A Study of Mode Choice Model Improvement Considering Age Grouping

Young-Hyun Seo, Hyunwoo Park, Dong-Kyu Kim, Seung-Young Kho

Abstract—The purpose of this study is providing an improved mode choice model considering parameters including age grouping of prime-aged and old age. In this study, 2010 Household Travel Survey data were used and improper samples were removed through the analysis. Chosen alternative, date of birth, mode, origin code, destination code, departure time, and arrival time are considered from Household Travel Survey. By preprocessing data, travel time, travel cost, mode, and ratio of people aged 45 to 55 years, 55 to 65 years and over 65 years were calculated. After the manipulation, the mode choice model was constructed using LIMDEP by maximum likelihood estimation. A significance test was conducted for nine parameters, three age groups for three modes. Then the test was conducted again for the mode choice model with significant parameters, travel cost variable and travel time variable. As a result of the model estimation, as the age increases, the preference for the car decreases and the preference for the bus increases. This study is meaningful in that the individual and households characteristics are applied to the aggregate model.

Keywords—Age grouping, aging, mode choice model, multinomial logit model.

I. INTRODUCTION

BECAUSE of increasing traffic congestion and air pollution in Seoul, South Korea, the city considers various ways to solve these problems. As a result, there is growing interest in converting demand from private cars to public transportation demand.

In order to shift the demand for public transportation, it is essential to predict and identify the share of each origin and destination for each means. The modal split model can be applied to various models, but the most widely used model is the logit model. The travel cost and travel time by modes are used as the main variables of the model. It is necessary to estimate the model parameters based on the data of the passenger's mode choice behavior to calculate the more accurate share of the mode choice. In this respect, disaggregate data, rather than zone-based aggregate data, can be used to express relatively better trip characteristics [1].

The most common reason for errors in forecasting the mode choice is that the actual traffic behavior is dominated by other factors besides traffic costs and time. Therefore, in order to establish appropriate transportation policy, it is necessary to consider various variables such as local characteristics

Young-Hyun Seo, Hyunwoo Park, and Seung-Young Kho are with Seoul National University, Seoul, 08826 Republic of Korea (e-mail: yhseo@snu.ac.kr, woo6260@snu.ac.kr, sykho@snu.ac.kr).
Dong-Kyu Kim is with Seoul National University, Seoul, 08826 Republic of Korea (corresponding author to provide phone: +82-2-880-7348, fax: +82-2-880-2684, e-mail: dongkyukim@snu.ac.kr).

including the characteristics of individuals and households. Among the personal and household characteristics that have a significant effect on the mode choice, there are age, sex, spouse, possession of driving license, presence of preschool child, and possession of car [2]. However, considering all the variables has limitations in terms of time and cost. There is also a practical problem in applicability when using individual trip data because the mode share is presented based on zone-based data.

Korea is entering the aging society very quickly. Fig. 1 represents that the proportion of people over 65 years old or more compared to the total population was 5.9% in 1995, up from 12.8% in 2015. The National Statistical Office predicts that the pace of aging will accelerate to 28.7% in 2035 [3]. Therefore, the purpose of this study is to construct a mode choice model considering the age as the regional characteristics, as well as the variables of travel time and travel cost. The 2010 Household Travel Survey data were used to estimate the parameters by modes. Using the estimated results, it is judged whether the sign of parameter and the value of time is proper and statistically significant. It is also an important goal of the study to understand how the age affects the mode choice model.
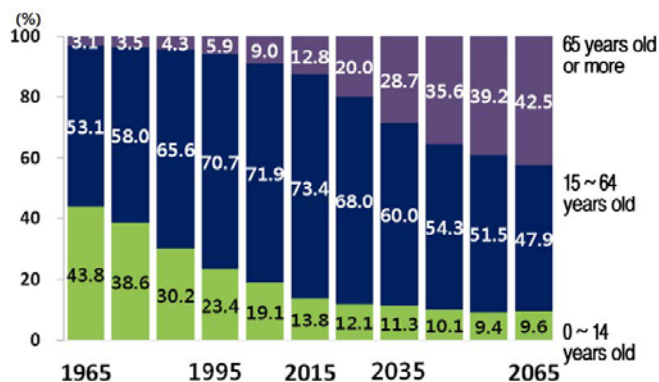

Fig. 1 Composition of the population by age group (1965~2065)

The composition of this paper is as follows. Chapter II explains the theory and statistical validation of the logit model and examines previous studies. Chapter III explains how to construct and supplement the aggregate data based on the household data. In Chapter IV, we set the parameters for travel time, travel cost, and the ratio for each age and verify the model. Finally, the conclusion of this study is presented.

