PREDICTION OF TRIP DISTRIBUTION BY DISAGGREGATE BEHAVIORAL MODEL

S. Morichi[1] and T. Yai[2]

1) S. Morichi
Tokyo Institute of Technology
2-12-1 Ookayama, Tokyo, Japan

2) T. Yai
Tokyo Institute of Technology
2-12-1 Ookayama, Tokyo, Japan

1. Introduction

It has been more than a decade since the disaggregate behavioral models based on random utility theory were developed. With these models it becomes possible to analyze and to predict travel behavior. Especially in mode choice studies, many good results have been presented. Theorefore it is considered highly possible that the aggregate modal-split models in the conventional procedures could be replaced by disaggregate ones.

In Japanese cities of which population is more than 300 thousands, person-trip surveys are carried out once a decade, so that present OD tables are available for urban transportation planning in these cities. Using the data, four-step procedure could be employed to forecast future travel demand. Although for the step of the modal split disaggregate models are often utilized, the principal method for the step of trip distribution is usually the aggregate models because of a certain difficulties in estimation with disaggregate destination choice models.

Ben-Akiva (1973), McFadden (1974) and others show strong points of disaggregate destination choice models, but important issues from the view point of practical sense are remained unclear. The main purpose of this paper is to establish the disaggregate demand forecasting techniques for OD tables.

In Section 2 a methodology to predict trip distribution with disaggregate models is discussed and in Section 3 we describe the sampling methods of choice alternatives, which reduces the time for the estimation of parameters in the case where the size of choice set is very large.

In the following three chapters we examine the practicality of destination choice models for the prediction of trip distribution. Section 4 presents the case studies of a new methodology developed in Section 2. In Section 5 we examine the tolerance to the reduction of choice alternatives by considering the stability of parameters. Further, in Section 6 a sample size which is enough to estimate trip distribution is discussed in the case that disaggregate destination choice model is utilized to predict OD tables.

Finally, Section 7 presents conclusions of this paper.

2. New Approach to Forecast Trip Distribution with Disaggregate Model

The aggregation method which we develop in this paper is to predict trip distribution so as to maximize total utilities in a population under the various restrictions.

Firstly, we consider the random utility $U_{jn|i}$ of destination j for trip maker n in orign i; it can be written as,

$$U_{jn|i} = V_{jn|i} + \varepsilon_{jn|i} \tag{1}$$

In the above, if the stochastic term $\varepsilon_{jn|i}$ follows Gumbel distribution as,

$$\mathrm{Prob}[\varepsilon_{jn|i} \leqq \varepsilon] = e^{-e^{-\lambda(\varepsilon + \alpha_{jn|i})}} \tag{2}$$

the expected value of maximum utility $MU_{n|i}$ finally can be expressed as follows,

$$MU_{n|i} = E[\max_{j\in D} U_{jn|i}] = \frac{1}{\lambda} \ln \sum_{j\in D} e^{\lambda V_{jn|i}} \tag{3}$$

$MU_{n|i}$ represents the choice situation of trip maker n, and also is reffered to as Consumer's Surplus (Williams (1977)), Satisfaction function(Daganzo (1979)), or Accessibility (Ben-Akiva & Lerman (1979)).

Secondly, we consider the expected value in a population. For this derivation, some aggregation techniques are needed and we adopt classification method which is one of the most efficient techniques; it has some generalities so that it includes Naive and Enumeration methods as its special cases. The validity of this method has been proved in many recent studies.

Now, we classify a population by G segments in each origin zone i. In this case, the utility $MU_{g|i}$ is defined as,

$$MU_{g|i} = \frac{1}{\lambda} \ln \sum_{j \in D} e^{\lambda V_{jg|i}} \tag{4}$$

where $V_{jg|i}$ is a strict utility of destination j for segment g in origin i. Thus, MU_i (the utility in origin i) is,

$$MU_i = t_{i\cdot} \sum_g w_{ig} \cdot MU_{g|i} \tag{5}$$

where w_{ig} is the share of segment g in origin i ($\sum_g w_{ig} = 1$), and $t_{i\cdot}$ is trip generation of zone i. In eq.(5), it is assumed that the utility function has additive property.

Further, under this assumption a total utility can be shown that,

$$MU = \sum_i MU_i \tag{6}$$

In addition in eq.(4) we can obtain choice probability $P_{jg|i}$ from the property of $MU_{g|i}$.

$$P_{jg|i} = \frac{\partial MU_{g|i}}{\partial V_{jg|i}} = \frac{e^{\lambda V_{jg|i}}}{\sum_{j' \in D} e^{\lambda V_{j'g|i}}} \tag{7}$$

Since we assumed that the random term follows the Gumbel, Logit Model is derived here.

Now transforming eq.(7) into,

$$\ln P_{jg|i} = \lambda V_{jg|i} - \ln \sum_{j \in D} e^{\lambda V_{jg|i}} \tag{8}$$

further, calculating the expected value of both sides of eq.(8) with $P_{jg|i}$, then

$$\begin{aligned} \sum_{j \in D} P_{jg|i} \ln P_{jg|i} &= \sum_{j \in D} P_{jg|i} (\lambda V_{jg|i} - \ln \sum_{j \in D} e^{\lambda V_{jg|i}}) \\ &= \lambda \sum_{j \in D} P_{jg|i} V_{jg|i} - \ln \sum_{j \in D} e^{\lambda V_{jg|i}} \end{aligned} \tag{9}$$

By substituting (9) into (6), MU is rewritten as follows,

$$MU = \sum_{i \in D} t_{i\cdot} \sum_g w_{ig} (- \frac{1}{\lambda} \sum_{j \in D} P_{jg|i} \ln P_{jg|i} + \sum_{j \in D} P_{jg|i} V_{jg|i}) \tag{10}$$

If parameters of destination choice model have already been estimated, MU is derived by summing up utilities, which are calculated with the average values of characteristics of each segment.

This is the aggregation technique mentioned before, however it should be noted that calculated distribution of trips does not always coincide with observed one. For example, when the trip attractions are also given, those calculated from the estimated OD trips are generally different from observations.

