Application of an Algorithm for Estimating Freeway Trip Tables

ROBERT W. STOKES and DANIEL E. MORRIS

ABSTRACT

The application of an existing computer program that uses only the marginal totals of a freeway trip table to estimate ramp-to-ramp traffic flows is described. The flows estimated by this program [synthetic origin-destination matrices (SYNODM)] were compared with origin-destination data obtained from a postcard survey and the distribution and magnitude of the estimation errors across the trip table were documented. A modified version of the basic SYNODM algorithm that incorporates limited origin-destination data into the estimation method is also discussed. Relative to the postcard data, the estimates of trip interchanges obtained from the modified version of SYNODM exhibited a substantially smaller average trip error than those obtained from the basic SYNODM program. However, given the limited scope of the study, no generalizations regarding the adequacy of the basic SYNODM program could be made.

A large variety of freeway planning and management activities require information about the pattern of vehicular traffic flows between entry and exit ramps. Traditionally the collection of such information has involved elaborate, time-consuming, and expensive special-purpose surveys such as roadside interviews. It has been suggested that instead of special-purpose surveys being relied on to gather this information, more readily available information such as mainline and ramp traffic volumes could be used to estimate traffic distribution patterns (1).

Reliable procedures for synthesizing trip distribution patterns offer a number of advantages over strictly empirical approaches. First, synthetic trip distribution data would certainly be less expensive and more easily obtained than data obtained from field studies. Second, the use of available (or in any case easily obtainable) information on flows on the system could greatly reduce the lead time between the data collection and model calibration phases of the planning process. Third, the advantages of low cost and quick turnaround time could lead to broader applications of analytic procedures that, because of their extensive data requirements, have in the past been restricted to use in large-scale studies. Finally, reliable procedures for estimating freeway traffic distributions from readily available, longitudinal data, such as observed flows, could alleviate many of the problems attributable to the cross-sectional nature of empirically derived origin-destination (OD) data.

The objective of this research was to evaluate an existing algorithm for estimating freeway traffic distributions and assess the extent to which the algorithm could be used in lieu of more traditional empirical approaches.

BACKGROUND

In recent years the Texas State Department of Highways and Public Transportation and the Texas Transportation Institute have made extensive use of the FREQ simulation programs (2) in evaluating proposed design configurations and control strategies for freeways in Houston and San Antonio (3-6). One of the basic data inputs required by the FREQ programs is an OD table for each of the time periods being simulated. Fortunately, much of the necessary OD data had been collected as part of earlier planning studies. However, as use of the FREQ programs was extended to more and more freeways it became apparent that more efficient and economical procedures for generating reliable OD data were needed.

A preliminary literature review revealed several OD estimation methods. Nihan (7), for example, has presented a summary of several proposed estimation procedures and their limitations. The computer program available to users of the FREQ simulation programs, synthetic origin-destination matrices (SYNODM) (8), was chosen for preliminary evaluation. SYNODM was designed to be internally compatible with the FREQ simulation programs and therefore represented a logical starting point in the search for an acceptable OD estimation method. The basics of the SYNODM algorithm are presented in the following.

DESCRIPTION OF THE ALGORITHM

Basic Approach

Consider the case of a complete OD matrix (i.e., a matrix in which all cells may have nonzero entries). For such a matrix, with r on ramps and c off ramps, estimates of trip interchanges could be obtained from (1)

$$T^*_{ij} = (O_{i.} D_{.j})/S \qquad (1)$$

where

T^*_{ij} = estimate of trips from origin i to destination j,
$O_{i.}$ = total entering volume for origin i,
$D_{.j}$ = total exiting volume for destination j, and

$$S = \sum_{i=1}^{r} O_{i.} = \sum_{j=1}^{c} D_{.j}.$$

Equation 1 is, of course, the familiar maximum-likelihood estimate (MLE) of individual cell values for statistically independent objects in a two-way contingency table (see Table 1). Although this formulation does not account for any observed regularities in travel behavior nor take into consideration the spatial separation between origins and destinations, it does provide a convenient measure for assessing the precision of the estimation method and the adequacy of the underlying assumption of statistical independence of origins and destinations (1,8). Specifically, under the assumption of independence, the statistic