II. LITERATURE REVIEW

It is common to use a logit model when analyzing mode

World Academy of Science, Engineering and Technology
International Journal of Transport and Vehicle Engineering
Vol:11, No:7, 2017

choice. McFadden used a random utility model (multinomial logit model) initially in the field of transportation research [4]. In this study, the logit model is used as a basic model for the analysis. The logit model is based on utility maximization theory. If one of the modes is selected, then the utility of the chosen mode is much more perceived than the other modes not selected.

The utility of these alternatives is assumed to be composed of observable systematic utilities and random utilities [5].

$$U_i = V_i + \varepsilon_i \qquad (1)$$

where U_i = total utility, V_i = systematic utility, ε_i = random utility.

If a traveler chooses the i-th mode, then the utility of the mode i is greater than the utility of the other alternatives.

$$P(i \mid C_n) = P(U_{in} \geq U_{jn}, \forall j \in C_n) \qquad (2)$$

where C_n = choice set, U_i = utility of mode i, U_j = utility of alternative j.

This probability is applied to the logit model.

$$P(i) = \frac{e^{U_i}}{\sum_{j}^{n} e^{U_j}} \qquad (3)$$

where n = number of modes.

In general, the utility function is estimated by taking into account the travel time, access time, fare, and the dummy expressing characteristics of each mode.

Reference [2] finds the variables that influence the mode choice among the data obtained from the Household Travel Survey data. As a result, in the regional characteristics model, the characteristics such as population density, employment density, presence of subway station affect the mode choice.

In order to include the independent variables in the model, it is necessary to investigate data thoroughly, and the statistical significance and the sign should be logical in the process of estimating the parameter value of the model. However, there are too many factors to consider when choosing a mode.

Reflecting the various variables increases the accuracy of the model, but it becomes complicated and less compatible. In order to introduce socio-economic and regional characteristics, utility functions should be constructed as disaggregated data of individual trips. However, models that use disaggregated data are ineffective because the modal share is predicted based on zone-based statistics.

In this paper, we try to improve the model by adding the age ratio variable based on the aggregate data in order to increase the realistic usability. We used the data from the Passenger Travel Survey to analyze the Seoul metropolitan districts' trip data. Limdep was used to construct the model. Even if the goodness-of-fit of the model is high and the sign and chi-square statistics of each variable are significant, the

final decision on the significance of the model and its application to reality is based on whether the value of travel time is socially acceptable, and the results yield similarly with the actual results [5].

III. METHODOLOGY

We used data on the Passenger Travel Survey by the Metropolitan Transportation Authority (MTA) in 2010, and constructed the data on 25 districts of Seoul. Since the trip in the inner zone includes a short distance, it is difficult to calculate the average speed for each mode and excluded from the travel distance and travel time calculation.

Among the 14 modes surveyed, the modes used in this analysis were simplified by passenger car (passenger car and motorcycle), bus (commuting, urban, seat, express, and feeder), and subway. Pedestrian, bicycle, and truck are excluded.

In order to obtain valid data, this study excluded unreasonable values from the errors of the respondents in the data input process, errors in logical verification. We removed the values that contain errors or negative values, and removed the outliers.

The following assumptions were made to estimate the travel cost. In the case of passenger cars, only the direct payment directly affects the modal choice, so only the fuel cost and the parking cost are considered. The fuel cost of a passenger car is determined by the following calculation, and the fuel cost is based on 2010.

