HIGHWAY RESEARCH RECORD

Number 191

Origin and Destination Characteristics

5 Reports

	Subject Area
55	Traffic Measurements
81	Urban Transportation Administration
84	Urban Transportation Systems

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C., 1967 Publication 1522

Department of Traffic and Operations

Harold L. Michael, Chairman
Associate Director, Joint Highway Research Project
Purdue University, Lafayette, Indiana

HIGHWAY RESEARCH BOARD STAFF

E. A. Mueller, Engineer of Traffic and Operations

COMMITTEE ON ORIGIN AND DESTINATION
(As of December 31, 1966)

Alan M. Voorhees, Chairman
Alan M. Voorhees and Associates, Inc., McLean, Virginia

David K. Witheford, Secretary
Transportation Planning Consultant
West Haven, Connecticut

Willard H. Armstrong, Bureau of Planning Survey, Ohio Department of Highways, Columbus
John E. Baerwald, Department of Civil Engineering, University of Illinois, Urbana
Robert D. Bedwell, Highway Economist, California Division of Highways, Sacramento
Glenn E. Brokke, U. S. Bureau of Public Roads, Washington, D. C.
Nathan Cherniack, Economist, The Port of New York Authority, New York City
Donald E. Cleveland, Associate Professor, Department of Civil Engineering, University of Michigan, Ann Arbor
Francis E. Coleman, Highway Associate Engineer, Connecticut Highway Department, Wethersfield
Roger L. Creighton, Director, Transportation Planning and Programming Subdivision, Division of Finance and Planning, New York State Department of Public Works, Albany
John A. Dearinger, Department of Civil Engineering, University of Kentucky, Lexington
Thomas B. Deen, Vice President and Associate, Alan M. Voorhees and Associates, McLean, Virginia
George Ferguson, Southeastern Pennsylvania Regional Planning Association, Pittsburgh, Pennsylvania
Thomas J. Fratar, Partner, Tippetts-Abbett-McCarthy-Stratton, New York, N. Y.
W. L. Grecco, Purdue University, Lafayette, Indiana
Harold W. Hansen, Senior Planning Engineer, Portland Cement Association, Chicago, Illinois
Walter G. Hansen, Alan M. Voorhees and Associates, McLean, Virginia
Kevin E. Heanue, Highway Engineer, Urban Planning Division, U. S. Bureau of Public Roads, Washington, D. C.
Donald M. Hill, Traffic Research Corporation, Ltd., Toronto, Ontario, Canada
G. H. Johnston, Assistant Planning Studies Engineer, Traffic and Planning Studies Division, Ontario Department of Highways, Downsview, Canada
Louis E. Keefer, Consultant for Transportation Planning and Research, Woodmont, Milford, Connecticut
Norman Kennedy, Associate Director, Institute of Transportation and Traffic Engineering, University of California, Richmond
Warren B. Lovejoy, The Port of New York Authority, New York City
John T. Lynch, Washington, D. C.
Brian V. Martin, Assistant Chief Engineer (Research), Department of Highways and Transportation, Greater London Council, London, England
James J. McDonnell, U. S. Bureau of Public Roads, Washington, D. C.
W. L. Mertz, Technical Director, Tri-State Transportation Committee, New York, N. Y.

John K. Mladinov, Director, Puget Sound Regional Transportation Study, Seattle, Washington
Robbie W. Parker, Vogt, Ivers and Associates, Consulting Engineers, Jeffersonville, Indiana
William S. Pollard, Jr., Memphis, Tennessee
Lloyd A. Rivard, Chief Engineer, D. C. Department of Highways and Traffic, Washington, D. C.
James J. Schuster, Assistant Professor, Civil Engineering Department, Villanova University, Villanova, Pennsylvania
Arthur Schwartz, Mass Transportation Planner, Tri-State Transportation Committee, New York, N. Y.
Billy J. Sexton, Chief Planning Engineer, Division of Planning, Kentucky Department of Highways, Frankfort
Paul W. Shuldiner, Consultant to the Director, Transportation Systems Planning Division, Office of High Speed Ground Transportation, U. S. Department of Commerce, Washington, D. C.
Bob L. Smith, Professor of Civil Engineering, Kansas State University, Manhattan
Vergil Stover, Assistant Research Engineer, Texas Transportation Institute, Texas A & M University, College Station
Anthony R. Tomazinis, Transportation Planning Consultant, Delaware Valley Regional Planning, Philadelphia, Pennsylvania
George V. Wickstrom, Program Director, New Castle County Program, Wilmington, Delaware
Martin Wohl, National Science Foundation Fellow, University of California, Berkeley
J. E. Wright, Traffic Manager, Planning Survey Division, Texas Highway Department, Austin
F. Houston Wynn, Wilbur Smith and Associates, New Haven, Connecticut

Contents

Foreword

Highly sophisticated aspects of urban transportation planning are very much in the ascendancy, as the papers in this RECORD point out. Since conduct of these studies is mandatory in urban areas over 50,000 population, the transportation and planning literature is constantly being increased by the research undertaken by the agencies performing the planning.

Proper transportation planning requires detailed study and identification of the origin-and-destination characteristics of the urban area and application of these characteristics to a future point in time so that a transportation system may be furnished, presumably adequate for these future characteristics. In order to arrive at this rather monumental achievement, great use is made of mathematical modeling and high-speed computers and, of course, research is continually being performed in order to work through the multifaceted structure that such tasks present.

The five papers in this RECORD provide valuable insight into some of the complex procedures inherent in undertakings such as the comprehensive planning process. Survey planners and administrators will find much of value and interest. All of the research relates directly to some phase of the transportation planning process.

The first paper, derived at the Chicago Area Transportation Study, sets forth an alternative method of estimating trips. A complete and statistically valid mathematical model was developed that would measure trip generation by purpose of trip.

The second paper, developed at the Tri-State Transportation Commission, indicates a new statistical criterion which objectively chooses the best variable or combination of variables to forecast person trips and auto trips per household from examination of the variables used in the trip generation process.

The next paper, by a Bureau of Public Roads investigator, evaluates the accuracy of the traffic assignment process when capacity restraints are applied. Using data from ten transportation studies, the evaluation was performed using five tests for measurement. Indications of apparent accuracy were found, as were methods of improving accuracy.

Another BPR researcher has investigated the inherent time savings possible in using computers more extensively. Computer applications are described that could perform functions such as the editing of O-D data, coding trips to blocks in survey zones, identification of land use, and computation of expansion factors, and that allow for the management of large volumes of data and generally provide more flexibility in using data. Two planning officials, skilled in the use of survey data, discuss the pros and cons of the proposed computerization, and the author provides a closure to their discussions.

The last paper concerns a BPR research project that attempts to evaluate the calibration and testing techniques of the gravity and intervening opportunities trip distribution models. Two approaches used in calibrating the intervening opportunities model were investigated and the calibration of the gravity model is briefly described. The results of a comprehensive series of analytical and statistical tests to each model are reported.

Methods for Estimating Trip Destinations by Trip Purpose

NATHALIE GEORGIA SATO, Chief Urban Planner,
Chicago Area Transportation Study

•DURING the preparation of its 1980 transportation plan, CATS developed a method, based on land use, for estimating and forecasting internal person trip destinations. Theoretically, trips were forecast from trip rates to six generalized land uses—residential, manufacturing, commercial, transportation, public buildings, and public open space. There were, however, a number of basic modifications.

The land-use method was used in establishing one control total for all 1980 person trips, but it was replaced by a second control total based on population, car ownership, and net residential density. Even the control totals for the distribution of future person trips by land use were based on the trip-making propensities of the population, economic forecasts of employment, and dollar output of economic activities. The distribution of trips to zones was a compound of methods using population, employment, and land use.

The land-use method has a number of limitations in trip forecasts. The basic limitation, of course, is that land does not make trips. Rather, it is people who make trips. Land use and trip rates to land uses are convenient ways of describing or measuring the volume of trips, but the land-use method is not always appropriate for forecasting trips. One reason is that land use has not been defined in transportation terms, but rather in terms borrowed primarily from the field of city planning. City planning definitions of land are static—type, amount, location, and perhaps some indication of density, or intensity of use. The definitions do not ordinarily identify and describe those features or activities of land use which attract people and trips.

Because of the limitations of the land-use method, studies to develop an alternative method were pursued. The alternative method was premised on the fact that it is people who make trips, and that their motivations, expressed as trip purposes, can be used to estimate and forecast the number of trip destinations. The 1956 home interview survey of trips had elicited information about the purpose of the trip, so that, in general, there were sufficient data to analyze trips by purpose and destination.

Trip purposes were classified into nine categories—home, work, shop, school, social-recreation, personal business, eat meal, serve passenger, and ride. In its analysis of trip generation characteristics, CATS found that the number of trips per dwelling place or family was related to net residential density and automobile ownership. CATS also found that the distribution of trips among the nine purposes varied, depending on the total number of trips per dwelling place. As the number of trips per family increased, the proportion of trips to home declined slightly, and the proportion of work trips decreased significantly, while the proportion of trips for all other purposes increased (Table 1 and Fig. 1).

From the same 1956 survey of trips, CATS estimated that there were 9,930,681 internal person trips. The purpose of these trips at destination is shown in Table 2. This table also shows the percentage distribution by purpose, average trip length in miles, and the mode of travel used for each trip purpose. It will be observed that, except for school trips, the automobile was the most important mode of travel.

It was the intent of the analysis of trips by purpose to isolate and define the causal factors which best explained the number of trip destinations by purpose. When the causal factor could not be identified, because of the complexity of human motivations, then factors which had a logical relationship to the trip purpose were examined. In

---

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.

TABLE 1

PERCENTAGE DISTRIBUTION OF TRIPS BY PURPOSE RELATED TO TRIP-MAKING PER DWELLING PLACE

Trips Per Dwelling Place	Purpose						
	Home	Work	Shop	School	Social-Recreation	Personal Business	Eat, Ride, and Serve Passenger
2	49.9	37.2	2.1	0.5	3.7	6.1	0.5
4	47.7	27.1	4.9	1.4	9.2	7.6	2.0
6	45.9	20.5	5.9	2.3	11.8	10.2	3.4
8	44.2	17.2	5.9	2.4	14.7	10.2	5.4
10	43.0	15.7	6.9	2.7	16.0	10.0	5.7
12	41.5	14.0	6.7	1.8	12.9	14.0	9.1
14	39.5	12.6	7.1	2.0	17.3	12.7	8.8
16	41.1	13.0	5.2	3.1	18.2	10.4	8.9
18	39.8	11.8	6.4	2.1	17.7	11.6	10.7
Mean	43.7	20.3	5.5	2.0	13.0	10.1	5.4

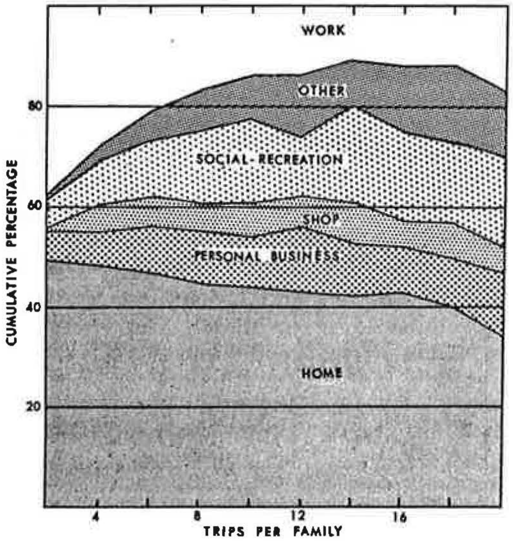


Figure 1. Percentage distribution of trips by purpose related to trip-making per family.

addition to having a logical relationship to the trip purpose, the factors had three other constraints placed on them. Data for the factor either had to be readily available or collectible without undue expenditure of time and effort. There had to be acceptable and reliable forecast methods for the factor. Insofar as possible, the factor had to indicate future changes in the destination of the trip purpose under consideration.

MAJOR FINDINGS

The derivation of acceptable estimating methods of trips by purpose is feasible, although the results of this study are uneven in deriving estimating methods for each and every one of the nine trip purposes. The estimating methods are best for total, home, and work trips. The methods are good for school, social, and personal business trips. Methods for estimating shopping trips do require further study. No methods were derived for the minor trip purposes, since they were combined with a major

TABLE 2

SELECTED CHARACTERISTICS OF INTERNAL PERSON TRIPS BY PURPOSE

Purpose	Total Trips	Percent Distribution	Average Trip Length (airline miles)	Percent Distribution by Mode				
				Auto Driver	Auto Passenger	Suburban Rail	Subway + Elevated	Bus
Total	9,930,681	100.0	4.3	48.5	27.2	2.5	4.8	17.0
Home	4,328,569	43.6	4.4	45.9	28.0	2.8	5.3	18.0
Work	2,033,035	20.5	5.3	55.0	11.2	5.1	8.6	20.1
Shop	550,215	5.5	2.6	54.5	27.9	1.1	3.1	13.4
School	204,794	2.1	2.6	9.4	33.1	0.7	4.3	52.5
Soc. -recreation	1,251,785	12.6	4.0	41.6	44.7	0.6	1.5	11.6
Eat meal	210,983	2.1	2.9	54.3	39.5	0.4	1.2	4.6
Pers. business	1,009,362	10.2	3.5	51.9	28.2	1.0	3.0	15.9
Serve passenger	234,752	2.3	2.1	96.1	3.9	—	—	—
Ride	107,186	1.1	3.0	0.1	99.6	—	0.1	0.2

trip purpose. Statistically, the methods are best at the district level. The poorer statistical results at the zonal level do not invalidate the method in the case of logical, causal factors, but rather indicate the probability of data input error. Poor statistical measures of reliability at both the district and zonal level do indicate the need for further study.

The number of total trips and trips home to the place of destination can be estimated and forecast from knowledge of the number of automobiles owned and net residential density. Total automobiles owned is a measure of the total number of trips involved. It also, together with net residential density, indicates the propensity of people to make trips. Furthermore, of the possible modes of travel, the automobile is used most frequently. Net residential density is an index of the availability and accessibility of trip destination places. With high net residential density many activities usually are close at hand, and it is possible to reach many destinations by walking. Congestion, which high net residential density often implies, also acts as a deterrent to making many trips by automobile. Low net residential density, on the other hand, usually is associated with fewer nonresidential activities. Travel, however, usually is easy because streets and highways are not ordinarily congested. More trips are required to fulfill the objectives, but it also is easier to make the trip. There are acceptable substitutes for automobiles owned and net residential density, but either they re-express these resident characteristics or show a high degree of covariance with automobiles owned and net residential density.

Trips to work are related directly to employment, for such trips are not made unless there are jobs. But because of illness, vacation, and other types of absences, the number of work trips will be between 80 and 90 percent of total employment.

Like work trips, trips to school have a compulsory element. They are related directly to school enrollment, but again an absenteeism factor must be applied. Enrollment data may be approximated from school size or capacity, or from the percentage of the school age population enrolled in school. Since many school trips will be made on foot, it is necessary to estimate the proportion of walking trips. Most trips to the urban neighborhood elementary school and resident college will be pedestrian. Because of the larger service area, trips to high schools and day institutions—academic, business, trade—will have transit and automotive modes.

Approximately half of the social-recreation trips are social, that is, trips to visit friends and relatives. Automobile ownership is a good estimator of these trips, for it not only reflects trip-making propensity, but also the tendency of families to exchange visits with other families of like economic and social characteristics. Recreation trips are affected by the season of the year, weather, and scheduling of events. Approximate estimates of many of these trips can be made from automobiles owned, because there are many local, neighborhood indoor and outdoor recreation facilities to which trips are made. Where the destination was a major spectator facility, such as a major-league ball park or race track, there was not only a small sample from which to derive the estimating method, but also there were several unpredictable elements, such as weather and season of year. Like horse betting, it was a gamble, and the odds were against the trip estimator. A more serious problem in recreation trip forecasts is the impact of the shorter work week or working day on recreation travel. The resultant recreation trips may be similar to weekend recreation trips, but it was not possible to derive an estimating method for weekend recreation trips, since the trip survey data were limited to weekday travel.

Most personal business trips were local in nature, and for this reason it was possible to use a resident characteristic in the district or zone of destination to estimate the total number of trips. The number of automobiles owned was used, but the numbers of dwelling units or population were acceptable alternatives. At the zonal level it was found desirable to add net residential density. For the nonlocal personal business trips, estimated employment in several of the service trades and professions were appropriate factors on which to base an estimate. Because the present distribution of employment in the service trades and professions exhibits covariance, employment in one of the service trades and professions gave a good estimate of all nonlocal personal business trips. Of these trades and professions, the employment in medical service was recommended because it represented one of the more important destination places, current data on the number and location of this employment were obtainable, and because it was believed that forecasts were feasible. There were acceptable alternatives.

A reliable method for estimating shopping trips was not derived. There was logical evidence that shopping trips could be estimated from retail sales. While reasonably current data on sales were available, the techniques of forecasting retail sales by small areas left much to be desired. Other measures of retail activity, such as any measure of retail space, revealed that much of the existing retail inventory was built to serve a pedestrian shopper, and that while it continued to exist, it was inefficient and did not attract shopping trips in the same manner as the newer, competitive postwar shopping centers with their parking lots for the shopper who now arrives by car. A qualitative measure of retail activity is called for, but isolating and defining it requires further study.

The minor trip purposes—eat meal, ride, and serve passenger—were combined with a major trip purpose. The combinations did not impair the estimating method of the major trip purpose. Conceivably, all should have been combined with personal business trips, since this group of trips included a number of different trip purposes. Only trips to eat a meal were so combined. There was evidence that many trips to serve a passenger had the objective of taking a child to school. From the principal type of destination of trips to ride, it was hypothesized that a child or friend was taken along for "a ride" on a shopping trip. On the basis of these theories, trips to serve a passenger were combined with school trips, and trips to ride were combined with shopping trips. The justification of other possible combinations is not denied, but they were not tested. It is strongly recommended that these trip purposes be consolidated with a major trip purpose in future trip surveys. The trip purpose of the person serving a passenger should be the same as the person served, and the trip purpose of the person riding should be the same as the person taking the rider along. Trips to eat a meal should be combined with personal business, unless there is a recreational element in the trip purpose.

Summarizing, the factors which are required to estimate the number of person trip destinations by purpose are as follows:

Factor	Alternative Factors
Total number of automobiles owned by residents	Population, dwelling units, licensed automobile drivers
Net residential density	
Total employment in all industries, manufacturing and nonmanufacturing	
Estimated employment in medical service	Estimated employment in government service, estimated employment in other service trade or profession
School enrollment by type of school	School size or capacity, school-age population
Qualitative measure of retail activity	Retail sales, classification of retail

In addition, it is desirable to have knowledge of special trip generators. These facilities and places are few in number and the number of trips may be high in small localized areas. Trip estimating methods are difficult to derive because of the small number of such facilities, and because the generation of trips may be subject to such unpredictables as season of the year, weather, and scheduling of events at the facility.

Finally, it should be remembered that these estimating methods are premised on the urban environment and traffic that we know today. Radical departures from these conditions would invalidate many of the methods. For example, those methods which are based on automobile ownership are not applicable to conditions fifty years ago when the automobile was in its infancy, nor a future condition where the helicopter or other airborne vehicle has replaced the family car. Radical changes and shifts in the location of urban facilities will also affect the pattern and number of trip destinations.

The principal tool used in deriving the estimating equations for the number of person trips to the district or zone of destination was regression analysis. With existing computer hardware it is possible to put in any number of independent variables and with appropriate manipulations derive complex estimating equations with impressive coefficients of multiple correlation. The practical use of such estimating equations is of a dubious nature. First, while the computer may be accurate to the fourth decimal place, the input data derived from trip and land-use surveys may contain considerable error as the result of human fallibility. As a consequence, the resulting high correlation coefficients may be spurious. Second, even assuming accuracy in the original data and reliability of the derived estimating equation, use of a complex equation probably will necessitate considerable time and effort in the collection of data which do not contribute significantly to the reliability of the estimate.

Because of these considerations, the following principles were followed in the selection of independent variables for the regression analysis of trips:

1. The number of independent variables was held to a minimum, and the addition of second independent variables had to be justified by a significant increase in the statistical measures of correlation.
2. Insofar as possible, the independent variable had a logical, causal relationship with the dependent variable.
3. Data for the independent variables were obtainable from existing data collections, or from relatively simple surveys.
4. There were reasonably reliable methods for forecasting the independent variable under consideration.
5. The independent variable not only explained the existing trip destinations, but also explained future changes in the pattern of these trips. If it did not completely, then possible corrective factors were described.

Reliability of Estimating Equations

There are various statistical tests for the reliability of estimating equations derived from regression analyses, of which the standard error of estimate is one of the more important. While the standard error of estimate does indicate the reliability of the estimate, it should also be compared with the amount of error which can be tolerated in the subsequent use of the estimate. In general, it may be said that the estimating equations are more accurate for large areal destinations and a large number of trips. With small areas and a small number of trips, there is a greater chance of error. The estimating equations were derived in a large area, that is, the CATS study area, and for a variety of "average" conditions, but they do not include factors explaining the variation caused by a condition unique to one small locality.

In deriving the estimating equations, the highest correlation coefficient and lowest standard error estimate was the objective, of course. But even where there was a logical causal relationship between the dependent and independent variables, the standard error of estimate was frequently high in comparison to the mean number of trips.

When the relationship between the independent and dependent variables was not directly causal, there was a corresponding decrease in the measures of correlation. Obviously, this was due to the inability to isolate the causal variable and define it. This was particularly true where the trip purpose was of a noncompulsory nature, and where many complex human motivations were involved.

In the regression analysis of trip destinations to the districts, correlation coefficents of 0.95 or better, and standard errors of estimate, which were 25 percent of the mean number of trips to the district of destination, were obtained. At the zonal level, where smaller areas and a smaller number of trips were involved, the correlation coefficients dropped to the 0.85 level, while the standard error of estimate was approximately one-half of the mean number of trip destinations (Table 3).

In some cases, the poor statistical relationships can be explained by error in the input data. That is, there existed a logical relationship between the dependent and independent variables which hardly needed statistical proof, but in the statistical analysis, the measures of correlation were not perfect. It was known that error in the original data

TABLE 3

RELIABILITY OF ESTIMATING METHODS

Trip Purpose and Place of Destination	Mean Number of Trips	Method	Correlation Coefficient	Standard Error of Estimate
		(a) District		
Total	225,719	Autos owned D. U. per acre	0.91	51,796
Home	98,416	Autos owned Pop. per acre	0.98	9,353
Work (mfg.)	16,600	Employment (in mfg.)	0.98	4,600
Social-recreation (all districts except 01)	28,200	Autos owned	0.95	4,400
Personal business	22,940	Med. serv. emp. Autos owned	0.94	4,979
		(b) Zone		
Total	20,913	Autos owned D. U. per 10 acres	0.86	9,507
Home	9,188	Autos owned D. U. per 10 acres	0.97	1,971
Social-recreation	2,620	Autos owned D. U. per 10 acres	0.75	1,505
Personal business	2,133	Med. serv. emp. Autos owned D. U. per 10 acres	0.85	1,334
Personal business and Eat a meal	2,580	Med. serv. emp. Autos owned D. U. per 10 acres	0.86	1,495

existed, but it could not be corrected easily. In the case of such trip purposes as personal business and social-recreation, a number of minor trip purposes was included. The selected independent variable explained the primary component of the trip purpose, but it did not always explain the minor components. The addition of independent variables to explain the minor trip components was assayed, but without notable success.

Another approach, breaking the major trip purpose into its component elements, was feasible to a limited extent at the district level but at the zonal level the number of trip destinations was frequently less than 1,000 at which point error in the trip survey data increased significantly. Actually, at the zonal level it appeared more appropriate to combine minor trip purposes, which constituted less than 5 percent of the total trips, with a major trip purpose. Generally, there was no clear-cut, logical independent variable to explain the minor trip destinations. Further, it was known that data on the trip destinations contained considerable error both in the coding of the trip purpose and in the factoring of the sample.

There is ample room to improve the derived estimating equations, notably those for shopping, recreation, and personal business trips. Primarily, it is a problem of identifying the independent variables which explain these trip destinations. Obviously, there were other variables which should have been examined, but while the CATS files were a rich source of data, they did not contain all. There were, of course, other sources in the Chicago area, but the data had to be coded to the CATS zone and district system before they could be used. Where there was no conversion program, the data were essentially useless and unavailable.

Sample Size

The regression analysis of trip destinations by purpose was carried out at the district level and at the zonal level. Much of the preliminary analysis of district trips was done on a desk calculator. Subsequent district analyses and all of the zonal analyses were computerized.

There were 44 districts in the CATS study area. These in turn were subdivided into 582 zones. All of the districts were used, although the CBD district 01 was frequently omitted in the regression analyses to eliminate the effect of the CBD, or to isolate the independent variable which was needed to explain the variation created by the CBD.

Of the 582 zones, 113 were eliminated completely from the zonal analysis. In these 113 zones the total number of trip destinations was less than 1,000, the number of trips to home was less than 500, or there was no residential land. The sample size of the zonal analysis was thus 469. The four zones in the CBD district 01 were again identified and the analysis was made with and without the four zones. Where the independent variable involved a measure of floor area in a particular land use, zones in which floor area was inventoried were eliminated for that particular regression analysis.

In the analysis of specific purposes, it was known that certain zones contained a major trip-generating facility. For example, the estimating equation for social-recreation did not explain the variation in trips caused by the presence of a major-league ball park in a specific zone. As in the case of the four CBD zones, the regression analysis was made with and without the selected zones having a major trip-generating facility.

In the analysis of shopping trips, the total sample size was reduced to 367 zones by eliminating zones in which the total number of shopping destinations was less than 100. Much of the shopping trip analysis was also made with 204 zones where retail floor area had been inventoried.

Data Sources

With a few exceptions, data for the regression analysis were taken from the CATS 1956 trip and land-use surveys. The independent variable was the number of internal person trip destinations to the district and zone by trip purpose. The method for making the trip survey and factoring the sample is explained in CATS Final Report, Volume I, Survey Findings, and various supplementary reports on the trip survey. The sampling variability in the survey also is estimated there, and this was used as a basis for selecting zones in the zonal regression analysis.

The trip survey data were originally recorded on home interview cards. Subsequently, in 1964, these data were transferred to tape, but in the 8-year interval, 27 cards representing 957 trips either were badly damaged or were lost. The regression analysis was based on data from old tabulations made from the home interview cards. Many of the trip characteristics, particularly the land-use destinations and the district interchange of trips, were obtained from the tapes.

The home interview of trips produced data on population and automobile ownership which were used extensively as independent variables. There also were other characteristics of the population, such as dwelling units, licensed automobile drivers, school age population, and automobile ownership ratios, which were examined as independent variables.

In the 1956 land-use inventory, the amount of land area in six major categories of land use—residential, manufacturing, commercial, transportation, public buildings, and public open space—were inventoried. Floor area was inventoried for the inner built-up areas of the study area and summarized in detail. These data were used extensively in the shopping trip analysis and to varying degrees in the analysis of the other trip purposes.

At the time of the 1956 survey, there were no available data on employment by small area within the CATS study area. Since that time, the Illinois State Employment Service began publication of biennial reports on employment covered by the Unemployment Compensation Act in the postal zones of Chicago and the larger municipalities of Cook and Du Page Counties. These data were used to obtain estimates of manufacturing, retail trade, wholesale trade, finance, and construction employment by district in the CATS study area. The problems of time, geographical area, and definition of employment activity in these estimates of employment are outlined briefly in a later section.

The other data sources which were utilized are noted under the appropriate trip purpose. None was used extensively. The suggestions for further study indicate that greater utilization of school enrollment and retail sales data would have been appropriate.

TOTAL TRIPS AND TRIPS HOME

From the 1956 trip survey, CATS estimated a total of 9,930, 681 internal person trips which had destinations in the 44 districts of the study area. Of the nearly ten million total trips, 4,328,569 were for the purpose of returning home. Total trips are the summation of all trip purposes, and the destinations are many and varied. Trips home have only one type of destination, residential land, but these destinations are located throughout the study area. Because 469 zones out of 582 were used in the zonal analysis, the total internal person trip destinations was 9,808,353, while the number of trips home was 4,308,996.

Because trips home constitute nearly half of the total trips, and because the trip home was frequently the return from another trip purpose, the average trip lengths and modes of travel were approximately the same for total and home trips. Nearly 75 percent of both types of trips were made by automobile, either as a driver or as a passenger. Bus was the second most important mode. The mean airline trip length was slightly over 4 miles.

