Effects of Small Sample Origin-Destination Data on Transportation Study Results

Ralph E. Foster, Northeastern Oklahoma State University, Tahlequah
Vergil G. Stover and Jimmie D. Benson, Texas Transportation Institute, Texas A&M University

This paper discusses the effects of using small-sample origin-destination survey data as the basis for urban transportation planning models. The study was based primarily on home-interview survey data collected in the 1969 San Antonio-Bexar County Urban Transportation Study. The analysis demonstrates the ability of small sample origin-destination data to produce travel estimates that are in close agreement with model results obtained by using traditional large-sample survey data. The survey data for 12 477 dwelling units (i.e., a 5 percent sample) were used as a data base from which repeated geographically stratified random samples of 6400, 3200, 1600, 800, 400, and 200 observations were drawn. Two samples for each sample size, representing the 10th and 90th percentiles of distributions of sample estimates of total automobile-driver travel, were selected for evaluation. This procedure provided a 0.8 probability that samples of similar size will produce travel estimates as good as or better than those obtained here. The selected samples were used to develop inputs to trip-generation, trip-distribution (gravity), and traffic-assignment models. Samples of 400 or more dwelling units were found to produce acceptable results.

Traditional urban transportation studies invest a significant amount of time and effort in data collection. The home-interview survey generally accounts for a major portion of these costs, and reductions in the sampling rates used in these surveys will obviously result in significant cost savings. This paper reports an investigation to determine the reduction that could be made in the size of dwelling-unit samples without producing significantly different travel predictions after applying forecasting techniques. Data from the San Antonio-Bexar County, Texas, urban area transportation study were used as the data base.

PREVIOUS RESEARCH

There has been a limited amount of research directed toward the determination of minimum sample-size requirements, and much of this has been directed toward sample-size requirements for the calibration of trip-generation and distribution models. Little research has been devoted to determining the smallest acceptable sample for the entire transportation process; i.e., from trip generation through traffic assignment.

Sosslau and Brokke (1) used one-half, one-third, and one-tenth subsamples from the 1-in-15 dwelling-unit sample collected in the 1957 Phoenix origin-destination (O-D) survey to estimate the root mean square error (RMSE) as a function of sampling rate. The results were extrapolated for a range of sampling rates.

Heanue, Hamner, and Hall (2) tested cluster sampling as a way to reduce sample size in a 1960 study of the Pittsburgh urban area. They concluded that cluster sampling was unsatisfactory because of the bias introduced by the location of the clusters.

Parsonson and Cribbins (3) used systematic subsamples from the Raleigh, North Carolina, Urban Transportation Study to compare several approaches to trip generation. They observed differences in the responses of different trip-generation models to decreasing sample size and concluded that models based on small samples estimate the full origin-destination (O-D) data better than do models based on O-D data expanded from the small samples.

By using 100 percent survey data from three zones in San Antonio, Stover, Benson, and Ringer (4) found that large variances of estimates can be expected when traditional sampling rates are used to estimate both trips per dwelling unit and total trip ends for a zone. This research is significant because the sample estimates could be compared to known population values. They concluded that regression models or cross-classification rates provide better estimates of the total number of trip ends by zone than do expanded O-D survey data.

Further analysis of the San Antonio data by Benson, Pearson, and Stover (5) showed that, with the traditional sampling rates, a large majority of the interchange volumes of 1 to 10 trips were undetected while those detected were substantially overestimated. Sampling rates of more than 25 percent would generally be required to estimate nonzero interchange volumes of fewer than 50 trips within ±100 percent at 95 percent confidence.

In an investigation of the sensitivity of traffic assignment, Stover, Benson, and Buechler (6) showed the power of the assignment process to mask major inaccuracies in estimates of zonal trip ends and zonal interchange volumes. They concluded that reasonably accurate assignment results may be anticipated if the preceding modeling phases produce reliable estimates of the total trips and trip-length frequencies. Only reasonable (i.e., relatively coarse) estimates of the geographic distribution of trip ends are needed. They also found (7) that the trip-length frequency distribution is functionally related to the mean trip length and the maximum interzonal separation and developed a procedure for estimating the trip-length frequency distribution.

