Development of a Modal Share Estimation Method Considering Latent Passenger Preference for Transportation Modes

Munenori SHIBATA, Dr. Eng.
Assistant Senior Researcher,
Transport Planning and Marketing Laboratory, Signalling and Transport Information Technology Division

Masai MUTO, Dr. Eng.
General Manager,
Strategic Research Section, Planning Division

Daiki OKUDA
Administration Division

The study has attempted to develop a method for estimating modal share of inter-city express trains focusing on mode choice behavior in inter-regional trips. Firstly, the paper indicates that most travelers are mode captive, who recognize only one transportation mode among various alternatives in their mode choice behavior. Thus, an attempt was made in this paper to develop a PLCS (Parameterized Logit Captivity and Selectivity) model to describe mode choice behavior more appropriately. This study also showed that the method developed to estimate modal share with the PLCS model was sufficiently accurate.

Keywords: inter-regional transportation, modal share, mode choice behavior, latent preference

1. Introduction

In Japan, most analyses and project evaluations for inter-regional transportation planning employs Inter-regional Net Flow Survey (hereafter INFS) data [1], [2]. This data is based on a nationwide trip surveys conducted every five years since 1990. Collected trip data and trip based Origin-Destination (hereafter OD) tables are officially disclosed by the Ministry of Land Infrastructure, Transport and Tourism (hereafter MLIT). The traditional four step method and disaggregate behavior models (e.g. disaggregate logit model) based on INFS data are usually applied to demand analyses and estimation. Most of these analyses usually assume that traveler preference, which is usually explained by some observable Level Of Service (hereafter LOS) factors such as travel time, travel cost, frequency of transportation modes, and impedance of transfer etc, is constant in the applicable scope of these demand models.

In particular, there are some assumptions when applying a disaggregate behavior model based on consumer behavior theory such as disaggregate logit model to traveler behavior: (1) every trip maker has a perception of multiple transportation modes as choice alternatives, (2) every trip maker chooses a transportation mode among multiple alternatives so as to maximize his utility referring to some sort of information about LOS of these alternatives. However, it seems that there are varieties of traveler preference segments which cannot be properly explained by the abovementioned LOS factors alone, because inter-regional trips, which are different from inner-regional daily trips such as commuter journeys etc, are an occasional experience for most travelers and the amount of information in the case of inter-regional trips seems to be less than for daily trips. In other words, there is a possibility that decision-making is influenced not only by LOS, but also by the magnitude of importance for intangible factors, such as punctuality, ease of loading luggage (etc e.g. [3]). Additionally, potential traveler preference for inter-regional transportation modes, such as a like or dislike of a specific transportation mode itself, seems to have an effect on mode choice behavior when the trip purpose is leisure.

This study, therefore, focuses on the various abovementioned traveler intuitive subjective factors, hereafter referred to as “latent preference factors,” and their influence on the mode choice behavior of inter-regional travelers. The study attempted to calibrate the PLCS (Parameterized Logit Captivity and Selectivity) model with a consideration of a mode alternative screening process in mode choice behavior. Finally, the study demonstrates the developed modal share estimation method with using the PLCS model for inter-regional trips.

2. Survey of inter-regional trips

For the purposes of this study, therefore, a new questionnaire based survey was conducted to collect trip data for every representative inter-regional transportation mode, set of inter-regional express train passengers, inter-city express bus passengers, airplane passengers, and automobile users in November 2008. The survey aimed to gather not only trip attribute data (e.g. origin, destination, the result of choice (=revealed preference) etc), but also mode alternatives recognized by the trip maker, and some magnitudes of importance of intangible factors. In addition, the study also focused on potential traveler preference for inter-regional transportation modes, such as like or dislike of each transportation mode itself, since these seem to affect the mode choice behavior in the case of leisure journeys. The magnitude of potential preference for each mode was therefore also observed for each trip maker independent of the section of trip survey. The investigation was conducted through a web survey system, and respondents were driving license holders having made a domestic inter-regional
journey within the previous two months.

