.0. The market share of ETC 2.0-equipped vehicles has gradually increased. Our target was an 80km×80km area including the TMA. We chose 70 downtown destination points associated with heavy traffic flow, as shown in Fig.1. The points were selected as follows: (1) five points from each red

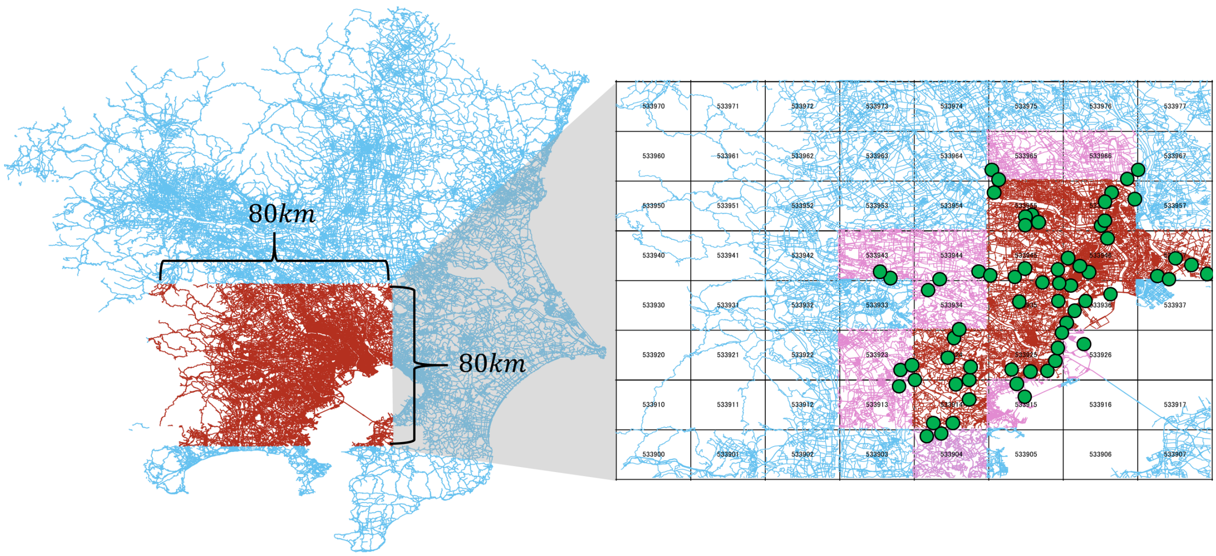

Fig. 1. 80km×80km target area and the destinations

[1] http://www.drm.jp
[2] https://www.go-etc.jp/english

mesh (the 10 most-congested meshes); and (2) two points from each pink mesh (the 20 most-congested meshes). The details are shown in Table.1.

Here, we analyze data from Oct. 16, 2016 (a Sunday); most trips were thus leisure trips. Also, departure points were taken to be at the edges of the target area if the trips commenced outside the target area. Thus, the data will include both long- and short-distance trips. We divided all trips into long, medium, and short trips (> 30, ≥ 10 but < 30, and < 10km, respectively) Fig.3 shows trip length distributions and Table.2 shows trips by length. The area features various links (Fig.2). Almost all roads were of Type 9 (minor roads). Also, Fig.4 and Fig.5 show that link length distributions depended on road type. The higher the rank of road type, its link length tends to be longer. For example, the average link length for Road Type 2 was 403 m, whereas that for Road Type 9 was 172 m. The standard deviation of road length for Road Type 2 was 488 m, and that for Road Type 9 was 176 m.

Table 1. Trip data specifications

Date for analysis	Oct. 16, 2016 (Sun.)
Target area	The right chart (80km×80km) in Fig.1
Number of trips	13,622
Number of links	228,387
Number of destinations	70
Average number of link selections	75.2