To develop estimating and forecasting methods for total trips and trips home, the methods used previously by CATS were readily suggested. The definitions of total trips were the same. Trips home and trips to residential land were approximately the same. Trips to residential land included all trips home, and a few other trips where the purpose was social, domestic work, or personal business. Trips home constituted 80 percent of the trips to residential land.

The CATS forecast equation expressed total trips as a ratio—specifically, trips per dwelling unit. Similarly, the equation for residential trips was expressed in another ratio, trips per capita. In this analysis, the derived equations were expressed in terms of total trips to the place of destination. This difference may be considered minor, for it is possible to perform the necessary algebra and convert either equation. What is relevant are the factors included in each equation.

CATS selected cars per dwelling unit and the log of net residential density (dwelling units per acre) as the variables from which to estimate total trips. Residential trips were estimated from cars per person and the actual net residential density (persons per residential land).

The variables used by CATS and a few other alternative expressions for residential characteristics were tested in the district analysis. The tested resident characteristics were population, dwelling units, automobiles owned, licensed automobile drivers, residential land, and ratios expressing net residential density, car ownership rate, and the percentage of the population licensed to drive. At the zonal level, the independent variables were limited to the number of dwelling units, automobiles owned, dwelling units per acre, and autos per dwelling unit.

The regression analysis at both the district and zonal level reaffirmed the CATS method, and that automobile ownership and net residential density were appropriate independent variables for estimating total trip destinations and trips home. In the district analysis, where additional independent variables were tested, it was found that there were acceptable alternatives for automobile ownership and net residential density (Tables 4 and 5). The alternatives, however, were similar to the variables used by

TABLE 4

STATISTICAL MEASURES OF CORRELATION FOR TOTAL PERSON TRIP DESTINATIONS AND SELECTED INDEPENDENT VARIABLES

Sample Size	Independent Variables	Intercept	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate	Mean Number of Trips
44 districts	Autos owned	-14,171	7.036	0.711	86,431	225,719
	Dwelling units per acre		587.026	0.909	51,796	
44 districts	Population	-189,310	1.874	0.763	79,415	225,719
	Percent of pop. lic. drivers		4,589.487	0.946	40,126	
44 districts	Licensed auto drivers	-5,179	4.552	0.773	77,925	225,719
	Dwelling units per acre		532.962	0.934	44,574	
44 districts	Dwelling units	-122,579	4.705	0.785	76,181	225,719
	Percent of pop. lic. drivers		4,004.331	0.931	45,300	
469 zones	Autos owned	1,855	5.811	0.710	13,314	20,913
	Dwelling units per 10 acres		9.101	0.865	9,507	

TABLE 5

STATISTICAL MEASURES OF CORRELATION FOR TRIPS TO HOME AND SELECTED INDEPENDENT VARIABLES

Sample Size	Independent Variables	Intercept	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate	Mean Number of Trips
44 districts	Autos owned	-5,424	3.268	0.979	10,211	98,416
	Population per acre		49.915	0.983	9,353	
44 districts	Licensed auto drivers	4,533	2.068	0.991	6,842	98,416
	Dwelling units per acre		-13.384	0.991	6,756	
44 districts	Autos owned	4,338	3.193	0.979	10,211	98,416
	Dwelling units per acre		211.887	0.164	10,195	
44 districts	Population	-63,600	0.845	0.913	20,444	98,416
	Autos per dwelling unit		68,782.134	0.956	14,842	
44 districts	Dwelling units	58,275	2.197	0.858	25,697	98,416
	Autos per dwelling unit		81,323.555	0.934	18,055	
469 zones	Autos owned	-114	3.240	0.964	2,001	9,188
	Dwelling units per 10 acres		0.351	0.966	1,971	

CATS, and can be considered as redefinitions of essentially the same thing. There was also a high degree of covariance between these alternatives (Table 6).

The analyses were made with the central business district and without. The statistical measures of correlation were approximately the same. There were, however, marked differences between the district and zonal measures of correlation. The coefficient of multiple correlation decreased, while the standard error of estimate as a percent of the mean number of trips doubled.

In conclusion, it may be stated that total trip destinations and trips home may be estimated from the number of automobiles owned and residential density. There are acceptable alternatives which can be recommended. Among the two or three best, the selection may be guided by such practical considerations as the availability of data and the ability to forecast the selected independent variables. That one method appears better from the statistical measures of correlation probably is due to chance and error in the basic data.

TABLE 6

SIMPLE CORRELATION COEFFICIENTS BETWEEN INDEPENDENT VARIABLES DEFINING RESIDENT CHARACTERISTICS

Variable	Population	Dwelling Units	Autos Owned	Licensed Auto Dr.	Percent Pop. Lic. to Drive	Auto/ D. U.	D. U./ Acre	Pop./ Acre
(a) All Districts								
Population	1.000	0.976	0.862	0.915	-0.394	-0.367	-0.147	0.231
Dwelling units		1.000	0.800	0.868	-0.330	-0.464	-0.040	0.347
Autos owned			1.000	0.980	-0.202	0.011	-0.270	-0.147
Licensed auto drivers				1.000	-0.188	-0.077	-0.205	-0.035
Percent pop. licensed to drive					1.000	0.110	0.735	0.167
Autos per D. U.						1.000	-0.560	-0.909
D. U. per acre							1.000	0.767
Population per acre								1.000
(b) All Districts Except 01								
Population	1.000	0.979	0.851	0.908	-0.323	-0.537	0.580	0.527
Dwelling units		1.000	0.796	0.865	-0.338	-0.594	0.668	0.594
Autos owned			1.000	0.979	0.094	-0.125	0.121	0.053
Licensed auto drivers				1.000	0.952	-0.204	0.237	0.165
Percent pop. licensed to drive					1.000	0.890	-0.760	-0.829
Autos per D. U.						1.000	-0.896	-0.920
D. U. per acre							1.000	0.984
Population per acre								1.000

The relationships of automobiles owned and net residential density to total trips and trips to home or residential land have been described in CATS reports and in other transportation studies. This hardly needs repetition, except to mention that these two variables are better measures of trips home than of total trips. Total trips include other trip purposes which are not related to home, except as a place of origin. The non-home component of total trips is related to many nonresidential characteristics, such as the place of work.

The estimating equations derived from the 1956 trip survey in the Chicago area indicated that there was a total of approximately 7 person trips per automobile owned, 2 per person, or 4½ per licensed automobile driver. In the case of trips home, there were approximately 3 person trips per automobile owned and 2 per licensed automobile driver. These ratios then were augmented by net residential density in the district of destination. Forecasts should reconsider these ratios and any conditions which may alter them.

WORK TRIPS

Of the trip purposes, those for the purpose of working can be ranked among the most important. In 1956 there were two million internal person work trips within the CATS study area. Approximately 20 percent of the total person trips occurring on a typical weekday were work trips.

Only trips to home exceeded the number and proportion of work trips, but approximately one-third of such trips home were return trips from work. The concentration of work trips during a couple of hours in the morning and their convergence on a small number of places, and their counterpart return trips home, created the highest weekday traffic volumes. Problems created by these daily work trip volumes make it imperative to have accurate estimates of work trips.

It is obvious that the work trip would not be made unless a job existed, and that the best measurement of the number of jobs is total employment and its distribution by location. However, during its survey and analysis phase, CATS did not have access to employment data by small areas such as the districts into which the study area was divided. Only for the entire study area were estimates of employment prepared. Estimates of manufacturing employment density for small areas were made, but they were based on CATS own trip and land-use surveys. These estimated manufacturing work trip and employment densities subsequently were used to forecast trips to manufacturing land, because work trips constituted 90 percent of the total trips to manufacturing land.

The estimates and forecasts of total employment in the Chicago SMA and CATS study area were prepared during the years 1957-58. Since then, the Illinois State Employment Service began to publish a series of reports on employment covered by the Unemployment Compensation Act by industry and postal zone in Chicago and the larger municipalities and sections of Cook and Du Page Counties. As an independent estimate of employment by small areas in the CATS study area, it is the best, although there are a number of limitations involving time, geographic area, and definition of employment.

The Problem of Time

When the Employment Service began the compilation of covered employment data by postal zone, it was done biennially for the first quarter of the odd years. The 1956 CATS trip survey fell between the first two reporting periods, specifically March 1955 and March 1957. Covered employment data for 1957 were selected as the best approximation of the 1956 employment picture, because estimates of total employment in the Chicago SMA showed little change between 1956 and 1957. There was a greater difference between the 1955 and 1956 totals.

The Problem of Area

While CATS districts and postal zones were approximately the same in number and size, there was little agreement on boundaries. It appeared to be only by accident that the same street was used as a boundary by both CATS and the Post Office. Because of the almost complete lack of coterminity in boundaries, it was necessary to develop a method for converting employment by postal zone to CATS district. The method which was developed assumed that employment in a postal zone was distributed in the same manner as the land or floor area of the land use which corresponded to the industrial employment under consideration. The CATS coordinate system used in the land-use inventory made it possible to compile reasonably accurate totals of land use in the postal zones, which then could be apportioned to the overlapping CATS district. (Only one major and unavoidable "error" is known. It affects the Central Business District, and for this reason any employment total for CATS districts 01 and 11 should be added together.) These portions or ratios then were used to subdivide postal zone employment. In converting manufacturing employment from postal zone to CATS district, it was found that there was no significant statistical difference between using land area or floor area in the conversion method.

The Problem of Definition

By definition, covered employment is not total employment; it is approximately three-fourths of total nonagricultural employment in the Chicago area. In the manufacturing industries, 99 percent of the employees are covered. In the construction, communication and public utilities, wholesale and retail trade, and mining industries, coverage is over 90 percent; 83 percent of total finance employment is covered, while 50 to 60 percent is covered in transportation and service trades. Government is one notable nonagricultural industry where there is no coverage.

TABLE 7

1956 INTERNAL FIRST WORK TRIPS AND EXTERNAL WORK TRIPS TO MANUFACTURING LAND, AND TWO ESTIMATES OF 1957 COVERED MANUFACTURING EMPLOYMENT BY CATS DISTRICT

(in thousands)

CATS District	Work Trips to Manufacturing	Estimated Manufacturing Employment	
		Land Area Method	Floor Area Method
01 & 11	151.9	176.4	182.7
21	15.7	8.7	10.4
22	27.9	43.0	39.9
23	21.4	26.9	26.7
24	11.0	9.2	11.1
25	22.3	26.9	24.5
26	20.7	24.7	25.6
27	6.7	7.6	6.7
31	8.0	8.5	8.6
32	28.4	33.9	35.3
33	40.3	45.2	45.0
34	37.0	28.7	37.4
35	26.2	34.0	33.4
36	15.8	19.1	16.3
37	3.4	1.8	2.1
41	11.2	8.6	
42	15.4	13.3	
43	16.1	13.4	
44	25.6	32.7	
45	10.4	30.8	
46	12.2	11.1	
47	9.1	8.9	
51	7.9	4.7	
52	15.2	15.1	
53	20.8	18.9	
54	12.0	24.2	
55	24.2	29.5	
56	2.4	2.7	
57	23.4	18.0	
61	0.6	2.0	
62	3.7	8.3	
63	8.7	9.9	
64	0.8	1.4	
65	3.7	6.4	
66	11.5	17.2	
67	18.2	27.7	
71	1.3	0.5	
72	1.9	7.5	
73	1.5	2.3	
74	1.1	1.0	
75	0.8	0.2	
76	8.6	6.5	
77	7.7	10.3	
Total	712.7	827.6	

The conversion of employment by postal zone to CATS district involved matching SIC industry definitions with the CATS definitions of land use. At the two-digit level of each, it was possible to obtain a reasonable matching without serious discrepancies between definitions.

A more serious problem arose in comparing definitions of employment and work trips. Covered employment is all covered jobs; it does not indicate the number of multiple job holdings of an individual worker, nor does it indicate normal absenteeism due to illness or vacation. Work trips were defined as first work trips. Trips to a second job or trips made subsequent to the first—trips of the traveling salesman—were omitted in this definition. By implication, persons staying home because of illness or other reason did not make a work trip.

Estimated Employment and Work Trips

Estimates of employment by CATS district were made for the selected industries of manufacturing, retail trade, wholesale trade, finance, and

TABLE 8

ESTIMATES OF MARCH 1957 COVERED EMPLOYMENT FOR SELECTED NONMANUFACTURING INDUSTRIES AND 1956 INTERNAL FIRST WORK TRIPS TO CATS DISTRICTS 01 THROUGH 37

(in thousands)

CATS District	Retail Trade[a]		Wholesale Trade[b]		Finance[c]		Construction[d]	
	Employment	Work Trips	Employment	Work Trips	Employment	Work Trips	Employment	Work Trips
01	38.8	39.4	21.4	8.8	62.5	50.7	6.5	2.9
11	36.8	34.4	68.7	15.3	13.7	4.7	11.9	3.1
21	4.0	7.5	2.2	1.0	1.7	1.0	2.2	0.4
22	3.0	4.9	3.4	1.1	0.8	0.3	2.5	0.9
23	3.8	5.0	3.0	1.2	0.5	0.5	2.3	0.7
24	10.4	11.4	1.2	0.3	0.2	0.4	0.9	0.3
25	0.9	3.4	4.1	1.2	0.2	0.1	0.4	0.5
26	3.0	5.5	3.9	1.6	0.3	0.2	0.2	0.4
27	2.0	3.7	2.3	0.7	0.7	0.5	0.7	0.9
31	3.2	5.8	1.2	0.5	3.4	2.0	1.8	0.5
32	2.9	6.3	2.0	1.1	0.4	0.5	2.1	0.7
33	5.7	8.4	4.0	1.2	1.0	1.1	3.7	1.2
34	4.2	5.3	2.4	0.8	0.4	0.4	1.5	0.4
35	2.2	3.1	3.5	1.4	0.4	0.3	1.0	0.8
36	6.4	2.9	1.1	1.0	0.7	0.5	0.9	0.3
37	3.9	7.5	0.6	0.3	1.6	1.3	0.6	0.3
Total	131.2	154.6	126.0	37.5	88.5	64.5	39.2	14.3

[a] SIC Code 53-59; CATS Code 50, 51, 52, 53, 54, 55, 56, 57, 58, 59.
[b] SIC Code 50-52; CATS Code 70, 71, 73.
[c] SIC Code 60-67; CATS Code 60.
[d] SIC Code 15-17; CATS Code 74, 75, 77.

construction. Estimates for other industries were not prepared because of the low coverage in these industries. Because estimates for retail and wholesale trade, finance, and construction had to be based on the distribution of floor area, the estimates could only be prepared for the districts in the inner rings where floor area had been inventoried. The estimates are given in Tables 7 and 8.

The estimates of employment were correlated with first work trips. In all cases the correlation coefficient was 0.95 or better. The regression coefficient, b, in the estimating equation, was a reflection of both normal absenteeism and the degree to which employment in the industry was covered (Table 9).

The standard error of estimate was relatively high. It can be explained partially by the problems of matching dates, industry definitions, and the conversion of employment data from postal zones to CATS districts. Another source of error was in the CATS survey of trips. A comparison of manufacturing work trip densities in CATS districts and manufacturing employment densities by postal zones revealed some evidences of underreporting of trips and overfactoring of the sample. While the location of these possible errors can be described in considerable detail, they cannot be proved conclusively.

The statistical analyses support the common-sense conclusion that work trips can be estimated from employment. While no specific estimating equation is offered, it can safely be assumed that work trips will be 80 to 90 percent of total employment. CATS used employment or rather employment density in estimating work and total trips to manufacturing land. The statistical analyses confirm the validity of this method. It is further recommended that not just manufacturing work trips but all

TABLE 9

STATISTICAL MEASURES OF CORRELATION BETWEEN ESTIMATED EMPLOYMENT AND WORK TRIPS

Industry	N or Number of Districts	Mean Number of Trips[a]	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate[a]
Manufacturing	43 (01 and 11; 21 thru 77)	16.6	0.84	0.98	±4.6
Retail and wholesale trade	16 (01 thru 37)	12.0	0.51	0.95	±4.2
Finance	16 (01 thru 37)	4.0	0.80	0.99	±1.4

[a] Thousands of trips.

work trips be estimated from employment. The implication of this recommendation is the necessity of preparing forecasts of future total employment by small areas. This is no easy task, but it should be noted that an industrial breakdown by small area is not required, although it might be obtained in the employment forecasting procedures as a check on reasonableness.

The implication of this recommendation in land-use definitions appropriate to transportation is the identification of work places by number of employees and location. Work places mean all, not only manufacturing, but also commercial, government, and any other place where people work.

While the employment estimates and statistical analyses were made at the district level only, there is no reason to question that similar relationships and conclusions could not be found for CATS zones.

A Footnote on Employment Density and Work Trips

During the preparation of the employment estimates by CATS districts, considerable data on employment density became available. Since employment density is an alternate to total employment in estimating work trips, these derived densities were examined with the view of developing a second estimating method. Without going into the details of the specific analyses, it was concluded that the employment method was preferred because of its directness, although employment density could be used to supplement or to check the total employment method. There are a number of reasons for this:

1. More work is entailed in collecting and forecasting land or floor area on which employment occurs.
2. Forecasting employment density involves several pitfalls. Any forecast based on current employment density patterns will encounter such problems as employment definition (average annual or seasonal, total or main shift), possible greater variation of employment densities within an industry than between industries, and differences in the utilization of space, such as an establishment making do in old cramped quarters, or an establishment holding reserve space for future expansion.
3. Forecasts of future employment density must consider the effects of automation and other similar changes which will affect employment density. Automation will, naturally, also be a factor in any forecast of total employment.

SOCIAL-RECREATION TRIPS

Of the 9,930,681 weekday internal person trips surveyed by CATS in 1956, 1,251,785 or 12.6 percent were social-recreation trips. Social-recreation trips included two basic, although related, types of trips—social trips to visit friends and relatives, and recreation trips to go to the ball game, to attend the theater, or to play golf. If a residential land use destination were reported, it was assumed that a social trip was made. The destinations of recreation trips were many and were dependent on the individual recreational objectives of the trip-maker, season of the year, and the weather. Summer suggested baseball, swimming at the beach, and escape from the heat to an air-conditioned movie. Cooler autumn weather suggested the opening of the new football and theater season. Winter snow and ice curtailed most outdoor recreation, except activities involving snow and ice. Social trips to visit friends, however, were a year-round activity.

These social and recreation activities were reflected in the land-use destinations, which during the six warmer months of 1956 received the following proportion of trips:

Residence	55.1 percent
Indoor amusement	12.3
Eating and drinking places	5.8
Outdoor amusement, n. e. c.	5.4
Public parks	4.7
All other public buildings (primarily Y's and similar institutions)	2.5
Schools	2.4
Swimming pools	1.5

Golf courses	1.5 percent
Race tracks & stadia	1.5
Churches	1.1
Museums, art galleries, zoos, arboretums	1.0
All other uses	5.2

In the CBD district 01, 56 percent of the social-recreation trips had indoor amusement destinations, that is, first-run movies, legitimate theater, concerts, night shows, and similar facilities. No other land-use destination had a significant proportion of the total recreation trips.

As might be expected, weekday recreation trips occurred primarily during the evening hours, and Friday was the most popular day. The automobile was by far the most important mode of transportation in the social-recreation trips. The number of social-recreation trips per family increased with each increase in the total number of trips made by the family.

Among these general characteristics of the social-recreation trips, there were three which had relevance to the derivation of estimating methods for these trips:

1. As a noncompulsory type of trip, the number of social-recreation trips depended on the availability of leisure time and income. During the week, work and school hours effectively reduced the time available for making social-recreation trips. Since most social-recreation activity also costs money—admission tickets, special clothing and equipment, or carfare—the number of social-recreation trips also was governed by family or individual income. It should be noted that the number of leisure hours during the week and on the weekend differed greatly. Since the CATS trip survey was made during the week, any estimating method derived from the data is applicable only to weekday social-recreation travel. The other constraint on social-recreation trips, income, can be approximated by automobile ownership.
2. In the social trips, which constituted approximately half of the total social-recreation trips, there was a strong tendency to exchange visits with friends in the same socio-economic level. This characteristic was particularly evident at the district level where some equilibrium in the district interchange of trips was observed.
3. Some districts and zones had a unique recreational facility which attracted a large number of trips. The first-run movies, theaters, and night shows of the CBD district 01, and the major-league ball parks, race tracks, amusement parks in other zones and districts could be identified as spectator facilities which drew large crowds. Not to be overlooked were the museums, zoos, beaches, some schools and Y's which attracted trips from outside the immediate vicinity. Size of the recreation facility, measured in terms of land area or floor area, was not always relevant. The basement floor area of a night club was not easily compared with the 100 acres of a zoo. The hundreds of acres of a forest preserve do not attract as many trips as the opening game at the ball park located on one city block. Measures of capacity, such as seating capacity,

TABLE 10

STATISTICAL MEASURES OF CORRELATION FOR SOCIAL-RECREATION TRIPS AND SELECTED INDEPENDENT VARIABLES

Trip Purpose	Sample Size	Independent Variable	Inter-cept	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate	Mean Number of Trips
Soc. -rec.	43 districts	Autos owned	0.075	0.90	0.95	4.4	28.2
Social	44 districts	Autos owned	-0.021	0.52	0.94	2.7	15.7
Soc. -rec.	469 zones	Autos owned	408	0.739	0.733	1,559	2,620
		Net res. density		0.402	0.754	1,505	
Soc. -rec.	449 zones	Autos owned	268	0.759	0.842	1,101	2,456
		Net res. density		0.180	0.843	1,100	

43 districts: CBD district 01 omitted; all trip data expressed in thousands of trips.

44 districts: all districts; all trip data expressed in thousands of trips; social trips are all social-recreation trips to residential land, i.e., visiting friends.

469 zones: all zones used in zonal analysis.

449 zones: all zones except 20 with major recreational facility; these are: 01001, 01002, 01003, 01004, 11013, 11015, 21023, 27045, 31051, 32055, 34071, 37087, 45150, 53216, 54225, 62294, 65329, 71420, 72454, and 76540.

appeared to be more appropriate, but the CATS files did not contain such data. Furthermore, methods for measuring the capacity of recreation facilities were still in the preliminary stages of development in the field of recreation planning.

Estimating Equations

In the regression analysis of social-recreation trips, the first two characteristics—propensity to make social-recreation trips and the equal exchange of social trips—were utilized. In the absence of appropriate variables to explain trip destinations to major recreation facilities, the regression analyses were made by omitting those trips at the district level, or by omitting entire zones with these recreation facilities at the zonal level. The more important findings are summarized in Table 10. A number of observations can be made from these findings:

1. Statistically, the better equations are those for the district. At the zonal level, there was greater chance of data error, because of the smaller area and smaller number of trips.
2. Automobiles owned was a good index of income. It was, however, a residential characteristic of the district or zone of destination, and not necessarily a characteristic of the person making the trip. But because half of the trips were social, and because people had the tendency to visit others of a similar economic level, atuomobile ownership in the district of destination could represent the characteristics of the trip-maker. The district estimating equation for all social-recreation trips stated that, on an average weekday, there was approximately one person trip per one automobile owned. With nearly four resident persons per automobile, the average person made 1¼ social-recreation trips per week (Monday through Friday). This was not unreasonable. Not every automobile was used on the social-recreation trip, for half of those using the automobile on the trip rode as a passenger.
3. The computer program used in the zonal analysis was not set up for simple correlation. It was evident, however, that the addition of a second independent variable, net residential density, contributed little or nothing to the reliability of the estimating equation, and that one independent variable, automobiles owned, was sufficient. This is not to say that a second independent variable was not needed, but that the second independent variable should not be another residential characteristic.
4. The equations are best for estimating social and neighborhood recreational trips. They are most inadequate for estimating trips to a major recreation facility.
5. While automobile ownership provided a good approximation of social-recreation trip destinations to large areas, use of automobile ownership in trip forecasts should be re-examined in view of two developments. First are the changes in the relationship between automobile ownership ratios and the number of social-recreation trips. Second, and more important, are changes in the number of leisure hours, such as would occur with a shorter working day or week. A 6-hour working day or a 4-day work week could easily mean more recreation trips whose number approached the scale of weekend recreation trips.
6. It was evident that the equations were unsatisfactory in estimating trips to major recreation facilities. In transportation problems which involve estimates of trips to a particular major recreation facility, it is recommended that specific studies be made on that facility. There is no question of the need for additional analyses of recreation trips to the major facilities. Use of independent variables expressing the facility's capacity appears to be an appropriate approach, but the definition of capacity measures have not been developed fully.

PERSONAL BUSINESS TRIPS AND TRIPS TO EAT A MEAL

In the 1956 CATS survey of internal trips, there were 1,009,362 personal business trips and 210,983 trips to eat a meal. They constituted respectively 10 and 2 percent of the total internal person trip destinations. Both trip purposes were analyzed separately at the district level, but preliminary studies to combine the two trip purposes were made at the district level, and continued at the zonal level.

As a group, the one million trips for personal business were a catchall which included such objectives as obtaining medical service, attending church, getting a haircut, applying for a job, and paying bills. The group also included all other trip purposes

TABLE 11

PERCENTAGE DISTRIBUTION OF PERSONAL BUSINESS TRIP DESTINATIONS BY LAND USE

Land Use	Percent
Residence	16
Medical service	14
Retail trade	10
Personal service	10
Finance, insurance, real estate	9
Church	7
Hospital	7
Business, auto, and miscellaneous service	6
School and nonprofit institutions	5
Federal, state, and local government	5
All other uses	11

not classified elsewhere, and trips for which an inadequate response on purpose had been obtained.

The nature of the personal business trip can be deduced from the land-use destination; that is, there is a strong presumption that a person was visiting a doctor or dentist if the destination were medical service and that a person was obtaining auto servicing if the destination were a filling station or repair garage. All of the land uses can be a place for personal business, but no one land use was dominant. The variety of objectives and land-use destinations is illustrated in Table 11.

Because of the variety and multiplicity of objectives and destinations, personal business trips cannot be characterized by one or two objectives or destinations. However, most personal business trips were short, because many of the destination places were relatively ubiquitous. There were, however, some objectives and destinations which were few in number and specialized in character. Among these latter were offices of medical specialists, transportation terminals, and government offices.

For these reasons it was theorized that most personal business trips were local in nature and that the number of trip destinations could be estimated from knowledge of the resident characteristics in the district or zone of destination. This theory was valid for the bulk of the personal business trips where the origin and destination were in the same district or zone, or where the interchange of trips was nearly equal. For the remaining nonlocal personal business trips, a description of the nonresidential destinations was required.

Trips to eat a meal, as defined in the trip survey manual, were for the specific purpose of eating. If, however, there were social or recreational aspects, the trip was classified social-recreation. Trips to a friend's for dinner, or "to eat out" at the new restaurant should have been coded social-recreation. Because this distinction was not always clear, miscoding of the trip purpose was high. The noon-hour meal of the worker or school child, which involved travel, was a legitimate trip to eat a meal. So, also, was the businessman's two-hour luncheon. The primary land use at destination was retail, or, more specifically, eating and drinking places. The secondary destination was home.

It was possible to make some conjectures about the characteristics of the person making the trip to eat a meal, and thus to select appropriate independent variables to be used in the analysis. Workers and school children can make trips to eat a meal, but many workers—clerical, sales, skilled, and most school children—brought a lunch, ate at the cafeteria in the building, or walked to the nearest restaurant or lunch stand. By the CATS definition of trips, no trip was made. On the other hand, executives, managers, and professional workers may have the time and means for making a trip to eat a meal. It thus was logical to relate trips to eat a meal with this group of workers, or an index of income.

Analysis

Except for keeping a medical appointment or luncheon engagement, personal business trips and trips to eat a meal do not have an obligatory element. The motivations and objectives for making these trips, however, were many and not easily defined. It appeared that the total number of trips was related to total population rather than personal income of the trip-maker. The type of personal business trip may have been related to income.