For the reasonable solution of such a problem as above and to make estimation of trip distribution precise, we consider the following method.

The amount of OD trips can be estimated at the point that MU is maximized under the restrictions of total trips (ex. trip attraction).

For example, we consider the following maximization problem using (10),

$$\max_{P_{jg|i}} MU = \sum_{i\in D} t_{i\cdot} \sum_{g} w_{ig} \left(-\frac{1}{\lambda} \sum_{j\in D} P_{jg|i} \ln P_{jg|i} + \sum_{j\in D} P_{jg|i} V_{jg|i} \right) \tag{11}$$

$$\text{s.t.} \quad \sum_{i\in D} t_{i\cdot} \sum_{g} w_{ig} P_{jg|i} = t_{\cdot j} \quad \forall_{j\in D} \tag{12}$$

In this case, eq.(12) is the restriction on the trip attraction. In eq.(11), MU is maximized by $P_{jg|i}$ to estimate a new distribution pattern, thus $P_{jg|i}$ need not be in Logit form.

By the introduction of multipliers, γ_j and η_{ig}, Lagrangean L is expressed as,

$$L(p,\gamma,\eta) = \sum_{i\in D} t_{i\cdot} \sum_{g} w_{ig} \left(-\frac{1}{\lambda} \sum_{j\in D} P_{jg|i} \ln P_{jg|i} + \sum_{j\in D} P_{jg|i} V_{jg|i} \right)$$

$$+ \sum_{j\in D} \gamma_j \left(t_{\cdot j} - \sum_{j\in D} t_{i\cdot} \sum_{g} w_{ig} P_{jg|i} \right) + \sum_{i\in D} \sum_{g} \eta_{ig} \left(\sum_{j\in D} P_{jg|i} - 1 \right) \tag{13}$$

From eq.(13), the following partial differential equations are obtained.

$$\frac{\partial L}{\partial P_{jg|i}} = t_{i\cdot} w_{ig} \left[-\frac{1}{\lambda} \ln P_{jg|i} - \frac{1}{\lambda} + V_{jg|i} - \gamma_j + \eta_{jg} \right] = 0 \tag{14}$$

$$\frac{\partial L}{\partial \gamma_j} = t_{\cdot j} - \sum_{i\in D} t_{i\cdot} \sum_{g} w_{ig} P_{ig|i} = 0 \tag{15}$$

$$\frac{\partial L}{\partial \eta_{ig}} = \sum_{j\in D} P_{jg|i} - 1 = 0 \tag{16}$$

From (14) and (16), $P_{jg|i}$ is finally obtained as,

$$P_{jg|i} = \frac{e^{\lambda(V_{jg|i} - \gamma_j)}}{\sum_{j'\in D} e^{\lambda(V_{j'g|i} - \gamma_{j'})}} \tag{17}$$

write MU, $P_{jg|i}$, γ_j where MU is at its maximum as MU*, $P^*_{jg|i}$, γ^*_j is expressed as,

$$\frac{\partial MU^*}{\partial t_{\cdot j}} = \gamma_j^* \tag{18}$$

In eq.(18) γ^*_j is the first derivative of maximum total utility MU* about the observed trip attraction $t_{\cdot j}$. Therefore γ^*_j has the dimension of marginal utility, and the choice probability $P^*_{jg|i}$ is estimated according to the utility $V_{jg|i}$ and marginal utility γ^*_j which is dependent on observed trip attraction and independent of segments or origins.

From equation (15), we can evaluate γ^*_j by introducing,

$$f_j(\gamma) = \sum_{i\epsilon D} t_{i\cdot} \sum_g w_{ig} \frac{e^{\lambda(V_{jg|i} - \gamma_j)}}{\sum_{j'\epsilon D} e^{\lambda(V_{j'g|i} - \gamma_{j'})}} - t_{\cdot j} \qquad \forall_{j\epsilon D} \tag{19}$$

and solving $f_j(\gamma) = 0$ ($\forall_{j\epsilon D}$) iteratively with Newton-Raphson method.

After such evaluation of γ_j^* , we can obtain choice probability $P^*_{jg|i}$ which maximizes the total utility under the restriction of amount.

Finally the estimated trip distribution $\hat{t}_{ij}$ as

$$\hat{t}_{ij} = t_{i\cdot} \sum_g w_{ig} P^*_{jg|i} \tag{20}$$

satisfies the restriction of the volume of observed trip attraction.

The zone size in which the restriction of the volume is introduced should be determined according to the reliability of data. When the trip distribution of larger zone partition is available, introduce the following instead of (12),

$$\sum_{i\epsilon D_k} \sum_{j\epsilon D_\ell} t_{i\cdot} w_{ig} P_{jg|i} = t_{k\ell} \tag{21}$$

and finally $P^*_{jg|i}$ is obtained as,

$$P^*_{jg|i} = \frac{e^{\lambda(V_{jg|i} - \mu^*_{\ell(j)|k})}}{\sum_{j'\epsilon D} e^{\lambda(V_{j'g|i} - \mu^*_{\ell(j')|k})}} \tag{22}$$

where D_k and D_ℓ are the sub-sets of smaller zones of needed scale belonging to larger zones k and ℓ.

In the same way as discussed before, $\mu^*_{\ell|k}$ can be estimated, and we can predict the trip distribution in needed scale of zone partition.

For example, when the inter-municipality distribution of commuter trips is available from the national-wide census, this method can be employed effectively to predict OD table of finer zone partition with disaggregate data by another smaller survey.

Further, it is clear that this methodology can easily be generalized to fit many other phenomena.

3. Sampling of Destination Alternatives

When using a large number of destination alternatives, it can be prohibitively expensive to prepare the Level-of-Service data and to estimate model parameters.

For the solution to the problem as above, McFadden(1978) gave the method to estimate the parameters with smaller choice set sampled from given choice set, showing that the parameters have consistency with those estimated with full choice set. And Ben-Akiva (1984) summarized several ways of sampling.

These studies are based on the IIA property of Logit model. Outline of these methodology is as following.