METHOD OF STUDY

This investigation used data for automobile-driver trips from the San Antonio-Bexar County Urban Transportation Study (SABCUTS). The study covered an area of 3230 km^2 (1247 $miles^2$), which was divided into 778 internal zones with 24 external stations. The population of the area at the time (1969) was 825 800 and comprised 256 640 households. Samples of 200, 401, 803, 1606, 3212, and 6425 observations (for convenience, these sample sizes are rounded to the nearest hundred when referred to in the text) were drawn from the 12 477 home interviews completed and collected in the nominal 5 percent home-interview survey. These samples represent sampling rates ranging from 0.08 to 2.56 percent.

The complete processing and evaluation of even a single sample is a costly and time-consuming process;

the evaluation of a large number of samples is prohibitively expensive. A study design was required that would make possible valid conclusions from the results of evaluating a limited number of samples for a given sample size. Previous research (6) has shown that the total number of trips and the trip-length frequency distribution are the dominant variables in terms of the results obtained from the traffic-assignment procedure. The product of the two variables—total number of trips times mean trip length—is an estimate of total travel. This estimate of total travel (in vehicle-minutes) was selected as the indicator of the traffic-assignment results that a given sample should produce when used as an input to urban transportation planning models.

The procedures for sampling from the full data set were designed so that the various samples selected would reflect, to the extent possible, the observations that would have resulted if the various small samples had been selected by using traditional sampling procedures. Therefore, the number of observations to be made at each sampling rate was calculated by multiplying the desired nominal sampling rate (e.g., ⅛ × 5 percent = 0.625 percent) times the number of households (256 640) in the study area. The specific households in each sample were then selected in a manner similar to the traditional sampling procedures used in home-interview surveys. This ensured that the geographical distribution of the observations selected would be proportional to the distribution of households in the study area.

One thousand samples were drawn at each sample size to assess the sampling distributions of mean trip length, total trips, and total travel for automobile-driver trips. A sample selected near the mean of the distribution of the total-travel estimates would be expected to produce good assignment results because it would provide a good estimate of total travel. The selection of a sample at random, however, would not allow analysis of the results that might be expected from some other sample of the same size. Therefore, at each sample size, the samples representing the 10th and 90th percentiles of the expanded estimates of total automobile-driver travel were selected for use in the modeling procedures. These two samples can provide a basis for estimating the probability that other samples of the same size will perform as well as or better than the samples evaluated. If both samples produce acceptable travel estimates, other randomly selected samples of the same size will have an approximately 0.8 probability of producing travel estimates as good as or better than the samples evaluated. If either sample for a given sample size fails to produce acceptable travel estimates, the sampling level should be considered unsatisfactory.

The full set of survey data and the sample sets of data were processed independently to develop inputs for trip-generation, trip-distribution, and traffic-assignment models. The trip-generation analysis used disaggregated trip-generation rates. Rates were developed for three trip purposes (home-based work, home-based nonwork, and non-home-based) by using income and automobile ownership as independent variables. A minimum of 25 observations in any cell were used as the criterion for combining cells before the calculation of trip-generation rates.

Because the estimates of truck and taxi travel, external-local travel, and external-through travel are based on surveys other than the home-interview survey, the trip generation and trip distribution for these trip purposes were performed once, and the resulting trip tables were merged with the gravity-model trip tables developed for each set of home-interview data. Gravity-model trip distributions and all-or-nothing assignments were performed by using the Federal Highway Administration battery of programs.

The relative zonal attractions used were those developed in the urban transportation study. The implicit assumption is that small-sample home-interview data are not used for estimating zonal attractions. Instead, attraction rates are developed by using special surveys or secondary data sources.

Data from the SABCUTS study provided a dwelling-unit count for each analysis zone. However, because very small samples are not an adequate basis for estimating zonal distributions of dwelling units by income or by automobile-ownership level, tract data from the 1970 census of population and housing were used to estimate these. These population estimates were used when applying the trip-generation rates to the calculation of the total zonal trips by purpose for the full data set and all sample sets.

When samples of a given size give satisfactory results, it is reasonable to assume that larger samples will also give satisfactory results, and unsatisfactory results indicate that smaller samples will also be unsatisfactory. Samples of 1600 observations (approximately one-eighth of the full data set) were evaluated first to determine whether further processing should use larger or smaller samples. Because the results using samples consisting of 1600 observations were satisfactory, further processing of larger samples was discontinued.