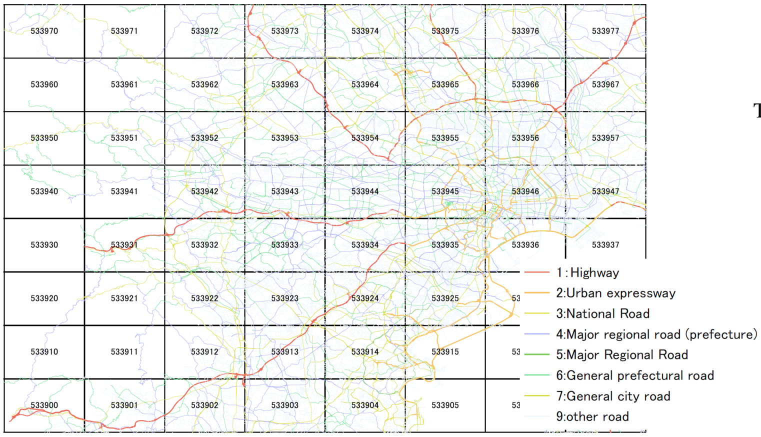

Fig. 2. Road type distribution

Table 2. Four types of trips by trip length

Group name	Target	Number of trips
Group 1	All trips	13,622
Group 2	Long trips (30km∼)	4,128
Group 3	Medium trips (10km∼30km)	5,999
Group 4	Short trips (∼10km)	3,495

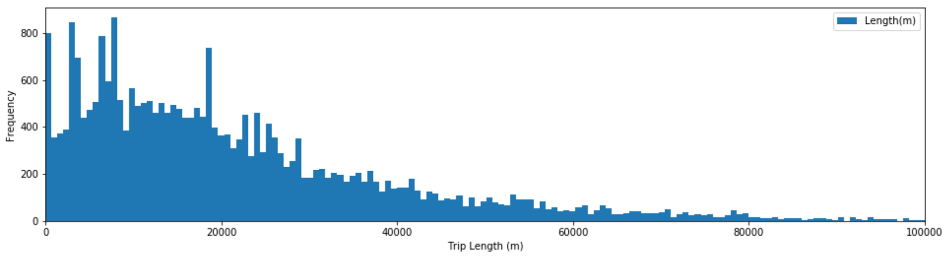

Fig. 3. Trip length distributions

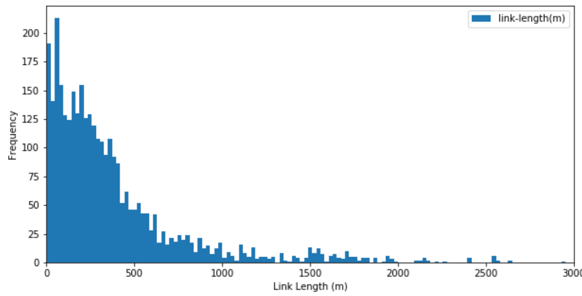

Fig. 4. Link length distributions for Road Type 2 (urban expressways)

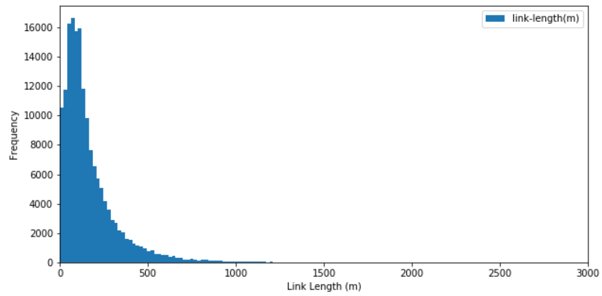

Fig. 5. Link length distributions for Road Type 9 (minor roads)

3. Modeling Methodology

3.1. A recursive logit model

In this section, following Fosgerau et al. (2013), we briefly introduce the RL model and its use for estimations. In network $G = (A, v)$, we define v as a set of nodes, A as a set of links $(k, a \in A)$, $A(k)$ as a set of outgoing links from k, and n as a traveler. In each state k, a traveler chooses link a (from the set of outgoing links $A(k)$) that maximizes the sum of instantaneous utility and expected downstream utility. Instantaneous utility is defined as the sum of deterministic and stochastic utility, as shown in Eq.(1). The random terms $\varepsilon_n(a)$ are assumed to be of i.i.d. extreme-value type 1, thus with zero means, and are independent of all other model components; μ is a scaling parameter.

$$u_n(a|k) = v_n(a|k) + \mu\varepsilon_n(a) \tag{1}$$

The downstream utility is the expected utility at the destination, as given by Eq.(2). This is obtained from the Bellman equations (Rust (1987)) as Eq.(3).