Let A be the number of elements, of full choice set D, and SD be the sub-set of alternatives (SA elements), and P(SD|j) be the probability that choice set SD is formed when the alternative j has already been chosen. Assume the positive conditioning property as

$$P(SD|j) > 0 , \qquad \forall j \in D \tag{23}$$

is satisfied, and assume that in the following conditional probability,

$$P(j|SD) = \frac{P(SD|j)\cdot P(j:D)}{\sum\limits_{j'\in SD} P(SD|j')\cdot P(j':D)} \tag{24}$$

$P(j:D)$ has the Logit form as,

$$P(j:D) = \frac{e^{V_j}}{\sum\limits_{j'\in D} e^{V_{j'}}} \tag{25}$$

then $P(j|SD)$ becomes

$$P(j|SD) = \frac{e^{V_j + \ln P(SD|j)}}{\sum\limits_{j'\in SD} e^{V_{j'} + \ln P(SD|j')}} \tag{26}$$

McFadden (1978) proposed the method to estimate parameters constructing likelihood function from (26).

As a method to estimate $P(SD|j)$, Ben-Akiva (1984) showed Simple Random Sampling and Inportance Sampling. The latter is to take much of important information avoiding the inefficiency of the former. And Independent Importance Sampling, which is a family of the latter, postulates that the probability R_j of sampling of alternative j is independent of the probability for other alternatives. So it expresses $P(SD|j)$ as

$$P(SD|j) = \prod_{\substack{j'\in SD \\ (j'\neq j)}} R_{j'} \prod_{j'\notin SD} (1 - R_{j'}) \tag{27}$$

And then equation (26) is rewritten as

$$P(j|SD) = \frac{e^{V_j - \ln R_j}}{\sum\limits_{j'\in SD} e^{V_{j'} - \ln R_{j'}}} \tag{28}$$

To determine R_j, Ben-Akiva (1984) proposed a method which introduces the preliminary estimates of the choice probabilities. And for this method, two factors of distance and size are usually used.

In this paper, we propose another method of determining R_j. In this method, observed share is directly used. Let i be the subscript of origin, and $S_{j|i}$ be the observed share of destination j in the trips from origin i. Then, we describe $R_{j|i}$ as

$$R_{j|i} = a + (1 - a)bS_{j|i}/S_{m|i} \tag{29}$$

where $$S_{m|i} = \max_j S_{j|i} \tag{30}$$

$$\sum_j S_{j|i} = 1 \tag{31}$$

$$0 \leqq a \leqq 1 \tag{32}$$

$$0 \leqq b \leqq 1 \tag{33}$$

$$a + b \neq 1 \tag{34}$$

where coefficient a represents a fixed value of the probability and coefficient b is concerned with an upper limit of the variable part which is composed of observed shares.

The sub-set of each person SD_n is determined by using Monte-Calro simulation with the probability $R_{j|i}$ in eq.(29). And consistent estimators of the model parameters are finally obtained by the maximization of the following conditional likelihood function.

$$L = \sum_{n} \sum_{i \in SD_n} \delta_{in} \ln P(i|SD_n) \tag{35}$$

In this method, values of the both coefficients can vary a size of the subset. However up to now it is not clear how much effect on the parameters arises from the reduction of the number of alternatives.
We discuss this issue in Section 5.

4. Estimation results of trip distribution

An application of the methodology presented in Section 2 is carried out in order to examine its effectiveness. We use the person-trip-survey data which was obtained in Maebashi metropolitan area in 1977. In the following study, "to work trips" only in Maebashi city are applyed; the number of observation is 4723 and the area is drawn in Fig. 1. The area is divided into 11 B-zones and each B-zone is divided into several C-zone; the number of C-zone is 41.

Alternatives of destination choice models are the latter zones and aggregation to produce O-D tables is carried out with B-zones.

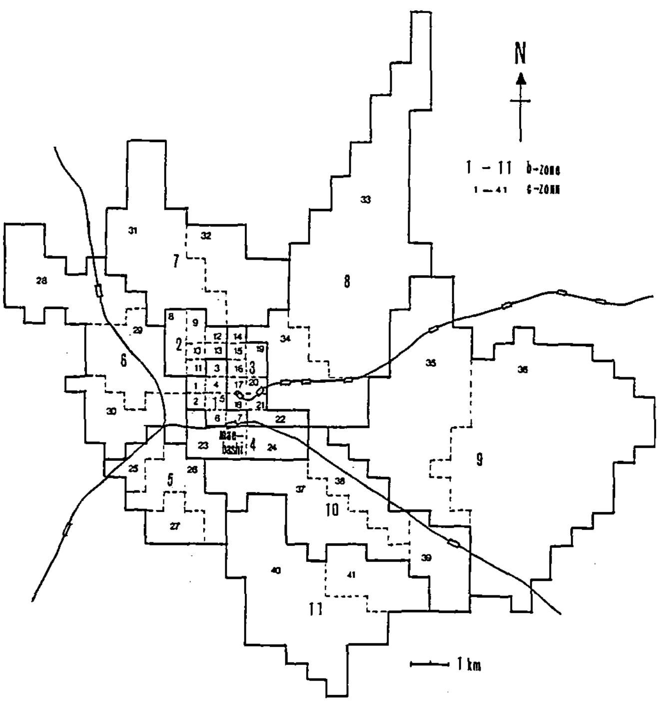

Fig. 1 Devided Zones in Maebashi-Metropolitan Area

Now define the utility of destination j by,

$$V_{jn|i} = V_{jn} + E(\max_{m} U_{mn|ij}) \tag{36}$$

where the inclusive price which represents the accessibility concerned with all modes is contained. The inclusive price is expressed by

$$E(\max_{m} U_{mn|ij}) = \frac{1}{\lambda} \ln \sum_{m' \in M} e^{\lambda V_{m'n|ij}} \tag{37}$$

and we can obtain this value from the following mode choice model,

$$P_{mn|ij} = \frac{e^{V_{mn|ij}}}{\sum_{m' \in M} e^{\lambda V_{m'n|ij}}} \tag{38}$$

Now first we mention about the result of model estimates. Mode-choice model is presented in Table 1. The number of observations is 2,000, which is sampled randomly from total ones.