TABLE 1 Typical Complete OD Matrix (Two-Way r x c Contingency Table)

Origin \ Destination	1	2	...	j	...	c	Total
1	T_{11}	T_{12}	...	T_{1j}	...	T_{1c}	$O_{1.}$
2	T_{21}	T_{22}	...	T_{2j}	...	T_{2c}	$O_{2.}$
⋮							⋮
i	T_{i1}	T_{i2}	...	T_{ij}	...	T_{ic}	$O_{i.}$
⋮							⋮
r	T_{r1}	T_{r2}	...	T_{rj}	...	T_{rc}	$O_{r.}$
Total	$D_{.1}$	$D_{.2}$	...	$D_{.j}$	...	$D_{.c}$	S

$$\sum_{i=1}^{r} \sum_{j=1}^{c} [(T_{ij} - T^*_{ij})^2/T^*_{ij}] \qquad (2)$$

where T_{ij} is the observed flow from origin i to destination j, follows a chi-square distribution with (r - 1) (c - 1) degrees of freedom (1).

The fundamental weakness in Equation 1 is that OD flows are estimated using only the total trips for each origin and destination. Consequently, for two OD matrices that may have completely different cell values but the same trip totals for each origin and destination, Equation 1 would produce two identical matrices. Intuitively, then, one would expect sizeable errors in the T^*_{ij}'s obtained from Equation 1.

Estimation Procedure

The estimation method given by Equation 1, though appealing in its simplicity, is not, unfortunately, directly applicable to simple linear systems with flow in one direction. For a typical freeway OD matrix for flow in one direction, such as that shown in Table 2, the basic contingency-table approach to estimating the T_{ij}'s requires modification. The need for this modification stems from the fact that some trip interchanges are not permissible (i.e., the resulting OD matrix is not complete). The matrix resulting from such an arrangement, then, is roughly upper triangular. The structure of such a matrix also implies that for those freeway sections where each on ramp is accompanied by an off ramp a few hundred yards downstream, few vehicles will enter the freeway just to exit immediately downstream (1).

TABLE 2 Typical OD Matrix for Freeway with Flow in One Direction

On Ramp \ Off Ramp	1	2	...	n-2	n-1	n	Total
1	▨	T_{12}	...	$T_{1,n-2}$	$T_{1,n-1}$	T_{1n}	$O_{1.}$
2	▨	▨	...	$T_{2,n-2}$	$T_{2,n-1}$	T_{2n}	$O_{2.}$
.	▨	▨	▨				
.	▨	▨	▨	...	...	...	...
.	▨	▨	▨				
n-2	▨	▨	▨	▨	$T_{n-2,n-1}$	$T_{n-2,n}$	$O_{(n-2).}$
n-1	▨	▨	▨	▨	▨	$T_{n-1,n}$	$O_{(n-1).}$
n	▨	▨	▨	▨	▨	▨	
Total		$D_{.2}$	...	$D_{.(n-2)}$	$D_{.(n-1)}$	$D_{.n}$	

Despite these minor complications, a solution to the matrix remains fairly straightforward. For example, for the matrix shown in Table 2, it is clear that $T^*_{12} = D_{.2}$ and $T^*_{n-1,n} = O_{(n-1).}$ Once T^*_{12} and $T^*_{n-1,n}$ have been found the corresponding row and column sums can be adjusted and a solution sought for the reduced matrix. Following this approach, $T^*_{n-2,n-1}$ and $T^*_{n-2,n}$ can be obtained from

$$T^*_{n-2,n-1} = [O_{(n-2).} D_{.(n-1)}]/[D_{.(n-1)} + D'_{.n}] \qquad (3)$$

$$T^*_{n-2,n} = [O_{(n-2).} D'_{.n}]/[D_{.(n-1)} + D'_{.n}] \qquad (4)$$

where $D'_{.n} = D_{.n} - T^*_{n-1,n}$. Except for minor deviations, Equations 3 and 4 follow the notation and formulation given by Hauer and Shin (1). By subtracting the T^*_{ij}'s for row n - 2 from the column sums $D_{.(n-1)}$ and $D'_{.n}$, one can proceed to obtain the T^*_{ij}'s for row n - 3 in the same fashion (1). Note in Equations 3 and 4 that only the second term in the numerator varies across the row. Consequently, the ratio of the current row sum to the current destination grand total can be stored in calculator memory and used as a distribution factor to calculate the T^*_{ij}'s from the corresponding current column sums. It is precisely these transition probabilities (8) that SYNODM uses to apportion entry volumes among the downstream ramps.