Fuel cost = vehicle distance × fuel consumption × fuel cost per liter. Fuel consumption: 9.42 km per liter. Fuel cost per liter: 1,582 Korean won per liter.

For buses except for wide area buses and subways, 900 won is applied within 10km, and 100 won for every 5 km over 10 km. In the case of a wide-area bus, it costs 1,800 won within 10 km, and 100 won every 5 km over 10 km.

IV. MODEL CONSTRUCTION AND PARAMETER ESTIMATION

A. Model Construction

In order to add the socioeconomic variable, age ratio, to the model, the following model was constructed.

First, total travel time and total travel cost were applied as generic variables, and n-1 alternative specific constants for n modes were used for the model.

Applying same coefficients to the utility function of all alternatives is called an alternative general variable, and when the different coefficients are applied to each alternative, it is called an alternative specific variable. In this study, we used alternative general variables for travel costs and travel time, and constructed the model using alternative specific variables. Age ratio of each alternative was added 45 to 55 years, 55 to 65 years, and over 65 years.

In order to verify the estimated coefficients according to these analytical methods, the values of travel time and statistical significance were examined.

World Academy of Science, Engineering and Technology
International Journal of Transport and Vehicle Engineering
Vol:11, No:7, 2017

TABLE I
RESULTS FOR MODEL PARAMETERS

Variable	Mode	Estimates ($\times 10^{-4}$)	$-2(\mathcal{L}(0) - \mathcal{L}(\beta))$
Ratio of people aged 45 to 55 years	Car	1.973	1.971
	Bus	-0.629	0.057
	Subway	-1.317	1.242
Ratio of people aged 55 to 65 years	Car	0.960	0.039
	Bus	2.449	1.002
	Subway	-2.401	0.946
Ratio of people over 65 years	Car	-0.761	0.230
	Bus	6.744	1.654
	Subway	-3.665	0.338
Travel time (10 minutes)	-	-0.0672	-
Travel cost (100 Korean Won)	-	-0.0361	-
Alternative specific constant for bus	-	-0.635	-
Alternative specific constant for subway	-	0.983	-

Number of observations: 625

B. Results

Table I shows the results of analysis of three modes such as car, bus, and subway.

First, since the signs of travel time and cost are both negative, it is consistent with the common sense that increasing travel time and increasing travel costs will reduce utility.

Second, the result of calculation of the travel time value based on the marginal replacement rate is 11,174 won, which does not deviate much from the value of average passenger time in the metropolitan area.

Third, the chi-squared statistics show that buses and subway passengers and subway passengers in the age group of 45 to 55 years are statistically significant at buses and subways at 65 years and over. It can be said that only these alternative characteristic variables are different from zero.

Fourth, as for the alternative characteristic variables that are significant, in the '45 to 55 years old ratio 'parameter, the car has a positive sign and the subway has a negative sign. It can be seen that they prefer to use the subway rather than other age groups. In the '55 to 65 years old' ratio, the bus was pumped up and the subway was negative. Bus is preferred to subway. For those aged 65 years and over who were categorized as elderly, only bus results were significant and their sign was positive. The absolute value of the bus parameters was 6.744, which is much higher than the other significant parameters.

Only the significant estimates are extracted from the results, and a new improved sharing model is established.

$$U(car) = \beta_c \times COST + \beta_T \times TIME + \beta_{car,1} \times AGE_1$$
$$U(bus) = \alpha_b + \beta_c \times COST + \beta_T \times TIME$$
$$+ \beta_{bus,2} \times AGE_2 + \beta_{bus,3} \times AGE_3 \quad (4)$$
$$U(sub) = \alpha_s + \beta_c \times COST + \beta_T \times TIME$$
$$+ \beta_{sub,1} \times AGE_1 + \beta_{sub,3} \times AGE_3$$

Table II shows the results of estimating the parameters

again through the improved mode choice model. A chi-square test was performed using the results of the newly estimated results. Since five variables are added to the existing model, the degree of freedom is 5. $-2(\mathcal{L}(0) - \mathcal{L}(\beta)) = 4.75$ and it means better performance than the existing model.