Further, Table 2 shows destination choice model which contains the inclusive price obtained from Table 1. The choice set of this model is composed by the sampling method presented in Section 3; coefficient a is fixed at 0.25, coefficient b is 1.0, and Log-Likelihood function is presented at eq.(35). Every parameter is enough significant and Log-likelihood ratio is much higher.

Secondly we aggregate these models using classification method with a segmentation by occupation; a job is at productive-industry or not.

Table 4 is the OD matrix obtained after the model-aggregation and Table 3 is the observed matrix. The correlation coefficient between these matrices is 0.926.

Third we estimate OD table introducing γ_j in equation (13). Table 5 is the estimation result of the marginal utility γ_j. (γ_{11} is fixed at zero because of the property of Logit model) OD matrix from this process is presented in Table 6. In this case, the correlation coefficient is 0.960 and it is higher than the former.

From the case study above, the effectiveness of introduction of γ_j is shown.

Table 1 Estimation Results of Mode-Choice Model

Independent Variable		Estimated Coefficient	T-Statistic
Total travel time(min.)	G	-0.05419	8.20
Access time to station(min.)	R	-0.1943	2.30
Num. of cars/persons in household	C	1.759	5.59
Sex ; male 1 , female 0	C	0.5059	3.72
Car license ; owned 1 , otherwise 0	C	2.738	17.25
Age ; 50-69 1 , others 0	2W	0.3207	2.20
Num. of 2Ws/persons in household	2W	1.953	10.05
Constant	R	-2.168	4.74
Constant	B	-1.674	13.42
Constant	C	-3.187	16.47
Constant	2W	-1.996	12.13
Chi-square (d.f.=11)		1790.0	
Log likelihood at market shares		-2322.3	
Log likelihood at convergence		-1427.3	
Likelihood ratio index		0.384	
Num. of observations		2000	

* 2Ws ; two wheelers

Table 2 Estimation Results of Destination-Choice Model
(41 choice alternatives)

No.	Independent Variable		Estimated Coefficient	T-Statistic
1	Inclusive price	G	1.398	30.94
2	Trip attractions	G	0.0003714	23.47
3	(Num. of people engaged	23-24	0.0005575	4.61
4	in manufacturing industry)	28-30	0.0003298	10.32
5	*(manufacturing industry	33-34	0.0004800	5.57
6	dummy)	35-36	0.002097	9.96
7		37-39	0.0003935	9.17
8	(Num. of people engaged	1-7	0.00008520	7.01
9	in tertiary industry)*	31-32	0.0002341	5.60
	(tertiary industry dummy)			
	Chi-square (d.f.=9)		6993.9	
	Log likelihood at zero		-8346.7	
	Log likelihood at convergence		-4849.7	
	Likelihood ratio index		0.419	
	Num. of observations		2000	

Table 3 OD volume from observed data

O \ D	1	2	3	4	5	6	7	8	9	10	11	tk.
1	1038	160	220	196	57	313	70	124	27	196	14	2415
2	1173	2083	653	256	146	718	310	222	31	259	28	5879
3	1825	548	1950	347	259	780	233	846	143	594	55	7580
4	1220	277	702	1128	99	615	105	308	142	729	118	5443
5	1155	245	227	204	998	1435	86	13	28	58	69	4518
6	1011	351	228	174	352	5421	176	195	62	117	33	8120
7	1444	1122	507	428	406	881	1439	378	152	446	54	7257
8	1182	372	971	547	167	839	236	2489	283	621	60	7767
9	937	285	513	301	134	498	86	655	1426	869	46	5750
10	1814	450	672	945	335	1000	317	451	503	3568	323	10378
11	364	77	133	183	99	286	14	101	68	346	388	2059
t.l	13163	5970	6776	4709	3052	12786	3072	5782	2865	7803	1188	67166

Table 4 OD volume aggregated with classification method

O \ D	1	2	3	4	5	6	7	8	9	10	11	$t_{k.}$
1	1100	191	366	189	46	276	42	94	18	89	6	2415
2	1783	1351	840	186	92	980	207	240	36	154	9	5879
3	2268	685	1800	437	126	794	170	794	88	400	19	7580
4	1701	284	792	778	176	663	100	300	64	558	27	5443
5	1071	267	437	383	606	1130	91	175	38	282	38	4518
6	1519	649	629	301	324	3877	209	276	64	253	19	8120
7	1682	889	958	306	127	1255	1266	429	67	262	18	7257
8	1758	583	1678	414	140	880	213	1299	209	565	27	7767
9	1090	294	767	280	90	556	108	694	997	831	45	5750
10	1901	465	1182	1024	274	1012	180	674	230	3290	146	10378
11	475	90	193	176	117	287	37	85	41	319	239	2059
$t_{.l}$	16347	5749	9640	4474	2116	11711	2623	5059	1851	7004	592	67166

Table 5 Estimated marginal utilities γ_j

Zone No.	γ_ℓ
1	-1.015
2	-0.760
3	-1.149
4	-0.714
5	-0.389
6	-0.659
7	-0.630
8	-0.635
9	-0.214
10	-0.611
11	0

Table 6 OD volume obtained from the present method (with estimated γ_j)

O \ D	1	2	3	4	5	6	7	8	9	10	11	$t_{k.}$
1	968	216	279	219	74	332	54	117	32	112	14	2415
2	1502	1458	618	207	143	1133	253	289	66	189	21	5879
3	1955	758	1347	494	199	941	213	973	160	495	45	7580
4	1404	299	564	850	269	752	120	350	112	664	59	5443
5	805	259	287	385	842	1193	101	191	63	314	78	4518
6	1170	643	422	307	463	4140	239	305	106	285	40	8120
7	1338	910	666	326	188	1397	1478	493	115	307	39	7257
8	1437	610	1201	448	212	989	256	1520	365	670	60	7767
9	813	281	494	272	124	559	118	726	1416	860	88	5750
10	1445	455	782	1040	389	1071	202	735	370	3590	299	10378
11	327	81	116	161	151	280	38	85	61	318	443	2059
$t_{.l}$	13163	5970	6776	4709	3052	12786	3072	5782	2865	7803	1188	67166

5. The size of a Choice Set

In Section 3 we described the technique of reducing the choice set used in parameter estimation. But the estimators in this method are valid only asymptotically and we have little knowledge of the sub-choice set size which does not affect the accuracy of the parameters. In this section, we discuss this issue by using 500 observations sampled randomly from the complete data in Maebashi city and 40 destinations excluding only one by which no trip in the sample is attracted.