While measures of population or income were adequate for estimating the total number of personal business trips and district trip destinations which were local in nature,

TABLE 12

MULTIPLE CORRELATION OF PERSONAL BUSINESS TRIPS
AND SELECTED FACTORS FOR ALL DISTRICTS
(Mean Number of Trips: 22,940)

Independent Variable	Intercept	Regression Coefficient	Coefficient of Multiple Correlation	Standard Error of Estimate
Licensed auto drivers	-2,484	0.480	0.643	11,208
D. U. acre		80.845	0.926	5,591
Population	-30,014	0.205	0.623	11,441
Percent pop. lic. to drive		679.874	0.935	5,259
Dwelling units	13,091	0.364	0.666	10,918
Autos per D. U.		-4,415.706	0.670	10,988
Autos owned	-3,010	0.729	0.561	12,109
D. U. per acre		86.109	0.896	6,578
Medical service emp.	2,203	24.020	0.842	7,894
Autos owned		0.413	0.942	4,979
State and local govt. emp.	-851	6.277	0.714	10,246
Autos owned		0.615	0.956	4,348
Retail trade emp.	805	0.148	0.828	8,199
Autos owned		0.423	0.935	5,266
Autos owned	6,601	0.421	0.561	12,109
Hospital emp.		3.887	0.695	10,640
Medical service emp.	...	...	0.842	7,894
Autos owned			0.942	4,979
State and local govt. emp.			0.957	4,368
Hospital emp.			0.960	4,264
Finance emp.			0.962	4,178

other independent variables were required for nonlocal trips and for the number of zonal destinations. The land-use destinations of personal business trips provided clues to the selection of appropriate independent variables which explained nonlocal destinations. Land areas in the six major land-use categories were not used because they were gross aggregations of a number of activities which did not correspond to the aggregation of the personal business objectives. Floor area on which these activities occurred was limited to the central rings of the study area. Estimated employment, that is, first work trips to selected service activities, was used because data could be used in detail or in aggregate in the entire study area. Estimates of employment from covered employment were not used because, in a number of cases, the selected employment category was not included under covered employment, or because the percentage of coverage was low. The data on first work trips, on the other hand, were readily available in the CATS data files on trips.

District Analysis

Personal business trips to the district of destination were correlated with several resident characteristics and with combinations of resident characteristics and estimated employment in selected destination places. The better combinations are summarized in Table 12. Independent variables defining resident characteristics in the district of destination produced rough estimates of the number of trips to the districts, except the CBD district 01. The combination resident characteristic and estimated employment in one of the service trades or professions gave better results, and it was not necessary to exclude the CBD from the analysis. The noteworthy finding at the district level was that two or more independent variables describing the destination places did not produce much improvement over automobiles owned and an independent variable describing one destination place. The reason was not difficult to discover—there was considerable covariance between the various places of destination. Table 13 shows the simple correlation coefficients between estimated employment in selected service trades and professions.

Because covariance between the places of destination existed at the district level, it is recommended that one place be selected. Since the places of destination have been

TABLE 13

SIMPLE CORRELATION COEFFICIENTS BETWEEN INDEPENDENT VARIABLES

Employment in:	Personal Business	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
(a) All Districts												
1. Medical service	0.842	1.000	0.928	0.787	0.498	0.694	0.736	0.858	0.621	0.864	0.921	0.955
2. State and local govt.	0.595		1.000	0.896	0.383	0.609	0.865	0.963	0.548	0.957	0.963	0.922
3. Finance	0.549			1.000	0.017	0.411	0.994	0.974	0.156	0.869	0.800	0.757
4. Hospital	0.555				1.000	0.652	-0.057	0.195	0.874	0.392	0.558	0.574
5. Education	0.758					1.000	0.348	0.490	0.596	0.569	0.680	0.702
6. Legal service	0.479						1.000	0.959	0.086	0.842	0.757	0.700
7. Offices	0.605							1.000	0.363	0.945	0.908	0.842
8. Post Office	0.586								1.000	0.549	0.705	0.667
9. Federal	0.608									1.000	0.956	0.884
10. Non-profit	0.722										1.000	0.941
11. Retail trade	0.828											1.000
(b) All Districts Except 01												
1. Medical service	0.829	1.000	0.867	0.818	0.828	0.696	0.540	0.816	0.824	0.683	0.828	0.911
2. State and local govt.	0.695		1.000	0.706	0.867	0.668	0.366	0.947	0.947	0.848	0.941	0.888
3. Finance	0.830			1.000	0.712	0.700	0.566	0.689	0.674	0.536	0.697	0.826
4. Hospital	0.664				1.000	0.718	0.360	0.891	0.884	0.830	0.924	0.861
5. Education	0.723					1.000	0.478	0.617	0.607	0.563	0.688	0.691
6. Legal service	0.559						1.000	0.327	0.328	0.184	0.347	0.451
7. Offices	0.621							1.000	0.996	0.892	0.985	0.859
8. Post Office	0.624								1.000	0.897	0.984	0.852
9. Federal	0.453									1.000	0.908	0.774
10. Non-profit	0.639										1.000	0.884
11. Retail trade	0.792											1.000

defined in terms of first work trips, which can be regarded as employment, the selection can be guided by such practical criteria as ease in collecting current employment data and the reliability of the forecasts.

Trips to eat a meal were correlated with automobiles owned, and with an estimate of professional and managerial workers. With a correlation coefficient of 0.83 and a standard error of estimate nearly one-half of the mean number of trip destinations to the district in each case, it can be said that a relationship existed, but that the estimating method was not particularly reliable. With the smaller number of trips and probably higher sampling error in the data, it was concluded that it would be futile to analyze trips to eat a meal at the zonal level, and that it was preferable to combine these trips with a major trip purpose. Combination with personal business trips was logical, for this group of trips included a miscellaneous group of trips. Combination with social-recreation trips was not illogical, if trips to eat a meal had a recreation element.

Both combinations were examined at the district level and at the zonal level. Although the difference in statistical measures of correlation showed no appreciable change after the combination was made, there was an insignificant improvement when trips to eat a meal were added to personal business trips, while the measures of correlation remained the same or showed an insignificant decrease when trips to eat a meal were added to social-recreation trips.

The combination of trips to eat a meal and personal business trips was based more on definition of the trip purposes. The statistical analysis did not justify the combination so much as it showed that the combination could be made without undermining the reliability of the estimating equations for personal business trips.

Zonal Analysis

The zonal analysis of personal business trips and trips to eat a meal was similar to the district analysis. The number of resident characteristics used was reduced, and the estimated employment in a few other destination places was examined. Tables 14 and 15 summarize the important findings from the regression analysis.

It will be noted that the measures of correlation are lower at the zonal level, but this had been anticipated. It had been expected that several independent variables would be required but, surprisingly, three variables produced the best measures of correlation,

TABLE 14

STATISTICAL MEASURES OF CORRELATION FOR PERSONAL BUSINESS TRIPS IN 469 ZONES

(Mean Number of Trips 2,133)

Variable[a]	Intercept	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate
Per. ser. emp.			0.729	1,737
Med. ser. emp.			0.796	1,537
Auto owned			0.828	1,426
State and local govt. emp.			0.874	1,234
Finance emp.			0.899	1,164
Med. ser. emp.	121	11.206	0.678	1,863
Auto owned		0.489	0.788	1,563
State and local govt. emp.		1.757	0.860	1,297
Finance emp.		0.319	0.882	1,199
Hosp. emp.		0.954	0.890	1,164
Pub. bldg. emp.	92	0.914	0.674	1,874
Auto owned		0.549	0.817	1,465
Comm. ser. emp.		0.271	0.860	1,297
Med. ser. emp.	92	14.091	0.678	1,863
Auto owned		0.479	0.788	1,563
Net res. den.		0.841	0.851	1,334
Med. ser. emp.	641	13.898	0.678	1,863
Dwelling units		0.245	0.781	1,586
Net res. den.		0.664	0.820	1,453

[a] All employment data are estimated employment, specifically first work trips to the land use corresponding to the employment activity.

while the addition of fourth and fifth variables could not be justified by the improvement in the measures of correlation. The desirability of a good fourth or fifth variable is not denied if significant improvements in the reliability of the equation can be made. But to collect data on a variable which contributes little to the reliability of the equation is pointless.

It will be noted also that there are alternative methods. Until a much better method is derived, the actual selection of one of the alternatives can be guided by such criteria as practicality and feasibility in gathering and forecasting data on the independent variables. The resident characteristics are those which are needed in the estimates of the other trip purposes. As to employment data in the various service trades and professions, some are more available than others. Current estimates of medical employment are obtainable readily from directories of physicians, dentists, and related professions. Because there is a relation between total medical employment and population, a reasonable forecast of total medical employment can be derived. Future zonal location of this employment is more difficult to forecast. At present the location of medical offices is undergoing change as more clinics for group and individual practice are established.

It is possible to obtain data on current public employment, although there are problems in determining the place of employment of such categories as patrolling police, inspectors, and street maintenance crews. Again there is a relation between public employment and total population. Except for employees who work in the field, the future location of public employees can probably be forecast with a reasonable accuracy.

TABLE 15

STATISTICAL MEASURES OF CORRELATION FOR THE COMBINATION OF PERSONAL BUSINESS AND EAT MEAL IN 469 ZONES

(Mean Number of Trips 2,580)

Variable	Intercept	Regression Coefficient	Correlation Coefficient	Standard Error of Estimate
Med. ser. emp.	194	17.694	0.706	2,092
Auto owned		0.555	0.811	1,730
Net res. den.		0.898	0.863	1,495

While the use of employment in the personal service trades is logical, the problems of data collection and forecast are complex. Identifying the number and location of personal service workers not covered by unemployment insurance compensation is difficult, and the problem of forecasting is greater. Retail trade was one of the important destinations of personal business trips, and there was correlation between personal business trips and various measures of retail activity. Ninety percent of the retail employment is covered by unemployment compensation insurance and data on the current location of this employment is available by postal zone. The problem of forecasting retail employment is similar to the problem of forecasting employment in the personal service trades.

Trips to eat a meal were combined with personal business trips and correlated with the same set of independent variables. Minor differences in the measures of correlation were noted, but the differences were of a nature to justify the combination on the basis of expediency and practicality, rather than any significant improvement in the correlation measures. On the other hand, the differences did not indicate that separation was justified to maintain the reliability of the personal business trip estimate.

It may be concluded that:

1. Personal business trips and trips to eat a meal should be combined.
2. A simple causal variable explaining personal business trips is difficult to isolate because a number of objectives are involved. Breaking this group of trips down into the component purposes is feasible only at the district level and where there is some assurance that input data errors are minimal. In aggregating minor trip purposes, the effect of data error is minimized. Aggregation is feasible also because of the covariance between the independent variables, which are required to explain the component trip destinations.
3. A relatively simple estimating equation is recommended in preference to an equation including definitions of all of the destination places. In these analyses, a complex equation did not give a more reliable estimate.
4. Among the acceptable methods for forecasting personal business trips and trips to eat a meal, the methods which are based on resident characteristics and employment in medical service are recommended. Data on the resident characteristics will be available if trip estimates for the other trip purposes are made. It is believed that fewer problems will be encountered in obtaining data on current and forecast employment in medical service by location. Medical service is also one of the more important objectives of personal business trips. Use of the alternatives, that is, those using employment data in another service trade or profession, can be justified if the employment data are easier to obtain and to forecast.

SCHOOL TRIPS AND TRIPS TO SERVE A PASSENGER

There were 204,794 estimated school trips in 1956, or 2 percent of the total. These figures are deceptive in stating the relative importance of school trips. The estimates probably were low because the trip survey was made primarily during the summer months when most schools were not in session. By definition, the figures did not include walking trips to school which in 1956 were estimated at approximately 400,000. Of the 200,000 trips using some form of vehicular transportation, approximately half used the bus. Because school trips are made during the morning rush hour, the impact of the school trips and, particularly, those having a bus mode, is important.

School trips have a common characteristic with work trips: both are of a compulsory nature and are regular in their occurrence. The worker makes a trip to perform certain functions at the place of destination in return for a paycheck; the school child is required to attend school between the ages of 7 and 16 (in Illinois). The hours of commencing work or school are usually fixed, so that the time of the work or school trip can be ascertained readily. While there are requirements to be at work or in school, there are absences, and thus no trip, because of illness, excused absence, vacation, or truancy.

There were 234,752 trips to serve a passenger. Trips to serve a passenger were those trips where the automobile driver took a passenger to the passenger's place of destination—public buildings (22 percent of the total land-use destinations), transportation (20 percent), residence (18 percent), retail (12 percent), or public open space (11 percent). While the trip purpose of the passenger is unknown, much can be conjectured from the land-use destinations. In the case of public buildings destinations, there was logic in assuming that the passenger was a minor unable to drive and was chauffeured to a public building, or, more specifically, to a school by a parent. Trips to public open space could have involved the same type of passenger, although the purpose was different. Trips to transportation land probably included such objectives as taking a passenger to an airport or rail station where the passenger then left the area, or taking a passenger to a commuter station at which point the passenger changed modes and continued to work. Retail destinations may have involved taking the wife on the weekly grocery shopping trip.

It is logical to assume that the trip would not have been made except to fulfill the passenger's objective, and that any method developed for estimating trips to serve a passenger should be based on the trip purpose of the passenger. However, it was not possible to obtain information on the passenger's trip purpose from the existing CATS trip tabulations, except by conjectures based on the land use at the place of destination. At the district level, trips to serve a passenger showed a relationship—that is, there was a simple correlation coefficient of 0.80—with those school trips having an automobile passenger mode, residential land, and trips to ride. Because of these relationships, particularly the relation to school trips, and because schools were an important destination, it was decided to test the combination of trips to serve a passenger and trips to school.

Methods for Estimating School Trips

Methods for estimating the total number of school trips are obvious. There is no question that the number of trips with a school purpose to the district or zone of destination is the average daily attendance of the school at the place of destination. In lieu of average daily attendance, the number of school trips may be estimated as being approximately 85 percent of the total enrollment. Variation in this percentage is due to absence, whether because of illness or truancy.

The potential number of school trips having an origin in a zone or district is equivalent to the school-age population resident in the zone or district and enrolled in school. The proportion enrolled in school is related to age. In the older age brackets, above 17, the proportion is not only related to age, but sex and race. There are standard demographic techniques for estimating the school-age population, and there is a wealth of data on the proportion of the school-age population enrolled in school. While no significant change will occur in the proportion enrolled in school among the compulsory school-age groups, change can be expected in the older school-age population groups, particularly in the minority racial groups. Absenteeism will also reduce the total potential number of trips originating in a zone. In addition, a small and probably insignificant number of trips will start from a non-home origin.

While these relationships are obvious and are easy to demonstrate, there remains one problem. Many of the school trips (approximately two-thirds) were pedestrian. To be counted as a trip in the CATS inventory, an automotive or transit mode was requisite. While one-third of the average daily attendance or school enrollment will yield an approximate estimate of the number of vehicular trips in large areas, this method is not satisfactory in small areas such as zones.

School Systems and Service Areas

The nature of the school system and its service area offer clues to the mode of the school trip, that is, whether a vehicular or pedestrian mode was used. In the Chicago area, there are two major school systems—the public school system and the Catholic parochial system. There are in addition the various colleges, universities, and technical schools, and a few non-Catholic private elementary and secondary schools. Approximately 30 percent of the elementary and secondary school enrollment of Cook County and Chicago is in Catholic parochial schools. Enrollment in other private schools is an insignificant portion of the total. In the 1964-65 school year there were 455 public elementary schools and 75 general and technical high schools in the city of Chicago; the corresponding number of parochial schools in Chicago was 279 and 69. Many of the Catholic high schools were segregated by sex.

Both the public and parochial systems have used the same general principles of school location. Elementary schools have been located in residential areas within reasonable walking distance of the schoolchild's home. High schools have been located to serve several neighborhoods, and because of the larger service area, many are unable to walk to school. Schoolchildren making the trip to and from school by public bus ride at reduced fare.

In recent years the neighborhood school has been challenged in many northern metropolitan areas because of de facto segregated schools in racially segregated residential areas. Proposals have been made to change the attendance areas, to pair schools, or to establish schools on the Princeton plan or school village. The arguments pro and con stated in meetings of school boards, parent-teacher associations, and various civic organizations are not the concern of the transportation planner. The implication of some of the plans in bussing a greater proportion of school children does, however, lie within the province of the planner. Here, specifically, it has the implication of changing the proportion of vehicular trips.

It should be noted in passing that rural neighborhood schools, i.e., the little, red, one-room school, has long since been replaced by the consolidated school and a system of buses. Many outlying suburban areas also provide bus transportation for elementary schools where the schools cannot be located within walking distance of all.

From the transportation viewpoint, there are two types of institutions of higher learning, be they academic, business or technical. The two are resident and commuter. Resident institutions are those colleges and universities which provide dormitories and married student quarters for most of the student body. In general, students walk to class and may use a car only for nonschool purposes. Campus restrictions on student use of cars usually will reinforce these trip patterns. Resident institutions also may have another characteristic which should be noted in forecasts of school-age population—frequently, the student body may be drawn from outside the study area under consideration.

Commuter, or day institutions—academic, business, technical—do not provide living accommodations. The number of such institutions is small, but their service area is the entire community rather than a school district or a segment of the area. Trips thus are drawn from the entire area and, because of the age of the student body, many will drive automobiles.

Estimating Vehicular School Trips

This general description of school systems and service areas suggests that most trips to elementary schools and resident colleges are pedestrian, and that trips to high schools and day institutions have some vehicular mode of transportation. This theory could not be tested with the CATS trip and land-use data. Neither the trip data nor the land-use data indicated the level of schooling at the destination.

At the district level, all school trips—vehicular and pedestrian—were correlated with an estimated school-age population. Considering that the estimate of school-age population was approximate, the measure of correlation was satisfactory. The analysis demonstrated that most of the trips had an origin and destination in the same district, which was logical since two-thirds of the trips were pedestrian. It did not solve the problem of estimating the number of nonpedestrian trip destinations.

At the zonal level, vehicular school trips at place of destination were correlated with school floor area. There was a correlation coefficient of 0.64 and a standard error of estimate which equaled the mean number of trips to the zone. It can be shown that school floor area is related to school enrollment, although many examples of overcrowding or underutilization can be pointed out in specific schools. However, total school floor area does not indicate the proportion of vehicular school trips adequately. Further studies are indicated. It is suggested that these studies utilize data on average daily attendance or enrollment, and school trips by type of school—specifically, elementary, secondary, and college.

Estimating Trips to Serve a Passenger

No attempt was made to derive a method for estimating trips to serve a passenger as a separate trip purpose. They were combined with school trips, although the land-use destinations of the trips to serve a passenger indicated that perhaps only 20 percent had school destinations, and could be combined logically with school trips. The combination was feasible, however. It should be noted that trips to serve a passenger continued to show a relationship to residential land at the zonal level, that is, there was a simple correlation coefficient of 0.67 between the trips and residential land. If this correlation was not due to chance, then a logical explanation is not easy to produce.

Therefore, the following conclusions may be reached:

1. Estimates of the number of trips with a school purpose to the zone or district of destination are best made from average daily attendance. Average daily attendance can be derived from school enrollments. Approximations of enrollment can be made from various measures of school size and capacity. If forecasts of school enrollments are not available from the school authorities, there are acceptable methods for preparing enrollment forecasts.
2. Determining the proportion of total enrollment or attendance using vehicular modes remains a problem. It is doubtful that a method can be derived from the 1956 CATS data in their present form. Net residential density or ring of destination are not satisfactory methods at the zonal level. It is believed that the level of school is a relevant factor, but this cannot be tested with the 1956 CATS data.
3. As a rough-and-ready method of estimating the proportion of vehicular trips, it might be assumed that elementary school trips are pedestrian, and that all other school trips are vehicular. Other common-sense rules of thumb may be used until better methods are developed.
4. Without a great deal of effort—far more than the relative importance of trips to serve a passenger warrants—it is doubtful that a logical and satisfactory method for estimating trips to serve a passenger can be found. This is especially true at the zonal level.
5. Trips to serve a passenger should be broken up and reclassified according to the trip purpose of the person being served. This was not possible to do with the 1956 CATS data, except in a crude manner. The combination of trips to serve a passenger with school trips cannot be justified completely, but it does demonstrate that the method derived to estimate the major trip purpose can be retained without loss of reliability, and that the problem of estimating trips to serve a passenger as a separate trip purpose is eliminated.

SHOPPING TRIPS AND TRIPS TO RIDE

From the 1956 trip survey, CATS estimated that there were 550,215 shopping trips. In another survey made during the same year, CATS estimated that there were an additional 618,592 pedestrian shopping trips. There was another group of trips, those with a purpose to ride, which were believed to be related to shopping trips. There were 107,186 such trips in 1956.

The vehicular shopping trips constituted 5.5 percent of the total trips, while trips to ride were 1.1 percent. In total number, both trip purposes may be considered minor. However, because shopping trips have destinations to retail land which is located in the central business district, outlying shopping areas, or on major traffic arteries, the traffic generation of shopping areas is important to transportation planning. Shopping streets are noted for their congestion, and the approaches to the planned shopping centers have their traffic bottlenecks. The transportation planner is not the only one interested in the shopper and his (or her) trip; there are also the retail market analysts, land planners, and merchants.

Approximately half of the vehicular shopping trips had an origin and destination in the same district, but the net district interchange of shopping trips indicated that considerable choice and selection were made in the destinations. As might be expected, the CBD district 01 had the highest net gain in shopping trips, but the net gain of a suburban CBD, such as Oak Park (district 43), or of a major shopping center such as 63rd and Halsted (district 46), or Irving Park and Cicero (district 42), or Evergreen (district 56) were high. At the other extreme were districts 23, 34, and 37 where the

TABLE 16

PERSON SHOPPING TRIP DESTINATIONS BY MODE, NET DISTRICT INTERCHANGES AND RETAIL USE AT DESTINATION

Dist. of Dest.	Pedestrian[a] Shopping Trips	Vehicular Shopping Trips[b]				
		Total	Net Interchange (Dest. -Orig.)	Dest. & Orig. in Same District	Trips to Food & Drug	Trips to General Merchandise
01	22,229	46,568	41,315	1,104	1,505	39,693
11	45,162	12,740	-5,961	4,650	2,592	5,534
21	20,945	9,357	396	3,157	1,929	5,418
22	13,762	5,016	-833	1,654	1,945	1,730
23	16,615	3,445	-3,989	1,009	1,613	345
24	13,700	4,353	191	795	1,337	2,373
25	9,693	1,727	-1,334	313	883	313
26	8,115	1,760	-1,503	632	851	281
27	11,619	3,073	-1,915	1,272	1,796	219
31	19,641	7,171	-2,249	2,679	2,993	2,692
32	27,016	10,005	-2,729	1,331	4,914	2,138
33	28,316	12,785	886	1,362	4,371	5,095
34	15,055	2,245	-6,298	863	1,213	190
35	10,532	5,256	442	1,385	2,161	2,031
36	17,170	4,928	-973	1,565	1,686	2,530
37	29,994	3,863	-5,262	1,826	1,820	853
41	26,597	13,078	-3,787	8,195	7,249	2,464
42	23,690	23,086	6,391	8,335	6,481	10,709
43	17,171	46,165	14,956	22,035	17,602	24,020
44	11,608	15,757	-2,397	10,070	9,704	2,316
45	12,886	6,014	-4,362	2,970	4,523	218
46	25,958	29,050	9,493	11,849	9,596	12,467
47	41,690	17,879	-4,733	11,370	10,317	3,808
51	13,556	15,926	114	8,043	4,383	8,030
52	13,777	16,571	-899	8,651	12,144	841
53	12,568	12,970	-7,036	7,221	9,297	984
54	3,000	8,618	-2,854	5,463	6,209	307
55	4,214	8,688	-1,604	4,446	4,898	1,953
56	11,172	27,905	5,314	13,753	12,477	7,808
57	15,515	16,550	-2,278	10,891	9,981	2,665
61	4,439	15,397	740	9,272	8,369	3,127
62	6,010	12,256	-2,616	8,246	6,776	2,061
63	2,432	14,882	-2,815	11,465	10,463	1,891
64	3,645	15,823	889	10,202	9,691	2,407
65	2,968	4,739	-2,907	2,266	3,434	184
66	10,399	16,559	-5,635	11,622	12,186	1,207
67	14,330	15,838	2,001	8,621	6,646	6,707
71	3,657	8,895	-1,043	7,304	5,153	1,169
72	1,799	9,811	-2,442	8,873	6,457	706
73	1,255	6,880	-989	4,997	5,249	698
74	1,216	6,792	-1,929	5,317	5,003	643
75	2,552	762	-498	762	500	60
76	12,587	25,435	2,643	20,443	15,380	5,948
77	8,337	3,597	-1,901	2,758	2,635	156
Total	618,592	550,215	0	271,037	256,412	176,989

[a] Source: Walking Trip Survey. Total excludes shopping trips from shop.

[b] Source: Tape prepared from home interview cards. District of origin defined as district from which person started the shopping trip.

net loss of shopping trips exceeded the number of trip destinations to the district. The sum of these net district interchanges, signs disregarded, was 171,542, or approximately one-third of the total vehicular trips.

Since walking trips averaged two or three blocks, it can be assumed that the origin and destination of the pedestrian shopping trips were in the same district.

With the exception of a negligible number of trips, all shopping trips had destinations on retail land. However, there appeared to be a distortion in the reported retail destinations; 47 percent of the respondents reported a grocery or drug store destination, while 32 percent gave a general merchandise destination. Other types of retail destination received small mention. Trips to other retail establishments were either forgotten or combined with the primary shopping purpose.

Where there was a sizable proportion of general merchandise destinations in the district, there was also a large number of shoppers originating in that district and another large group entering the district from a second district. Conversely, a high proportion of grocery and drug store destinations usually indicated a net loss in shopping trips (Table 16).

TABLE 17

SIMPLE CORRELATION COEFFICIENTS BETWEEN SHOPPING TRIPS AND SELECTED CHARACTERISTICS OF POPULATION AND RETAIL FACILITIES

Sample Size	Characteristic	Shopping Trips by Mode	
		Vehicular	Pedestrian
44 districts	Population	0.238	0.828
	Dwelling units	0.214	0.877
	Autos owned	0.450	...
24 districts	Autos owned	...	0.659
24 districts	Total retail floor area	0.483	0.612
	Floor area in:		
	Food and drug	0.005	0.509
	Eat and drink places	0.343	0.601
	General merchandise	0.466	0.234
	Apparel	0.694	0.379
	Furniture	0.536	0.667
	Car sales	0.172	0.736
	Gas sales	-0.054	0.100
	Hardware	0.108	0.519
	Liquor	0.476	0.891
	Miscellaneous	0.527	0.529
24 districts	Total retail establishments	0.309	0.855
	Number of establishments:		
	Food and drug	0.047	0.796
	General merchandise	0.336	0.848
	Apparel	0.655	0.775
	Gas sales	0.396	0.628
	Liquor	0.203	0.805
	Miscellaneous	0.596	0.785
367 zones	Autos owned	0.172	
	Number of department stores	0.708	
	Commercial land	0.204	
204 zones	Autos owned	-0.033	
	Number of department stores	0.761	
	Commercial land	0.059	
	Total retail floor area	0.458	
	Food and drug floor area	0.031	
	General merchandise floor area	0.494	
33 zones	Autos owned	-0.107	
	Number of department stores	0.603	
	Total retail floor area	0.232	
	Retail sales	0.808	

44 districts: all districts.
24 districts: districts 01 through 47, and 57.
367 zones: all zones used in analysis of shopping trips, includes the four zones of the CBD district 01.
204 zones: all zones where floor area was inventoried; with a few exceptions, the 204 zones are the same zones in the 24 districts; the four zones of the CBD district 01 are included.
33 zones: selected zones used to analyze retail sales and shopping trips; no zone is located in the CBD.
There are no data on the zonal destination of pedestrian shopping trips.

Trips To Ride

A trip to ride indicated that the person had no other purpose in mind. Nearly half of these trips had a retail trade land-use destination, from which it was hypothesized that the shopper took along her child, who thus became a rider, but not necessarily a shopper. There were two other important land-use destinations, residential land and public open space. In the case of public open space destinations, there may have been miscoding of the trip purpose. There were several hot days during the survey period when families took a ride in the parks to escape the heat. In response to the survey question on purpose of that trip, the family may have said, "to ride," where in fact the trip purpose was social-recreation.

At the district level, with the CBD district 01 omitted, there was a simple correlation coefficient of 0.88 between trips to ride and shopping trips. Under the same conditions, trips to ride also had a simple correlation coefficient of 0.82 with trips to serve a passenger, and of 0.72 with residential land.