ANALYSIS

Analyses using 100 percent survey data (4) have shown that much of the difference between observed and estimated trip ends is due to sampling error in the number of observed trips. This indicates that expanded O-D data will be of limited value in evaluating results obtained from small samples. Therefore, the results obtained from models generated or calibrated by using the full set of survey data for trip generation, distribution, and assignment were used as the standard of comparison for evaluating sample results. Each sample was evaluated as to its representation of the full data set and its performance in trip-generation, trip-distribution, and traffic-assignment models. Traffic-assignment results were also compared with actual traffic counts.

Samples

As expected, the variation in the distribution of estimates of total travel, total trips, and mean trip length increased as the sample size decreased. The 80 percent probability limits for sample estimates of total automobile-driver travel, total trips, and mean trip length are given below.

	Percentage of Full-Data Values		
Number of Observations	Total Travel	Total Trips	Mean Trip Length
6400	±1	±1	±1
3200	±2	±2	±1
1600	±3	±3	±2
800	±4	±4	±2
400	±6	±6	±5
200	±8	±9	±8

Thus, for each sample size, there is a 0.8 probability that a randomly selected sample will estimate the parameter within the indicated range of the population value as estimated from a sample of 12 477 observations. (ranges were rounded to the nearest integer value).

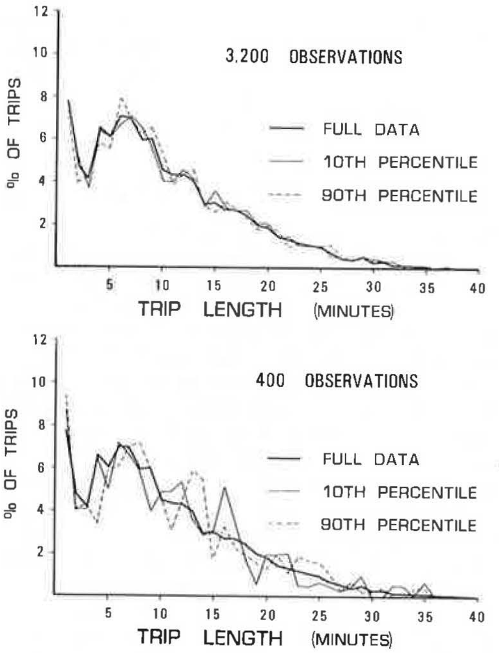
For example, a sample consisting of 800 dwelling-unit observations will estimate the total automobile-

Table 1. Sampling means and standard deviations by sample size.

Sample Size	Total Travel (vehicle·min)		Standard Deviation (% of mean)	Total Trips		Standard Deviation (% of mean)	Mean Trip Length (vehicle·min)		Standard Deviation (% of mean)
	Amount	Percent[a]		Amount	Percent[a]		Amount	Percent[a]	
12 477	13 922 025		–	1 265 923		–	11.00		–
6 400	13 775 250	98.95	1.21	1 251 225	98.84	1.20	11.01	100.09	1.10
3 200	13 780 320	98.98	1.69	1 250 041	98.75	1.65	11.02	100.18	1.28
1 600	13 759 140	98.83	2.31	1 243 970	98.27	2.29	11.06	100.55	1.53
800	13 822 820	98.29	3.31	1 240 522	97.99	3.36	11.14	101.27	2.01
400	13 813 390	99.22	4.65	1 239 805	97.94	4.67	11.14	101.27	2.57

[a] Based on amount for full-data set.

Figure 1. Comparison of origin-destination trip-length frequency distributions.



driver trips within 4 percent of the population value at the 80 percent probability level. These data indicate that the dispersion of extreme values does not become significant until the sample consists of 400 or fewer total dwelling-unit observations. As shown by the data in Table 1, the means of the sampling distributions for each of the three estimates (total travel, total trips, and mean trip length) for all of the sample sizes investigated were within 2 percent of the full data values. For samples consisting of 6400 observations, the standard deviations for the estimates of total travel, total trips, and mean trip length were about 1.2 percent of the mean values. Beyond 800 observations, further decreases in sample size resulted in significantly greater increases in the standard deviation expressed as a percentage of the mean. The results indicate that a relatively small sample has a high probability of yielding good estimates of study-area travel parameters.