We analyze the property of the estimators with the sampling technique described in eq.(29). Monte Carlo Simulation is carried out to obtain the sub sets of choice alternatives. In eq.(29) fixing b at 1.0 and varying a from 0 to 0.75 (4 cases of a = 0, 0.25, 0.5, 0.75), sample $N \cdot R_{j|i}$ observations for each destination and origin randomly from 500. With this simulation, we can obtain a subset of alternatives for every observation. Trying this process five times for each a, 20 data sets (4 cases and each 5 data sets) are finally obtained and are utilized to estimate parameters. For the comparison we estimate the parameters with the subset sampled by using b fixed at 0.0, and we call this process Random sampling. Now we use the following index as a substitutive value of CPU time,

$$\text{alt.} = \sum_n SD_n \tag{39}$$

and mean value and coefficient of variation of the estimated parameters as,

$$\bar{\theta}_k = \frac{1}{T} \sum_t \hat{\theta}_{kt} \tag{40}$$

$$CV_k = \frac{1}{\bar{\theta}_k} \sqrt{\frac{1}{T} \sum_t (\hat{\theta}_{kt} - \bar{\theta}_k)^2} \tag{41}$$

and the relative error as,

$$ERR_k = |\bar{\theta}_k - \theta_k^0| / \theta_k^0 \tag{42}$$

In above equations, $\hat{\theta}_{kt}$ is an estimated parameter, θ_k^0 is a parameter estimated with the complete choice set, and T(=5) is the number of trials.

Table 7 and 8 show the values of CV_k and ERR_k.

Table 7 Coefficients of variation of each parameter

(R : Random Sampling , w : Weighted Sampling)
(alt.1 : a=0 , alt.2 : a=0.25, alt.3 : a=0.5
alt.4 : a=0.75)

alt.		parameter 1	2	3	4	5	6	7	8	9
1	R	0.031	0.022	1.760	0.185	0.350	0.141	0.284	0.187	0.181
	W	0.015	0.043	0.521	0.026	0.118	0.064	0.061	0.075	0.063
2	R	0.016	0.026	2.448	0.065	0.174	0.060	0.096	0.096	0.056
	W	0.002	0.012	0.058	0.055	0.070	0.022	0.037	0.055	0.027
3	R	0.010	0.015	0.780	0.077	0.048	0.035	0.091	0.067	0.038
	W	0.002	0.005	0.304	0.039	0.050	0.025	0.012	0.021	0.032
4	R	0.008	0.004	0.211	0.017	0.046	0.023	0.043	0.012	0.030
	W	0.002	0.002	0.188	0.009	0.033	0.011	0.015	0.024	0.021

Table 8 Relative Errors of each parameter

(R : Random Sampling , w : Weighted Sampling)

alt.		parameter 1	2	3	4	5	6	7	8	9
1	R	0.121	0.398	0.190	0.072	0.020	0.186	0.159	0.162	0.427
	W	0.106	0.201	0.626	0.074	0.139	0.050	0.085	0.028	0.104
2	R	0.046	0.165	0.450	0.017	0.244	0.011	0.034	0.109	0.119
	W	0.014	0.002	0.497	0.037	0.072	0.006	0.034	0.021	0.035
3	R	0.014	0.056	1.274	0.018	0.064	0.042	0.059	0.021	0.066
	W	0.001	0.003	0.165	0.023	0.010	0.006	0.010	0.017	0.004
4	R	0.014	0.021	0.118	0.002	0.003	0.003	0.034	0.023	0.036
	W	0.001	0.001	0.101	0.003	0.018	0.001	0.009	0.010	0.009

While we should note that the value of alt. is slightly different between the cases of b = 1 (Weighted sampling) and b = 0 (Random sampling) for the same category in these tables, but we can say that the Random sampling gives generally worse estimation. Further, this tendency is clearer with samller value of alt.. In Fig. 2 and 3, CV_m and ERR_m are mean values of CV_k and ERR_k as,

$$CV_m = \frac{1}{K}\sum_k^K CV_k \tag{43}$$

$$ERR_m = \frac{1}{K}\sum_k^K ERR_k \tag{44}$$

where K = 9 is the number of parameters.

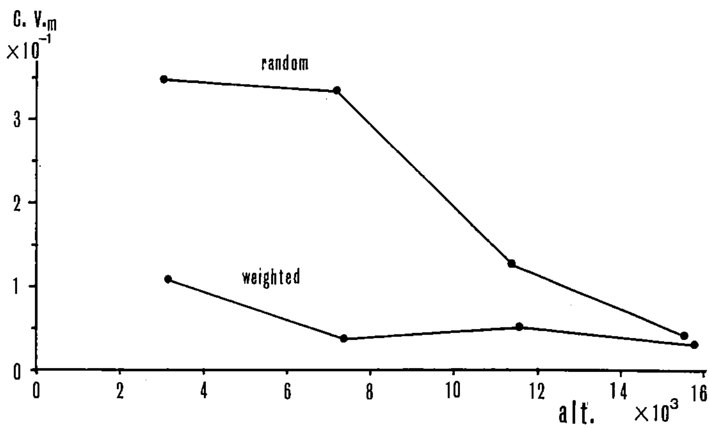

Fig. 2 Mean value of coefficients of variation and the number of choice-alternatives (alt.)