Combination of trips to ride with shopping trips is a tenuous one and other combinations may be offered with equal validity. The objective was the elimination of trips to ride as a separate trip purpose, for an estimating method for the number of trip destinations was not readily apparent. The time which would have been required to find a method was not justified by the relative importance of the trip purpose. Essentially, it can be said that the combination with shopping trips was not illogical. Subsequently, it was found that the regression analysis of shopping trips alone and in combination with trips to ride were similar. The influence of trips to ride in the regression analysis of the combined trips was relatively minor, since the trips to ride were approximately one-sixth of the total. The subsequent discussion is related primarily to shopping trips.

The Pedestrian Shopper

As mentioned above, there were two types of shoppers, the shopper who used some vehicular form of transportation, usually the automobile, and the shopper who walked. The two are different in their trip patterns and destinations.

The trip destinations of the pedestrian shopper were definitely local in character and could be related to a number of factors describing local conditions within the district of destination. For example, it was found that the number of pedestrian shoppers was related to population or number of dwelling units in the district. The number of pedestrian shoppers was also related to the number of retail establishments or floor area in the district (Table 17).

The number and area of retail facilities also were related to population. This relationship was not surprising when it was remembered that much of the retail space inventory was old and retained many characteristics of the pre-automobile days when retail facilities were constructed to serve shoppers who were almost all pedestrian. The only important nonpedestrian shopper was the trolley rider for whom shopping facilities were provided at major street intersections and transfer points.

These interrelationships between population, shopping facilities, and the trip destinations of the pedestrian shopper led to the finding of a high correlation between the pedestrian shopper and liquor stores. In turn this led to a question, one of the mathematician's favorite problems, involving whether the drunk's random walk home might be reversed. Specifically, what were the chances of the pedestrian shopper arriving at the tavern? Apparently, they were good.[1]

The Vehicular Shopper

The shopper who has access to a car was not as predictable. Even at the district level, the total number of trips could not be related to measures of population or crude estimates of income derived from automobile ownership. Nor could the district destinations be related to gross measures of retail activity such as floor area and number of establishments.

The regression analysis was taken a step further by using several independent variables. At the district level these variables included various resident characteristics and indexes of retail activity defined in terms of floor area and number of establishments by type. At the zonal level the measures of retail activity were reduced and two ratios of retail activity were tested.

One ratio, derived by dividing general merchandise floor area by the floor area in grocery and drugs, gave an indication of the character of the shopping facilities in the zone. A high ratio indicated proportionately more general merchandise facilities, while a low ratio indicated proportionately more convenience retail facilities. The second ratio was retail floor area per dwelling unit. This ratio was used in both the district and zonal analyses.

At both the district and zonal levels, combinations of independent variables defining resident characteristics and gross or quantitative measures of retail activity failed to produce estimating equations which could be recommended on the basis of a high correlation coefficient and a low standard error of estimate.

It was evident that the shopper used discrimination and that the basis of discrimination had qualitative rather than quantitative aspects. Ask a woman where she shops and she may reply, Jewel for meats, National for canned goods and trading stamps, Sears for children's clothes, and Marshall Field for a dress. Shoppers may not be aware of gross measures of retail activity, but they are certainly aware of shopping areas, at least within their section of the city. They may be vocal in stating a preference for Shoppers World to downtown Evanston because of the variety of purchases which can be made on one stop, or a preference for downtown Oak Park to downtown Chicago because branches of the leading stores are located in Oak Park. These qualities, however, are not easily reduced to definable terms which can be utilized in a regression analysis.

There were, however, several studies of retail areas made in the Chicago area. In one of these, Commercial Structure and Commercial Blight, Brian J. L. Berry surveyed the various shopping areas in Chicago thoroughly and gave each center a numerical rank. The rank and classification of the shopping centers were based on a highly

---

[1] But not in Evanston and a few other suburban municipalities, which had dry ordinances. By conicidence, districts containing these municipalities were not included in the sample, because floor area had not been inventoried.

complex, sophisticated factor analysis which used data on the composition, employment, and sales of the shopping center, and data on the population and personal income of the trade area. A sample of 32 of these ranked centers was selected where it was believed that most of the shopping trips to the CATS zone in which the center was located probably had destinations in the centers. There was a rank correlation of 0.76 between the centers and the shopping trips. From this it can be said that there was some agreement between the statistical geographer and the shopper on the relative rank of the shopping areas. It also can be said that classification of shopping areas was a step toward developing qualitative measures of shopping facilities. But with all due respect to Berry, his classification system is not one that can be recommended. Not only is the data collection a formidable task, but forecasting of the shopping center's future rank by this method is an even greater task. If classification of shopping centers or all shopping facilities in a zone is an appropriate consideration in trip estimates, then the classification system should be one where data collection and analysis can be made within the normal time and budget limitations of the transportation agency.

Another type of qualitative measure was tried in the zonal analysis. This measure was the number of major department stores in the zone. Number of establishments is a quantitative measure, but the method of collecting data had discriminatory aspects. Rather than use such criteria as the SIC four-digit code classification system, the selection was based on what was believed to be the consensus of opinion of what stores were the leading major department stores in the Chicago area. The selected stores were Fields, Carsons, Sears, Wards, Goldblatts, and Wieboldts. The stores ranged from those with quality merchandise to those with the more inexpensive, bargain merchandise, but each carried a full line of merchandise. In 1956, there were 39 main and branch stores, which were located by their zonal location. In the zonal regression analyses, those analyses which used number of department stores as an independent variable were the best. Some of the findings are shown in Table 18. Again, this may

TABLE 18

STATISTICAL MEASURES OF CORRELATION BETWEEN SHOPPING TRIPS AND SELECTED INDEPENDENT VARIABLES

Sample Size	Mean Number of Trips	Independent Variables	Correlation Coefficient	Standard Error of Estimate
44 districts	12,505	Dwelling units per acre	0.455	9,228
		Autos owned	0.749	6,950
43 districts	11,713	Autos owned	0.737	6,085
		Dwelling units per acre	0.820	5,221
24 districts	12,549	Retail floor per D. U.	0.522	10,958
		Autos owned	0.852	6,878
24 districts	12,549	Apparel floor area	0.694	9,253
		General merchandise floor area	0.698	9,413
		Furniture floor area	0.701	9,603
367 zones	1,483	Number of department stores	0.708	1,647
		Autos owned	0.714	1,636
		Gross population density	0.741	1,570
204 zones	1,740	Number of department stores	0.761	1,869
		Autos owned	0.762	1,870
204 zones	1,740	Number of department stores	0.761	1,869
		General mdse./food floor area	0.798	1,741
		Distance from CBD	0.806	1,714
33 zones	2,983	Retail sales	0.808	1,747
		Number of department stores	0.838	1,646

44 districts: all districts.
43 districts: district 01 omitted.
24 districts: districts 01 through 47, and 57.
367 zones: all zones having more than 100 shopping trip destinations.
204 zones: zones in which floor area inventoried, includes CBD zones.
33 zones: zones for which retail sales data obtained, does not included CBD zones.

be said to be a step toward definition of qualitative measures, but additional analysis is required before a reliable estimating method is derived.

In addition to the independent variables defining the quantitative and qualitative characteristics of retail facilities, there was another characteristic—retail sales. After eliminating mail order and similar types of sales, it can readily be demonstrated that sales represent shopping trips, and that, on the average, one shopping trip represents a dollar's (or some multiple) worth of sales. With respect to current estimates of retail sales, there were the quadrennial Census of Business Reports on retail sales by census tracts, although there were, of course, technical problems in adjusting discrepancies between time periods and geographical areas. Some of these adjustments had already been made by Berry in his study, so that it was possible to draw a small sample from his study and test the relationship between sales and shopping trips. The sample was composed of 33 of Berry's shopping centers which lay completely within a CATS zone. Sales data for the center were assumed to be the retail sales of that zone. With such a small sample and method of data collection, the chance of data error was high. Nevertheless, the simple correlation between trips and sales was 0.81. While no estimating equation is recommended from this small sample, it can be recommended that further consideration be given to the relationship between sales and trips. However, before extensive study is started, attention should be given to the thorny problem of forecasting retail sales by location. Simple projection of trends may be satisfactory for short periods of time into the future. Twenty-year forecasts are hazardous. Furthermore, it is conceivable that the same types of problems encountered in seeking a trip estimation method will be found in retail sales forecasts.

Conclusions and Recommendations

1. Trips to ride should be combined with the purpose of the driver. In this analysis they were combined with shopping trips, since there appeared to be some relationship between the two trip purposes. Other combinations, such as a combination of trips to ride with personal business trips, are not ruled out, but these other combinations were not tested.

2. Reliable estimating methods for shopping trips were not derived from resident characteristics or such quantitative measures of retail facilities as floor area and number of establishments. A good portion of the existing retail inventory was built with the pedestrian shopper in mind and not the shopper who arrives by car.

3. It is apparent that a qualitative measure of retail activity is needed. The two qualitative measures—classification of shopping facilities and number of major department stores—indicate that further analysis along these lines may be fruitful. Both measures require additional study before they can be utilized in the derivation of a reliable estimating method. Other qualitative measures should be examined and tested for their appropriateness in estimating shopping trips. The major problem encountered in these analyses was the problem of defining and obtaining qualitative data on shopping facilities. The CATS land-use data files were lacking in such information. Other data sources were not readily adapted to the CATS division of the Chicago area into districts and zones. Even much of Berry's data could not be used directly, despite the fact that he had identified the shopping centers' location by the CATS coordinate system.

4. Retail sales appear to be a logical independent variable from which to estimate trips. It is possible to obtain current data on sales, but forecast of sales by small areas such as zones is no easy problem.

Appendix A

ESTIMATING EQUATIONS

From the 1956 CATS data on trip destinations by purpose and the analysis of these trips, the following equations may be recommended:

a. Total trips to place of destination.

District = -14, 171 + 7. 036(Autos Owned) + 587. 026(Dwelling Unit per Acre)

Zone = 1, 855 + 5. 811(Autos Owned) + 9. 101(Dwelling Unit per 10 Acres)

These equations may be compared with the CATS equation for trips per dwelling unit in the zone:

= 682. 84 + 3. 8109(Autos per 100 DU) - . 1939 log(DU per 10 Acres)

b. Trips to home.

District = -5, 424 + 3. 268(Autos Owned) + 49. 915(Population per Acre)

Zone = -114 + 3. 340(Autos Owned) + . 351(Dwelling Unit per 10 Acres)

c. Work Trips.

No specific equation was derived. However, work trips can be estimated to be approximately 85 per cent of the total employment in the district or zone of destination. The percentage may range from 80 to 90 per cent of total employment. There are several definitions of employment of which average annual is appropriate for most trip estimating purposes. Seasonal and main shift employment may be required for special trip estimates.

d. Social-recreation trips.

The following equations are recommended for estimating social and local recreation trips:

1, 000 trips to all districts except 01 = . 075 + . 90(1, 000 Autos Owned)

Zones except those with major recreation facility = 268 + . 759(Autos Owned) + . 180(DU/10 Acres)

Additional analyses are required before an equation can be recommended for all social-recreation trips or trips to major recreation facilities. The equation for all trips to all zones can be listed, if it is understood that the reliability is poor.

Zone = 408 + . 739(Automobiles Owned) + . 402(Dwelling Unit per 10 Acres)

e. Personal business trips and trips to eat a meal.

There are several acceptable equations for estimating personal business trips to the district of destination. They are of about equal reliability, and the selection may be based on the availability of data.

District = 2, 203 + 24. 020(Med. Serv. Emp.) + . 413(Autos Owned)
= -851 + 6. 277(Local Govt. Emp.) + . 615(Autos Owned)
= -2, 434 + . 480(Lic. Auto Driver) + 80. 845(DU per Acre)

At the zonal level, the following equation is recommended for personal business trips:

Zone = 92 + 14. 091(Med. Serv. Emp.) + . 479(Autos Owned) + . 841(DU/10 Acres)

For the combination of personal business trips and trips to eat a meal, the recommended equation is:

Zone = 194 + 17. 694(Med. Serv. Emp.) + . 555(Autos Owned) + . 898(DU/10 Acres)

At the zonal level, there are acceptable alternatives which substitute other types of service trade and professional employment for employment in medical service.

f. School trips and trips to serve a passenger.

The total number of school trips to the district or zone of destination is average daily attendance or approximately 85 per cent of total enrollment. It is believed that most elementary school trips and many trips to resident educational institutions are pedestrian, and that all other school trips have a vehicular mode of transportation, bus, or automobile

It is recommended that trips to serve a passenger be coded according to the trip purpose of the passenger. There was evidence that many of thes trips were for the purpose of taking a child to school. Combination of all trips to serve a passenger with school trips is a feasible alternative.

g. Shopping trips and trips to ride.

Additional analyses are required before a reliable equation can be recommended. There is evidence that shopping trips can be estimated from retail sales, and this method would be recommended, if sales forecasts for small areas can be made. Vehicular shopping trips cannot be estimated from the number or amount of retail facilities, most of which were built to serve pedestrian shoppers. It appears that a qualitative measure of retail facilities is required. Classification or ranking of shopping facilities is one approach.

Trips to ride should be classified according to the purpose of the person giving the ride. There was evidence that the purpose of many trips to ride was going for "a ride" with the shopper. Combination of shopping trips and all trips to ride was feasible.

Appendix B

SOURCES OF ERROR IN EMPLOYMENT ESTIMATES AND WORK TRIP DATA

From a statistical viewpoint, the regression analysis of work trips and employment was not entirely satisfactory. There was no doubt that there was a logical causal relationship between employment and work trips. This relationship was borne out by the high correlation coefficient, but there was also a relatively high standard error of estimate. Rather than casting doubt on the validity of the relationship between work trips and employment, input error was indicated. Extensive studies were made to determine the location of the error. Possible error in the employment estimates were relatively easy to locate. Possible error in the work trip data was not so easy to find, for the information had been assembled several years before and many of those who had dealt directly with the trip survey and the factoring of the sample had left CATS employ. However, conjectures could be made about the location and type of errors in the work trip data.

It appeared that error existed in both the estimates of employment and the work trip data. Both series of data, however, were probably the most reliable of any of the various sets of data used in the analysis of trips by purpose. The errors in the employment estimates and work trip data merely illustrate the underlying problem which existed in the analysis of the other trip purposes. Where there was a logical relationship between the trip purpose and the selected independent variable, poor measures of correlation could be discounted by saying that input data error existed. Where the relationship was secondary and not directly causal, there was the problem of determining whether the resultant measures of correlation were due to poor selection and definition of the factor, or data error.

There were three basic types of error in the employment and work trip data: time, definition, and estimating method. Errors resulting from the difference in dates were unavoidable. Errors resulting from definition could be controlled to some extent in the employment data. There was evidence of definition errors, such as miscoding of the work trips, but the data could not be corrected. The method of estimating employment by CATS district was probably the best, given the time and personnel limitations, but error was possible.

PROBLEM OF TIME.

The problem of time arose from the fact that the CATS survey of trips was made during the second and third quarters of 1956, while the Employment Service data on covered employment was for the first quarter of 1955 and 1957. Since dates could not be matched, a compromise solution of selecting 1957 to represent 1956 was made.

The Illinois Department of Labor estimated that total employment in the Chicago standard metropolitan area was 2, 565, 300 in 1955; 2, 642, 700 in 1956, and 2, 645, 700 in 1957. Total employment in 1957 was higher than in 1956, but the difference between these two years was less than the difference between 1955 and 1956. It was for this reason, primarily, that 1957 data on covered employment were selected. The poor shape of the Employment Service's records for 1955 was a secondary reason for selecting the 1957 data.

The amount of error resulting from time cannot be quantified accurately. One check on the reasonableness of the data could be made, however, and that was a comparison of the total 1957 covered employment data with the estimates of 1956 total employment prepared by Dr. Irving Hoch for the CATS study area. With the exception of a few industries, the two sets of employment data are similar. More important is the comparison of the work trip data with either set of employment data. While some differences may be due to differences in dates, it appears that many of the differences may be due to definition.

PROBLEMS OF DEFINITION.

The problem of definition and the resultant error existed in both the employment and the work trip data. There were several problems.

a. Definition of employment.

The employment estimates were based on covered employment, which is not total employment. However, in the manufacturing industries, 99 per cent of the total employees were covered. For all practical purposes, it may be said that covered employment was total manufacturing employment. In construction, communications and public utilities, wholesale and retail trade, and mining the coverage was over 90 per cent. Eighty-three per cent of the total finance employment was covered, while approximately 50 to 60 per cent was covered in transportation and the service trades. Government employment was one notable nonagricultural industry where there was no coverage. On the average, for all nonagricultural employment, 77 per cent of the total was covered. No attempt was made to adjust covered employment to total.

District estimates of employment were made only for those industries where the percentage of covered employment was high. Since the conversion method of postal zone employment data to CATS district also involved matching definitions of employment industry with the CATS land use classifications, further limitations were placed on the number of employment estimates made. In general, it was possible to match the CATS two digit land classification with the SIC two digit industrial classification, but it should be noted that CATS land area was summarized by a one digit code, and that the floor area was summarized by the two digit code in the inner built up areas. Estimates of manufacturing employment by district could be made for the entire CATS study area, but estimates of other types of employment were made for only districts 01 through 37.

b. Multiple job holding.

The covered employment figures represented number of jobs. A worker holding two jobs would thus be counted twice. First work trips, on the other hand, represented one worker and one job. The magnitude of multiple job holding is unknown.

c. Absenteeism.

It is obvious to expect that the number of work trips will be lower than the number of employed workers because of absenteeism for vacation, illness, and other reasons. The average rate is unknown, although it appeared to be around 10 to 15 per cent of employment, with the possibility of being as high as 20 per cent, or as low as 5 per cent. Comparison of the employment and work trip data in Table 19, page 56, shows variation in the rate of absenteeism between industries. Normal expectation would lead one to believe that the rate of absenteeism should be approximately the same from industry to industry. Scheduling of summer vacations, strikes, and layoffs may have affected the rate of absenteeism, but without appropriate data it is not possible to estimate the effect of such causes on the rate of absenteeism.

d. Miscoding of work trip by industry.

Because of the similarity of titles of some pairs of industry groups, and because one industry in the pair had a high proportion of work trips, and the other a low proportion, it is believed that some misinterpretation and miscoding occurred in the work trip data. The paired industries are:

Non-electrical machinery and Electrical machinery

Textile and Apparel

Printing and Paper

Transportation manufacture and Transportation industry

Retail trade and Wholesale trade

e. Sampling error in the work trip data.

Work trips to the manufacturing industries of mining, tobacco, lumber petroleum, and rubber were less than 6, 000. With such a low number of trips, it was possible for sampling error in the work trip data to be high. Such sampling error would explain discrepancies between the number of work trips and employment.

f. Place of work.

Errors in determining the place of work can occur in both the employment data and work trip data. The Illinois Employment Service had problems in assigning employment to the appropriate postal zone in the case of establishments with several locations and of industries where the employees had no fixed place of work. This problem was of critical importance in the employment data for small geographical areas. The problem should be minimal in the total covered employment data for the metropolitan area.

TABLE 19 -- 1956 Internal First Work Trips in CATS Study Area; Total March 1957 Covered Employment in Chicago, Cook County, and Du Page County; and 1956 CATS Estimate of Total Employment in the Study Area

Industry	Internal First Work Trip in Study Area	1957 Covered Employment		1956 CATS Estimate of Employment (thousands)
		Total	% Covered	
Residential	88, 983	...	0	88. 4
Agriculture	3, 411	...	0	7. 9
Lumber	5, 313	8, 458	97. 1	8. 4
Furniture	16, 519	18, 659	99. 0	19. 4
Stone	9, 851	12, 384	99. 1	13. 1
Primary Metal	60, 005	65, 114	99. 9	62. 7
Fabricated Metal	78, 971	93, 778	99. 4	96. 3
Machinery	61, 025	119, 681	99. 1	128. 5
Elect. Machinery	122, 790	128, 818	99. 8	133. 9
Transportation	43, 430	42, 534	99. 7	39. 5
Professional	13, 744	22, 867	99. 4	22. 8
Miscellaneous	34, 721	36, 552	98. 8	39. 7
Total Dur. Mfg.	446, 401	548, 845	99. 4	564. 3
Food	81, 832	91, 076	99. 6	97. 7
Tobacco	921	486	90. 2	0. 5
Textile	7, 168	7, 251	99. 6	42. 3
Apparel	21, 380	33, 734	98. 6	
Paper	20, 403	22, 555	99. 6	23. 1
Printing	61, 707	79, 015	98. 7	81. 3
Chemical	29, 919	31, 210	99. 1	29. 3
Petroleum	5, 553	4, 294	99. 8	3. 5
Rubber	5, 047	4, 485	99. 6	4. 7
Leather	9, 372	9, 567	99. 0	10. 3
Total Nondur. Mfg.	243, 648	283, 673	99. 2	292. 7
Rail and Transit	61, 283	...		71. 9
Trucking	29, 766	...		50. 0
Warehousing	17, 966	...		10. 5
Air Transportation	9, 015	...		8. 4
Water Transportation	1, 662	...		2. 9
Highway Transportation	2, 852	...		28. 5
Total Transportation	122, 544	68, 900	57. 1	172. 2
Utilities & Comm.	45, 293	50, 739	96. 2	55. 1
Mining	558	2, 935	98. 8	2. 5
Retail Trade	275, 244	260, 074	90. 8	391. 9
Finance	88, 132	113, 670	83. 4	149. 3
Service	205, 808	175, 795	52. 1	250. 5
Wholesale	51, 332	168, 688	95. 3	171. 6
Construction	51, 525	89, 733	94. 5	153. 3
Public Buildings	163, 859	...	0	249. 1
Public Open Space	17, 364	...	0	...
Grand Total	1, 804, 102	1, 763, 052	76. 7	2, 548. 8

With respect to the work trip data and place of work, there was another problem. An employee in a given industry may not necessarily work at a place with a corresponding land use. Itinerant service personnel and door-to-door salesmen are obvious examples. Under this line of reasoning, the seemingly underreported work trips to construction may be explained if construction workers went to a construction site, which was classified as residential or commercial.

g. Metropolitan area.

The Employment Service defined the Chicago metropolitan area as the City of Chicago and the Counties of Cook and Du Page. The boundaries of this area were not coterminous with the CATS study area, but the difference probably was immaterial with respect to nonagricultural employment.

While problems of definition can be described as a source of error, it was not possible to quantify these explanations and make appropriate corrections. Many of the discrepancies which can be seen in the total work trips and total employment data probably are due to these problems of definition. These problems persisted in the subareas of the metropolitan area and could have been magnified.

PROBLEM OF ESTIMATING METHOD.

There were several ways of obtaining employment estimates for the CATS districts, the most accurate being examination of the Employment Service files on the individual records of some 50, 000 firms, posting the appropriate CATS zone and district location, and tabulating the total covered employment by CATS district. Because of the considerable time and personnel this procedure would have involved, this method was not used.

The Service's tabulations of covered employment by postal zone and industry could not be used directly. Within the Chicago metropolitan area there were some fifty postal zones in the City of Chicago and nearly forty suburban municipalities for which covered employment was reported by the Employment Service. Neither the postal zones nor the suburban municipalities corresponded to the CATS districts. In the case of the municipalities, this was expected. Despite the fact that many of the Chicago postal zones were approximately the same size as the CATS districts and used some of the same boundary streets, there was no postal zone or combination which encompassed the same area as a CATS district. It was not possible even to assemble groups of postal zones and CATS districts because of the extensive overlapping of postal zones and CATS districts. Even CATS zones were split by the postal zones. It was not until the next lower level, quarter mile sections that postal zones could be defined in terms of the CATS division of the Chicago area.

The conversion method assumed that employment was distributed in the same manner as the corresponding land use measured in terms of either land area or floor area. This required a matching of employment and land use definitions, and the assumption of a uniform employment density in the postal zone.

Manufacturing was the only industry where it was possible to prepare district employment estimates based on both land area and floor area. Estimated employment in the other industries was based on floor area. A statistical test of variance was made with the two sets of manufacturing employment estimates to determine whether there was any significant difference in using land area or floor area. A cursory examination of the two sets of employment estimates showed great similarity between the two sets. See Part IV, Work Trips, Table 7. And this indeed was the case as the statistical analysis of variance demonstrated.

With two bases of classification, district and estimation method, the following data were assembled:

$$\sum X_T^2 = 42,522.60$$

$$\sum X_R^2 = 42,447.87$$

$$\sum X_K^2 = 3.99$$

$$\sum X_D^2 = 70.74$$

The estimates of variance were:

$$S_R^2 = 42,447.87/15 = 2,829.86$$

$$S_K^2 = 3.99/1 = 3.99$$

$$S_D^2 = 70.74/15 = 4.72$$

The variance ratio for districts was:

$$F_R = \frac{2829.86}{4.27} = 662.73$$

and for the estimating methods:

$$F_K = \frac{3.99}{4.72} = .845 \qquad F_{.05,\ 1,\ 15} = 4.54$$

$$t = \sqrt{.845} = .92 \qquad v = 15 \text{ and } \alpha .10$$

$$t\alpha - t_{.10} = 1.341$$

There was variance in employment between districts. The variation between the two estimating methods was not significant.

In estimating employment in the CATS district from postal zone employment, the basic assumption was made that the employment density was uniform throughout the postal zone. It was obvious, however, that there were variations in employment density, and that error could have occurred, if there were extreme variations in the employment density among firms in the postal zone whose employment was assigned to two or more CATS districts. There was no way of checking the location and amount of this possible error without pinpointing the exact employment location of some 50, 000 firms. It was possible to compute the average postal zone employment density, but this did not point out errors which might have resulted from the assumption of uniform density. A comparison of the postal zone employment densities with the work trip densities in overlapping CATS districts revealed some possible under and over reporting of work trip destinations.

TECHNIQUES FOR MATCHING AREAS IN POSTAL ZONES AND CATS DISTRICTS.

Of interest to those familiar with the CATS land use inventory, is the actual method of allocating land use data to the postal zones. The problem is familiar to anyone who has been confronted with a metropolitan area which has been divided into as many ways as there are agencies collecting and publishing data on metropolitan characteristics. Full use of much data is lost when a second agency cannot use the data without first converting the data to its own division of the metropolitan area.

The land use information used to distribute employment was the CATS 1956 survey of land use. The CATS survey inventoried floor area by its two digit land use classification in the City of Chicago and selected suburban municipalities. Land area was inventoried in the entire study area, but it was generalized by six uses: residential, manufacturing, transportation and public utilities, public buildings, and public open space.

While the block was used for measurement, the smallest unit of area for which tabulations were made was the quarter square mile. From the quarter square mile, summaries were made for the zone, district, ring, sector, and total area.

As a step in distributing postal zone employment to the CATS districts, the quarter square mile totals were used to find the total amount of a particular land use in a postal zone. In a few cases the postal zone boundary split a quarter square mile, and it was then necessary to assign the entire quarter square mile to one postal zone. In most instances this arbitrary assignment was not serious. It did, however, affect the final estimates of employment in CATS districts 01 and 11; 01 being low and 11 being high. The four quarter square miles in CATS district 01 were split by postal zones 1, 2, 3, 4, 5, 6, 7, 10, and 11. These quarter square miles were assigned to the sum of postal zones 1, 2, 3, and 4, but this meant that a portion of the employment in postal zones 5, 6, 7, 10, and 11, properly belonging to CATS district 01, was assigned to CATS district 11. An easy method for correcting this error was the addition of the employment in CATS districts 01 and 11.

After the CATS land use survey was completed, several corrections were made on the land area tabulations at the zonal level. No records were left, however, on how these corrections affected the quarter mile totals on land area. In some 25 CATS zones it was necessary to estimate the manufacturing land area in the quarter mile square before an allocation of land area could be made. The estimate was made arbitrarily by dividing the total land area into equal quarters and assigning the quarters to the appropriate postal zones.

Establishing a Statistical Criterion for Selecting Trip Generation Procedures

HAROLD D. DEUTSCHMAN, Transportation Engineer,
Tri-State Transportation Commission

•TRIP generation procedures as used by regional transportation studies involve the systematic explanation of the relationship between the dependent variables (person trips per household and autos available per household) and the independent measures of social, economic, and household activities. Predictive equations relating travel characteristics to the behavior of the household (measuring the travel demands of the households) are part of the basis for the systematic planning of the network of highways and mass transit facilities to meet this demand.

The intent of this study is to examine the independent variables used in the trip generation process in the context of a developed criterion in order to choose the best single variable or combination of variables to most efficiently forecast the dependent variables of person trips per household and autos per household. Much attention is devoted to the establishment and development of a criterion to measure how closely the prediction approximates reality. Sources of error examined in the forecasting process include (a) errors in estimating the independent variables, (b) errors in the simulation of the dependent variables for present-day conditions, and (c) errors in the forecasting equation using the independent variables. It is the joint effect of these three sources of error that produces the actual error of estimate. A minimization of this joint error is established as the standard for selecting the "best" trip generation procedure.