Application of the algorithm, then, involves the repetition of two basic steps. In the first step, the T^*_{ij}'s in the current bottom row of the matrix are obtained by multiplying the corresponding current row and column totals and dividing by the current destination grand total. In the second step, the bottom row with the newly obtained estimates is deleted from the matrix and the affected column totals are adjusted accordingly (1).

To summarize, SYNODM distributes the total trips of each origin by working from upstream destinations to downstream destinations using the currently unassigned trips at upstream origins to satisfy the current destination totals (8). That is, total entry volumes are apportioned to downstream ramps in accord with flows leaving the downstream ramps (1,8). Basically, SYNODM estimates the T_{ij}'s by solving a portion of a complete matrix. For those situations where $\Sigma O_{i.}$ does not equal $\Sigma D_{.j}$, SYNODM allows the user to specify whether the distribution matrix is to be balanced with respect to input or output volumes (8).

Although the SYNODM computer program does not (in its present form) permit the user to specify any known T_{ij}'s, such as might have been obtained from a limited field survey, it is possible to manually incorporate such information into the estimation method. Known T_{ij}'s can be incorporated into the estimation method by simply adjusting the appropriate current row and column totals before applying the algorithm. A manual application of the basic SYNODM algorithm that incorporates limited OD data into the estimation method is presented later in the paper.

APPLICATION

Field Data Collection

OD data for the Katy Freeway (I-10W) in Houston, Texas (see Figure 1), were used to evaluate the SYNODM algorithm. The OD data for the Katy Freeway (9) were collected by the driver postcard survey

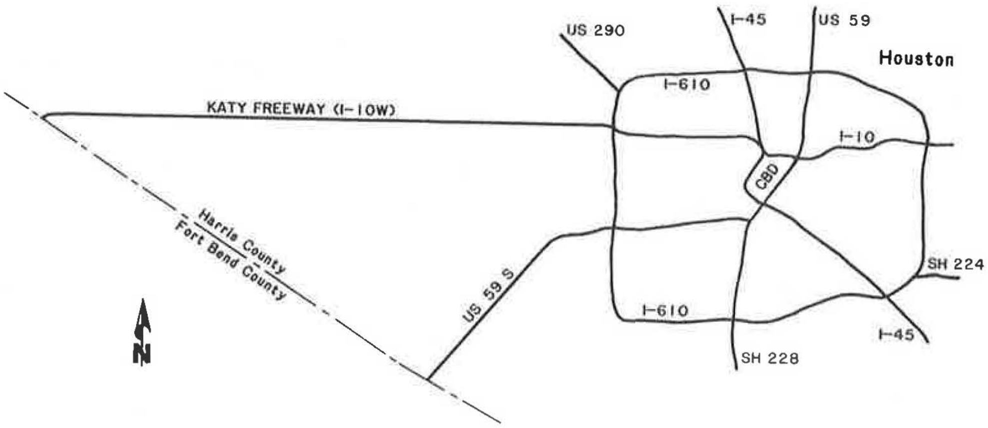

FIGURE 1 Freeways in the Houston region.

method. The survey was designed to sample 50 percent of entrance-ramp traffic for the period 6:00 a.m. to 6:30 p.m. The ramp survey stations from which the samples were drawn are shown in Figure 2.

To estimate total traffic volumes and OD distributions from the sample data, the sample percentages of ramp-to-ramp flows were applied to automatic counts of entrance-ramp volumes made on days just before the field survey. These presurvey volumes were used to compensate for possible diversions attributable to the presence of the survey crews. The sample percentages (distribution factors) that were applied to the presurvey ramp volumes are, of course, analogous to the transition probabilities used by SYNODM.