According to the improved model, 45 to 55 years old people prefer to use the car, 55 to 65 years old and over 65 years old people prefer to bus. This means the higher the age, the lower the rate of economic activity and the less time value. In addition, both groups, 45 to 55 years old and 55 to 65 years old people showed a tendency not to prefer the subway. This is due to the influence of the under 45 year old group not reflected in the model. Because the alternative specific constants of the subway are higher about 1.6 than the bus, the subway preference is much higher than the bus. However, compared to other age groups, 45 to 55 years old and 55 to 65 years old people have a low preference for subway, so the parameters for subway are negative.

V. CONCLUSIONS

The current mode choice model uses common parameter values for each region. In this case, since the average value is used, it does not respond sensitively to individual trip data. Therefore, in this study, the method of incorporating the age into the mode choice model was examined. We usually use the disaggregate model when we include individual characteristics in the model. In reality, however, the mode choice model is considered with zone-based data. In order to solve this problem, the age characteristics were considered in the aggregate model as the age group ratio.

In order to construct the mode choice model, we used data on 2010 Household Travel Survey by the Metropolitan Transportation Authority (MTA). Generic variables such as travel cost, travel time, and alternative characteristics, the ratio of the age group, were tested by each alternative. As a result, the 45 to 55 years old variable has a positive sign on the car, the negative sign on the subway. The 55 to 65 years old variable has a positive sign on the bus and the subway and over 65 years old variable has a positive sign on the bus.

As a result of the model estimation of this study, the following suggestions can be drawn. First, as the age increases, the preference for the car decreases and the preference for the bus increases. In Korea, which is aging rapidly, demand for cars is expected to be converted to buses. Therefore, the authorities should expand the bus system to the zone with high aging rate or establish appropriate transportation infrastructure plan.

Second, the possibility of reflecting the individual characteristics in aggregate models was examined. In previous studies, there was an academic trend that focused on individual activities rather than zones. The analysis of the individual or household characteristics and local variables influencing the mode choice was made. In reality, however, it is inadequate to apply individual trip because demand is predicted based on zone-based data. In this respect, this study is meaningful in that the individual and household

World Academy of Science, Engineering and Technology
International Journal of Transport and Vehicle Engineering
Vol:11, No:7, 2017

characteristics are applied to the aggregate model and meaningful results are obtained.

Future research topics are as follows. In this study, since age is classified without any specific criteria, a method to optimize the boundary and interval of age variable should be suggested. In addition, by applying the aggregate model to other variables besides the age variable, a more accurate and realistic model can be presented than this study. This paper suggests the direction of model estimation from existing disaggregate data that reflects individual, household, and local characteristics to aggregate model estimation with realistic applicability.

TABLE II
RESULTS FOR IMPROVED MODE CHOICE MODEL

Variable	Constant	Time (10 minutes)	Cost (100 Korean won)	Ratio of people aged 45 to 55 years	Ratio of people aged 55 to 65 years	Ratio of people over 65 years
U(car)	-			1.973	-	-
U(bus)	-0.635	-0.0672	-0.0361	-	2.499	6.744
U(sub)	0.983			-1.317	-2.401	-

REFERENCES

[1] I. K. Kim, K. S. Han, and H. J. Bang, "Suggesting a Demand Forecasting Technique Explicitly Considering Transfers in Light Rail Transit Project Analysis (Translation Journals style)," *Journal of Korean Society of Transportation*, Vol.24 No.3, May, 2006, pp. 197-205.
[2] S. N. Kwon, "A study on the factors influencing traveler's mode choice (Translation Journals style)," Dissertation of Master Degree, Yonsei University, 2007.
[3] Statistics Korea, "Population Projections for Korea (2015~2065)," (http://kostat.go.kr/portal/eng/pressReleases/1/index.board?bmode=read&aSeq=359108), 20/02/2017 reviewed.
[4] D. McFadden, "Conditional logit analysis of qualitative choice behavior," 1973.
[5] M. Ben-Akiva and S. R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, *The MIT press*, Cambridge, Massachusetts, 1987.