Preliminary findings from the Tri-State Transportation Commission are used to analyze the three types of errors and to point out the sensitivity of selection and the decision-making process of the analyst in selecting the most effective variables for trip generation purposes. In addition, auto registration and census data for the New York Metropolitan Area for the years 1950 and 1960 are used as a data base to analyze the error in the trip generation equation when used as a predictive device.

DEVELOPING THE CRITERION

A systematic approach in a trip generation process is to select an equation and/or model of n number of independent variables vs the dependent variables of person trips and autos per household. These equations are then tested against survey data to see how well they reproduce the data, the figure of merit usually consisting of the standard error of estimate and the coefficient of correlation. The next step is to examine the procedure to see if it logically may be used as a predictive device. The values of the independent variable must, of course, change over time at approximately the same rate as the dependent variable. It is not unusual to have a condition whereby an independent variable reproduces the survey data very well but fails completely when used for forecasting purposes. A third step in the trip generating process is to estimate how accurately and to what geographic level of detail the independent variables may be estimated. There would be a trade-off, of course, between (a) choosing an independent variable yielding an excellent correlation with the dependent variable but being difficult to estimate and (b) selecting a variable which is easy to estimate but which has only a fair-to-good correlation with the dependent variable. A fourth step in the procedure is to compute the joint error of the three sources of error described, with the statistical

---

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.

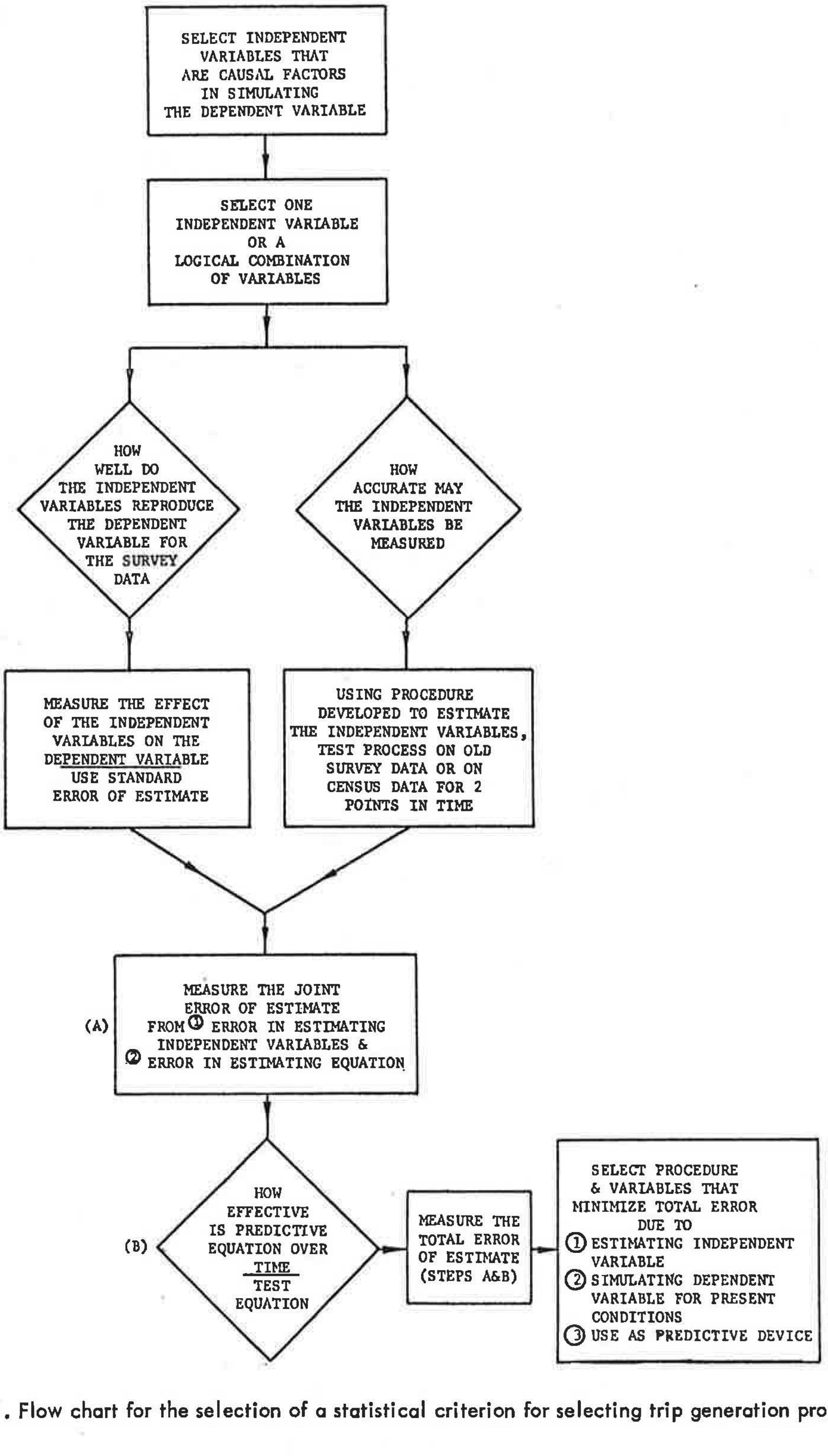


Figure 1. Flow chart for the selection of a statistical criterion for selecting trip generation procedures.

criterion for selecting a procedure simply consisting of a minimization of this joint error of estimate. A schematic diagram detailing this procedure is shown in Figure 1. The formula for calculating the joint error of estimate is as follows:

$$\text{ERROR joint} = \sqrt{\begin{matrix}\text{error}^2\\ \text{in}\\ \text{estimation}\\ \text{of}\\ \text{independent}\\ \text{variables}\end{matrix} + \begin{matrix}\text{error}^2\\ \text{in}\\ \text{estimation}\\ \text{equation to}\\ \text{simulate de-}\\ \text{pendent variable}\end{matrix} + \begin{matrix}\text{error}^2\\ \text{in}\\ \text{predictive}\\ \text{power of}\\ \text{equation}\\ \text{over time}\end{matrix}}$$

ERROR IN SIMULATING DEPENDENT VARIABLES FOR PRESENT-DAY CONDITIONS

Reproducing the Survey Data

Transportation studies spend much attention in determining how well their independent variables reproduce the survey data's dependent variables. The measurement of the intensity of the effect of the independent variables on the dependent variables of person trips per household and autos available per household is usually made by the statistical measures of the standard error of estimate and the coefficient of correlation. For the purposes of this analysis, the standard error of estimate will be used as the "error" measurement.

Preliminary findings from the Tri-State Transportation Commission are used to examine the variables for the error generated in reproducing the survey data. In expanding the Tri-State home interview survey, the study area was divided into 278 expansion areas which are composed of groups of census tracts, municipalities, and groups of municipalities. These areas are used as zones for trip generation equations in which data are available on person trips, auto availability, household characteristics, and density measures.

Trip generation rates were derived by using linear relationships between the dependent and independent variables. When necessary the variables were transformed to obtain this linear relationship (i. e., logarithm of gross density). Sets of equations were developed with the following independent and dependent variables:

Dependent Variable	Independent Variable (s)
(X_1) Person trips/household	(X_2) Vehicles/household (X_4) Median household income (X_6) Percent single unit structures (X_7) Log of gross residential density (X_3) Persons 5 years and older/household
(X_2) Vehicles/household	(X_4), (X_6), (X_7), (X_3)

In addition, linear combinations of the independent variables were tested against the two dependent variables. The results of this analysis are described below.

Dependent Variable—Person Trips per Household

The best single independent variable for estimating person trips for the survey data is vehicles per household, with the density measure of percent single unit structures ranking second while yielding a slightly greater error. The ranking of the independent variables, using the standard error of estimate divided by the mean of the dependent variable as a criterion for ranking, is as follows:

Independent Variable	S/$\overline{X}_1$ (Expressed as Percent)
Vehicles per household	15
Percent single unit structures	17
Log of gross residential density	21
Persons 5 years and older per household	31
Median household income	32

When vehicles per household is linearly combined with percent single unit structures, the standard error of estimate for estimating person trips is reduced to 14 percent, and when persons 5 years and older is added to these two variables, the standard error of estimate is reduced to 13 percent.

Dependent Variable—Vehicles per Household

The most efficient sole determinant of vehicles per household is a measure of residential density. Percent single unit structures and the logarithm of gross density, both approximations of residential density, yield the same magnitude of error (22 percent), while indicators of income and persons per household each produce about 2 times this error. The linear combination of (i. e., logarithm of) gross density and median household income reduces this error of estimate to 19 percent while the inclusion of persons per household with these two independent variables yields an error of 18 percent.

The ranking of the independent variables by their associated standard error of estimate in estimating vehicles per household is as follows:

Independent Variable	S/X_1 (Expressed as Percent)
Percent single unit structures	22
Log of gross residential density	22
Median household income	43
Persons 5 years and older per household	45

The equations for the regression lines and corresponding errors of estimate are detailed in Tables 1 and 2 of the Appendix for the dependent variables of vehicles per household and person trips per household.

ESTIMATION OF INDEPENDENT VARIABLES

There has not been a great deal of analytical work published by transportation planning groups on determining the error in estimating independent variables or predictors. The ideal case would be to set up a zonal system equivalent to the one planned for use in the forecasting process, and then make use of census data for 1950 and 1960 or previous surveys taken in the area to serve as the test of how well the independent variables may be estimated. Testing of this estimating process should be initiated even though data may be scarce or available only on a coarse geographic level.

It is generally agreed by analysts in the transportation field that the independent variables may be ranked by the ease and efficiency of estimation as follows: (a) population-related data; (b) density-related data, i. e., persons per residential area; (c) combination of population and density data, i. e., persons per household; and (d) income, i. e., median household income.

For the purposes of this paper, a hypothetical structure is created to measure the sensitivity of the error in estimating the independent variables on the efficiency of reproducing the survey data. The joint error of estimate from these two sources is calculated for an array of assumed errors of estimation for the independent variables.

The hypothetical structure used for analysis is based on a perturbation of the data from the Tri-State Transportation home interview survey in which (for purposes of expansion) the study area was divided into 278 zones. An error of estimation was applied to the independent variables (income, density, and vehicle measures), this error being a constant percentage of the actual zonal value, its sign (plus or minus) generated by a random number index. The following variables served as a basis for testing selected sets of equations:

Dependent Variable	Independent Variables
(S_1) Person trips/household	(S_3) Vehicles/household (0% error)
(S_2) Vehicles/household	(S_4) Vehicles/household (15% error)
	(S_5) Percent single family units (0% error)
	(S_6) Percent single family units (10% error)
	(S_7) Median household income (0% error)
	(S_8) Median household income (10% error)
	(S_9) Median household income (20% error)
	(S_{10}) Median household income (25% error)

Illustration of Perturbation Procedure

	Independent Variable—Median Household Income, 10% Error (S_8)		
Zone	Actual (Survey) Value of Income (S_7)	Sign of Error (Generated by Random Number Index)	New Value of Income With 10% Error (S_8)
1	$5000	+	$5500
2	$5500	+	$6050
3	$6000	-	$5400
4	$4000	+	$4400
5	$4500	-	$4050
6	$9000	-	$8100
7	$8200	+	$9020

The (joint) error of reproducing the survey and in estimating the independent variables was determined in a single calculation by running the (same) regression analysis with fixed errors in the independent variables.

Vehicles per Household

Assuming that a realistic figure for the error in estimating percent single family units is ±10 percent and the error in estimating median household income is in the range of ±20-25 percent, then either the use of (a) density as a sole variable for estimating vehicles per household, or (b) the linear combination of density and income to estimate vehicles per household, would yield equivalent results. If income may be estimated to within ±10 percent, then the joint effect of income and density would be a better estimator of vehicles than would density alone (standard error of estimate of 22% vs 24%). A more detailed tabular description of the results is shown in Table 3, Appendix.

Person Trips per Household

Assuming that the error in estimating the dependent variable of vehicles per household is ±15 percent (with the errors in estimating density and income previously described), then the use of either (a) vehicles per household as a sole determinant, or (b) the linear combination of vehicles and income, yields (approximately) the same standard error of estimate in estimating person trips per household (20%). A perturbation of a ±15 percent error in median household income (as a sole independent variable) causes the error in estimating person trips to rise from 15 percent to 20 percent (when compared to the theoretical case of vehicles estimated without any error, i. e., for known values of vehicles). See Table 4, Appendix, for a complete description of the results.

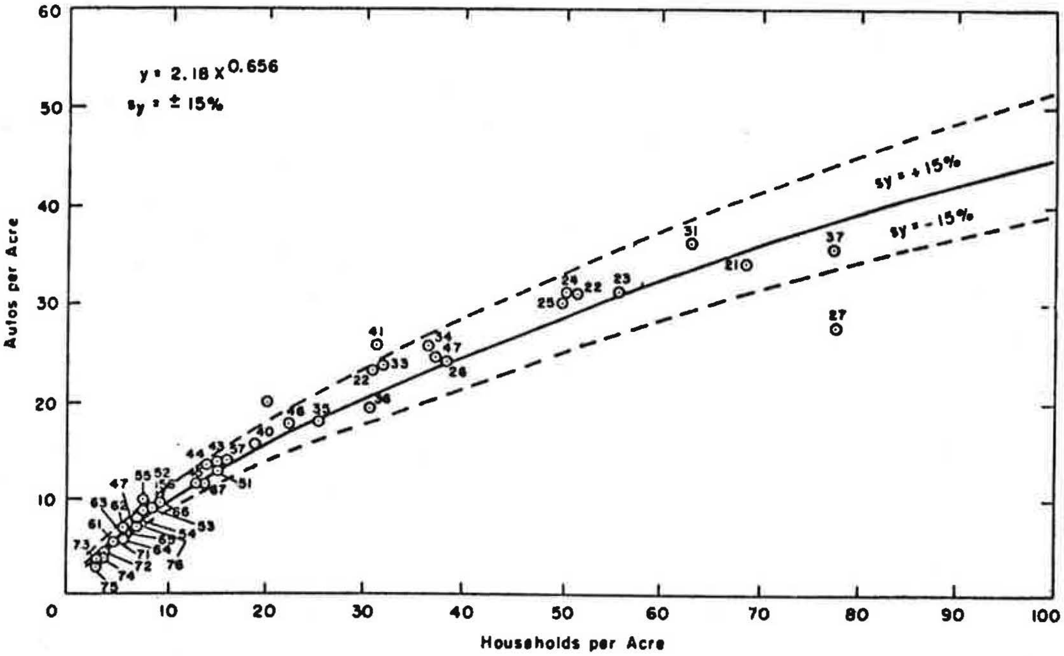


Figure 2. Relation between autos per acre and households per acre for 77 districts in Chicago, 1956-1957.

The joint error incurred in estimating the independent variables and reproducing the survey data (for the dependent variables) does not in itself yield a single clear-cut choice of trip generation equation (or procedure). It does narrow the list of variables and equations, however, to a few which now must undergo the test of forecasting efficiency.

USE AS A PREDICTIVE DEVICE

The most important test of the effectiveness of a variable or a group of variables in a trip generation equation is a test of its use as a predictive device. First, the relationship between the independent variables and the dependent variable should be visually displayed with the analyst studying the display to insure that the dependent variable is sensitive to changes in the independent variable. Hypothetical cases may be developed as a check on the predictive logic of the relationship. To illustrate, the relationship between households per acre (a measure of residential density) and autos per acre is cited; these data were derived by Cherniack (1) from data supplied by the Chicago Area Transportation Study (Fig. 2). This relationship (as interpreted by this author) in effect shows an excellent correlation for the present (or survey year), but probably needs a time parameter factor for use in future estimates. A graph of vehicle availability vs household income stratified by number of housing units in the structure (Fig. 3) reveals that vehicle availability will be higher (considering two areas in the same density configuration) where the household income is higher. To illustrate, a difference in income between \$6000 and \$8000 in high-density (apartment house) type residences would yield an average difference in vehicle availability per household of approximately 35 percent (0. 46 to 0. 62), while this difference for an area of single family units would be 16 percent (1. 18 to 1. 37).

There are also obvious limitations in using income as a sole measure for predicting autos. Income may only be used as a predictor of autos if the density configuration of an area remains constant since the rate of vehicles available differs significantly by density classifications. For example, the average rate for household in single family

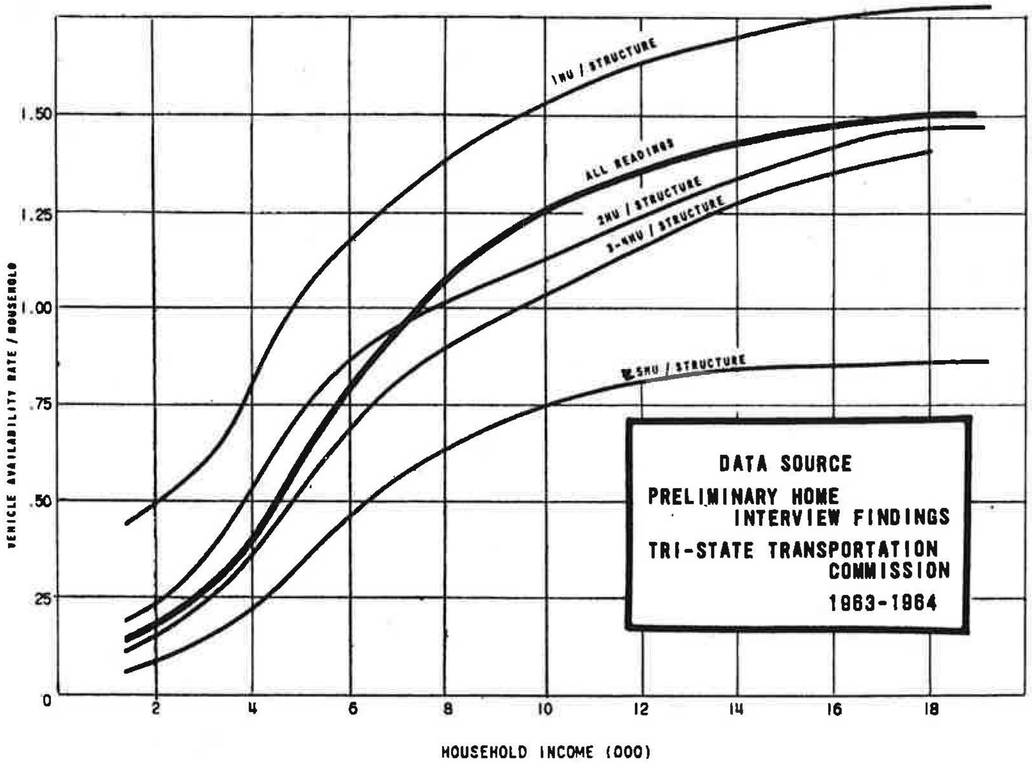


Figure 3. Vehicle availability vs household income, stratified by number of housing units in the structure.

units earning $10, 000 per year is more than twice as great as the households living in 5 or more units per structure (apartment houses) and earning a similar income.

In order to derive an objective figure of merit for the effectiveness of the independent variables in predicting the dependent variable, the trip generation equation had to be tested over two points in time. To accomplish this for the dependent variable autos per household, auto registration data were abstracted from state vehicular records with census data describing the density and income variables. The base years for this analysis were 1950 and 1960, with counties in the New York Metropolitan Area as the geographic area (or zones) used. This geographic detail is much too coarse but, because auto ownership rates were not available from the 1950 census, county totals from auto registrations were the best available source. In the future (as in 1960), the census will report autos available to households on the census tract level such that a similar test of the predictive power of the independent variable may be conducted on a fine-grained geographic level.

The strategy involved in this analysis of evaluating the predictive power of the variables for forecasting autos included deriving the best-fit linear regression line for 1950 for each of the independent variables and selected combinations of variables. These relationships were then used to estimate autos per household for 1960, and compared to the known actual figure of autos per household for 1960. The root-mean-square error was chosen as the figure of merit for comparing the results, which are shown in Table 5 in the Appendix.

TOTAL ERROR—BASIS FOR SELECTING THE TRIP GENERATION EQUATION

The procedure for calculating the total error of estimate for forecasting the dependent variable vehicles per household is illustrated below. It must be remembered that the joint error of reproducing the survey and error of estimating the independent variables was derived from a zonal scheme of 278 zones while the error in predictive power was based on a coarse zonal scheme of only 17 zones. The total error of estimate, therefore, should not be used as an absolute value but as a figure of merit to compare the various trip generation equations.

Dependent Variable—Vehicles per Household

Equation	Independent Variables Used	E_1 Error in Reproducing Survey (e_1) and Error in Estimating Independent Variables (e_2) ($E_1 = e_1 + e_2$)	E_2 Error in Use as Predictive Device	$\sqrt{E_1^2 + E_2^2}$ Total Error
1	Percent single family units	24%	20.0%	31.0%
2	Percent single family units and median household income	23%	8.5%	24.5%
3	Median household income	49%	20.0%	53.0%

The results indicate that the linear combination of median household income and percent single family units yields a significantly lower error of estimate (and is thus the recommended procedure) for forecasting vehicles per household when compared to the next best two equations. (It should be noted that for this sample analysis, percent single family units was given a ± 10% error and median household income a ±20% error of estimation.)

Unfortunately, data on person trips per household for two points in time were not yet available so that a similar analysis of the "total" forecast error could not be made for this dependent variable.

CONCLUSION

It is the purpose of this paper to present a methodology (or philosophy) to objectively select those independent variables that yield the best prediction of the dependent variables of vehicles and person trips per household. The procedure has been illustrated by sample calculations, using the Tri-State Transportation home interview survey as the primary data source. In the near future, Census Bureau publications will describe the number of vehicles available to the households on a small-area basis along with measures of income and density. This rich source of data will make it possible to use a single consistent zonal scheme in calculating the three different sources of errors described in this paper in forecasting vehicles per household. It may also be possible to make similar calculations for the dependent variable of person trips per household if the original home interview surveys are updated so that household trip information is available for two periods of time.

In the next decade, many of the large metropolitan transportation studies will reevaluate the travel demands to be generated by the households, update their travel surveys, and adjust their forecasts. It seems imperative to develop a criterion that objectively tests and reevaluates the forecasting procedures to select that one which will most effectively describe the future. It is hoped that this paper stimulates more thought in this area of concern for transportation planners.

REFERENCE

1. Cherniack, Nathan. Critique of Home-Interview Type O-D Surveys in Urban Areas. HRB Bull. 253, pp. 166-188, 1960.

Appendix

TABLE 1

REGRESSION LINES—DEPENDENT VARIABLE, PERSON TRIPS PER HOUSEHOLD

Dependent Variable	Equation No.	Independent Variable(s)		Equation	R (Coefficient of Correlation)	S Std. Error of Estimate	$S/\bar{X}$. (Expressed in Percent
		No.	Description				
(X_1) Person Trips/ H.H. ($\bar{X}_1$=5.37)	01	X2	Total Vehicles/H. H.	$X_1 = 4.784X_2 + 1.578$	0.94	0.83	15%
	02	X3	Persons 5 years & Older/H. H.	$X_1 = 3.978X_3 - 5.668$	0.72	1.67	31%
	03	X4	Median Household Dollars	$X_1 = 0.0008156X_4 -0.06941$	0.68	1.74	32%
	04	X5	Gross Density (living Quarters/Sq. Mile)	$X_1 = -0.000099X_5 + 6.736$	0.68	1.75	33%
	05	X6	Percent Single Unit Structures	$X_1 = 0.06597X_6 + 3.292$	0.93	0.90	17%
	06	X7	Log Gross Density	$X_1 = -2.938X_7 + 16.318$	0.88	1.15	21%
	07	X2 X3	Total Vehicles/H.H. Persons 5 years & Older	$X_1 = 4.205X_2 + 0.9567X_3 -0.6180$	0.95	0.77	14%
	08	X2 X3 X4	Total Vehicles/H.H. Persons 5 years & Older/H. H. Median Dollars/H.H.	$X_1 = 3.502X_2 + 1.227X_3 + 0.0001845X_4 - 2.042$	0.95	0.73	14%
	09	X2 X3 X4 X7	Total Vehicles/H.H. Persons 5 years & Older/H.H. Median Dollars Household Log Gross Density	$X_1 = 3.200X_2 + 1.200X_3 +0.0001955X_4 - 0.2040X_7 -1.040$	0.95	0.73	14%
	10	X2 X3 X4 X6	Total Vehicles/H.H. Persons 5 years & Older/H.H. Median Household Dollars Percent Single Unit Structures	$X_1 = 2.057X_2 + 1.012X_3 + 0.000177X_4 + 0.02396X_6 -1.007$	0.96	0.67	12%
	11	X3 X6	Persons 5 years & Older/H. H. Percent Single Unit Structures	$X_1 = 0.9217X_3 + 0.05802X_6 + 0.985$	0.93	0.85	16%

Table 1 (Continued)

Dependent Variable	Equation No.	Independent Variable(s) No.	Independent Variable(s) Description	Equation	R (Coefficient of Correlation)	S Std. Error of Estimate	S/$\bar{X}$, (Expressed in Percent)
(X) Person Trips/H.H. ($\bar{X}_1$ = 5.37)	12	X2 X6	Total Vehicles/H.H. Per Cent Single Unit Structures	$X_1 = 2.854X_2 + 0.0290X_6 + 2.195$	0.95	.75	14%
	13	X2 X7	Total Vehicles/H.H. Log Gross Density	$X_1 = 4.50X_2 - 0.20503X_7 + 2.570$	0.94	.83	15%
	14	X2 X4	Total Vehicles/H.H. Median Household Dollars	$X_1 = 4.501X_2 + 0.000097X_4 + 1.156$	0.94	.82	15%
	15	X2 X5	Total Vehicles/H.H. Gross Density (Living Qtrs/(mi)^2)	$X_1 = 5.017X_2 + 0.000009X_5 + 1.2725$	0.94	.83	15%
	16	X4 X6	Median Household Dollars % Single Unit Structures	$X_1 = 0.000206X_4 + 0.0583X_6 - 2.157$	0.94	.84	16%
	17	X4 X7	Median Household Dollars Log Gross Density	$X_1 = 0.000358X_4 - 2.3966X_7 + 11.909$	0.91	.98	18%
	18	X3 X4 X6	Persons 5 years and Older/H.H. Median Household Dollars % Single Unit Structures	$X_1 = 1.267X_3 + 0.000272X_4 + 0.0449X_6 - 1.374$	0.95	.74	14%
	19	X2 X3 X6	Total Vehicles/H.H. Persons 5 yrs. & Older H.H. % Single Unit Structures	$X_1 = 2.691X_2 + 0.746X_3 + 0.0246X_6 + 0.390$	0.95	.71	13%
	20	X2 X4 X6	Total Vehicles/H.H. Median Household Dollars. Percent Single Unit Structures	$X_1 = 2.507X_2 + 0.000107X_4 + 0.0295X_6 + 1.74$	0.95	.73	14%
	21	X3 X4	Persons (5 yrs. & Older/H.H.) Median Household Dollars	$X_1 = 3.1768X_3 + .00063X_4 - 7.644$	0.88	1.15	21%
	22	X2 X4 X7	Total Vehicles/H.H. Median Household Dollars Log Gross Density	$X_1 = 4.1028X_2 + 0.000074X_4 - 0.3403X_7 + 2.887$	0.94	.82	15%