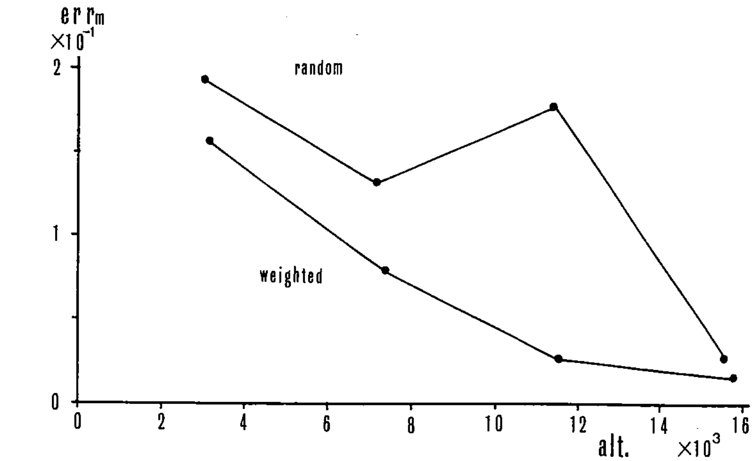

Fig. 3 Mean value of Relative Errors and the number of choice-alternatives (alt.)

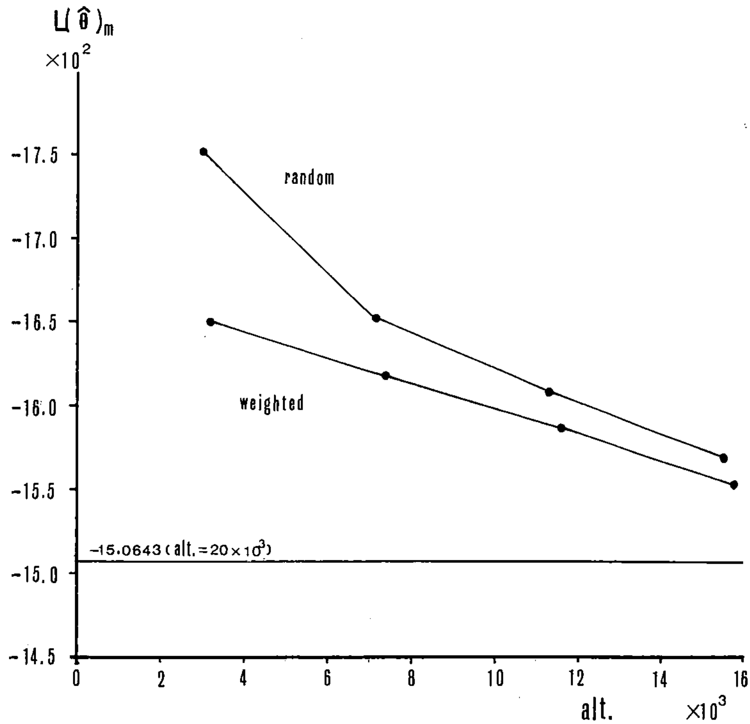

Fig. 4 Mean value of Log-likelihood and the number of choice-alternatives (alt.)

About the Weighted sampling, CV_m is less than 0.1 for each value of alt. while about the Random sampling, CV_m is greater than 0.3 when alt is small. And also the Weighted sampling has smaller values of ERR_m than those of Random Sampling for every value of alt.

Further, we show, in Fig. 4, the expected value of Log-likelihood $L(\hat{\theta})_m$ as

$$L(\hat{\theta})_m = \frac{1}{T} \sum_t L(\hat{\theta}_t) \tag{45}$$

$$L(\hat{\theta}_t) = \sum_n \sum_{j \in D} \delta_{jn} \ln P_{jn}(\hat{\theta}_t) \tag{46}$$

$$P_{jn}(\hat{\theta}_t) = \frac{e^{V_{jn}}}{\sum_{j' \in D} e^{V_{j'n}}} \tag{47}$$

$$V_{jn} = \sum_k \hat{\theta}_{kt} X_{jkn} \tag{48}$$

In the figure, $L(\hat{\theta})_m$ and alt. seem to have some linear correlation, and also here, Weighted Sampling has a definite advantage especially when alt is small.

From the above, we can say that the Weighted sampling (by observed share) gives better estimates than the Random Sampling, and also it is the result of Weighted sampling's better function of reducing the amount of information than the Random sampling's. The former reduces information according to observed share, and the latter reduces uniformly without the knowledge of share.

Judging from the CV_k and ERR_k, parameter estimation has adequate stability even when a = 0.25, i.e. we can reduce the number of alternatives to about 37% of the complete set. And similar analysis about the data set of 2.000 trips shows that the Weighted sampling gives adequately stable estimates even when a = 0.0.

In summary, we have found a way to reduce considerably the time for parameter estimation.

6. Sample Size needed to Predict OD Tables

In this Section, we discuss how much observations are required to calibrate desaggregate destination choice models and to predict OD tables with the models. Although many studies about the sample size problem are restricted to of the models with the small choice set, we discuss the case with large choice set of which the number is 41, investigating the stability of estimators and the prediction errors of trip distributions resulting from aggregation.

First , sampling randomly from the total observations of 4,723, we generate 10 data sets for each size (from 250 to 2,000 observations).

Second , we estimate the model parameters for each data set and aggregate the model to forecast OD table; the aggregation method proposed in Section 4 is utilized.

Figure 5 shows the coefficients of variation of estimated parameters against each sample size. The numbers 1-9 in the figure correspond to parameters in Table 2. The generic variables of inclusive price (1) and trip attraction (2) have sufficiently small values of less than 0.1 when the sample size is greater than 750 for the former and 500 for the latter. As for alternative specific variables 3-9, the CV of a variable changes according to which alternatives are specified, and the CV of the variables 4, 6, 7 are relatively stable against the changes of sample size, while those of 3, 5, 9 are greater than 0.1 even when sample size is equal to 2,000.

