USE OF ON-OFF COUNTS FOR ORIGIN-DESTINATION MATRICES ESTIMATION AN APPROACH TOWARDS MORE COST-EFFECTIVE BUS SURVEY*

Terdsak RONGVIRIYAPANICH**, Fumihiko NAKAMURA*** and Izumi OKURA****

1. Introduction

Bus operators, particularly in developed countries, have witnessed the declining number of patronage due to the rapid motorization in past decades. In Japan, despite of its advanced rail-based transit system, patronage of the bus system has been decreasing continually. More efficient operation is therefore required in order to make bus transit as much attractive as possible. It is clear that more up-to-date and in-depth data is important to the operation planning. Conventional on-board survey, which is conducted in Tokyo Metropolitan area every five year to estimate the route OD matrix by employing two surveyors on each bus for a whole weekday, is rather labor and capital intensive. In addition, this estimation scheme cannot deal with day-to-day as well as seasonal variation, which is a nature of the OD matrices.

Recent attempts by some local bus operators, for instance in Tokyo or Yokohama, to develop advanced management system such as automatic vehicle location system, automatic passenger counter system have helped broaden the database. This system offers us an abundant source of aggregate data, which cannot be used alone to estimate OD matrix, yet can capture the variation in travel demand. However, lack of capable data analysis framework has impeded the integration of such automatically collected information with the survey data. As a result, it is of interest to address the following questions. First is "can data obtained automatically through the bus data collection system (BDCS) be utilized in bus planning procedures?". Another question is "is it possible to replace the costly conventional OD survey by use of the BDCS data in conjunction with other readily available data?". These two questions are among motivations of this study.

OD matrix at the route level, as suggested by several researchers, is an important information for bus service planning and management[1]. Its applications are for instance, patronage forecasting or predicting the effect of change in level of service. Ben-Akiva et al.[2], among others, recommended the use of on-board survey in conjunction with ride-check data as an alternative method to estimate route-level OD matrix. It is revealed in their study that the simple expansion of on-board survey gives less accurate result than that obtained by the intervening opportunity method. However, comparisons of performances of the methods are still limited and mostly based on simulated data or test networks.

2. Objectives

This paper is aimed to evaluate performance of the methodologies to estimate OD matrices by applying them to two bus routes in Tokyo Metropolitan. Thanks to the availability of actual daily OD matrices of two years for the two bus routes, comparison of the models in terms of accuracy of the estimates for large-scale data is made possible. In spite of obvious usefulness of on-off counts for estimating OD matrices, in practice the operator has still relied solely on OD matrices obtained by the conventional survey. In this paper, we therefore attempt to show that by using the available information efficiently, the expenses for service planning can be considerably cut. Effects of the sample size on accuracy of the estimates are also investigated. Based on these results, potential rooms for further improvement of the estimation technique can be discussed.

3. OD estimation

Estimation problems for highway traffic OD and route-level bus OD are different in sense that the latter does not require any assumption on the route choice probability. Thus it could be estimated in more simple and accurate manner. To our knowledge, the most widely used methods to estimate OD matrices from traffic counts can be classified into two major families[3]. First, the entropy maximization method (ME), or sometimes known as the iterative proportional fitting method (IPF), is based on the concept of information minimization. The other family of OD estimator is known as the statistical inference method, including the maximum likelihood method (ML), the general least square method (GLS) and the Bayesian method. It is also possible to estimate OD matrices when ride-check data alone is available, by using the intervening opportunity method (IO). In general,

* Keywords: Public transportation demand

** M.Eng., Graduate Student, Department of Civil Engineering, Yokohama National University,
(79-5 Tokiwadai, Hodogaya-ku, Yokohama, Japan, TEL 045-339-4039, FAX 045-331-1707)

*** D.Eng., Associate Professor, Department of Civil Engineering, Yokohama National University,
(79-5 Tokiwadai, Hodogaya-ku, Yokohama, Japan, TEL 045-339-4033, FAX 045-339-4033)