The trip-length frequency distributions for non-home-based trips, shown as examples in Figure 1, compare two samples of 3200 observations and two samples of 400 observations with the distribution from the 12 477 observations in the full data set. As expected, the distributions become less smooth as the number of observations is reduced; however, they are considered satisfactory approximations. Fitting a smoothed curve to the data points for the full data set or to data points for the smaller samples produces essentially identical curves. Similar results are observed for the other trip purposes.

The percentage distributions of dwelling units by income and by automobile-ownership levels for each sample were compared with those for the full data set. None exactly matched the full data-set distributions, but all of the samples of 400 or more observations produced acceptable comparisons. The samples with only 200 observations exhibited distributions that were judged to be significantly different from the full data-set distributions. This agrees with the previous research (5) that found that at least 400 observations were necessary to estimate mean trip length for automobile-driver trips with acceptable accuracy.

The income-level and automobile-ownership distributions estimated by the samples of 400 observations or more agreed closely with the distributions from the full data set. Consequently, in view of the previous research (4), it was concluded that small samples selected in the traditional manner can provide reliable estimates of the socioeconomic variables used in transportation studies.

Trip Generation

The dwelling-unit records from the full and sample data sets were cross-classified by using four automobile-ownership and five income levels. After all of the dwelling-unit records were assigned to the appropriate cells, the cells were grouped as necessary so that no cell would contain fewer than 25 observations by combining cells that, a priori, might be expected to exhibit similar trip-generator characteristics. The basic criteria for cell combination were to

1. Make the fewest possible combinations;
2. Consider that, at zero automobile ownership, income is the less important variable and make combinations across income levels; and
3. Create combined cells that were rectangular rather than L-shaped.

Each of the three cross-classification matrices of trip productions (one for each trip purpose) for each sample was compared to the four by five matrices generated from the full data set by using the calculated chi square (χ^2) and RMSE values summarized in Table 2. These calculated values tended to increase with decreasing sample size. However, the pattern has several exceptions. For example, the RMSE values of the 90th percentile sample for 1600 observations are less than those for the 90th percentile sample for 3200 observations.

Because a very small difference may make a large contribution to the calculated χ^2 value, this statistic is of questionable practical significance. Moreover, the total trip matrices differ to a much lesser degree because errors by individual trip purpose tend to cancel out. As shown in Table 3, the range in total number of trips between the 10th and 90th percentiles increases significantly with fewer than 400 observations. With two exceptions (the samples consisting of 200 and 400

Table 2. Chi square and root-mean-square error comparison of trip generations for three trip purposes.

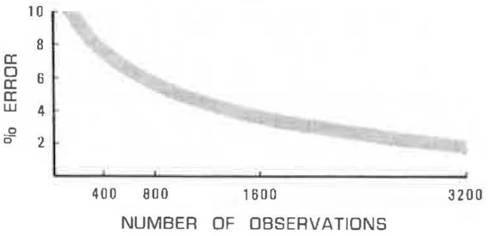
Sample		χ^2			RMSE[a]		
Size	Percentile	Home-Based Work	Home-Based Nonwork	Non-Home-Based	Home-Based Work	Home-Based Nonwork	Non-Home-Based
3200	10th	83.6	155.3	81.3	347	549	233
	90th	97.7	70.1	234.7	277	2333	477
1600	10th	102.9	353.3	447.7	334	990	670
	90th	36.1	215.6	274.4	241	792	481
800	10th	251.4	581.7	768.1	602	1181	1019
	90th	352.0	649.8	1604.9	637	1462	1758
400	10th	1466.7	513.1	315.4	1961	1268	371
	90th	351.2	644.5	1592.7	860	1365	2348

[a] Trips in each cell for full data were used as expected; trip generation rates estimated from each sample data set times total dwelling units in corresponding cells for full data were used to calculate observed trips.

Table 3. Trip productions generated for study area by different size samples.