In order to present a simplified illustration of the SYNODM algorithm, the OD tables developed from the field survey were compressed into a 6 x 6 table representing a 3-mile section of the freeway for the morning peak period (6:00 to 9:00 a.m.)

Summary of Results

A comparison of observed OD distributions (i.e., postcard data) and the OD distributions estimated by SYNODM from row and column sums only is presented in Table 3. As can be seen from Table 3, SYNODM has overestimated the trip interchanges for those off ramps immediately downstream of each on ramp. This tendency to overestimate is particularly evident for

TABLE 3 Observed and Estimated OD Distributions, Eastbound Katy Freeway (6:00 to 9:00 a.m.)

Off-Ramp / On-Ramp	1 Wilcrest	2 West Belt	3 Gessner	4 Bunker Hill	5 Blalock	6 "Farther East"	TOTAL
1 "Farther West"	822 (822) 0.000	1713 (1428) 56.880	1358 (1048) 91.698	536 (443) 19.524	501 (445) 7.047	7256 (8000) 69.192	12186
2 Wilcrest	▨	22 (307) 264.577	78 (226) 96.920	84 (95) 1.274	139 (96) 19.260	2123 (1722) 93.380	2446
3 West Belt	▨	▨	3 (166) 160.054	51 (70) 5.157	80 (70) 1.429	1437 (1265) 23.387	1571
4 Gessner	▨	▨	▨	18 (81) 49.000	26 (81) 37.346	1578 (1460) 9.537	1622
5 Bunker Hill	▨	▨	▨	▨	9 (62) 45.306	1166 (1113) 2.524	1175
6 Blalock	▨	▨	▨	▨	▨	1997 (1997) 0.000	1997
TOTAL	822	1735	1439	689	755	15557	

XXX	Observed
(XXX)	Estimated (SYNODM)
XX.XXX	Cell Chi-Square

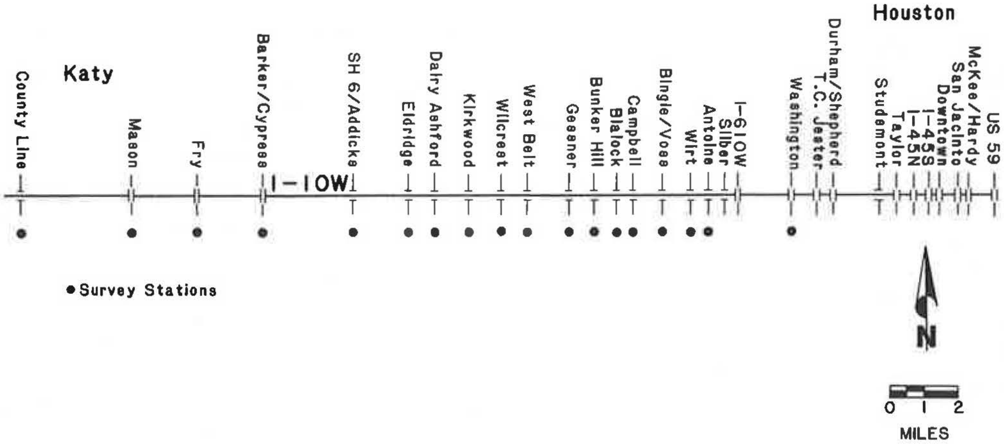

FIGURE 2 OD survey location on Katy Freeway.

those off ramps with low observed interchange volumes. As shown in Figure 3, SYNODM has generally overestimated the number of short trips and underestimated the number of long trips.

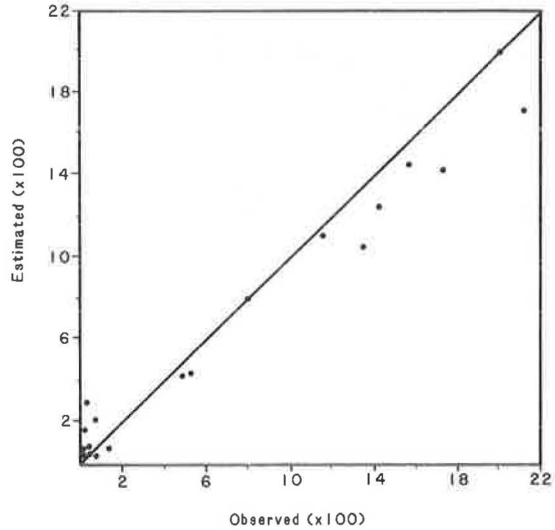

FIGURE 3 Correspondence of observed and estimated T_{ij}'s.