**** D.Eng., Professor, Department of Civil Engineering, Yokohama National University,
(79-5 Tokiwadai, Hodogaya-ku, Yokohama, Japan, TEL 045-339-4032, FAX 045-331-1707)

the OD estimation problem can be formulated as follows[3)]

$$\min f(T,N)$$

subject to $\sum_{j} t_{ij} = a_i$ and $\sum_{i} t_{ij} = b_j$

where t_{ij} is the number of passengers boarding at i and alighting at j, while a_i, b_j are the passenger count constraints on origin i and destination j respectively. N is an outdated OD matrix, usually obtained by on-board survey. Form of the objective function, $f(T,N)$, varies with the statistical distribution assumed for the sample, *N*. For the sake of simplicity, multinomial and Poisson distribution are usually assumed, although multivariate normal distribution is also adopted sometimes. As a result, computation algorithm for the OD estimation varies with the assumed distribution.

Table1: Characteristics of the OD estimation techniques

Method	Data required		Assumption
	OD survey	Ride checks	
IPF	y/n*	ns**	Maximum entropy (Information minimization) concept
ML	ns**	ns**	Sampling distribution of the base year OD is either Poisson or multinomial
GLS	ns**	ns**	Base year OD matrix is the unbiased mean of the true OD
IO	-***	ns**	Equal alighting probability for every on-board passenger

* necessity depends on the approach taken, ** necessary and *** not required

In their study, Ben-Akiva et al.[2)] compared the results, which are obtained from the above mentioned methodologies, by using the IPF estimators as a basis. Their conclusion is that, among the approaches included in the study, the iterative proportional fitting (IPF) method is the most effective method if an on-board survey and deterministic ride-check data are to be used. However, drawback of this technique has long been recognized by M.G.H. Bell[4)], namely a resulting OD matrix, obtained from the method, is not invariant to the application of uniform scaling to the prior estimates. Hence application of this technique is limited to only a corridor, in which there is no significant growth/decrease of the travel demand. Moreover, in the IPF method the underlying assumption of deterministic constraints can not be relaxed[2)].

4. Data and methodology

Data, utilized in this study, is available from routine annual bus surveys of the Tokyo Metropolitan bus operator. It is a main purpose of this study that data to be used should be readily available. The five-year bus survey provides us an actual daily OD matrix of that year. In this study we use the OD matrices in 1991 and 1996 of two bus routes in the Tokyo Metropolitan, namely number 1 and 8 respectively. The daily OD matrices in 1991 are used as base year matrices, while the daily OD matrices in 1996 are used as the target matrices for estimations. Figure 1 and 2 show the share of cells with different number of trips in the OD matrices in 1991 of route 1 and 8 respectively.

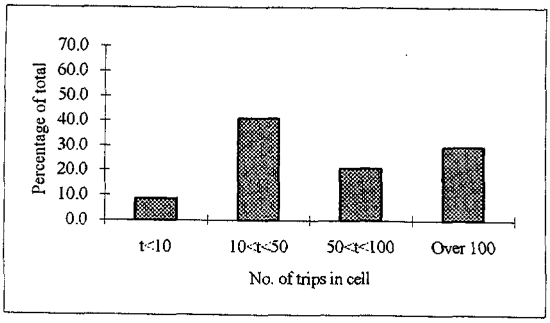

Fig. 1: Classification of cells in the OD matrix in 1991 of route 1based on the number of trips

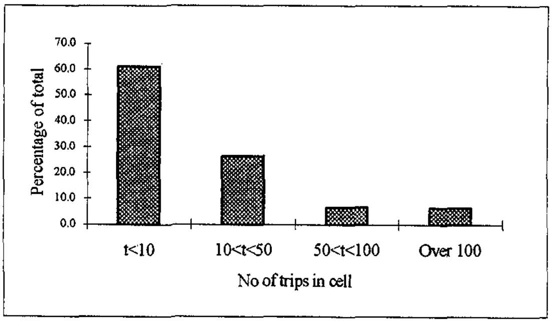

Fig. 2: Classification of cells in the OD matrix in 1991 of route 8 based on the number of trips