Sample		Productions			
Size	Percentile	Home-Based Work	Home-Based Nonwork	Non-Home Based	Total
12 477	—	413 705	664 473	343 165	1 421 343
3 200	10th	421 030	658 222	359 674	1 439 926
	90th	408 557	682 849	363 468	1 454 874
1 600	10th	423 212	686 262	318 740	1 428 214
	90th	409 722	679 531	365 792	1 455 045
800	10th	413 490	712 069	356 537	1 482 096
	90th	418 386	682 837	395 211	1 496 434
400	10th	382 460	673 825	325 383	1 381 668
	90th	395 493	692 481	399 240	1 487 214
200	10th	461 431	605 431	324 277	1 391 178
	90th	441 511	757 019	362 445	1 560 975

Figure 2. Expected error in number of total trips (0.8 probability level).



observations at the 10th percentile), the total trips generated by the samples slightly overestimated total driver trips as compared to total trips generated by the full data.

The frequency distribution of trip productions by zone for each sample was compared to those for the full data set. Samples of 400, 800, 1600, and 3200 observations had distributions essentially identical to that for the full data set. Distributions produced by samples of 200 observations were significantly different from that for the full data set.

Trip productions by trip purpose and by income levels were compared using 14 geographic sectors. The comparisons by geographic sectors indicated no significant differences between the full data set and the samples of 400 or more observations in either patterns of trip making or geographic patterns of travel.

Based on the combined analysis of the several sample data sets and previous research (4), the maximum expected error for total trips at the 0.8 probability level was determined as shown in Figure 2. The average error is estimated to be slightly less than 7 percent for a sample of 400 dwelling units and decreases to about 3 percent for a sample of 1600. Increasing the sample size from 1600 to 3200 observations decreases the maximum error to approximately 1.0 percent, which indicates that larger sample sizes contribute only marginally to the accuracy of travel estimates. Therefore, samples of 400 or more observations are adequate to produce acceptable trip-generation results.

The presample determination of cross-classification cells was compared to the use of a more detailed matrix and postsample combining of cells to provide a minimum of 25 observations per cell. The two samples of 400 observations, the smallest sample size that produced acceptable trip-generation results, were used in the analysis. The results achieved by the procedure of reducing the number of cells after the sample had been selected and the observations assigned to the appropriate cross-classification cell (four automobile-ownership and five income level) were compared to the following cross-classifications, which were established prior to the sample selection: (a) two automobile-ownership levels (zero and one plus) and no breakdown by income level, (b) three automobile-ownership levels (zero, one, and two plus) and no breakdown by income level, and (c) two automobile-ownership levels (zero and one plus) and two income levels (low and medium to high). To achieve at least 25 observations per cell, the four by five matrix was reduced to the following five cells: (a) zero automobiles by all income levels, (b) one plus automobile by low income, (c) one plus automobile by medium to low income, (d) one plus automobile by medium income, and (e) one plus automobile by medium to high and high-income levels. All of the presampling classification schemes produced results that were inferior to those of the cross-classification involving postsampling reduction in the number of cells. Furthermore, the analysis indicated that better results were produced as the number of cross-classification cells increased. This suggests that combining cells to achieve a minimum number of observations per cell after the sample has been selected is the better procedure to follow when a small sample is selected by using traditional sampling procedures.

Trip Distribution

The trip-length frequency distributions by trip purpose obtained from the survey data were used as gravity-model calibration criteria. The calibrated mean trip lengths were all within 3 percent of the target O-D data values; most differed from the O-D mean trip length by less than 1.0 percent. The calibrated trip-length frequency distributions for the samples were in close agreement with the comparable distributions for the full set of survey data.

The trip table is a more deterministic measure of the adequacy of models calibrated from small samples. Analysis of the zonal-interchange-volume distributions indicated no significant differences between the results

Figure 3. Limits of expected error for traffic-assignment parameters (0.8) probability level) (a) total VKT, (b) screen line volumes, (c) cutline volumes, and (d) link volumes.