TABLE 2

REGRESSION LINES—DEPENDENT VARIABLE, VEHICLES PER HOUSEHOLD

Dependent Variable	E-quation No.	Independent Variable(s)		Equation	R (Coefficient of Correlation)	S Std. Error of Estimate	S/X̄, (Expressed in Percent)
		No.	Description				
Vehicles/H.H.(X_2)	23	X3	Persons 5 years & Older/H.H.	$X_2 = 0.7185X_3 - 1.2009$	0.66	0.35	45%
($\bar{X}_2$) = 0.79	24	X4	Median Household Dollars	$X_2 = 0.0001597X_4 - 0.2724$	0.68	0.34	43%
	25	X5	Gross Density* (Living Quarters/Sq. Mi.)	$X_2 = -0.0000214X_5 + 1.089$	0.76	0.31	39%
	26	X6	% Single Unit Structures	$X_2 = 0.01296X_6 + 0.3843$	0.93	0.17	22%
	27	X7	Log Gross Density	$X_2 = -0.60795X_7 + 3.0579$	0.92	0.18	23%
	28	X3 X4	Persons 5 years & Older/H.H. Median Household Dollars	$X_2 = 0.5567X + 0.0001271X_4 -1.600$	0.84	0.27	34%
	29	X4 X7	Median Household Dollars Log Gross Density	$X_2 = 0.00006141X_4 - 0.515X +2.302$	0.95	0.15	19%
	30	X3 X4 X7	Persons 5 years & Older/H. H Median Household Dollar Log Gross Density	$X_7 = 0.1123X_3 + 0.000065X_4 - 0.4641X_7 + 1.780$	0.95	0.14	18%
	31	X3 X4 X6	Persons 5 years & Older/H.H. Median Household Dollars % Single Unit Structures	$X_2 = 0.1236X_3 0.000046X_4 + 0.0102X_6 - 0.178$	0.94	0.16	20%
	32	X4 X6	Median Household Dollars % Single Unit Structures	$X_2 = 0.0000397X_4 + 0.01149X_6 + 0.1663$	0.94	.16	20%
	33	X3 X6	Persons (5 yrs. & Older/H.H.) % Single Unit•Structures	$X_2 = 0.065X_3 + 0.0124X_6 + 0.221$	0.93	.17	22%
	34	X3 X7	Persons (5 years & Older)/H.H. Log Gross Density	$X_2 = 0.060X_3 - 0.5833X_7 + 2.800$	0.93	.18	23%

TABLE 3

ESTIMATING VEHICLES PER HOUSEHOLD

DEPENDENT VARIABLE	INDEPENDENT VARIABLE (01)	ERROR IN ESTIMATING VARIABLE (01)	INDEPENDENT VARIABLE (02)	ERROR IN ESTIMATING VARIABLE (02)	STANDARD ERROR OF ESTIMATE
VEHICLES PER HOUSEHOLD	% Single Family Units	0%			22%
	% Single Family Units	10%	Median Household Income	10%	22%
	% Single Family Units	10%	Median Household Income	20%	23%
	% Single Family Units	10%	Median Household Income	25%	23%
	% Single Family Units	0%	Median Household Income	0%	20%
	% Single Family Units	10%			24%
VEHICLES PER HOUSEHOLD	Median Household Income	0%			43%
	Median Household Income	10%			46%
	Median Household Income	20%			49%
	Median Household Income	25%			51%

TABLE 4

ESTIMATING PERSON TRIPS PER HOUSEHOLD

DEPENDENT VARIABLE	INDEPENDENT VARIABLE (01)	ERROR IN ESTIMATING VARIABLE (01)	INDEPENDENT VARIABLE (02)	ERROR IN ESTIMATING VARIABLE (02)	STANDARD ERROR OF ESTIMATE
PERSON TRIPS PER HOUSEHOLD	% Single Family Units	0%			16%
	% Single Family Units	10%	Median Household Income	10%	17%
	% Single Family Units	10%	Median Household Income	20%	17%
	% Single Family Units	10%	Median Household Income	25%	17%
	% Single Family Units	10%			18%
PERSON TRIPS PER HOUSEHOLD	Vehicles per Household	0%			16%
	Vehicles per Household	15%			20%
PERSON TRIPS PER HOUSEHOLD	Vehicles per Household	0%	% Single Family Units	0%	14%
	Vehicles per Household	15%	% Single Family Units	10%	16%
PERSON TRIPS PER HOUSEHOLD	Vehicles per Household	15%	Median Household Income	10%	19%
	Vehicles per Household	15%	Median Household Income	20%	19%
	Vehicles per Household	15%	Median Household Income	25%	20%
	Vehicles per Household	0%	Median Household Income	0%	15%

TABLE 5

PREDICTED 1960 AUTOS PER HOUSEHOLD

Zone* No.	Actual 1960 Autos/H.H.	Predictive Variables Median Household Income	%(1 & 2) Unit Structures	Persons/ Household	Median Income & %(1 & 2) Unit Structs.	Median Income & Persons/Household	%(1 & 2) Unit Structs. & Persons/Household
1	1.26	1.17**	1.05	0.90	1.31	1.16	1.05
2	1.02	1.09	0.72	0.82	0.88	0.98	0.73
3	0.77	1.06	0.64	0.81	0.77	0.93	0.65
4	1.18	1.10	1.07	0.93	1.24	1.09	1.08
5	1.23	1.12	1.09	0.96	1.29	1.16	1.10
6	1.26	1.08	1.09	0.93	1.24	1.07	1.10
7	1.42	1.16	1.13	0.97	1.37	1.22	1.14
8	1.09	1.08	0.94	0.85	1.08	0.99	0.95
9	1.33	1.16	1.01	0.89	1.25	1.14	1.02
10	0.47	1.06	0.50	0.76	0.61	0.87	0.51
11	1.14	1.08	1.04	1.00	1.18	1.16	1.04
12	1.37	1.20	1.11	1.00	1.41	1.31	1.11
13	1.06	1.04	1.04	0.91	1.13	1.00	1.04
14	1.59	1.09	1.14	0.91	1.30	1.06	1.15
15	1.24	1.14	1.08	1.06	1.30	1.30	1.08
16	1.31	1.10	1.16	1.03	1.33	1.22	1.16
17	1.20	1.18	0.81	0.89	1.07	1.16	0.81
ROOT MEAN SQUARE ERROR COMPARISON = $\sqrt{\frac{\Sigma(\text{Actual-Predicted})^2}{N}}$ % ERROR		.23 (20%)	.23 (20%)	.33 (28%)	.10 (8.5%)	.19 (16%)	.23 (20%) % of mean of dependent variable

Mean of Dependent Variable (Autos/HH) = 1.17 for 17 zones

** To illustrate the use of this table, the number 1.17 refers to the number of autos per household predicted by the independent variable, Median Household Income

*Description of Zones

Zone

1 Bergen
2 Essex
3 Hudson
4 Mercer
5 Middlesex
6 Monmouth
7 Morris
8 Passaic
9 Union
10 New York City
11 Dutchess
12 Nassau
13 Orange
14 Putnam
15 Rockland
16 Suffolk

A Report on the Accuracy of Traffic Assignment When Using Capacity Restraint

THOMAS F. HUMPHREY, Urban Planning Division, U. S. Bureau of Public Roads

•THE purpose of this paper is to evaluate the accuracy of the traffic assignment process when using capacity restraint to calibrate an analysis network in the urban transportation planning process. The practical aspects of the capacity restraint theory are explained, followed by a presentation of some actual results obtained from 10 urban area transportation studies which used capacity restraint. Finally, the accuracy of the results obtained in these studies is evaluated.

Traffic assignment may be defined as the process of allocating a given set of trip interchanges to a specific transportation system. It is a reproducible, mechanical tool which allows the transportation planner to assign either present or future trips to alternative transportation systems, helps him evaluate the effects of these systems on the community, and aids in the determination of the transportation plan which will best serve the needs of the community.

The traffic assignment process requires that an "assignment model" be calibrated initially. This calibration simply means that the assignment model must be adjusted so that it can reproduce the vehicular travel that is taking place on the existing transportation network as accurately as possible. It is then assumed that the same type of assignment procedure may be used to allocate future trip interchanges to a future transportation system in a reasonable manner.

The Bureau of Public Roads "all or nothing" version of the traffic assignment process was developed on the assumption that a vehicle operator chooses a route of travel between his origin and destination on the basis of the least possible travel time between those two points. Thus, minimum time paths (called trees) are computed between zone pairs and the total trips destined between these zone pairs are assigned to the minimum time paths. Trips between zones are accumulated directionally on each link in the transportation network, and turning movements can be computed at each intersection.

The calibration of the assignment model refers to the process whereby the assigned volumes are adjusted until they match the existing traffic counts on all routes as closely as possible. This adjustment is accomplished by changing the speeds (and thereby the time) on links until a reasonable match is obtained. The adjustment of link speeds can be made either manually, by carefully examining assigned link volumes and corresponding ground counts and estimating the speed change that should be made on each link, or by using a computerized technique referred to as "capacity restraint."

A manual adjustment of speeds may be the most efficient means of calibrating the transportation network in a small urban area (under 100, 000). However, in the larger urban areas it is usually more desirable and efficient to make use of the capacity restraint technique. This is because there are many more links involved in the larger studies, and the area-wide effects of speed adjustments are accounted for when the latter technique is used. Also, a computerized model is developed which can be reproduced mechanically, thereby increasing the confidence of the results obtained when assigning future trips to a network.

Using the capacity restraint technique, speeds are adjusted according to the ratio of assigned volume to practical capacity on individual links. Thus, for a link having an assigned volume greater than its practical capacity, the speed existing on the link for that assigned volume is reduced (thereby increasing the travel time on that link) to make

---

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.

it less desirable in the route selection process. Conversely, for a link having an assigned volume which is less than its practical capacity, the speed existing on the link for that assigned volume is increased (thereby decreasing the travel time on that link) to make it more desirable in the route selection process.

This adjustment is made on each link in the network for which practical capacity has been computed until speeds are obtained that result in an assignment which matches the existing ground counts most reasonably. Usually, three or four adjustments using capacity restraint furnish the desired results. A more detailed discussion of the traffic assignment process as well as the capacity restraint theory is found in the Traffic Assignment Manual (1). A discussion of the practical aspects and significance of the results obtained when using capacity restraint is included in subsequent sections of this paper.

THE NEED FOR CAPACITY RESTRAINT

The usual procedure employed in calibrating the traffic assignment model starts with computing minimum time paths (trees) between all zones in the analysis area, and assigning the trips obtained in the O-D survey to each tree. The speeds used to compute these trees are the ones obtained from the travel time study. If they represent a realistic estimate of the speed of vehicles on the highway network, the assignments made on this basis may be reasonably adequate. However, in most cases it is extremely difficult to obtain speeds which actually represent "average" operating speeds on the network. Consequently, the network speeds must be adjusted either manually or by using the capacity restraint technique. In most studies, especially those over 100,000 population, the use of capacity restraint has been found to be the best means for making adjustments.

Sometimes the question is asked, Why not use a volume restraint rather than a capacity restraint in adjusting a network, since it is traffic volume rather than capacity that we are trying to match? A battery of volume restraint programs was written for the IBM 704 computer and used for several network calibrations; however, it was determined that capacity restraint would be more efficient for calibrating a network (2). It must be remembered that the purpose of calibrating a traffic assignment model is to develop a dependable tool which can be used to assign future trips to a future network with relative confidence. Traffic volumes are not available for calibrating a future network, whereas practical capacity can be computed and used for this purpose. Thus, it is reasonable to conclude that the traffic assignment model should be calibrated for existing conditions using the same relative basis for adjustment (i. e., practical capacity) that will be used to calibrate a future network. Consequently, the capacity restraint technique was programmed for the Bureau of Public Roads package of IBM 7094 computer programs.

THE APPLICATION OF CAPACITY RESTRAINT

As explained earlier, when the capacity restraint procedure is applied to the network, the speed on each link having a value for capacity available is examined and a new speed is computed according to the ratio of volume to capacity. A new set of trees must then be built to reflect the changes in speeds which have been made on the entire network. Next, another assignment of O-D trips is made to the new set of trees, and the ground counts can be compared to the assigned volumes on each link. This adjustment, i. e., the process of building a new set of trees and assigning O-D trips to those trees, is commonly referred to as an iteration. Usually, a maximum of three or four iterations is needed to obtain acceptable comparisons between ground counts and assigned volumes. The first iteration usually produces a large adjustment to the speeds, and each subsequent iteration usually changes the link speeds by a relatively smaller degree until a "balanced" network is obtained. In the following, an explanation of the physical changes which actually take place as each iteration of capacity restraint is applied will be made.

Figure 1 represents two zones in a study area, and a partial highway network connecting these two zones (nodes 100 through 106). Assume that 400 trips occur between

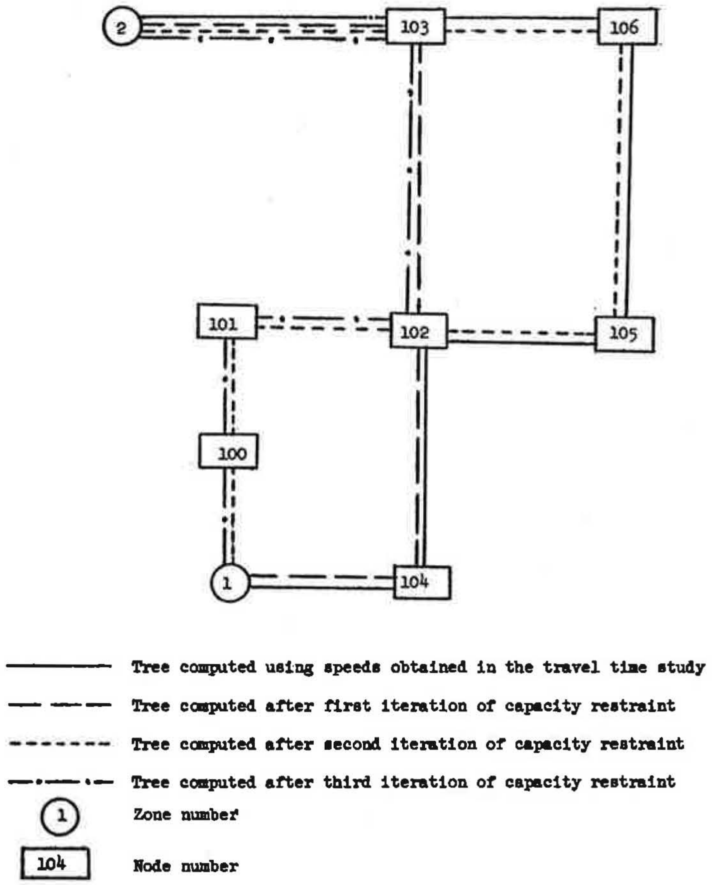


Figure 1. Illustration of the physical changes which take place on a highway network when using capacity restraint.

zones 1 and 2, and that three iterations of capacity restraint have been applied to the network. Thus, four trees have been computed between zones 1 and 2, each using a different route through the network as illustrated in the diagram. Each time a tree was built, 400 trips were assigned to that tree, the ratio of assigned volume to practical capacity was computed, a new speed was obtained for the next assignment, the next set of trees was built, and the process repeated for each iteration. The capacity restraint program has the ability to store the assignments made on each link for each iteration, and then compute the average assigned volume made on that link. For example, link 2 to 103 has been used four out of four times that trees were computed; therefore, it has an accumulated total volume of 4 × 400, or 1, 600 trips assigned to it. Similarly, link 100 to 101 has been used twice; therefore, it has accumulated a total volume of 2 × 400 = 800 trips assigned to it. A total of four assignments have been made to the network, each using a different set of trees. To obtain an average loading on the network, all accumulated volumes are then divided by four. This results in a final assigned volume on link 2 to 103 of 1, 600 ÷ 4, or 400 trips; on link 100 to 101 the final assigned volume is 800 ÷ 4, or 200 trips. The same procedure applies to all links in the network to which trips have been assigned, and the capacity restraint program makes all computations.

As seen from the example, a diversion type of assignment actually takes place during the capacity restraint process. Thus, trips are assigned to several different routes in the network and not just to those having the most desirable travel times initially, as

```
    RESULTS OF FIRST LOADING - ①


② TOTAL MEASURED VOLUME     1,777,278
③ TOTAL ASSIGNED VOLUME     1,515,252

④ AVERAGE PERCENT ERROR IN ASSIGNED VOLUME  -14.7          ERROR BREAKDOWN,    LATEST LOADING
      ⑤            ⑥          ⑦             ⑧             ⑨            ⑩             ⑪           ⑫
  VOL  GROUP  NO. SECTS   AV COUNT      AVE DIFF      STAN DEV    PC STAN DEV    PC OF TOTAL    WEIGHTED
   00-1/2          2          436       +   718          189         43.3            .0           .0
   1/2-01          3          686       +   333          303         44.1            .1           .0
    01-02          9        1,786       -   147        1,054         59.0            .9           .5
    02-03         11        2,512       - 1,000        1,897         75.5           1.6          1.2
    03-05         22        3,876       - 1,002        2,502         64.5           4.8          3.0
    05-10         56        7,318       - 1,501        3,948         53.9          23.1         12.4
    10-15         25       11,840       -   589        3,618         30.5          16.7          5.0
    15-20         24       17,501       - 1,090        6,136         35.0          23.6          8.2
    20-25          5       22,561       - 3,867        6,621         29.3           6.3          1.8
    25-30          2       27,407       -11,775       16,672         60.8           3.1          1.8
    30-up         10       35,186       - 6,230       11,964         34.0          19.8          6.7
    TOTAL        169       10,516       - 1,550        4,846         46.0         100.0         40.6
```

Figure 2. Summary statistics for the first free loading.

reflected in the travel time study. The diversion effect of this procedure is one definite advantage of the capacity restraint process, because it has the effect of distributing trips to several routes between zone pairs; in this way a better comparison with ground counts is usually obtained.

STATISTICS OBTAINED FROM THE CAPACITY RESTRAINT PROGRAM

As stated earlier, the purpose of calibrating a traffic assignment model is to develop a mechanical process which can be used to reproduce the vehicular travel that is taking place on the existing transportation network as accurately as possible. The ability of the model to perform adequately is measured by comparing the assigned volumes on each link to the ground counts which have been obtained for that link. (As explained in the Traffic Assignment Manual, assignments can be made for an ADT, a. m. peak, or p. m. peak network.) The BPR capacity restraint program has been written to provide summary statistics which can be used to measure the ability of the traffic assignment model to match ground counts after each iteration has been completed; these statistics can also be computed to show the effect of averaging the results of several iterations. Figure 2 shows the summary statistics obtained from the first free assignment (called the first loading) for a highway network used in an actual urban transportation study. An explanation of the values shown in Figure 2 will be given to provide a better understanding of their significance. A more detailed discussion of these statistics is provided elsewhere (3). The numbers enclosed in a circle in Figure 2 are keyed to the following explanation.

① Results of First Loading—This means that the statistics given on this page refer to the assignment of trips to the network using the travel times (or speeds) obtained from the original travel time study as the basis for computing the interzonal minimum time paths (trees). Thus, these trees are built using the time or speed that was coded in the link data cards (1). This loading is referred to as the "first free loading," and it is sometimes called the "first iteration."

② Total Measured Volume: 1, 777, 278—This number represents an accumulation of the total ground counts that have been coded on the link data cards used in the network (1).

③ Total Assigned Volume: 1, 515, 252—This number represents an accumulation of the total assigned vehicles to those links which have a ground count coded.

④ Average Percent Error in Assigned Volume: -14. 7—This number represents the ratio of total assigned volume (those links with ground counts) divided by total measured volume minus 100, or

$$\frac{1,515,252}{1,777,278} - 100 = -14.7$$

⑤ Vol. Group (volume group)—Into each row labeled 00-½, ½-01, 01-02, etc., is placed the data for all links having a coded ground count which falls into the range of 1 to 500, 501 to 1,000, 1,001 to 2,000, etc.

⑥ No. Sects. (number of sections)—The numbers in each row represent the number of sections (i.e., the number of links) which have the appropriate ground count coded. For example, in volume group 1 to 500, there are two links which have a ground count coded having a value within the range 1 to 500; there are three links having a ground count coded within the range 501 to 1,000; there are nine links having a ground count coded within the range 1,001 to 2,000, etc. The total given at the bottom of the tabulation (169) is the total number of links which have ground count data available. This is not the total number of links coded for the entire network.

⑦ Average Count—The number in each row represents the average ground count coded for all the links which fall within the appropriate range. For example, within the volume group 1 to 500 there are two links having an average ground count of 436; within the volume group 501 to 1,000 there are 3 links having an average ground count of 686; within volume group 1,001 to 2,000 there are 9 links having an average ground count of 1,786, etc. The total given at the bottom of this column is the average ground count for the entire network; this is computed by multiplying the number of sections in each volume group by the average count for that group, accumulating these products, and dividing that total by the total number of sections (169 in this example). This number is provided for relative comparative purposes only.

⑧ Ave. Diff. (average difference)—The numbers in each row represent the difference between the average ground count and the average assigned volume for each volume group. For example, for volume group 1-500, the average assigned volume is 718 vehicles greater than the average ground count on the two links which fall within that range; thus, the average assigned volume on these two links equals 436 + 718, or 1,154. For volume group 1001-2000 the average assigned volume is 147 vehicles less than the average ground count on the nine links which fall within that range; thus, the average assigned volume on these nine links equals 1,786-147, or 1,639.

The total value given at the bottom of this column represents the average difference between the average ground count computed for the entire network (10,516 in this example) and the average assigned volume for the entire network (which is computed in the same manner described under ⑦ describing the network average ground count). Thus, when considering the entire network (169 links), the average assigned volume equals 10,516 - 1,550, or 8,966. Again, this number is given for relative comparative purposes only.

⑨ Stan. Dev. (standard deviation)—The numbers entered in this column for each volume group represent, for all practical purposes, the standard deviation of the difference between the average ground count and the average assigned volume. Thus, for volume group 1-500, the value 189 recorded as the standard deviation means that the average difference between the average ground count and the average assigned volume falls between 718 ± 189, two-thirds of the time. A more detailed explanation of these computations is provided by Culp (3).

⑩ PC Stan. Dev. (percent standard deviation)—The numbers entered in this column are computed by dividing the standard deviation by the average ground count. Thus, for volume group 1-500, 189 ÷ 436 = 43.3 percent.

⑪ PC of Total (percent of total)—To obtain a weighted error, a computation must be made to determine what percentage the total volume within a particular volume group represents, as part of the total volume on all links used in the network. For example, in the volume group 2001-3000, there are 11 links having an average ground count of 2,512; therefore, there are 11 × 2,512 = 27,632 total vehicles counted on these 11 links. From item ② it is seen that there are 1,777,278 total vehicles counted on the entire network; therefore, the ground counts for volume group 2001-3000 represent 27,632 ÷ 1,777,278 = 1.6 percent of the total volume counted on the network.

⑫ Weighted (weighted error)—This column contains the weighted error computed for each volume group. It is computed by multiplying item ⑩ by item ⑪ . The total weighted error appearing at the bottom of the column (40.6 in this example) is the summation of the individual weighted errors computed for each volume group.

```
RESULTS OF SECOND LOADING -

TOTAL MEASURED VOLUME     1,777,278
TOTAL ASSIGNED VOLUME     1,502,184

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME -15.5  ERROR BREAKDOWN,     LATEST LOADING

     VOL GROUP  NO. SECTS    AV COUNT  AVE DIFF    STAN DEV  PC STAN DEV    PC OF TOTAL  WEIGHTED

      00-1/2          2         436    +   688        303       69.4            .0          .0
      1/2-01          3         686    +   571        631       91.9            .1          .0
       01-02          9       1,786    +   445      1,774       99.3            .9          .8
       02-03         11       2,512    -   261      2,005       79.8           1.6         1.2
       03-05         22       3,876    +   135      2,739       70.6           4.8         3.3
       05-10         56       7,318    -   824      3,459       47.2          23.1        10.9
       10-15         25      11,840    - 1,822      3,939       33.2          16.7         5.5
       15-20         24      17,501    - 3,754      6,748       38.5          23.6         9.0
       20-25          5      22,561    - 5,823      8,922       39.5           6.3         2.4
       25-30          2      27,407    - 6,649      9,494       34.6           3.1         1.0
       30-up         10      35,186    - 5,808     11,010       31.2          19.8         6.1
       TOTAL        169      10,516    - 1,628      4,601       43.7         100.0        40.2

TOTAL MEASURED VOLUME     1,777,278
AVGD ASGND VOLUME         1,508,718

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME -15.1  ERROR BREAKDOWN,      AVERAGED LOADING

     VOL GROUP  NO. SECTS    AV COUNT  AVE DIFF    STAN DEV  PC STAN DEV    PC OF TOTAL  WEIGHTED
      00-1/2          2         436    +   703        246       56.4            .0          .0
      1/2-01          3         686    +   452        465       67.7            .1          .0
       01-02          9       1,786    +   149      1,256       70.3            .9          .6
       02-03         11       2,512    -   631      1,683       66.9           1.6         1.0
       03-05         22       3,876    -   434      1,981       51.1           4.8         2.4
       05-10         56       7,318    - 1,162      3,298       45.0          23.1        10.3
       10-15         25      11,840    - 1,206      3,252       27.4          16.7         4.5
       15-20         24      17,501    - 2,422      5,064       28.9          23.6         6.8
       20-25          5      22,561    - 4,845      7,728       34.2           6.3         2.1
       25-30          2      27,407    - 9,212     13,068       47.6           3.1         1.4
       30-up         10      35,186    - 6,019     11,348       32.2          19.8         6.3
       TOTAL        169      10,516    - 1,589      4,260       40.5         100.0        35.4
```

Figure 3. Summary statistics for the second loading and for the average of two loads.

The value shown as the total weighted error is a number which has probably created more confusion than any other result obtained from the capacity restraint program. This number does not represent the true accuracy of the assignment process, or, in other words, the ability of the assignment process to adequately match the ground counts on the network. The value obtained for the total weighted error serves only as a relative index of the ability of the capacity restraint process to reduce the error in traffic assignment. To fully understand the significance of the total weighted error, it is important to continue the explanation of the output obtained from capacity restraint. A more detailed discussion of its significance is given in the next section of this paper.

Figure 3 shows the next portion of the summary statistics obtained from the capacity restraint program; the top portion of the page is labeled Results of Second Loading. The results presented here were obtained after actually applying capacity restraint to the network for the first time (second iteration), as described earlier in this paper. Thus, the speed on each link was changed according to the ratio of assigned volume to capacity, a new set of trees was built, an assignment was made using the new trees, and the summary statistics (top portion of Fig. 3) were accumulated so that an evaluation of the assignment could be made.