The value of the chi-square statistic for the trip table (i.e., 1,053) is clearly much higher than would have been expected if all trip interchanges were equally likely. There appears to be some regularity in travel behavior or some unique attributes of the trip makers that are not accounted for in the algorithm. Intuitively, this conclusion seems entirely reasonable. There are several major employment centers downstream of the portion of the freeway depicted in Table 3. Because the trips represented by Table 3 are primarily work trips, one would expect the majority of the trips originating at the upstream end of the table to exit at or near the downstream end of the table. Approximately 64 percent of the chi-square statistic comes from the five cells associated with ramps upstream of the Bunker Hill exit, which is consistent with this reasoning.

In addition to the cell chi-square statistics shown in Table 3, an average absolute trip error, defined as follows (7), was also calculated:

$$\text{Avg trip error} = \sum_{i=1}^{r} \sum_{j=1}^{c} |T^*_{ij} - T_{ij}|/N \qquad (5)$$

where N is the number of error values.

Whether the resulting average trip error of 147 is excessive or not cannot be answered in general. However, when compared with the low ramp volumes observed at the upstream end of the trip table (Table 3), the average trip error illustrates the relative magnitude of SYNODM's tendency to overestimate the number of short trips.

SYNODM's tendency to overestimate short trips suggests that it might be worthwhile, in terms of improving the precision of the estimation method, to conduct a field survey of the first few upstream ramps. For example, suppose a lights-on survey of the distribution of farther-west traffic among the first three off ramps was conducted. Assume that the survey resulted in T^*_{12} and T^*_{13} values of 1,713 and 1,358, respectively (i.e., identical to those obtained from the postcard survey). Utilizing the total exit volumes shown in Table 3 (i.e., assuming that these values represent ramp counts), estimates of T_{11}, T_{22}, and T_{66} could be obtained by inspection. Estimates of T_{44} through T_{56} would be unaffected by this additional information and would remain as given in Table 3. The trip table reflecting the incorporation of the T_{ij}^*'s obtained from the hypothetical field survey is shown in Table 4. The T^*_{23} through T^*_{26} and T^*_{14} through T^*_{16} were obtained by subtracting the known T_{ij}'s from the appropriate row sums and applying Equations 3 and 4. Incorporation of the hypothetical limited field data into the estimation method reduced the average absolute trip error from 147 (for the basic SYNODM program) to 42.

TABLE 4 Observed and Estimated OD Distributions, Eastbound Katy Freeway (6:00 to 9:00 a.m.): Incorporation of Limited Survey Data

Off-Ramp / On-Ramp	1 Wilcrest	2 West Belt	3 Gessner	4 Bunker Hill	5 Blalock	6 "Farther East"	TOTAL
1 "Farther West"	822 ---	1713 ----	1358 ----	--- (413)	--- (416)	---- (7464)	12186 8293[a]
2 Wilcrest		22 --	-- (71)	--- (117)	--- (118)	---- (2118)	2446 2353[a]
3 West Belt			-- (10)	-- (78)	-- (78)	---- (1405)	1571
4 Gessner				-- (81)	-- (81)	---- (1460)	1622
5 Bunker Hill					-- (62)	---- (1113)	1175
6 Blalock						1997 ----	1997
TOTAL	822	1735	1439	689	755	15557	
			81	608	612	10987	12288[b]
				530	534	9582	10646[c]

[a] "Adjusted" row sums.

[b] O'(.n) for row 3.

[c] O'(.n) for row 2.

XXX	Observed
(XXX)	Estimated

CONCLUSIONS

Given the limited scope of the study presented in this paper, the question of whether the accuracy of SYNODM is sufficient in some specific case cannot be answered in general. However, a few points deserve note. These observations are offered because they may provide some direction for future research in this problem.