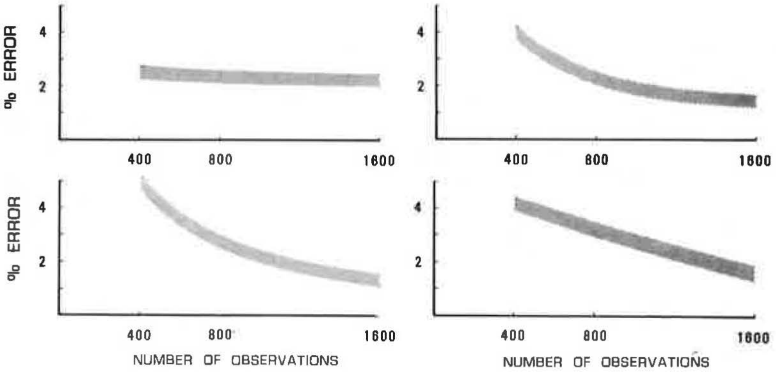


Table 4. Cutline volumes.

Sample		RMSE (comparison to counted volumes)	Comparison to Full-Data Set	
Size	Percentile		RMSE	Percent RMSE
12 477	–	11 084	–	–
1 600	10th	11 489	1653	3.6
	90th	10 806	734	1.6
800	10th	10 668	1456	3.1
	90th	10 643	1638	3.5
400	10th	11 138	3999	8.6
	90th	11 945	2314	5.0

Table 5. Link-volume differences.

Sample		Mean Volume Difference	Standard Deviation	RMSE
Size	Percentile			
1600	10th	-28	185	187
	90th	39	88	97
800	10th	68	116	135
	90th	85	131	156
400	10th	-78	214	407
	90th	102	407	420

using the small samples and the full data. In each comparison, more than 50 percent of the cells in the total trip table were estimated as having exactly the same interchange volume, and approximately 95 percent of the cells were within ±2 trips of the full data results. Although there was some tendency for the variations in the trip tables to increase as the sample sizes decreased, the differences in the trip tables were not significantly affected by the sample sizes used in the calibration. (The trip table that most closely resembled that using the full data set was produced by the 10th percentile sample of 400 observations, while the 90th percentile sample of 400 observations provided the poorest comparison.)

Traffic Assignment

Traffic assignments that used the modeled trip tables based on samples of 400, 800, and 1600 observations were compared to the traffic assignments based on the full set of survey data and to the counted traffic volumes. The expected errors (at the 0.8 probability level) for the total vehicle kilometers of travel (VKT), screen-line volumes, and cutline volumes are shown in Figure 3.

Estimates of the total VKT obtained with the models calibrated by using the small samples agreed with that using the full data set within 2.7 percent. When tabulated by the 14 geographic sectors and compared to the full data-set value, the RMSEs for the samples were less than 1.0 percent. Thus, a sample consisting of 400 observations is adequate to produce acceptably accurate estimates of the total VKT and the geographical distribution of the VKT.

A rail right-of-way that essentially bisects the study area was used as a major screen line. The screen-line volumes given in the table below are in close agreement with those for the full data.

Sample		Volume	
Size	Percentile	Value	Percent of Full Data
12 477	–	444 339	100.0
1 600	10th	436 368	98.2
	90th	448 684	101.2
800	10th	452 566	101.9
	90th	454 549	102.3
400	10th	428 647	96.5
	90th	462 154	104.3

Although the expected error begins to increase for samples of less than 800 observations, the samples of 400 observations produced estimates of screen-line volumes that are within 4.5 percent, or less, of the estimated volume based on the full data set. Therefore, sample sizes of 400 observations or more produce acceptably accurate estimates of screen-line volumes.

Twenty-eight cutlines were used to compare the assigned volumes in various travel corridors. The mean differences from the full data-set assignment were 1.2, 2.6, and 4.1 percent for 1600, 800, and 400 observations respectively. Although the maximum expected error tends to increase with decreasing sample size (Figure 3), the magnitude of the error in cutline volumes with samples of 400 observations is considered to be within acceptable limits. As indicated in Table 4, when the assigned volumes are compared to the counted volumes, the samples of 400 or more observations produced RMSEs that are not appreciably larger than that resulting with the full data set. All of these assigned cutline volumes were within 10 percent of the counted volume, and three-quarters were within 5 percent.

Comparisons of the individual cutline volumes indicate that the small samples produce results that are frequently as good as or better than that produced by the full data set. For example, when the 90th percentile sample of 800 observations is used, 17 of the 28 cutlines have assigned volumes that are closer to the counted volume than that produced by full data set. These analyses suggest that factors other than the number of dwelling-unit observations have equal, or greater, impacts on the cutline results.