$$V_n^d(k) = \mathbb{E}[\max_{a \in A(k)} (v_n(a|k) + V_n^d(a) + \mu\varepsilon_n(a))] \tag{2}$$

$$\boldsymbol{z} = \boldsymbol{M}\boldsymbol{z} + \boldsymbol{b} \Leftrightarrow (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{z} = \boldsymbol{b} \tag{3}$$

Here, we define $\boldsymbol{I}(|\tilde{A}|\times|\tilde{A}|)$ as the identity matrix, $\boldsymbol{M}(|\tilde{A}|\times|\tilde{A}|)$ as the incidence matrix defining instantaneous utilities, $\boldsymbol{z}(|\tilde{A}| \times 1)$ as a vector with elements $z_k = \exp\left(\frac{1}{\mu}V(k)\right)$, $\boldsymbol{b}$ as $(|\tilde{A}| \times 1)$ a vector with elements $b_k = 1$ (if $k \neq d$), $b_d = 1$. When the matrix $\boldsymbol{I} - \boldsymbol{M}$ has an inverse matrix, the Hawkins-Simon condition (Hawkins and Simon (1949)) should be satisfied, and $V(k)$ can then be described quantitatively. When a traveler $n = 1, ..., N$ is characterized by his/her link choices $\sigma_n(n = 1, ..., N)$, where σ_n consists of a set of links from the origin to the destination represented by $k_0, k_1, k_2, ..., k_{I^n-1}, k_{I^n}$,the log-likelihood function used to estimate parameters of interest is given as Eq.(4)

$$\begin{aligned}\max_{\boldsymbol{\beta}} LL_n(\boldsymbol{\beta}) &= \frac{1}{N}\sum_{n=1}^{N} \ln P(\sigma_n; \boldsymbol{\beta}) \\ &= \frac{1}{\mu}\sum_{n=1}^{N}\left\{\left[\sum_{i=0}^{I^n-1} v(k_{i+1}^n|k_i^n)\right] - V(k_0^n)\right\}\end{aligned} \tag{4}$$

3.2. Introducing link awareness

We restrict the link sets available for choice in a stochastic manner by introducing the IAP concept. The likelihood that link a is included in the choice set $A(k)$ of outgoing links from k is given by $\lambda_{A(k)}(a)$ $(0 \leq \lambda_{A(k)}(a) \leq 1)$. Then, the instantaneous utility can be reformulated as Eq.(5); thus, the sum of deterministic and stochastic utility, and link awareness:

$$u_n(a|k) = v_n(a|k) + \ln \lambda_{A(k)}(a) + \mu\varepsilon_n(a) \tag{5}$$

When $\lambda_{A(k)}(a) = 0$, the probability of awareness is 0%. Likewise, when $\lambda_{A(k)}(a) = 1$, the probability of awareness is 100%. The probability that a given alternative j will be chosen may be expressed as Cascetta et al. (2001). Here, $\overline{\lambda}$ is the average value of λ.

$$p^i(j) = Pr(U^i_j \geq U^i_h) \tag{6}$$

$$= Pr\left(\sigma^i_h - \sigma^i_j \leq V^i_j + \ln \overline{\lambda}^i_c(j)) - \frac{1 - \overline{\lambda}^i_c(j)}{2\overline{\lambda}^i_c(j)} - V^i_h - \ln \overline{\lambda}^i_c(h) + \frac{\overline{\lambda}^i_c(h)}{2\overline{\lambda}^i_c(h)}\right) \tag{7}$$

Under the simple assumption that the random residuals σ^i_j are independent and identically Gumbel$(0, \alpha)$ distributed, the choice probability can be expressed using a simple, multinomial logit model.

$$p^i(j) = \frac{\exp\left[\alpha\left(V^i_j + \ln \overline{\lambda}^i_c(j) - \frac{1 - \overline{\lambda}^i_c(j)}{2\overline{\lambda}^i_c(j)}\right)\right]}{\sum_n \exp\left[\alpha\left(V^i_j + \ln \overline{\lambda}^i_c(j) - \frac{1 - \overline{\lambda}^i_c(j)}{2\overline{\lambda}^i_c(j)}\right)\right]} \tag{8}$$

In Cascetta et al. (2001), the scale parameter α was set to 1. However, here, this was not the case; the parameter is estimated, as are all other parameters.