First, for the simple example of an urban freeway during a peak time period with fairly stable OD patterns, the SYNODM algorithm produced reasonably accurate estimates of freeway ramp-to-ramp traffic flows. For freeway planning studies employing macroscopic simulation models, SYNODM's estimates of OD flows may suffice in the absence of any actual OD data. Eldor (8), for example, has shown that the operational measures of effectiveness output by the FREQ simulation programs are not overly sensitive to errors of the magnitudes typically encountered in SYNODM estimates.

Second, incorporation of limited OD data into the estimation method produced a substantial reduction in the average trip error. This suggests that even with the fundamental weaknesses that characterize simplistic algorithms like SYNODM, they could be useful in expanding limited survey data.

In any case, additional research on OD estimation procedures is needed. In this regard, Nihan's (7) incorporation of a trip impedance factor into the estimation method has produced some encouraging results.

ACKNOWLEDGMENT

The research reported in this paper was conducted as part of an ongoing research project entitled West Loop Authorized Vehicle Lane Conceptual Planning and Design sponsored by the Texas State Department of Highways and Public Transportation.

REFERENCES

1. E. Hauer and B.T. Shin. Estimation of Origin-Destination Matrices from Observed Flows: Simple Systems. Research Report 65. Department of Civil Engineering, University of Toronto, Toronto, Canada, March 1980.
2. P.P. Jovanis, W.-K. Yip, and A.D. May. FREQ6PE: A Freeway Priority Entry Control Simulation Model. Research Report UCB-ITS-RR-78-9. Institute of Transportation Studies, University of California, Berkeley, 1978.
3. G.H. English, H. Joiner, and W. Neyland. Traffic Study of Southwest Freeway (US 59) Between Shepherd Drive and Bissonnet Street in Houston, Texas Using the FREQ3CP Simulation Model. Texas State Department of Highways and Public Transportation, Houston, Oct. 1978.
4. R.W. Stokes, J.M. Mounce, D.E. Morris, and R.L. Peterson. Simulation Analyses of Proposed Improvements for the Southwest Freeway (US 59), Houston; Part 1: Long-Term, Capital-Intensive Improvements. Final Report. Texas Transportation Institute, Texas A&M University System, College Station, Aug. 1982.
5. K. Banning, R. Stokes, J. Mounce, and D. Morris. Simulation Analyses of Proposed Improvements for the Eastex Freeway (US 59N), Houston. Texas Transportation Institute, Texas A&M University System, College Station, Feb. 1983.
6. L.G. Nungesser, D.L. Christiansen, and A.J. Ballard. Priority Treatment for High-Occupancy Vehicles on I-10 and I-35 in San Antonio, Texas: A Needs Assessment. Texas Transportation Institute, Texas A&M University System, College Station, March 1983.
7. N.L. Nihan. Procedure for Estimating Freeway Trip Tables. In Transportation Research Record 895, TRB, National Research Council, Washington, D.C., 1982, pp. 1-5.
8. M. Eldor. Integrated Real-Time Freeway On-Ramp Control Strategies. Ph.D. dissertation. Institute of Transportation Studies, University of California, Berkeley, 1976.
9. R.W. Stokes, D. Morris, and J.M. Mounce. Katy Freeway (I-10W) Origin-Destination Study. Texas Transportation Institute, Texas A&M University System, College Station, Jan. 1982.

Estimation of Origin-Destination Matrices with Constrained Regression

CHRIS HENDRICKSON and SUE McNEIL

ABSTRACT

The use of constrained generalized least-squares (CGLS) regression to estimate origin-destination travel matrices from aggregate data is described. The CGLS method does not require general surveys but allows any available data to be included. Variances of matrix entry estimates can be estimated and used as measures of uncertainty or to suggest additional sampling strategy. Two case studies are described from applications to data from Portland, Oregon. The first involves expanding a matrix of transit work trips to all transit trips. Second, a gravity-type model of trip distribution for all work trips is estimated. Comparisons are made with other estimation methods with respect to accuracy, computational effort, and the use of uncertainty measures.

Origin-destination (OD) matrices representing the number of trips between zones or locations in a particular time period are widely used in transpor-