The main contents of the trip survey are shown in Table 1. The questionnaire mainly aimed to gather three types of content: (1) level of potential preference (+1 ~ +5) such as like or dislike of each mode independently of the following trip attributes, (2) not only the result of mode choice, but also alternative modes recognized by the trip maker, (3) magnitude of seven levels (+1 ~ +7) of importance given to intangible factors (21 items) in relation to mode choice behavior.

Sample data was extracted according to the following criteria: (1) journeys made by inter-regional express train users, automobile users and inter-city express bus users formed the object of the study, (2) trips in which an airplane was recognized as an alternative by the trip maker were excluded, (3) trips in which an inter-city express bus could not be used because of unavailability of express bus lines were excluded. The above criteria brought the final sample size down to 1,611. Free address matching services on the web operated by the Center for Spatial Information Science, the University of Tokyo and National Integrated Transport Analysis System (NITAS) developed by the MLIT were applied to create some LOS data (e.g. travel time, cost, frequency etc.) for each sample.

3. Initial findings

One of the important features of this study’s trip survey was to observe which alternative modes were rec-
recognized by traveler. On the basis of the abovementioned collected mode alternative data, this section analyzes the relationship between mode alternative and latent preference. In the following analyses, a mode captive traveler is defined as a traveler who recognizes the chosen transportation mode as sole alternative. For simplicity of analysis and modeling, a selective traveler is defined as a traveler who recognizes more than two modes including the one actually chosen. The study included four mode alternative categories: mode captive traveler for express trains (Rcap), mode captive traveler for automobiles (Acap), mode captive traveler for inter-city express buses (Bcap), and the selective traveler (sel).

Figure 1 shows the share of mode alternative categories aggregated by each mode user. The share of mode captive traveler usually retains a high share rate of around 90% for every mode user group. This suggests that analyzing the screening process of mode alternatives should be used as a premise for modeling mode choice behavior.

Figure 2 indicates the mean value of relative potential preferences of each alternative category. The relative potential preference is defined as the difference between potential preference for one mode and average of potential preference for all transportation modes, calculated per sample. The values of relative potential preference were averaged for each category. Each captive category had a relatively high potential preference for its transportation mode. Potential preferences of the selective group had average points around 0 in comparison with captive categories. Figure 3 shows the mean value of relative magnitude of importance for some of the 21 intangible factors. The relative magnitude of importance for each of the 21 mode characteristics is defined as the difference between the magnitude of importance of each intangible factor and the average of the magnitudes of all intangible factors, calculated per sample. The values of relative magnitude of importance were averaged for each alternative category. Regarding mode qualitative characteristics, the Rcap group placed particular weight on “rapid-ability of reaching to destination” and “punctuality.” The Acap group gave more importance to “mobility,” “ease of loading luggage” and “availability of private space.” The Bcap group put great priority on “low travel cost.” Each captive category placed significant weight on potential enjoyment to be drawn from the selected mode of transport. For example, the Acap group demonstrated a high magnitude of importance for “potential of enjoyment on driving automobile.” The selective category, however, had an average magnitude of importance of around 0 for most intangible factors compared with captive categories.

The abovementioned characteristics found in aggregate analyses suggest that travelers showing neutral tendencies both for potential preference and magnitude of importance for intangible factors tend to belong to the selective group, whereas travelers who put high importance on potential preference for a specific mode tend to belong to a specific mode captive category. Similarly, travelers putting high magnitude of importance on intangible factors of specific mode characteristics show an increased tendency to belong to the relevant mode captive category.

4. Development of the PLCS model

4.1 Assumptions of the decision-making process in the mode choice behavior

According to the abovementioned analyses, the principal characteristics of mode choice behavior among inter-regional leisure travelers can be summarized as follows.

A) Most travelers recognize the transportation mode actually chosen as the only mode option. In other words, most travelers are mode captive and the share of selective travelers is extremely small according to the collected data on recognition of mode alternatives.