However, it is difficult to measure the extent of awareness $\lambda_{A(k)}(a)$. Thus, here, we sequentially estimate awareness using a binary logit (BL) model. Awareness can be expressed as Eq.(9). The BL model is applied whether each link was actually used by drivers. The BL model of link awareness is then formulated as:

$$\overline{\lambda}^i_c(j) = \frac{1}{1 + \exp\left(-\sum_k \gamma_k Y^i_{kj}\right)} \tag{9}$$

where, Y^i_{kj} : variables correlated with awareness/availability of each alternative, γ_k : coefficient relative to Y^i_{kj} attributes. In the BL model, Road Types are used as variables explaining the extent of awareness. The DRM road Types are 1: Highway, 2: Urban expressway, 3: National arterial road, 4: Major regional road (prefecture), 5: Major regional road (city), 6: General prefectural road, 7: General city road, 9: Other road (i.e., small streets), and 0: not surveyed.

4. Estimation Results

4.1. Computing link awareness

4.1.1. Model Specification

In the BL model of link awareness, Road Type 4 [Major regional road (prefecture)] served as the baseline. As the number of roads of Type 5 [major regional road (city)] was small in our area,(< 100 of 1,118,027 links), this Type was excluded from the estimation. Also, we further considered that urban and suburban areas differ, especially in terms of Road Types 2 and 9. Thus, we introduce interaction terms for these road types and urban areas; the latter included only four meshes within the circular road of the metropolitan expressway.

Consequently, the IAP specified in Eq.(10) include dummy variables for road Types 1, 2, 3, 6, 7, and 9; interaction terms for urban areas and road Types 2 and 9; and an alternative-specific constant (ASC) for the explanatory variables. In this estimation, weights (the observed usage numbers of each link) were considered because the number of times that links are used by travelers may explain the extent of awareness.

$$\begin{aligned} -\sum_k \gamma_k Y^i_{kj} = & -\gamma_{type1} x_{type1Dummy} - \gamma_{type2} x_{type2Dummy} - \gamma_{type3} x_{type3Dummy} - \gamma_{type4} x_{type4Dummy} - \gamma_{type6} x_{type6Dummy} - \gamma_{type7} x_{type7Dummy} \\ & -\gamma_{type9} x_{type9Dummy} - \gamma_{type2_{urban}} x_{type2Dummy_{urban}} - \gamma_{type9_{urban}} x_{type9Dummy_{urban}} + ASC \end{aligned} \tag{10}$$

4.1.2. Estimation results of link awareness term (BL model)

Table.3 shows the outcomes of the BL model of Eq.(3). The signs of all parameters are intuitively consistent, and were statistically significant. Also, the relative magnitudes of the parameters are plausible: the higher the road grade, the greater the awareness. Particularly, both highways (Types 1 and 2) were more likely to be recognized by drivers. The interaction terms among road Types 2, 9, and urban areas were significantly positive. Fig.7 lists the IAP-computed probabilities for each link.

Table 3. Estimations of link awareness terms

Parameter	Estimates	t-value
Road Type 1 (Highway)	4.26	38.25
Road Type 2 (Urban expressway)	6.38	37.71
Road Type 3 (National arterial road)	1.08	68.33
Road Type 4 (Major prefectural road)	-	-
Road Type 6 (General prefectural road)	-0.87	-60.94
Road Type 7 (General city road)	2.76	25.5
Road Type 9 (Other road)	-3.03	-261
Road Type 2 × Urban dummy	0.80	2.06
Road Type 9 × Urban dummy	0.68	62.1
ASC (Utilized)	2.54	268
Sample Size		1,118,027
LL(0)		-774,957
LL(f)		-201,337
ρ^2		0.740