The bottom portion of Figure 3 contains the summary statistics for the average of the two loads obtained thus far (the results of the first free loading, and the results after applying capacity restraint for the first time). Unfortunately, the bottom portion of Figure 3 (which is a copy of the output of the capacity restraint program as it actually appears) is not labeled very clearly. The only way to distinguish between the results of the second loading and the average of all loads is by noting the label which was assigned to item ③ in the previous discussion. In Figure 2, item ③ is labeled Total Assigned Volume. In the top portion of Figure 3, item ③ is also labeled Total Assigned Volume; however, in the bottom portion of Figure 3, item ③ is labeled Average

```
RESULTS OF THIRD LOADING

TOTAL MEASURED VOLUME     1,777,278
TOTAL ASSIGNED VOLUME     1,491,940

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME  -16.1  ERROR BREAKDOWN,     LATEST LOADING

    VOL GROUP   NO. SECTS    AV COUNT   AVE DIFF    STAN DEV   PC STAN DEV     PC OF TOTAL   WEIGHTED

     00-1/2          2          436     +   632        267        61.2             .0          .0
     1/2-01          3          686     +   531        613        89.3             .1          .0
      01-02          9        1,786     +   349      1,175        65.7             .9          .5
      02-03         11        2,512     +   453      3,060       121.8            1.6         1.9
      03-05         22        3,876     - 1,093      2,416        62.3            4.8         2.9
      05-10         56        7,318     -   729      3,394        46.3           23.1        10.6
      10-15         25       11,840     -   955      3,493        29.5           16.7         4.9
      15-20         24       17,501     - 3,077      5,677        32.4           23.6         7.6
      20-25          5       22,561     - 3,100      4,985        22.0            6.3         1.3
      25-30          2       27,407     - 7,289     10,489        38.2            3.1         1.1
      30-up         10       35,186     -10,365     16,628        47.2           19.8         9.3
      TOTAL        169       10,516     - 1,688      4,883        46.4          100.0        40.1

TOTAL MEASURED VOLUME     1,777,278
AVGD ASGND VOLUME         1,503,184

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME -15.4  ERROR BREAKDOWN,      AVERAGED LOADING

    VOL GROUP   NO. SECTS    AV COUNT   AVE DIFF    STAN DEV   PC STAN DEV     PC OF TOTAL   WEIGHTED
     00-1/2          2          436     +   680        253        58.0             .0          .0
     1/2-01          3          686     +   478        492        71.7             .1          .0
      01-02          9        1,786     +   216      1,223        68.4             .9          .6
      02-03         11        2,512     -   269      1,834        73.0            1.6         1.1
      03-05         22        3,876     -   653      1,808        46.6            4.8         2.2
      05-10         56        7,318     - 1,017      2,995        40.9           23.1         9.4
      10-15         25       11,840     - 1,122      3,088        26.0           16.7         4.3
      15-20         24       17,501     - 2,639      4,872        27.8           23.6         6.5
      20-25          5       22,561     - 4,263      6,478        28.7            6.3         1.8
      25-30          2       27,407     - 8,571     12,196        44.4            3.1         1.3
      30-up         10       35,186     - 7,467     12,840        36.4           19.8         7.2
      TOTAL        169       10,516     - 1,622      4,195        39.8          100.0        34.4
```

Figure 4. Summary statistics for the third loading and the average for three loads.

```
RESULTS OF FOURTH LOADING

TOTAL MEASURED VOLUME     1,777,278
TOTAL ASSIGNED VOLUME     1,504,184

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME -15.4  ERROR BREAKDOWN,     LATEST LOADING

    VOL GROUP   NO. SECTS    AV COUNT   AVE DIFF    STAN DEV   PC STAN DEV     PC OF TOTAL   WEIGHTED

     00-1/2          2          436     +   722        345        79.1             .0          .0
     1/2-01          3          686     +   428        444        64.7             .1          .0
      01-02          9        1,786     +   648      1,974       110.5             .9          .9
      02-03         11        2,512     +   426      2,086        83.0            1.6         1.3
      03-05         22        3,876     -   821      2,337        60.2            4.8         2.8
      05-10         56        7,318     -   880      3,412        46.6           23.1        10.7
      10-15         25       11,840     - 1,983      4,325        36.5           16.7         6.0
      15-20         24       17,501     - 2,400      4,321        24.6           23.6         5.8
      20-25          5       22,561     - 3,804      5,651        25.0            6.3         1.5
      25-30          2       27,407     - 6,861     10,029        36.5            3.1         1.1
      30-up         10       35,186     - 7,910     12,632        35.9           19.8         7.1
      TOTAL        169       10,516     - 1,616      4,276        40.6          100.0        37.2

TOTAL MEASURED VOLUME     1,777,278
AVGD ASGND VOLUME         1,503,390

AVERAGE PERCENT ERROR IN ASSIGNED VOLUME - 15.4 ERROR BREAKDOWN,      AVERAGED LOADING

    VOL GROUP   NO. SECTS    AV COUNT   AVE DIFF    STAN DEV   PC STAN DEV     PC OF TOTAL   WEIGHTED
     00-1/2          2          436     +   690        276        63.3             .0          .0
     1/2-01          3          686     +   466        472        68.8             .1          .0
      01-02          9        1,786     +   324      1,382        77.3             .9          .6
      02-03         11        2,512     -    96      1,483        59.0            1.6          .9
      03-05         22        3,876     -   696      1,762        45.4            4.8         2.1
      05-10         56        7,318     -   983      2,966        40.5           23.1         9.3
      10-15         25       11,840     - 1,337      3,198        27.0           16.7         4.5
      15-20         24       17,501     - 2,580      4,652        26.5           23.6         6.2
      20-25          5       22,561     - 4,149      6,222        27.5            6.3         1.7
      25-30          2       27,407     - 8,144     11,632        42.4            3.1         1.3
      30-up         10       35,186     - 7,578     12,728        36.1           19.8         7.1
      TOTAL        169       10,516     - 1,621      4,104        39.0          100.0        33.7
```

Figure 5. Summary statistics for the fourth loading and the average for four loads.

,PR-61 PAR CARD, FIRST ITERATION FIRST ITERATION

DISTRIBUTION OF LINKS BY TRAFFIC VOLUME AND SPEED
NEW SPEED MINUS ASSIGNMENT SPEED, BY INTERVAL

	SPEED DECREASED										SPEED INCREASED						TOTALS		
VOL/CAP RATIO	10 UP	7-10	5-7	3-5	2-3	1-2	0.5-1.0	0.0-0.5	0.0-0.5	0.5-1.0	1-2	2-3	3-5	5-7	7-10	10 UP	NO SECT	AVG CAP	AVG LOAD
0.0-0.1	0	0	0	0	0	0	0	0	0	0	42	207	265	3	0	0	517	6475	224
0.1-0.2	0	0	0	0	0	0	0	0	0	0	14	97	242	6	0	0	359	5931	891
0.2-0.3	0	0	0	0	0	0	0	0	0	0	7	76	265	11	5	1	365	6844	1725
0.3-0.4	0	0	0	1	0	0	0	0	0	0	7	58	237	12	4	0	319	6626	2318
0.4-0.5	0	0	0	0	0	0	0	0	0	0	5	49	242	22	9	0	327	7361	3298
0.5-0.6	0	0	0	0	0	0	0	0	0	0	4	57	199	22	7	0	289	7722	4265
0.6-0.7	0	0	0	1	0	0	0	0	0	0	13	87	155	22	4	0	282	7727	5022
0.7-0.8	0	0	0	0	0	0	0	0	0	1	49	103	78	8	0	0	239	8246	6163
0.8-0.9	0	0	0	0	2	0	0	1	1	5	100	56	5	2	0	0	172	7878	6716
0.9-1.0	0	0	0	0	7	1	1	14	35	42	50	3	3	0	1	0	157	6756	6419
1.0-1.1	0	0	0	1	14	45	30	13	21	2	1	0	0	0	0	0	127	6705	7046
1.1-1.2	2	0	2	42	51	24	1	0	1	0	1	0	0	0	0	0	124	6768	7769
1.2-1.3	2	0	44	47	6	1	0	0	0	0	0	0	0	0	0	0	100	5450	6819
1.3-1.4	1	34	27	10	0	0	0	0	0	0	0	0	0	0	0	0	72	5512	7437
1.4-1.5	15	33	7	3	0	0	0	0	0	0	0	0	0	0	0	0	58	5461	7872
1.5-1.6	42	18	3	0	0	0	0	0	0	0	0	0	0	0	0	0	63	5716	8845
1.6-1.7	57	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	61	5252	8648
1.7-1.8	38	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	5481	9566
1.8-1.9	48	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	5738	10629
1.9-2.0	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	4448	8614
2.0-2.2	34	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	5824	12145
2.2-2.4	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	5120	11745
2.4-2.6	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	5090	12456
2.6-2.8	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	4328	11560
2.8-3.0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	4340	12622
3.0-3.5	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	4861	15781
3.5-4.0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	5711	21355
4.0 UP	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4295	20428
TOTAL	352	94	84	105	80	71	32	28	58	50	293	793	1691	108	30	1	3870	6765	4271
		↑											↑						

Figure 6. Summary of speed adjustments after first loading.

,PR-61 PAR CARD, SECOND ITERATION SECOND ITERATION

DISTRIBUTION OF LINKS BY TRAFFIC VOLUME AND SPEED
NEW SPEED MINUS ASSIGNMENT SPEED, BY INTERVAL

	SPEED DECREASED										SPEED INCREASED						TOTALS		
VOL/CAP RATIO	10 UP	7-10	5-7	3-5	2-3	1-2	0.5-1.0	0.0-0.5	0.0-0.5	0.5-1.0	1-2	2-3	3-5	5-7	7-10	10 UP	NO SECT	AVG CAP	AVG LOAD
0.0-0.1	0	0	0	0	0	0	0	0	1	0	105	226	64	3	3	27	429	6675	244
0.1-0.2	0	0	0	0	0	0	0	0	1	0	35	187	63	10	0	12	308	6301	941
0.2-0.3	0	0	0	0	0	0	0	0	0	0	37	187	115	12	2	16	369	6755	1715
0.3-0.4	0	0	0	0	0	0	0	0	0	0	19	146	135	15	3	4	322	6543	2305
0.4-0.5	0	0	0	0	0	0	0	0	0	0	18	141	119	22	7	6	313	7618	3422
0.5-0.6	0	0	0	0	0	0	0	0	1	0	23	130	132	20	7	7	320	7876	4338
0.6-0.7	0	0	0	0	0	0	0	0	0	6	35	120	97	16	0	8	282	7461	4822
0.7-0.8	0	0	0	0	0	0	0	0	6	5	108	91	44	9	3	3	269	7844	5901
0.8-0.9	0	0	0	0	0	0	0	0	21	27	104	40	22	11	9	5	239	7173	6112
0.9-1.0	0	0	0	0	0	4	4	7	67	24	33	8	5	5	1	4	162	7169	6801
1.0-1.1	0	0	0	0	11	43	36	7	32	6	11	3	1	1	7	2	160	6572	6875
1.1-1.2	0	0	1	26	47	34	8	0	6	4	4	3	3	5	1	2	144	5809	6677
1.2-1.3	0	0	15	48	12	6	6	1	4	3	3	4	2	1	0	0	105	5319	6648
1.3-1.4	0	12	34	16	5	0	0	1	2	2	2	1	0	0	0	0	75	5906	7972
1.4-1.5	7	18	21	6	0	2	3	0	1	0	0	0	0	1	0	0	59	5017	7287
1.5-1.6	30	19	15	7	1	3	0	0	0	0	0	0	1	1	0	0	77	5039	7828
1.6-1.7	17	9	8	1	2	1	3	0	0	0	0	0	0	0	0	0	41	4524	7487
1.7-1.8	17	13	3	3	1	1	0	0	0	0	0	0	0	0	0	0	38	6157	10739
1.8-1.9	22	5	2	4	0	0	0	0	0	0	0	0	0	0	0	0	33	5435	10050
1.9-2.0	16	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	19	5660	11052
2.0-2.2	22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	5224	11062
2.2-2.4	25	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	28	5362	12304
2.4-2.6	9	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	12	5218	13089
2.6-2.8	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	4981	13360
2.8-3.0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4901	14072
3.0-3.5	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	3245	10929
3.5-4.0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	3303	12691
4.0 UP	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	10	3701	15971
TOTAL	207	82	100	111	82	94	60	17	142	77	537	1287	803	132	43	96	3870	6765	4321

Figure 7. Summary of speed adjustments after second loading.

```
,PR-61 PAR CARD, THIRD ITERATION                          THIRD ITERATION
                       DISTRIBUTION OF LINKS BY TRAFFIC VOLUME AND SPEED
                         NEW SPEED MINUS ASSIGNMENT SPEED, BY INTERVAL

                      SPEED DECREASED                                 SPEED INCREASED                      TOTALS

VOL/CAP      10  7-10   5-7   3-5   2-3   1-2  0.5-  0.0-  0.0-  0.5-   1-2   2-3   3-5   5-7  7-10    10    NO   AVG   AVG
RATIO        UP                                 1.0   0.5   0.5   1.0                                   UP  SECT   CAP  LOAD

0.0-0.1       0     0     0     0     0     0     0     0     1     0   181   139    37     5     8    62   433  6623   232
0.1-0.2       0     0     0     0     0     0     0     0     0     0    82   160    39     7     6    18   312  6772  1020
0.2-0.3       0     0     0     0     0     0     0     0     1     0    70   185    61    16     3     9   345  7051  1728
0.3-0.4       0     0     0     0     0     0     0     0     0     0    66   141    72    21     4    10   314  6493  2284
0.4-0.5       0     0     0     0     0     0     0     0     1     0    72   133    87    17     6    10   326  7865  3550
0.5-0.6       0     0     0     0     0     0     0     0     1     1    76   110    81    11     8     5   293  7614  4180
0.6-0.7       0     0     0     0     0     0     0     0     3    11    93    90    64    13     8    10   292  7540  4927
0.7-0.8       0     0     0     0     0     0     0     0     4    18   118    48    26     7     3    16   240  7631  5720
0.8-0.9       0     0     0     0     0     0     1     3    36    46    82    23    15    10     4    10   230  6452  5486
0.9-1.0       0     0     0     0     0     6    13    18    69    29    33    14    19     5     2     5   213  7024  6648
1.0-1.1       0     0     0     1    14    38    19     7    36    13    16     2     7     2     4     7   166  6395  6663
1.1-1.2       0     0     2    17    33    34    12     3    11     1     8     3     6     2     3     3   138  5377  6159
1.2-1.3       0     3    11    32    20    11     8     3    12     3     6     1     2     2     3     4   121  5603  7018
1.3-1.4       0    19    29    23     6     4     4     1     2     1     1     1     2     0     0     0    93  6493  8774
1.4-1.5       4    18    20    19    10     3     1     2     1     0     0     0     2     2     1     0    83  5575  8083
1.5-1.6      10    18     9     7     4     2     0     1     0     1     0     1     3     0     0     0    56  4854  7499
1.6-1.7      18    11    11     4     4     2     1     0     2     2     0     0     0     0     0     0    55  5298  8760
1.7-1.8       9     6     7     5     0     2     1     0     0     0     0     0     0     0     0     0    36  4942  8640
1.8-1.9      14     0     3     5     3     0     0     0     0     0     0     0     0     0     0     0    25  5041  9295
1.9-2.0      14     5     4     3     1     0     0     0     1     0     0     0     0     0     0     0    28  5269 10279
2.0-2.2      20     3     2     2     1     0     0     0     0     0     0     0     0     0     0     0    28  5243 10979
2.2-2.4      13     1     2     1     0     1     0     0     1     0     0     0     0     0     0     0    19  4993 11483
2.4-2.6      11     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0    11  5599 14144
2.6-2.8       6     1     0     0     1     0     0     0     0     0     0     0     0     0     0     0     8  5128 13696
2.8-3.0       4     0     0     1     0     0     0     0     0     0     0     0     0     0     0     0     5  3714 10528
3.0-3.5       1     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     1  6161 20580
3.5-4.0       1     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     1  3352 13360
4.0 UP        3     0     0     0     0     1     0     0     0     0     0     0     0     0     0     0     4  3671 16550

TOTAL       128    85   100   120    97   104    60    38   182   126   904  1051   523   120    63   169  3870  6765  4269
```

Figure 8. Summary of speed adjustments after third loading.

```
,PR-61 PAR CARD, FOURTH ITERATION                         FOURTH ITERATION
                       DISTRIBUTION OF LINKS BY TRAFFIC VOLUME AND SPEED
                         NEW SPEED MINUS ASSIGNMENT SPEED, BY INTERVAL

                      SPEED DECREASED                                 SPEED INCREASED                      TOTALS

VOL/CAP      10  7-10   5-7   3-5   2-3   1-2  0.5-  0.0-  0.0-  0.5-   1-2   2-3   3-5   5-7  7-10    10    NO   AVG   AVG
RATIO        UP                                 1.0   0.5   0.5   1.0                                   UP  SECT   CAP  LOAD

0.0-0.1       0     0     0     0     0     0     0     0     1    20   245    62    25     4     7    62   426  6213   226
0.1-0.2       0     0     0     0     0     0     0     0     0    11   176    47    32     8     7     8   289  7310  1101
0.2-0.3       0     0     0     0     0     0     0     0     1     6   167    82    47     8     4     9   324  6622  1668
0.3-0.4       0     0     0     0     0     0     0     0     0     2   129    84    47    20     8     8   298  6942  2458
0.4-0.5       0     0     0     0     0     0     0     0     0     3   151    87    61    21     9    13   345  7599  3439
0.5-0.6       0     0     0     0     0     0     0     0     2     4   119    96    67     7     8    19   322  7431  4098
0.6-0.7       0     0     0     0     0     0     0     0     4    26   106    68    51    20     6     6   287  7428  4861
0.7-0.8       0     0     0     0     0     0     0     0    17    39    97    38    40    11     4     8   254  8018  6046
0.8-0.9       0     0     0     0     0     1     7     7    61    43    53    22    30     5     9    14   252  6659  5644
0.9-1.0       0     0     0     0     0    13    17    13    63    27    35    25    17    13     3    10   236  7398  7018
1.0-1.1       0     0     0     0    12    34    12     9    36    10    14    15    12     4     1     9   168  6738  7046
1.1-1.2       0     0     1    19    18    33    16     9    11     8     4     5     5     6     5     1   141  6303  7236
1.2-1.3       0     0     9    19    17    27     4     3     7     2     5     3     4     7     2     4   113  5694  7126
1.3-1.4       0     5    17    12    12    10     3     1     5     1     6     0     9     2     3     1    87  4717  6343
1.4-1.5       0     8     7    18     8     7     1     0     3     2     1     1     6     0     1     0    63  5019  7241
1.5-1.6       5    12     8    17     3     5     1     2     2     2     0     2     1     0     0     1    61  5251  8083
1.6-1.7       9    13     4     5     4     2     2     1     1     0     0     0     0     0     0     0    41  5018  8288
1.7-1.8      10     2     5     7     3     2     1     1     1     1     1     0     0     0     1     0    35  4229  7397
1.8-1.9       8     7     2     5     1     1     0     0     0     0     0     0     0     0     0     0    24  4217  7770
1.9-2.0      11     1     0     3     3     2     0     0     0     0     0     0     0     0     0     0    20  4978  9658
2.0-2.2      13     4     1     2     1     1     0     0     0     1     0     0     1     0     0     0    24  4585  9689
2.2-2.4      14     2     2     2     0     1     0     0     0     0     0     0     0     0     0     0    21  4124  9411
2.4-2.6       9     2     2     0     0     0     1     0     0     0     0     0     0     0     0     0    14  4259 10634
2.6-2.8       7     2     0     0     0     1     0     0     0     0     0     0     0     0     0     0    10  3071  8269
2.8-3.0       2     0     0     1     0     0     0     0     0     0     0     0     0     0     0     0     3  3452  9955
3.0-3.5       8     1     0     0     0     0     0     0     0     0     0     0     0     0     0     0     9  2961  9459
3.5-4.0       2     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     2  2866 10612
4.0 UP        0     0     0     0     0     0     1     0     0     0     0     0     0     0     0     0     1  4176 17800

TOTAL        98    59    58   110    82   140    66    46   215   208  1309   637   455   136    78   173  3870  6765  4265
```

Figure 9. Summary of speed adjustments after fourth loading.

TABLE 1

CITIES STUDIES

City	Population[a]	Accuracy of Screenline Check (%)[b]	Adjustment Made to O-D Data Based on Screenline Results
Salem, Oregon[c]	49, 142	91	No
Sioux Falls	66, 582	80	Yes (96%)
Green Bay[c]	97, 162	81	Yes (92%)
Madison[c]	157, 814	78	Yes (92%)
Tucson	227, 533	77	No
Salt Lake City	348, 661	90	No
Honolulu	351, 336	90	No
Portland, Oregon[c]	651, 685	84 and 79	Yes (95 and 86%)
Atlanta	768, 125	85 and 86	No
Denver	803, 624	84. 5 (control point)	Yes (93%) (control point)

[a] Population as reported in the 1960 census.
[b] Screenline data obtained from transportation study reports.
[c] Used volume restraint.

Assigned Volume. There should be little difficulty in distinguishing between the results of an iteration and the average loading, however, because the results of the average loading always appear at the bottom of the page of the computer output.

The top portion of Figure 4 shows the summary statistics for the next iteration (the second application of capacity restraint) and the bottom portion shows the results of averaging the three assignments. Similarly, the top portion of Figure 5 shows the summary statistics for the third application of capacity restraint, and the bottom portion shows the results of averaging the four assignments that have been made.

Another output from the capacity restraint program should be explained at this point. Figures 6 through 9 are titled Distribution of Links by Traffic Volume and Speed: New Speed Minus Assignment Speed, by Interval. These tables are produced after each iteration of capacity restraint has been applied to the network. They summarize the speed adjustments by showing the number of links, by assigned volume-to-capacity ratio, that had positive or negative speed changes of a specified amount. For example, Figure 6 shows that a total of 94 links had speeds decreased by 7 to 10 mph as a result of applying capacity restraint; a total of 1, 691 links had speeds increased by 3 to 5 mph as a result of applying capacity restraint, etc. (see arrows). These tables have been included as an output of the capacity restraint program to assist in the evaluation of the ability of the program to "balance" the speeds on the analysis network; that is, the amount of speed adjustment should become less and less after each iteration has been applied, so that the final speed adjustment table should have most of the links clustered near the center of the table (in the columns for "speed decreased 0. 0 to 0. 5" and "speed increased 0. 0 to 0. 5").

Figure 6 shows the degree of speed changes made after the first application of capacity restraint. An examination of the data recorded in Figure 6 shows that many rather

TABLE 2

COMPARISON OF VMT FOR ATLANTA

Facility Type	Vehicle-Miles of Travel		Ratio of Traffic Assignment to Ground Counts
	From Ground Counts	From Traffic Assignment	
Freeway and ramp	914, 900	853; 632	93%
Arterial	3, 505, 904	2, 893, 853	83%
Secondary	1, 065, 653	986, 122	93%
Total	5, 486, 457	4, 733, 607	86%

From "Atlanta Area Transportation Study Base Year Report."

large speed adjustments were made for this first trial. This was probably caused by large imbalances of assigned volumes. Figures 7, 8, and 9 show that the speed changes become smaller and smaller after each successive adjustment; this occurs because the network speeds and assignments become better balanced as the process takes place.

ACCURACY OF TRAFFIC ASSIGNMENT WHEN USING CAPACITY RESTRAINT

There are five ways to evaluate the accuracy of the traffic assignment process when using capacity restraint. Each will be discussed so that some conclusions can be made concerning the ability of a mechanical traffic assignment to match the actual travel taking place in an urban area. The five measures of accuracy are:

1. Total counted volume compared to total assigned volume.
2. Total vehicle-miles of travel (VMT) computed from ground counts compared to total vehicle-miles of travel assigned.
3. The "total weighted error" computed by the capacity restraint program (item ⑫ previously explained).
4. The root-mean-square (RMS) error computed by comparing ground counts and assigned volumes on each link in the analysis network.
5. A graphic comparison of ground counts and assigned volumes when plotted on a map.

The results of traffic assignments using capacity restraint were obtained from the 10 urban transportation studies listed in Table 1.

Accuracy Tests for Atlanta

The results obtained after applying capacity restraint to the network coded for Atlanta will be used first to illustrate the accuracy of the assignment process. The summary statistics for this city were shown earlier in Figures 2 through 9. It is important to note that the screenline analysis made for Atlanta revealed that the data collected in the O-D surveys accounted for about 85 or 86 percent of the ground counts made on the screenline (4). Although the results of the screenline analysis may not always be a true indication of the completeness and accuracy of the travel data collected, it has been assumed that the travel data in this situation are approximately 14 or 15 percent low. This assumption is based on the data given (4) and the analysis of the traffic assignment results obtained.

It is also important to realize that each of the five measures of accuracy discussed are based on the assumption that the ground counts are reasonably accurate. Thus, the importance of having accurate ground count data available cannot be overemphasized. The assumption will be made in the subsequent discussion that the ground counts reported for the highway facilities are reasonably accurate.

Test No. 1: Total Ground Counts vs Total Assigned Vehicles—An examination of item ④ in Figures 2 through 5 shows that the average percent error in assigned volumes is -15. 4 percent Although this is only a gross measure of accuracy, it does tend to support the assumption that the trip data in this example are about 14 or 15 percent underreported. This test also indicates that the assignment can match ground counts only as well as the O-D survey data represent the actual travel taking place within the urban area.

Test No. 2: VMT Counted vs VMT Assigned—A comparison between actual and assigned VMT is an excellent method that can be used to evaluate the accuracy of traffic assignment on an area-wide basis. Several urban transportation studies have used this method successfully. Table 2 contains a summary of the VMT check made for Atlanta.

Table 2 shows that traffic assignment has accounted for about 86 percent of the total VMT actually occurring on the highway network. This again supports the assumption that the trip data are approximately 14 or 15 percent underreported, assuming that the ground counts are relatively accurate. Note that the VMT assigned to the freeways and ramps account for 93 percent of the actual VMT, however. This indicates that the

TABLE 3

SUMMARY OF TOTAL WEIGHTED ERRORS FOR 10 CITIES[a]

City	ADT Volumes—Total Weighted Error After:									Avg. Percent Error in Assigned Volumes	Final Screenline Check (%)
	First Load[b]	Second Load	Avg. of 2 Loads	Third Load	Avg. of 3 Loads	Fourth Load	Avg. of 4 Loads	Fifth Load	Avg. of 5 Loads		
Salem	37. 1	40. 6	35. 7	38. 4	34. 9	36. 3	33. 9	—	—	-17. 1	91
Sioux Falls	47. 6	48. 4	46. 5	46. 2	46. 2	—	—	—	—	-12. 5	96
Green Bay	42. 9	47. 2	43. 7	45. 2	43. 9	44. 3	43. 8	—	—	-13. 5	92
Madison	33. 5	29. 9	29. 6	27. 3	27. 8	27. 9	26. 9	—	—	- 4. 3	92
Tucson	43. 6	43. 7	40. 2	40. 9	39. 4	—	—	—	—	-20. 1	77
Salt Lake City	NA	56. 1	40. 2	59. 1	36. 5	50. 8	33. 9	51. 0	32. 7	- 6. 5	90
Honolulu[c]	61. 1	68. 6	51. 1	66. 9	50. 4	58. 0	48. 1	—	—	-22. 6	90
Portland	56. 0	45. 0	47. 3	46. 5	47. 3	—	—	—	—	- 9. 5	95/86
Atlanta	40. 6	40. 2	35. 4	40. 1	34. 4	37. 2	33. 7	—	—	-15. 4	85/86
Denver	NA	NA	NA	83. 9	42. 1	—	—	—	—	-10. 5	93

[a] Data summarized from computer outputs obtained from each study.

[b] First free assignment.

[c] Trips loaded on network did not include truck trips; assignments are approximately 10 to 12 percent low.

NA = Not Available

longer trips, which represent the predominant type of trip using freeway facilities, are probably more fully reported than the shorter trips. Thus this test also indicates that the traffic assignment process has provided a reasonable match with the actual travel taking place on the transportation network, but only within the range of accuracy provided by the O-D surveys.

Test No. 3: Total Weighted Error—An examination of the total weighted error for Atlanta (item ⑫ in Figs. 2, 3, 4, and 5) shows the following results:

Loading	Total Weighted Error
First	40. 6
Second	40. 2
Average of two	35. 4
Third	40. 1
Average of three	34. 4
Fourth	37. 2
Average of four	33. 7

To make the significance of these results more meaningful, Table 3 was prepared to show similar results for 10 cities. Table 3 also shows the average ratio of total assigned to total counted vehicles (average percent error in assigned volumes) and the screenline check obtained after adjustment factors were applied for each of these cities.

As stated earlier, the total weighted error obtained after several iterations of capacity restraint does not represent the true accuracy of the assignment process (or, the ability of the assignment process to match the ground counts on the network). This number serves only as a relative index of the ability of the capacity restraint process to reduce the error in traffic assignment. There are probably several ways to explain the meaning of this relative index, but one interpretation can be expressed as follows.

Since the assignment process is based on the selection of one route between zones, the first time a route is selected (the first free loading) there may be a considerable overload or underload on that route. This is due, in part, to the difficulty in estimating an average operating speed for every facility on the highway network. Consequently, a poor match between ground counts and assigned volumes may be obtained. As each iteration of capacity restraint is applied, new sets of routes are

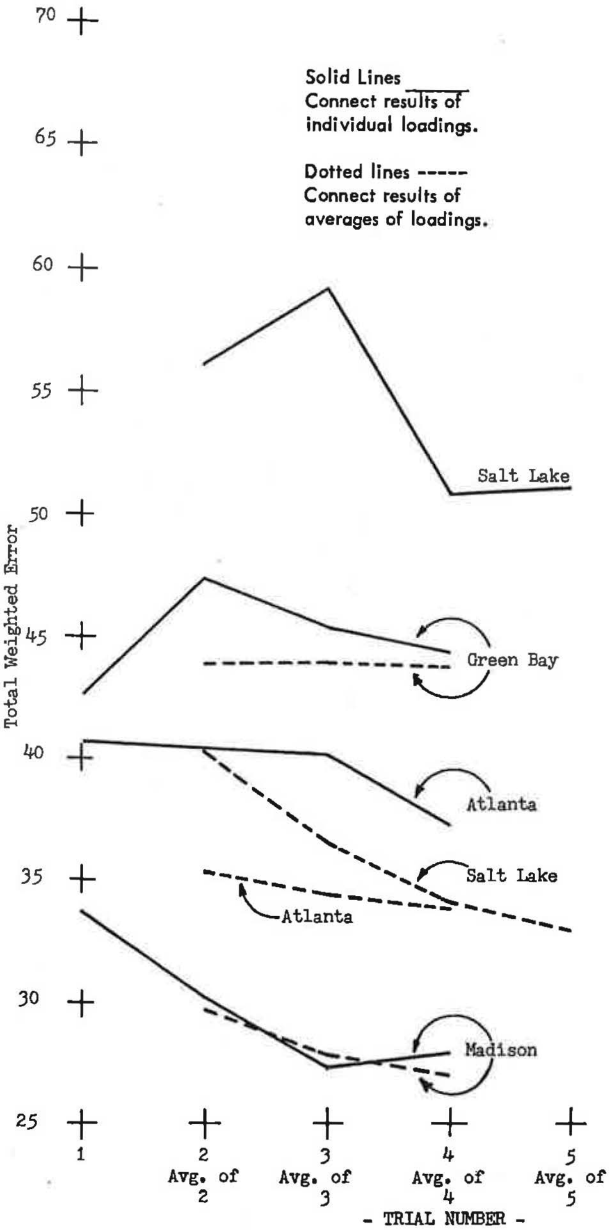


Figure 10. Plot of total weighted errors.

computed between zones. Again, each individual route may not be the best in terms of satisfying the goal of matching ground counts. However, as each assignment is averaged (Fig. 1) the assigned volumes tend to be distributed over an increasingly larger portion of the network. This tends to reduce the error on individual links as the averaging technique is applied after each iteration.