B) Latent preference factors such as magnitude of impor-
Fig. 4 Decision-making process assumptions in mode choice behavior for this study

Table

<table>
<thead>
<tr>
<th>Mode Alternative</th>
<th>Selective (Sel)</th>
<th>Express Train Captive (Rcap)</th>
<th>Automobile Captive (Acap)</th>
<th>Inter-city Express Bus Captive (Bcap)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Mode Alternative Screening Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2: Determining Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2 The PLCS Model

The general formula of the random utility model considering the mode alternative screening process is formulated in (1) [6].

\[
P_{in} = \sum_{C \in G_j} P_n(i | C) \cdot Q_n(C | G_j) \tag{1}
\]

where \( P_{in} \) : probability that individual \( n \) selects mode \( i \), \( P_n(i | C) \) : probability that mode \( i \) is included in choice alternative of individual \( n \), \( G_j \) : alternative set \( j \), \( Q_n(C | G_j) \) : probability that individual \( n \) opts for alternative set \( C \) screening among all alternative sets \( G \). In the study, step 1 “Mode alternative screening process” corresponds to \( Q_n(C | G_j) \) and step 2 “Determining process” corresponds to \( P_n(i | C) \).

4.2.1 Formulation of PLCS Model

This study deals with mode choice behavior among 3 alternatives ((i=1: express train, 2: automobile, 3: intercity express bus). For the sake of modeling simplicity, in this study, a selective traveler is defined as a traveler who recognizes more than two modes including the mode actually chosen. There are therefore four categories of mode alternative. Suffix \( j \) represents the number given to the mode alternative. \( G \) is defined as \( G_0 = \{1, 2, 3\} \) : “selective,” \( G_1 = \{1\} \) : “Rcap,” \( G_2 = \{2\} \) : “Acap,” \( G_3 = \{3\} \) : “Bcap.”

In this study, for step 1 “Mode alternative screening process,” the alternative category probability function is formulated as the commonly used logit probability model in (2) and (3).

\[
Q_n(G_j) = \frac{\exp(U_{jn})}{\sum_j \exp(U_{jn})} \tag{2}
\]

\[
U_{jn} = \sum_k \theta_{jnk} X_{jnk} \tag{3}
\]

where \( Q_n(G_j) \) : probability that individual \( n \) has alternative set \( G_j \) screening among all alternative sets \( G \), \( U_{jn} \): function of the alternative category \( j \) of individual \( n \), \( X_{jnk} \): each explanatory variable, \( \theta_{jnk} \): parameters to be estimated.

Step 2 “Determining process” is only applied to travelers in \( G_0 \) (selective category), since they select one transportation mode among several alternatives. In this case, the traditional disaggregate logit model can be applied because this process can be interpreted as a normal choice behavior with several alternatives. Step 2 is formulated in (4) and (5).

\[
P_n(i | G_0) = \frac{\exp(V_{in})}{\sum_i \exp(V_{in})} \tag{4}
\]
\[ V_{in} = \sum_k \beta_{nk} X_{nk} \]  

where \( P_n(i | G_0) \) : probability that individual \( n \) selects transportation mode \( i \) when the \( n \) belongs to \( G_0 \), \( V_{in} \) : utility function of mode \( n \), \( Y_{mk} \) : each explanatory variable, \( \beta_{nk} \) : parameters to be estimated.

The PLC (Parameterized Logit Captivity) model developed by Swait et. al. \[7\] can express the adherence to only one of each mode captive categories by functions composed of some attribute variables in step 1. However, the model formulated in the study is capable of considering factors which lead travelers to belonging not only to captive categories but also to the selective category in step 1. Hence, the authors of the present paper named the decision-making process model “PLCS” (Parameterized Logit Captivity and Selectivity). This model can be classed as an enhancement of the PLC model.

### 4.2.2 Abstraction of “attitude of importance for each mode” by factor analysis model

Many kinds of latent preference factors influence the alternative screening process in mode choice behavior. However, it is expected to be difficult to deal with so many latent factors in a decision-making process model. Hence, the study tries to represent these many factors into a set of abstracted factors. This step is expected to be useful for installing the latent factors into the PLCS model.

Factor analysis is applied to latent preference factor data in order to abstract some representative factors. The result of factor analysis is shown in Fig. 5. Latent preference data was not normalized and was inputted “as is” in each observed latent factor shown by a square box in Fig. 5. GFI is goodness-of-fit index of the estimated model, and AGFI is adjusted GFI. From the fitness shown by GFI and AGFI, the model can abstract some reasonable latent factors from observed latent preference factor data with sufficient accuracy.