As shown above in figure 1 and 2, cells in the OD matrices are classified into four groups according to the number of trips. It is apparent that structure of the two matrices is considerably different. The OD matrix of route 1 is dominated by cells with considerable number of trips, while that of route 8 mainly comprises of cells with few trips. We shall discuss later how this affects the OD estimation. In this study, three estimators namely the IPF, ML and IO estimators, obtained by using the based year OD matrix in 1991 in conjunction with the number of generating and attracting trips by each stop in 1996 as constraints, are tested against the actual OD matrix in 1996. It should be noted here that as an initial stage of the study, we neglect variation of the OD so that the available OD matrices in 1996 can be treated as the actual OD matrices. In addition, effects of sample size on the accuracy of the estimates are also examined by randomly sampling the base year OD at different sizes of 20, 25, 30, 40 and 50 percent of the original size respectively with 100 repetitions.

5. Results

In order to compare the performance of different models and OD matrices, we use the statistical measures shown below.

Root-mean square error (RMSE) $= \sqrt{\frac{1}{N}\sum_{ij}(\hat{t}_{ij} - t_{ij})^2}$

Root-mean square difference (RMSD) $= \sqrt{(1/K)\sum_{ij}\left((\hat{t}_{ij} - t_{ij})/t_{..}\right)^2}$

Average weighted fractional error (AWFE) $= \sqrt{\sum_{ij}(t_{ij}/t_{..})\left((\hat{t}_{ij} - t_{ij})/t_{ij}\right)^2}$

Where $\hat{t}_{ij}$ and t_{ij} are the estimated and actual number of passengers for particular OD pair. $t_{..}$ is the total number of trips in the actual OD matrix. N is the number of cells in the actual matrix, while K is the number of nonempty cells in the estimated matrix.

Table 2: Statistical measures of the OD estimates of route 1

Technique	RMSE	RMSD	AWFE
IPF	35.43	10.4E-4	0.182
ML	35.40	10.4E-4	0.183
IO	95.14	27.9E-4	1.333

Table 3: Statistical measures of the OD estimates of route 8

Technique	RMSE	RMSD	AWFE
IPF	8.89	7.5E-4	0.369
ML	9.36	7.9E-4	0.394
IO	15.61	12.7E-4	0.829

Table 4: Statistical measures of the IPF estimates of route 1 based on different sampling ratios.

Sampling Ratio (%)	RMSE	RMSD	AWFE
20	40.22	12.1E-4	0.238
25	39.89	11.9E-4	0.229
30	39.14	11.7E-4	0.222
40	38.57	11.4E-4	0.212
50	37.53	11.1E-4	0.206
100	35.43	10.4E-4	0.182

Table 5: Statistical measures of the IPF estimates of route 8 based on different sampling ratios.

Sampling Ratio (%)	RMSE	RMSD	AWFE
20	13.32	13.4E-4	0.612
25	12.36	12.1E-4	0.572
30	11.95	11.4E-4	0.544
40	11.36	10.5E-4	0.506
50	10.73	9.7E-4	0.478
100	8.89	7.5E-4	0.369

Based on the estimation results, comparison of the performance of the models, as expressed by the statistical measures, can be summarized in table 2 and 3. As we expected, the IPF and ML methods perform equally well even with the big sample data due to its similarity in mathematical formation. Since it has been concluded in the study by Ben-akiva et al.[2)] that the results obtained by the IPF and ML methods are comparable when the sample size is small, we shall generalize here that one may choose either IPF or ML technique to estimate the OD matrix. However, due to its computational simplicity, the IPF technique appears to be more preferable to the ML method, if the assumption of deterministic constraints on on-off counts is taken. Given that outdate OD matrix or on-board sample is available, there should be no reasons to use the IO method for OD estimation, as its estimates are clearly inferior to the methods, which utilize information of the priori probability.