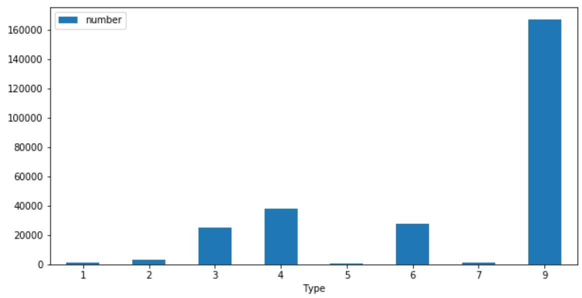

Fig. 6. Road type distribution

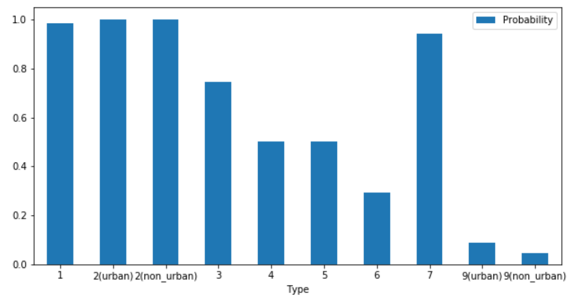

Fig. 7. Computed probability of IAP-mediated recognition of each link

4.2. *Estimating RL route choice model*

4.2.1. *Specification of the utility function*

The basic instantaneous utility from k to a $v(a|k)$ is defined as Eq.(11). We compared two models: model 1 considered link awareness whereas model 2 did not.

$$v_n(a|k) = \beta_{time}x_{time} + \beta_{length}x_{length} + \beta_{cost}x_{cost} + \beta_{width}x_{width} + \beta_{Uturn}x_{Uturn} + \beta_{Rturn}x_{Rturn} \tag{11}$$

where x_{time}: is the travel time of link a (sec), x_{length}: the length of link a (m), x_{cost}:the travel cost of link a (JPY), x_{width}:the width of link a (m), x_{Rturn}:a dummy variable when the angle between link pair $k-a$ is 70°∼175°, and x_{Uturn}:a dummy variable when the angle between link pair $k-a$ is 175°∼185°. As cars drive on the left in Japan, x_{Rturn} also served as an explanatory variable.

4.2.2. *Estimations of the route choice model*

The estimation results of the two models (model 1 with awareness and model 2 without) for each type of trip are shown in Table.4. In both models, almost all parameter signs were intuitively consistent, and all parameters are statistically significant. In all cases, the travel time (i.e., the ratio of the time to the cost parameter) are reasonable. Thus, the model yielded significant information.

However, the time parameter for Group 2 was small; the t-value was not significant. Thus, drivers who travel long distances tend to be strongly influenced by the awareness term. Some are not familiar with the area, and thus tend to use major roads including highways, not minor roads. Also, when traveling long distances, highways (road Types 1 and 2) are used to shorten travel time. On the contrary, for Group 4, no significant difference between models 1 and 2 was apparent; those who drive only within small areas know the roads well; the awareness term is not significant. This makes sense; such drivers do not use highways. The characteristics of Group 3 were intermediate between those of Groups 2 and 4; this is reasonable because Group 3 includes equal numbers of those who do and do not know the area, and those who want to travel faster or slower. However, model 1 was slightly better in terms of the Max LL data for Groups 1 and 3; awareness explained driver diversity.

In terms of computational time[3], that for the model including awareness was longer because more parameters were included. Thus, when drivers cover considerable distances, the awareness term is useful.

5. Conclusions and Future Works

We developed an RL model featuring link awareness for large-scale analysis of route choices, and evaluated its accuracy and efficiency using vehicle-trajectory data from the TMA and detailed road attributes. The method works well when evaluating long-distance trips, despite the large computational load. Sometimes, the conventional method (i.e., a model without link awareness) yields unstable estimations if the network is large. Under such circumstances, the awareness term is useful. However, the awareness is not important when drivers are familiar with the roads, and travel only short distances; the computational time is then excessive. Thus, awareness should be included only when appropriate. This study is the first to include a mental influence using the IAP method; this is a potentially important factor. Only the road type and interaction terms were used to define awareness. However, other factors (such as the number of lanes) may affect the term; further research is necessary.

Acknowledgements

We thank Kay W. Axhausen for suggesting how we might derive model estimations. The study was supported by the Committee on Advanced Road Technology (CART), Ministry of Land, Infrastructure, Transport, and Tourism, Japan.

References

Cascetta, E. and Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transportation Research Part C: Emerging Technologies, 9(4), pp.249-263.

Makoto Chikaraishi, Jun Tanabe, and Tanashi Oguchi.,2016. Analysis of freight traffic flow in the Tokyo metropolitan urban expressway networks using probe data. The 55th Conference on Infrastructure Planning and Management for JSCE,

Fosgerau, M., Frejinger, E. and Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological, 56, pp.70-80.