Three latent factors “Attitude of Importance (hereafter A.I.)” each of which corresponds to a transportation mode were abstracted through the model. The latent factor #1 was named “A.I. for express trains,” because it comprised “rapid-ability of reaching to destination,” “punctuality” and “potential preference for express trains.” The latent factor #2 consisted of “mobility,” “ease of loading luggage,” “availability of private space” and “potential preference for automobiles” and was therefore named “A.I. for automobiles.” Similarly, the latent factor #3 was called “A.I. for inter-city express buses,” because it was influenced by “low travel cost” and “potential preference for inter-city express buses.”
4.2.3 Result of parameter estimation of the PLCS model

The PLCS model has one important feature in comparison with the PLC model, which is that the element explaining the condition where a traveler tends to belong to the selective category is part of the function $U_0$ of "Alternative screening process ((3))." Regarding the abovementioned characteristics mentioned in the results of the aggregate analyses in chapter 3, travelers with neutral attitudes to both potential preference and magnitude of importance for intangible factors tended to belong to the selective category. Thus, a special "index of difference between each $A.I.$ (hereafter $dif\ A.I.$)" is included in the function in this study. The index $dif\ A.I.$ is defined by the equation (6).

$$dif\ A.I. = \frac{1}{\left|AI_{1n} - AI_{2n}\right| + \left|AI_{1n} - AI_{3n}\right| + \left|AI_{2n} - AI_{3n}\right|}$$  \hspace{1cm} (6)

where $dif\ A.I._n$: $dif\ A.I.$ index of individual $n$, $AI_i$: latent factor score ($A.I.$) of mode $i$ of individual $n$. When every $A.I.$ has the same factor score, $dif\ A.I.$ reaches an infinite value. On the other hand, the larger the $A.I.$ difference becomes, the smaller the calculated $dif\ A.I.$.

The calibrated parameters of the PLCS models are shown in Table 2. In the $Q_n(G_j)$ of model 1, $AI_i\; dif\ A.I.$ are included in the functions. In model 2, the "Min. difference of GC" is included. Here, GC means Generalized Cost which is calculated by the summation of travel cost and cost of travel time corresponding to the value of the time "42.3 yen/minute" derived from calibrated disaggregated logit model. The min. difference of GC is defined as the difference between a relevant mode’s GC and the GC of the most competitive mode. Here, the most competitive mode means the mode which has the closest GC to the relevant mode. For a concrete example, min. difference of AGC is calculated by automobile GC minus express train GC when express train GC is nearer to AGC than inter-city express bus GC. If AGC is larger than the most competitive mode’s GC, the probability of encountering Acap is expected to decrease. Therefore, the sign condition of each minimum (min.) difference of GC should be negative. The utility functions of $P_y(i|G_0)$ of both models had explanatory variables.

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step1</td>
<td>$Q_n(G_j)$</td>
<td>Acap</td>
<td>$Min.; difference; of; AGC$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bcap</td>
<td>$Min.; difference; of; BGC$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rcap</td>
<td>$A.I.; 1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acap</td>
<td>$A.I.; 2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bcap</td>
<td>$A.I.; 3$</td>
</tr>
<tr>
<td></td>
<td>Bcap</td>
<td>$sel; dif; A.I.$</td>
<td>$3.3287$</td>
</tr>
<tr>
<td>Step2</td>
<td>$P_y(i</td>
<td>G_0)$</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Travel cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10,000 yen/person)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Express train constant term</td>
<td>$-0.7937$</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>Automobile constant term</td>
<td>$-2.2061$</td>
</tr>
</tbody>
</table>

| $p^2$ | 0.550 | 0.568 |
| AIC (Akaike’s Information Criteria) | 1597.56 | 1535.77 |
| Hit ratio of mode choice result (%) | 81.4% | 82.1% |
| Hit ratio of alternative category (%) | 55.2% | 59.2% |
| Value of time (yen/minute) | 73.7 | 84.8 |
| Number of samples | 1,611 | 1,611 |

* ( ) : Statistical t-value
* Rcap : Express train captive travelers  Acap : Automobile captive travelers  Bcap : Inter-city express bus captive travelers  sel : Selective travelers
* C : Common variable  R: Express train specific variable  A: Automobile specific variable
of LOS and a mode constant. All parameters in both steps were estimated simultaneously.