The network links were classified for comparison into 15 volume groups on the basis of the full-data-set traffic assignment. Over 95 percent of the link volumes for the sample assignments were within the predetermined acceptable error ranges of the volume for the full-data-set assignment. The remainder were only slightly outside the acceptable range of error. The mean percent volume differences (based on the mean volume difference as a percentage of the midpoint of each volume range) were 1.0, 2.2, and 2.9 percent with samples of 1600, 800, and 400 observations respectively.

As indicated by Table 5, the calibration of models with vary small sample sizes does not contribute to a serious deterioration in overall results, although the mean differences and variances in assigned link volumes increase with decreasing sample size. A comparison of these parameters by volume group found, for example, that the 10th percentile samples of the 1600 and 400 observations produced nearly identical results. The best overall results by volume group were those of the sample of 800 observations. This, together with the observed pattern of trip generations by income and automobile-ownership levels, suggests that random variations are more significant in affecting the assignment results than is sample size. Therefore, although the maximum expected error (Figure 3) for assigned link volumes tends to increase with smaller sample sizes, assignments developed from samples of 400 observations are within the acceptable limits of error.

Examination and comparison of the posted traffic assignments did not identify any significant differences between the assignment based on the full data set and the assignments based on the samples. Similarly, comparisons based on 12 selected major routes did not identify significant differences in assigned volumes. The full data set resulted in overassignment on 7 routes and underassignment on 5 routes when compared to the counted volumes. For samples of 800 and 1600 observations, the assignments on each of the routes were within 2.5 percent of those produced by the full data set. The samples of 400 observations produced assignments that were within 6.0 percent of those from the full data set.

EVALUATION AND IMPLICATIONS

The results of this study indicate that urban transportation models calibrated for three trip purposes (home-based work, home-based nonwork, and non-home-based) will produce acceptably accurate travel estimates from data based on as few as 400 dwelling-unit interviews. The analyses indicate that there is only a modest decrease in the precision of the estimates as the sample size is reduced to 400 observations, but that thereafter the reliability of the estimates deteriorates rapidly. This indicates that the collection of the larger samples that results from traditional sampling rates is not cost-effective in the traffic-assignment results.

This and other research have established that it is the number of observations in the sample, rather than the percentage of dwelling units surveyed, that determines whether urban planning models produce acceptably accurate traffic assignments. Thus, the results of this study can be applied to study areas having both smaller and larger populations. The minimum sampling rate will vary inversely with the study-area population, but the minimum number of observations required is a constant.

Previous research has shown that the development of acceptably accurate traffic assignments requires relatively precise estimates only for total trips for the study area and for the mean trip length. Estimates of trip-length frequencies and the geographic (zonal) distribution of trip ends need only be reasonable approximations (and errors in the geographic distribution of trip ends will be offset).

Although samples consisting of 400 dwelling-unit observations are an adequate basis for the calibration of transportation models to produce acceptably accurate traffic assignments, this does not imply that 400 observations is an adequate sample size for defining other, more detailed travel characteristics. A larger number of observations or specially designed studies are necessary to define travel characteristics such as the number of trip attractions or the relative attractions for individual zones; travel patterns for trip interchanges between zones; the temporal stability of trip-generation rates; trip priorities, for example, which trips would not be made under adverse conditions such as fuel shortages; modal choice among alternative means of travel; measures of trip-generation characteristics of specific population segments, such as by dwelling-unit type or persons per dwelling unit; and the number of trip ends in a specific geographic area on the basis of O-D data expansion.

The modification of urban transportation study procedures to efficiently use small samples will permit significant cost savings in data collection and reduction without causing measurable effects on the assignment results. The effective use of small-sample survey data suggests the following modifications in transportation study procedures:

1. Disaggregate trip-generation techniques must be used—cross-classification with the number of cells to be used in the trip-generation analysis should be determined after the collection of the O-D data and
2. Stratified cluster sampling might be used to further simplify and reduce the cost of the dwelling-unit inventory and sample selection tasks.