In table 4 and 5, we show that by using the IPF technique based on on-off counts and sufficient OD sample we can obtain the result within acceptable region of confidence. As shown by the statistics, sampling ratio of 50% or so, when used with the on-off counts already gives adequately desirable results. Comparing the results obtained from route 1 and 8, we may see that the effects of sample size on the accuracy are more visible for the case of route 8, as reducing size of the sample by half from

100% to 50% of the original size results in 10% and 30% more error of the estimates for route 1 and 8 respectively. This is attributed to the property of IPF technique namely empty cells of the base year matrix always results in zero estimates. Hence optimal sample size is in turn dependent on the pattern of OD matrix. In general, sufficiently large sample size would be necessary for the matrix with relatively small number of trips in each cell like the OD matrix of route 8 in this study, as shown in figure 2, so that small but nonempty cells will not be missed out from the sampling.

6. Conclusion

This paper shows some empirical evidences on the performance of different OD estimation techniques based on on-off counts and the outdated OD matrices. The IPF method is proved to be the most practical technique for the OD estimation problem under the assumption of deterministic count constraints due to its computational simplicity and accuracy. Effects of sample size on the accuracy tend to be trivial given that most of the nonempty cells are sampled out. Decision on the approximate sample size has to be done on case by case basis. Nevertheless, at this point it is clear that given the availability of the on-off counts, the conventional daily OD survey can be replaced by an OD sample without losing much accuracy. Thus, expenses required for the service planning could be reduced in some extent.

At the next stage of this research, the generalized least square method (GLS) is expected to be investigated and compared to the above approaches. Variances of the estimates will also be computed, if feasible, or estimated so that it can be used as another criteria for evaluating the accuracy of each method. At the present, we have treated the daily OD survey as the actual OD matrices, hence neglecting inherent day-to-day variation nature of the OD. It is thus important to take into account of the dynamics of OD into our future study. This is an aspect, which can not be dealt with when the survey data alone is utilized. In addition, performance of the GLS and ML methods should be examined under the assumption of probabilistic count constraints.

Here let us mention some important aspects, which have not been touched on in this research. First as pointed out by, Jornsten and Wallace[5), it should be important in practice to know quantitatively how the estimates improve by an additional information. This could be used in a design of additional survey. Secondly, due to the expected variation of demand and OD pattern between peak and off-peak periods, OD matrix estimation of these two periods should be carried out separately. This would be useful for service planning to ensure that the level of service is responsive to variations of demand within day. Lastly, more robust estimation technique, in which information from various sources is utilized, is necessary so that changes in economic or demographic conditions of the system can be taken into account. The study by Carey, Hendrickson and Siddharthan[6) is among examples of effort to combine a direct demand model with the OD estimation.

7. References

1) Fielding P.: Obtaining Bus Transit Service Consumption Information through Sampling Procedures, in New Survey Methods in Transportation; 2nd international conference, E.S. Ampt, A.S. Richardson and W. Brog, Utrecht, the Netherlands, pp. 227-239, 1983.

2) Ben-Akiva, M.E., Macke, P.P. and Hsu P.S.: Alternative Methods to Estimate Route-Level Trip Tables and Expand On-Board Surveys, Transportation Research Record 1037, TRB, pp. 1-11, 1985.

3) Cascetta E. and Nguyen S.: A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts, Transportation Research Vol. 22(B), pp. 437-455, 1988.

4) Bell, M.G.H.: The Estimation of an Origin-Destination from Traffic Counts, Transportation Science Vol. 17, pp. 199-217, 1983.

5) Jornsten K. and Wallace, S. W.: Overcoming the Apparent Problem of Inconsistency in Origin-Destination Matrix Estimations, Transportation Science Vol. 27, pp. 374-380, 1993.

6) Carey M., Hendrickson C. and Siddharthan K.: A Method for Direct Estimation of Origin/Destination Trip Matrices, Transportation Science Vol. 15, pp. 32-49, 1981.