Hawkins, D. and Simon, H.A., 1949. Note: some conditions of macroeconomic stability. Econometrica, Journal of the Econometric Society, pp.245-248.

Mai, T., Fosgerau, M. and Frejinger, E., 2015. A nested recursive logit model for route choice analysis. Transportation Research Part B: Methodological, 75, pp.100-112.

Mai, T., Bastin, F. and Frejinger, E., 2016. A decomposition method for estimating recursive logit based route choice models. EURO Journal on Transportation and Logistics, pp.1-23.

Hideki Oka, Makoto Chikaraishi, Jun Tanabe, and Takashi Oguchi., 2017. An analysis of truck route choice behavior considering vehicle types. The 57th Conference on Infrastructure Planning and Management for JSCE

Rust, J., 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica: Journal of the Econometric Society, pp.999-1033.

[3] All calculations were performed using a Mac Pro fitted with a quad-core Intel Xeon (3.5 GHz) and 64 GB of DDR3 memory clocked at 1,866 MHz

Table 4. Model estimations for all trips

Results of Group 1				
	Results with awareness		Results without awareness	
Parameter	Estimates	t-value	Estimates	t-value
β_{time} (sec)	-8.73	-14.94	-8.06	-10.35
β_{length} (m)	-10.37	-14.94	-10.44	-4.40
β_{cost} (JPY)	-10.68	-12.54	-10.33	-11.65
β_{width} (m)	9.62	3.02	10.59	7.08
β_{Uturn}	-97.87	-24.04	-106.8	-112.7
β_{Rturn}	-109.8	-35.66	-104.6	-213.5
$\beta_{Awareness}$	10.08	7.88	–	-
Sample size	13,622		13,622	
Max LL	−572,150		−732,832	
Value of time (JPY/hour)	58.25		60.63	
Computational time (hour)	57.3		24.3	
Results of Group 2				
	Results with awareness		Results without awareness	
Parameter	Estimates	t-value	Estimates	t-value
β_{time} (sec)	-10.06	-3.20	-6.15×10^{-2}	-0.32
β_{length} (m)	-9.52	-2.65	-8.62	-10.55
β_{cost} (JPY)	-10.13	-4.49	-13.17	-17.37
β_{width} (m)	10.18	34.09	11.31	12.11
β_{Uturn}	-101.8	-225.6	-107.2	-2.01
β_{Rturn}	-101.2	-113.2	-161.9	-8.53
$\beta_{Awareness}$	10.14	5.67	–	-
Sample size	4,128		4,128	
Max LL	−679,804		−17,853	
Value of time (JPY/hour)	56.38		58.37	
Computational time (hour)	53.56		24.5	
Results of Group 3				
	Results with awareness		Results without awareness	
Parameter	Estimates	t-value	Estimates	t-value
β_{time} (sec)	-9.66	-9.97	-10.09	-6.83
β_{length} (m)	-10.33	-5.28	-9.97	-49.6
β_{cost} (JPY)	-10.44	-17.62	-10.24	-22.5
β_{width} (m)	9.83	4.08	9.93	150.6
β_{Uturn}	-104.3	-50.65	-102.1	-112.5
β_{Rturn}	-98.26	-102.6	-100.3	-29.5
$\beta_{Awareness}$	10.09	10.45	–	-
Sample size	5,999		5,999	
Max LL	−1,219,180		−1,503,239	
Value of time (JPY/hour)	59.36		60.89	
Computational time (hour)	42.29		31.03	
Results of Group 4				
	Results with awareness		Results without awareness	
Parameter	Estimates	t-value	Estimates	t-value
β_{time} (sec)	-10.68	-63.84	-9.81	-23.6
β_{length} (m)	-10.10	-7.75	-10.14	-7.86
β_{cost} (JPY)	-9.99	-9.03	-10.01	-10.23
β_{width} (m)	10.42	15.73	10.25	43.5
β_{Uturn}	-101.6	-35.75	-100.2	-10.25
β_{Rturn}	-100.0	-28.79	-102.0	-98.79
$\beta_{Awareness}$	9.57	5.60	–	-
Sample size	3,495		3,495	
Max LL	−1,062,143		−1,058,921	
Value of time (JPY/hour)	60.66		61.22	
Computational time (hour)	56.16		23.6	