As for goodness-of-fit shown by $\hat{R}^2$, each model had a significant $R^2$, larger than 0.2. However, model 2 was superior to model 1 due to the introduction of the min. difference of GC. Moreover, the AIC (Akaike’s Information Criteria) of model 2 was smaller than that of model 1. Accordingly, model 2 can be used to explain the 2-step mode choice behavior decision-making process more accurately than model 1. From the parameter significance of the explanatory variables point of view, each variable except for the express train constant term had a significant t-value. Difference of both GC and A.I. had an impact on the “Mode alternative screening process (step1)” with statistical efficiency. LOS, such as travel time and cost, also showed statistical efficiency in the utility functions. This result appears to prove that the assumption of a mode choice behavior decision-making process in inter-regional leisure travelers in this study (Fig. 4) was reasonable and appropriate.

5. Estimation performance of the PLCS model for modal split share in inter-regional trips

5.1 Estimation method

Latent preference factors for each transportation mode are important variables to describe traveler mode choice behavior in the developed PLCS model. These variables are indices compounded with some latent preference factors. When the PLCS model is applied to an estimate of modal share, these latent preference factors should be assumed in some way. In fact, however, the variances in these factors are very wide even in a particular individual attribute, for example among male travelers over sixty, there exist both those who have high potential preference for express trains and the others who do not. Hence, it is difficult to explain latent preference factors entirely through individual attributes in a statistical model. This study has therefore tried to develop an estimation method of modal share with the PLCS model (model 2 in table 2) with the assumption that latent preference factors emerge in accordance with a certain statistical distribution. The flow diagram of the developed estimation method implemented in relation to the Monte Carlo Simulation is shown in Fig. 6.

Firstly, with the assumption that distributions of 9 latent preference factors introduced into the latent analysis model shown in Fig. 5 are distributed according to statistical normal distribution [8], the correlation matrix of the 9 factors is calculated on the basis of trip survey data. The next step is to generate multidimensional random numbers based on the correlation matrix. These random numbers express a condition of latent preference factors among N persons. By substitution of these random numbers and LOS data for analyzed OD to the PLCS model, probabilities of mode choice for N persons are estimated. Finally, the modal share of the analyzed OD is calculated by averaging the probabilities obtained for N persons.

5.2 PLCS model estimation performance

In this study, the modal shares for some ODs related to Japanese high speed rail lines (Shinkansen) having opened recently, where competition between express trains, automobiles and inter-city express bus can be observed, are estimated in order to demonstrate estimation performance of the PLCS model as shown in Fig. 7. In this Fig., actual Shinkansen shares are calculated based on trip data derived from the INFS implemented by MLIT. Estimated shares from the PLCS model nearly correspond to actual shares in 4 cases, for trip distances (one way) longer than approx. 100km and shorter than approx. 300km. However, for OD between city G and city H, where trip distance is shorter than approx. 100km, the estimated share is less than the actual share. These case studies suggest that the suitable range of trips to which the developed method can be applied are those between approx. 100km to 300km. For OD involving shorter distances such as under 100km, demand estimation models for inner-regional Shinkansen
6. Concluding remarks

Finally, this study led to the following conclusions:

A) The mode captive share usually remains high for every inter-regional mode user. Thus, mode choice behavior should be treated as at least a 2-step decision-making process: “Mode alternative screening process” and “Determining process.”

B) The PLCS model proposed by the study, a kind of extended PLC model, was efficient for describing the 2-step decision-making process in mode choice behavior.

C) In particular, latent preference factors, such as “potential preference for each transportation mode” and “magnitude of importance for intangible factors,” greatly influenced the “Mode alternative screening process” suggested by the PLCS model.

D) The developed estimation method for modal share with the PLCS model demonstrated good performance in express train modal share estimation for ODs with a trip distance of between approx. 100km to 300km.

The following challenges have been identified for future work: A) more validation studies applying new 2010 INFS data, B) an improved estimation method to make application possible to longer distance trips, where severe competition usually exists between airlines and express trains.

References