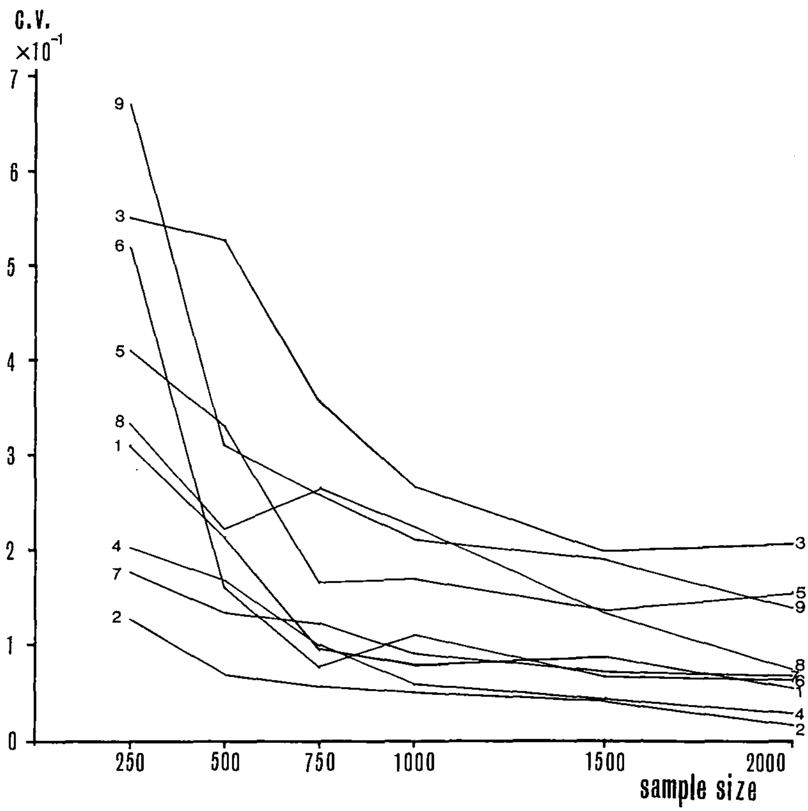

Fig. 5 Coefficient of variation of each parameter and calibration sample size

Further, we examine the difference between observed trip distribution and the estimated by aggregating the model. For aggregation, we took the method which we proposed in Section 4, so the restriction of trip attraction is satisfied by the introduction of γ_j. We define error indices in 2 directions of column and row of OD table. In column (generation) direction, we define,

$$mae = \frac{1}{ND} \sum_{i \in D} ae_i \tag{49}$$

$$ae_i = \sum_{j \in D} |s_{j|i} - \hat{s}_{j|i}| \tag{50}$$

In row (attraction) direction, we define,

$$mae = \frac{1}{ND} \sum_{j \in D} ae_j \tag{51}$$

$$ae_j = \sum_{i \in D} t_{i.} |s_{j|i}/t_{.j} - \hat{s}_{j|i} / \sum_{i \in D} t_{i.} \hat{s}_{j|i}| \tag{52}$$

where ND (=11) is the number of B-zones.

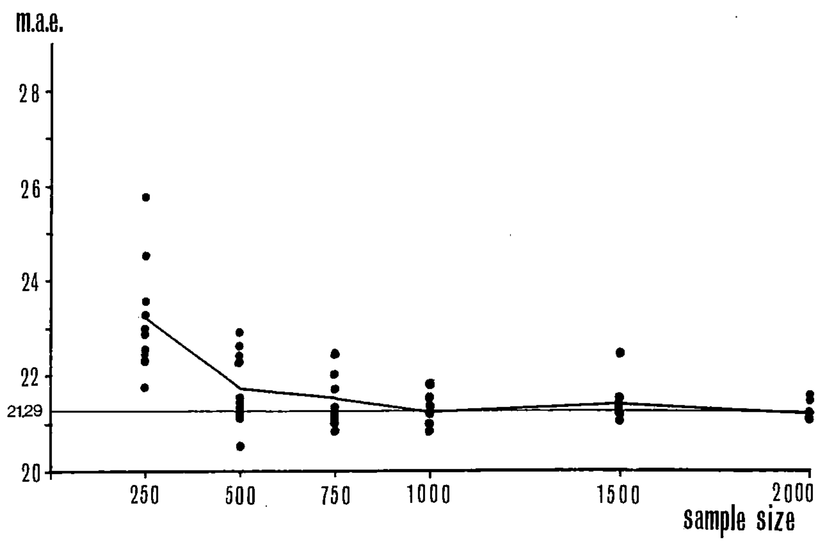

Fig. 6 Mean value of the absolute errors of estimated OD tables (generation side) and calibration sample size

Figures 6 and 7 show the values of these indices against the sample size. In these figures, line-graphs show expected values of 10 trials, and the decimals 0.2129 in figure 6 and 0.2226 in figure 7 are the mae of the estimation with full sample.

From these figures, we can say as following. The dispersion and the mean of mae get small as the sample size become larger. And above 1000 observations, the dispersion changes little against the changes of sample size, and the values of mae are little different from those with full sample in this range of the size. Therefore , from the result of estimated trip distribution, we can conclude that practically the same result with full sample is obtained by more than 1,000 trips sampled.

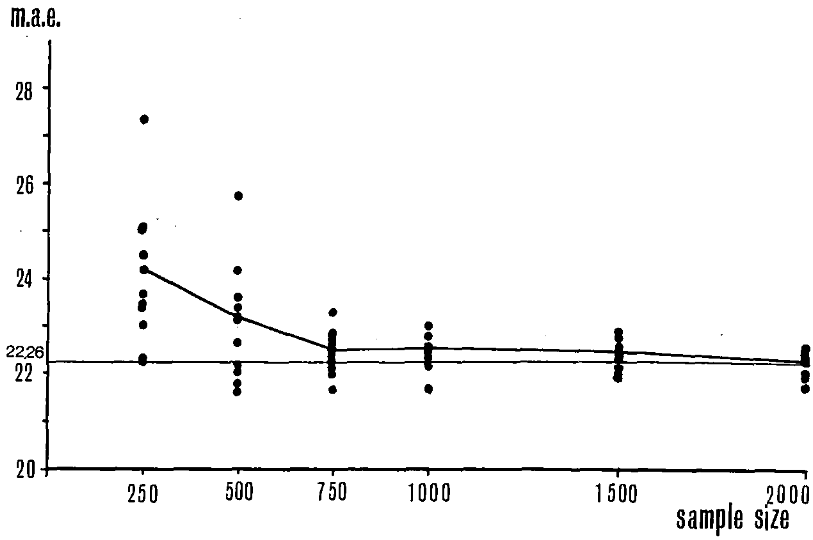

Fig. 7 Mean value of the absolute errors of estimated OD tables (attraction side) and calibration sample size

Figure 8 shows the coefficients of variation of γ_j in equation (19). In the figure, the numbers affixed to lines correspond to zones, and γ_j of zone 11 is fixed at 0.0 as in Section 4. At the sample size of 1000 trips, γ_j except of zones 6, 7, 9 are stably estimated to have the values of CV of less than 0.1. At this sample size, CV of zone 6, 7 are less than 0.2, while that of zone 9 is no less than 0.4 with no improvement within sample size of 2,000 trips. One of the reasons of the unstability of zone 9 is that the zone has a very little attraction.