Traditional record-keeping systems and procedures for developing the dwelling-unit inventory might be used for the geographic subdivisions not selected for data collection, for which only the number of dwelling units in each area would be required.

The urban transportation study has been a valuable tool for the transportation analyst in the development of transportation demand forecasts, the evaluation of land-use and transportation system alternatives, and the identification of major inconsistencies between proposed activity patterns and the transportation network. The use of a much smaller sample of home-interview data will permit the transportation study to continue to be a valuable, but more cost-effective tool.

ACKNOWLEDGMENT

The research leading to the development of this paper was sponsored by the Federal Highway Administration, U.S. Department of Transportation. The paper is based on a research report entitled Consequences of Small Sample O-D Data Collection in the Transportation Planning Process.

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

REFERENCES

1. A. B. Sosslau and G. E. Brokke. Appraisal of Sample Size Base on Phoenix O-D Survey Data. HRB, Bulletin 253, 1960, p. 114-127.
2. K. E. Heanue, L. B. Hamner, and R. M. Hall. Adequacy of Clustered Home-Interview Sampling for Calibrating a Gravity-Model Trip-Distribution Formula. HRB, Highway Research Record 88, 1965, pp. 116-136.
3. P. S. Parsonson and P. D. Cribbins. Estimating Urban Trip Production and Attraction. Journal of the Highway Division, Proc., ASCE, Nov. 1967, pp. 115-134.
4. V. G. Stover, J. D. Benson, and L. J. Ringer. Accuracy of Travel-Pattern Estimated From the Home-Interview Survey. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 167-7, Aug. 1973.
5. J. D. Benson, D. F. Pearson, and V. G. Stover. Accuracy of Travel-Pattern Estimates From the Home-Interview Survey. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 167-8, March 1974.
6. V. G. Stover, J. D. Benson, and J. Buechler. Sensitivity Evaluation of the Traffic Assignment Process. TRB, Transportation Research Record 610, 1976, pp. 50-52.
7. D. F. Pearson, V. G. Stover, and J. D. Benson. A Procedure for the Estimation of Trip-Length Frequency Distributions. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 17-1, May 1974.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

An Application of Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model

Daniel McFadden and Kenneth Train, University of California, Berkeley
William B. Tye, Charles River Associates, Inc., Cambridge, Massachusetts

Statistical tests are proposed to diagnose the validity of the independence from (of) irrelevant alternatives property of the multinomial logit model. Application of the tests is illustrated by the use of actual travel data representing urban modal choice in the San Francisco area. The property as it applies to travel demand forecasting is discussed, and the common misconception that the property holds for market shares in heterogeneous populations is shown by examples to be incorrect. The relation of the property to the basic assumptions of the model is described, and it is shown that the validity of the property in disaggregate modeling is an empirical issue that depends on the model specification and data in a particular application. A series of diagnostic tests for the property are developed and applied to actual travel data.

The most widely used functional form for choice probabilities in disaggregated transportation-demand analysis is the multinomial logit (MNL) model,

$$P(i|C) = \exp V(x^i, s) \Big/ \sum_{j \in C} \exp V(x^j, s) \qquad (1)$$

where

C = finite choice set,
$P(i|C)$ = choice probability for alternative $i \in C$,
x^i = vector of the observed characteristics of alternative i, and
s = vector of the observed characteristics of the decision maker and the choice environment.

The scale function $V(x^i, s)$ may be interpreted as the representative utility of alternative i and is normally assumed to be linear in the parameters. The MNL model has significant advantages over the available alternatives in terms of flexibility and computational efficiency and permits a simple behavioral interpretation of the parameters of the scale function.

The MNL model also has the property that the ratio of the probabilities of choosing any two alternatives

$$P(i|C)/P(k|C) = \exp V(x^i, s)/\exp V(x^k, s) \qquad (2)$$

is independent of the attributes or the availability of a third alternative (j), which is termed the independence from (of) irrelevant alternatives (IIA) property. This property greatly reduces the complexity of estimation and forecasting and in this respect is quite useful. However, it imposes restrictions on the structure of choice probabilities and cross elasticities; these restrictions may be invalid in some applications. Hence, tests of the validity of the IIA property should be made whenever a violation of the assumption is suspected.

This paper analyzes the IIA property and discusses