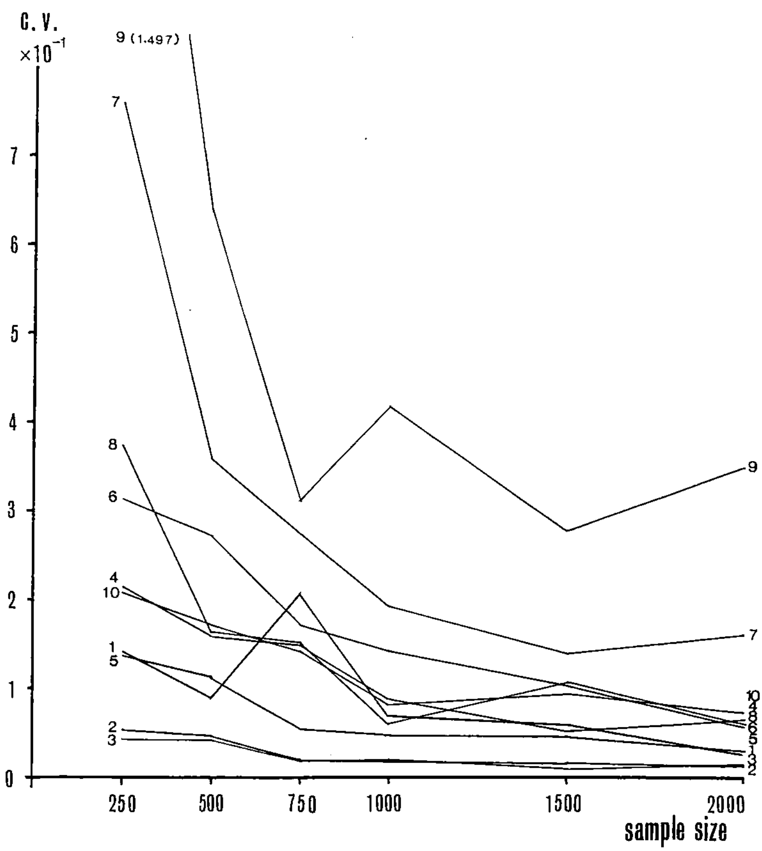

Fig. 8 Coefficients of variation of γ_j and calibration sample size

From the above examination, it has become clear that the necessary sample size for the caliblation of disaggregate destination choice model is 1000 trips from the point of view of the prediction of OD tables. But we should notice that from the point of view of the stability of parameters, 1000 trips is not sufficient for some parameters, while about the generic variables as the principal variables, 1000 trips is proved to give adequately stable estimates.

7. Conclusion

This paper has considered the problems of predicting trip distributions with disaggregate destination choice models. A new methodology to forecast OD table is developed. And some important issues are discussed to improve the practicality of destination choice models. The results of this study are summarized as follows.

(1) The aggregation scheme for producing OD tables with disaggregate models is developed in Section 2. It is now possible to introduce total control associated with trip volumes by modification of the utility functions of disaggregate models. An example of this application to a local area in Japan is focused in order to evaluate some of the features of the methodology in Section 4.

(2) In the case where the size of choice-set is very large, comparison of the sampling methods of choice alternatives which reduces the time for estimation of parameters are carried out in Section 3 and 5. The results indicate that to what degree the number of alternatives is reduced should be determined according to the sampling methods and the sample size.

(3) The sample size enough to calibrate destination choice model and to predict OD tables is discussed in Section 6. We found that it is not necessary to use much data (more than 1.000 observations), when we evaluate the prediction errors of OD tables. However, it requires the larger samples to satisfy the statistical reliability of all parameters.

REFERENCES

1. Ben-Akiva, M. (1973), "Structure of Passenger Travel Demand Models", Ph.D. Dissertation, Department of Civil Engineering, MIT.

2. McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior, in Zarembka, P. (ed.), Frontiers in Econometrics, Academic Press, New York.

3. McFadden, D. (1978), "Modelling the Choice of Residential Location, in Karlquvist", A. etal (ed.), Spatial Interaction Theory and Planning Models, North-Holland, Amsterdam.

4. Ben-Akiva, M. & S. Lerman, (1979), "Disaggregate Travel and Mobility-Choice Models and Measures of Accessibility", in Hensher, D. etal (ed.), Behavioural Travel Modelling, Croom Helm, London, pp. 279-318.

5. Daganzo, C. (1979), "Multinomial Probit : "The Theory and Its Application to Demand Forecasting", Academic Press.

6. Williams, H.D.W.L. (1977), "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", Environment and Planning, Vol. 9A, pp. 285-344.

7. Ben-Akiva, M., Gunn, H. and L. Silman, (1984), "Disaggregate Trip distribution Models", Proceedings of JSCE, Vol. 347/IV-1, pp. 1-17.

8. Morichi, S. and T. Yai, (1984), "Application of Disaggregate Travel Demand Model and Choice-based Sampling for Infrequent Trips", Proceedings of JSCE, Vol. 343, pp. 161-170. (in Japanese)

9. Morichi, S. and T. Yai, (1985), "Disaggregate Modelling Techniques of Predicting Trip-Distribution", Infrastructure Planning Review, No. 2, pp. 45-52, Japan, January. (in Japanese)

10. Morichi, S., Ishida, H. and T. Yai, (1984), "Comparison of Various Utility Functions for Behavioral Travel Demand Model", Proceedings of the World Conference on Transport Research, SNV, Hamburg.

11. T. Yai, (1985), "Travel Demand Forecasting Techniques with Disaggregate Behavioral Models", D.Eng. Dissertation of Tokyo Institute of Technology, January. (in Japanese)