The weighted errors obtained for Atlanta after the first, second, third, and fourth loads are 40. 6, 40. 2, 40. 1, and 37. 2, respectively. However, the weighted errors computed after averaging the first two loads, the first three loads, and the first four loads are 35. 4, 34, 4, and 33.7, respectively.

A more dramatic example of the reduction in the total weighted error that is obtained by averaging assignments is shown by the data for Salt Lake City. The total weighted errors computed after the second (the results of the first load were not available), third, fourth, and fifth loads are 56. 1, 59. 1, 50. 8, and 51. 0, respectively. The total weighted errors computed after averaging the first two loads, the first three loads, the first four loads, and the first five loads are 40. 2, 36. 5, 33. 9 and 32. 7, respectively.

To summarize the results for these two cities, the total weighted error for Atlanta was reduced from 40. 6 after the first free load to 33.7 after averaging four loads; the total weighted error for Salt Lake City was reduced from 56. 1 after the second load to 32. 7 after averaging five loads. Since the weighted error is computed by comparing the differences between the ground counts and assigned volumes on a link-by-link basis, it is obvious that the capacity restraint process can reduce the overall error in assignments by a considerable amount, especially if the average loads are used.

Examination of Table 3 also shows that capacity restraint does not always reduce the total weighted error. For example, the data obtained for Green Bay show that the weighted error was 42. 9 after the first free load and 43. 8 after averaging four loads. The weighted error did not change significantly after any of the individual assignments, nor did the averaging of the assignments have much significance. This may be caused by some peculiarity in the network, or by some other unknown characteristic. Peak-hour assignments might give better results. Unfortunately, there are not enough data available at the present time to evaluate this condition properly. Needless to say, more research is needed in order to understand it fully.

Some general observations concerning the data shown in Table 3 are worth mentioning at this point. Figure 10 has been plotted from the data to show how the "total weighted error" varies after each iteration has been applied, and how the "total weighted error" varies after averaging the results of several iterations. For the sake of clarity, only four of the ten sets of data shown in Table 3 were plotted.

Note that in every case the results of individual trials always produce an error greater than that obtained by averaging several trials. (There is one exception—where the third load and the average of three loads are about equal in the Madison data.) In some cases, Salt Lake City, for example, there is a large decrease in error resulting from the averaging technique. For other cities the error is reduced, but not so dramatically.

Notice also that in almost every case, the second loading for individual trials increases the total weighted error, but that each succeeding trial shows a reduction in the total weighted error. This probably occurs because the speeds on links are changed too drastically for the first application of capacity restraint. However, as the speed changes become less and less (Figs. 6 through 9) the errors tend to stabilize. Madison is a notable exception to this trend, probably because volume restraint rather than capacity restraint was used in this case.

The results of averaging the assignments show a continuous decrease in the "total weighted error," but not at a significantly decreasing rate. Thus, since the "total weighted error" after averaging two assignments is not significantly different from that computed after averaging three or four assignments, the logical conclusion might be that the use of the average of two loads is adequate for obtaining a desirable level of accuracy in the assignment process. However, because of the "diversion" effect of the capacity restraint process (Fig. 1), and because each successive iteration and averaging spreads the traffic over an increasingly larger portion of the network, it is desirable to use the average of three or four iterations to obtain the best assignment results. This is further verified by the values obtained for the total weighted error shown in Table 2.

Test No. 4: The Root-Mean-Square (RMS) Error When Comparing Ground Counts to Assigned Volumes—For the purposes of this evaluation, a computer program was written to compute the actual RMS error for each of the volume groups shown earlier (Figs. 2 through 5). The RMS error for each volume group was computed using the equation

$$\text{RMS} = \sqrt{\frac{\Sigma (X_{GC} - X_{TA})^2}{N - 1}}$$

TABLE 4

COMPARISON OF PERCENT STANDARD DEVIATION AND
RMS ERROR BY VOLUME GROUPS FOR ATLANTA

Volume Group	No. of Links	Percent Standard Deviation	Percent RMS Error
00-½	2	68.3	Not computed
½-01	3	68.8	Not computed
01-02	9	77.3	54.4
02-03	11	59.0	58.9
03-05	22	45.4	41.8
05-10	56	40.5	38.2
10-15	25	27.0	24.5
15-20	24	26.5	22.1
20-25	5	27.5	20.5
25-30	2	42.4	30.3
30-up	10	36.1	29.1
Total	169	39.0	35.7

in which:

X_{GC} = ground count on link L_i
X_{TA} = volume assigned to link L_i
N = total number of links in a particular volume group
i = 1 through N

More details concerning the use of this program are found elsewhere (3).

The percent RMS error computed for each volume group was compared to the value computed for the percent standard deviation obtained from the capacity restraint statistics (item ⑩ in Fig. 5). The results obtained for Atlanta, using the average of four trials, are shown in Table 4. The same statistics were summarized from the computer output obtained from Madison, and the results obtained for the average of four trials are shown in Table 5.

The reason for computing the RMS error by volume group was to determine which of the statistics obtained from the capacity restraint program (Figs. 2 through 5) provided a more reasonable indication of the error in assignment than that provided by the "total weighted error." As seen in Table 3, the total weighted error after averaging four trials for Atlanta was 33.7; the total weighted error after averaging four trials for

TABLE 5

COMPARISON OF PERCENT STANDARD DEVIATION AND
RMS ERROR BY VOLUME GROUPS FOR MADISON[a]

Volume Group	No. of Links	Percent Standard Deviation	Percent RMS Error
00-½	51	199.6	Not computed
½-01	72	99.4	Not computed
01-02	132	61.1	49.8
02-03	134	42.7	46.4
03-05	168	36.4	36.0
05-10	224	30.9	30.5
10-15	143	25.6	27.8
15-20	59	18.1	18.1
20-25	22	12.7	13.0
25-30	11	7.0	7.1
30-up	4	14.8	15.1
Total	1,020	30.9	32.4

[a] Volume restraint, rather than capacity restraint, was used in this study; however, this is not relevant to the discussion of RMS errors.

TABLE 6

PERCENT STANDARD DEVIATION BASED ON THE RESULTS OF AVERAGING FOUR LOADINGS

City	Percent Standard Deviation by Volume Group										
	00-½	½-01	01-02	02-03	03-05	05-10	10-15	15-20	20-25	25-30	30-up	Total
Salem	76.2	49.5	41.9	31.3	37.6	33.8	32.8	21.7	59.7	—	—	41.8
Sioux Falls[a]	132.8	156.8	107.5	56.4	54.6	38.7	36.6	—	—	—	—	49.1
Green Bay	62.9	55.4	62.2	71.6	58.8	46.7	38.9	29.3	15.2	—	—	49.4
Madison	199.6	99.4	61.1	42.7	36.4	30.9	25.6	18.1	12.7	7.0	14.8	30.9
Tucson[a]	138.0	91.6	124.3	79.1	44.3	36.6	34.2	39.6	21.3	21.7	—	47.7
Salt Lake[b]	485.4	212.9	88.0	50.7	43.6	34.0	20.8	32.0	30.3	15.9	12.2	38.0
Honolulu	141.4	120.0	98.6	51.1	63.1	54.4	52.1	25.0	35.1	34.0	36.7	53.5
Portland[a]	256.4	152.7	126.8	68.6	61.5	48.8	39.7	32.5	31.3	29.1	53.1	55.3
Atlanta	63.3	68.8	77.3	59.0	45.4	40.5	27.0	26.5	27.5	42.4	36.1	39.0
Denver[a]	817.4	315.4	174.9	97.5	75.6	53.1	31.8	23.3	26.3	25.5	16.0	44.4

[a] Average of 3 loadings.
[b] Average of 5 loadings.

Madison was 26.9. As explained earlier, these numbers serve only as a relative index of the ability of the capacity restraint process to reduce the error in traffic assignment. They are not a true measure of the accuracy of the assignments.

A comparison of the percent standard deviation and the percent RMS error for each volume group listed in Tables 4 and 5 for Atlanta and Madison show that the two numbers are in reasonable agreement. The conclusion drawn from this observation is that the percent standard deviation as computed in the capacity restraint program for each volume group may be used to estimate the average error resulting in the assignment process by volume group. However, it is further concluded that these values should not be used alone to determine the accuracy of the assignment process, because they are based on averages. The final test of the assignment process, a graphic presentation of the results on a map, must also be made to determine the adequacy of the process.

Table 6 shows the percent standard deviation computed by volume group from the capacity restraint program for the 10 cities. These values are based on the average of four loadings, except as noted. The standard deviations shown for volume groups up to about 3,000 to 5,000 vehicles are extremely high. However, these values start to decrease from that point on so that the results obtained are more reasonable as the volume increases. The errors indicated by percent standard deviation values are believed to be higher than the actual error in assignment. This will be expanded upon later in this paper.

Test No. 5: Graphic Comparison of Ground Counts and Assigned Volumes When Plotted on a Map—Figure 11 shows a comparison of assigned volume to ground counts on a portion of the existing freeway system in Atlanta; 19 links are included in this presentation. In this graphic display of the results, approximately two-thirds of the assigned volumes are within ±14.3 percent of the ground counts. The two-way ground counts on the links shown range from 14,668 to 81,146 ADT.

The assignments to these high-volume links are reasonably accurate; however, the assigned volumes are about 14 percent low on the average. This again seems to verify the assumption made earlier that the O-D data are about 14 percent low.

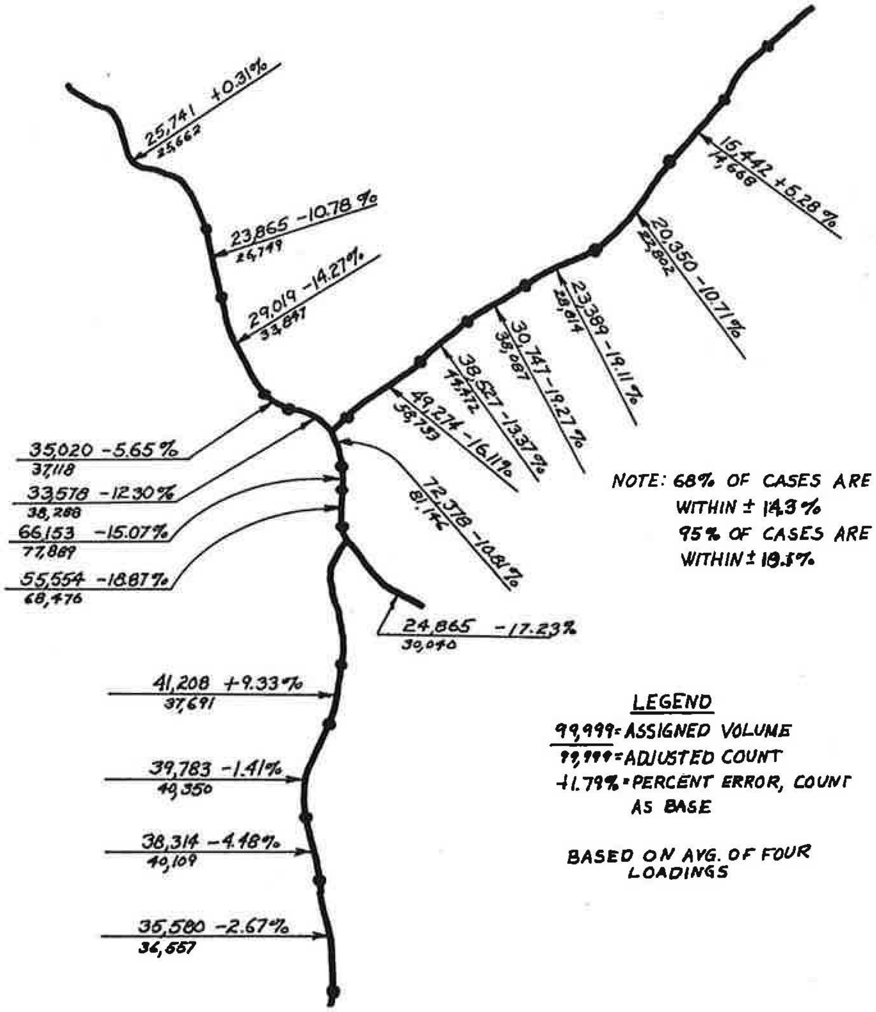


Figure 11. Comparison of assigned volume to ground counts on existing freeway system in Atlanta.

A graphic comparison of this type can be made quickly and efficiently at a reasonable cost by making use of automatic data plotters (1). The results obtained from a graphic presentation of this type are extremely valuable in the evaluation of the traffic assignment model.

Accuracy Tests for Madison

A limited amount of data was also available from the Madison study for the purposes of this report; some of the results obtained for Madison were mentioned earlier in connection with the evaluation of the Atlanta data.

Madison reported an initial screenline check of 78 percent. The O-D data were then adjusted by applying factors to the data, so that eventually a 92 percent screenline check was obtained (5).

The following tests were made using the Madison data to aid in the evaluation of the traffic assignment process. Wherever possible, the same tests were made as described earlier for Atlanta.

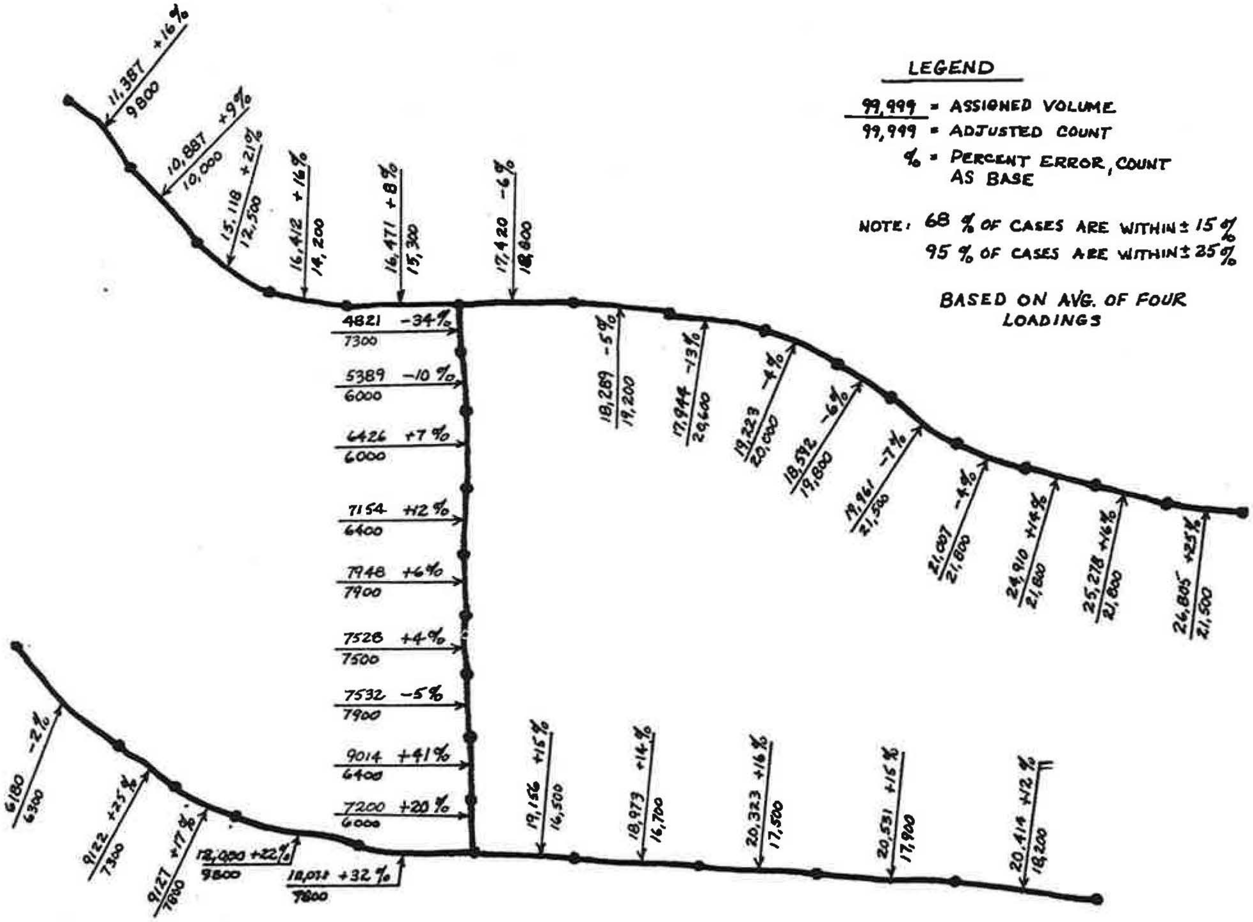


Figure 12. Comparison of assigned volume to ground counts on a portion of the Madison network.

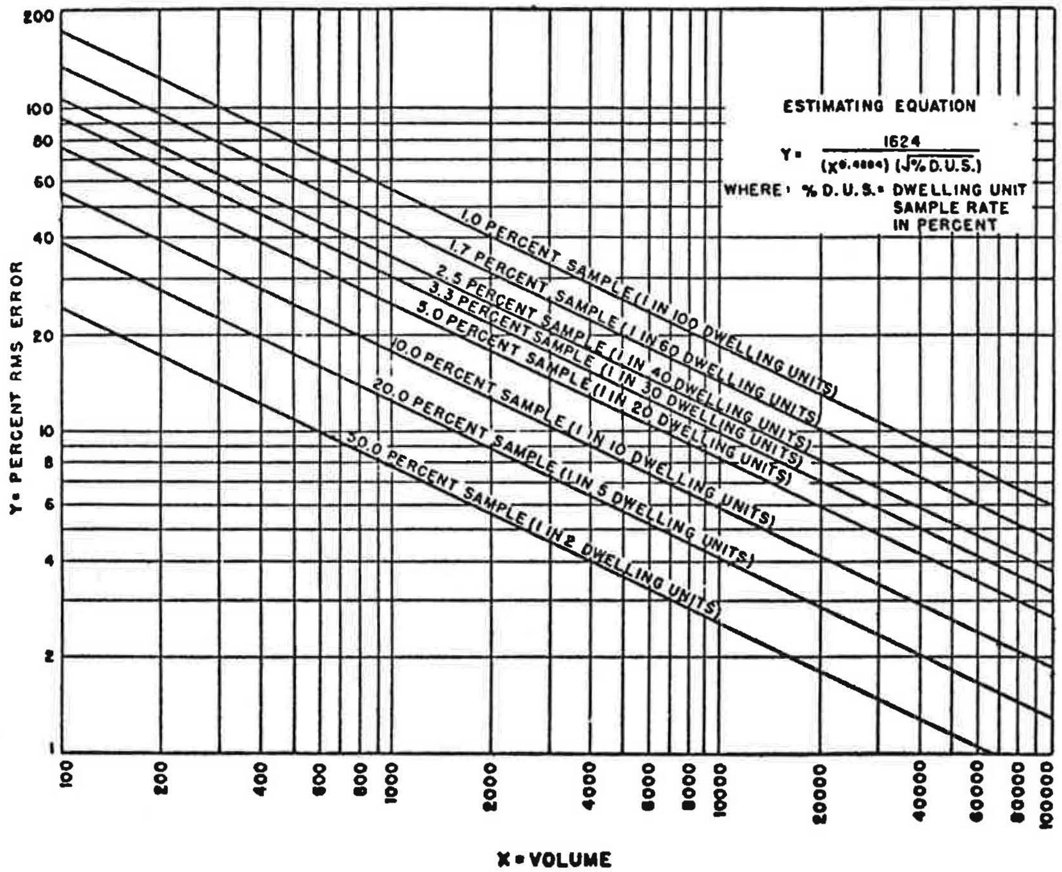


Figure 13. Relation of percent root-mean-square error and volume for various dwelling unit sample rates (6).

Test No. 1: Total Ground Counts vs Total Assigned Vehicles—An examination of Table 3 shows that the average percent error in assigned volumes is -4. 3 percent for Madison. This indicates that the assigned O-D data represent the total travel taking place in a reasonable manner. It also tends to verify the assumption that factoring the trip data on the basis of the screenline analysis provides a better traffic assignment, and more accurate overall results. Earlier it was shown that the O-D data accounted for an 85 percent screenline check in Atlanta. Since the data were not factored, the assignments were also about 15 percent low.

Test No. 2: VMT Counted vs VMT Assigned—Data were not available for making this comparison.

Test No. 3: Total Weighted Error—The total weighted errors computed for Madison are shown in Table 3 and Figure 10. These values are the lowest obtained from the group of 10 cities; this might have occurred because a volume restraint was used rather than a capacity restraint. The same comments apply to this test as stated earlier for Atlanta.

Test No. 4: The RMS Error—The RMS errors computed for Madison were shown earlier in Table 5. These values indicate that the traffic assignment process resulted in an excellent match with ground count data, especially for volumes exceeding 10,000 ADT.

Test No. 5: A Graphic Comparison—A small portion of the traffic assignment network used in the Madison study is shown in Figure 12. Values of assigned volumes, ground counts, and the percent difference between ground counts and assignments are recorded for 34 links. (More than 1, 000 links are contained in the complete network.) The data show that approximately two-thirds of the assigned volumes are within ±15 percent of the ground counts. The two-way ground counts on the links shown range from 6, 000 to 21, 800 ADT.

The section of the Madison network shown in Figure 12 was selected randomly, without knowing any details concerning the characteristics of the city or the network. It is interesting to note that the assignment error found in this portion of the network (two-thirds of the assigned volumes are within ±15 percent of the ground counts) is almost exactly the same as the error shown for the Atlanta data in Figure 11 (two-thirds of the assigned volumes are within ±14. 3 percent of the ground counts). It is difficult to explain why this similarity in the results occurred, since the RMS errors for the Madison data were less than those computed for the Atlanta data. However, both the Atlanta data and the Madison data indicate that the assignment process is reasonably accurate, since it matches ground counts within ±15 percent, two-thirds of the time.

ERRORS IN TRAFFIC ASSIGNMENT

The traffic assignment process is based on necessary decisions, which result in several inherent errors being "built into" the process. These decisions are as follows:

1. A route between zones is selected on the basis of one parameter only, either time, distance, or cost.
2. When speed is used as the route selection parameter, only one set of speeds (ADT or peak hour) is used to select the best route between zones.
3. The all-or-nothing concept is used to assign trips.
4. Trips start and end at one point in a zone.
5. Intrazonal trips are not loaded.
6. A limited street and highway system must be used, thereby eliminating some from consideration.

These errors are mentioned because they must be considered in evaluating the traffic assignment process. Capacity restraint tends to minimize the effects of some of the inherent errors in the process, but no technique available at the present time can eliminate all of them.

In addition to errors that are peculiar to the assignment process there are two other sources of errors which must be considered:

1. Sampling error in the collection of O-D data in the home interview surveys, and
2. Errors in the ground counts used to evaluate the assignment model.

Figure 13 shows the relation of percent RMS error and volume for various dwelling unit sample rates (6). The set of curves shown in this figure may be used to estimate the RMS error expected to occur from the sampling procedure alone on links having various volumes, for several different sampling rates.

It is possible to obtain area-wide traffic counts (coverage counts) which provide data having a ±7 percent error of estimate on the 68 percent confidence limit; this error may be reduced even further with the proper use of adjustment factors (7). It was stated earlier that accurate traffic counting procedures must be utilized if the traffic assignment model is to be calibrated properly.

In spite of all the errors that may be included in the traffic assignment process, the results of the data presented in this report indicate that it still provides reasonable assignments to a transportation network. The best accuracy is usually obtained on the high-volume routes, the accuracy decreasing with decreasing volumes. However, it is the traffic occurring on these high-volume routes that is of concern at the present time. Consequently, the traffic assignment process does seem to be a success in providing this important information.

SUMMARY AND CONCLUSIONS

An explanation of the capacity restraint technique used in the traffic assignment process was presented first to provide the background information needed to understand the results obtained when using this technique. An evaluation was then made to determine the adequacy of traffic assignment when using capacity restraint to match existing travel patterns for a highway network in an urban transportation study. Detailed data were available from the Atlanta Area Transportation Study and the Madison Area Transportation Study; more limited data were available from eight other urban transportation studies for making this evaluation.

Five different tests were used to evaluate the Atlanta data. The results of these tests indicated that:

1. The traffic assignments were within ±15 percent of ground counts on a study-wide basis, as well as on a link-by-link basis; this is verified by the results obtained in Tests 1, 2, and 5. In most cases the assigned volumes were lower than the ground counts.
2. The underassignments appeared to be caused by the fact that the O-D survey data were about 15 percent underreported, as indicated by the screenline results. Consequently, it is reasonable to assume that if the O-D data had been factored to account for 100 percent of the screenline crossings, the traffic assignments would have matched the ground counts much more closely.
3. Some of the errors occurring in the assignment process must be attributed to the sampling error in the O-D surveys. A 5 percent sample of dwelling units was interviewed in Atlanta. According to the curves shown in Figure 13 the sampling error could account for about a 12 percent RMS error on links carrying 5,000 vehicles, an 8 percent RMS error on links carrying 10,000 vehicles, a 6 percent RMS error on links carrying 20,000 vehicles, and a 5 percent RMS error on links carrying 30,000 vehicles. To account for this error in the evaluation of traffic assignments for Atlanta, the following tabulation was prepared:

Volume Group	Average Volume	Number of Links	① Assignment RMS Error (Table 3)	② Sample Rate RMS Error (Fig. 13)	($①^2$ - $②^2$)	$\sqrt{①^2 - ②^2}$ Effective RMS Error
03-05	0.4	22	41.8	15.0	1520	39.0
05-10	7.5	56	38.2	9.5	1424	37.7
10-15	12.5	25	24.5	7.5	544	23.3
15-20	17.5	24	22.1	6.5	447	21.2
20-25	22.5	5	20.5	5.5	391	19.7
25-30	27.5	2	30.3	5.0	885	29.7
30-up	30	10	29.1	4.8	826	28.7

The effective RMS error for each volume group was computed by taking the square root of the squared differences between the RMS error from assignment (as obtained from Table 3) and the RMS error resulting from the sampling rate (from Fig. 13). The values shown in the last column as effective RMS error indicate that the RMS errors computed in Test 4 are of the same order of magnitude as the errors shown by Test 5, which provides a link-by-link comparison of the results. Thus, the conclusion that traffic assignments obtained in Atlanta are within ±15 percent again seems to be reasonable.

Data obtained from the Madison Area Transportation Study were also evaluated on the basis of five tests; the results of these tests indicated that:

1. The traffic assignments were approximately 5 percent low on an area-wide basis.
2. A graphic evaluation of results on a randomly selected portion of the Madison network indicated that the assignments were within ±15 percent of ground counts on a link-by-link basis.
3. Some of the error occurring in the assignment process must be attributed to the sampling error in the O-D survey. A 10 percent sample of dwelling units was interviewed in Madison. According to the curves in Figure 13, the sampling error could

account for about an 8 percent RMS error on links carrying 5, 000 vehicles, a 6 percent RMS error on links carrying 10, 000 vehicles, a 4 percent RMS error on links carrying 20, 000 vehicles, and a 3. 5 percent RMS error on links carrying 30, 000 vehicles. To account for this error in the evaluation of traffic assignment for Madison, the following tabulation was prepared:

Volume Group	Average Volume	Number of Links	① Assignment RMS Error (Table 4)	② Sample Rate RMS Error (Fig. 13)	$(①^2 - ②^2)$	$\sqrt{①^2 - ②^2}$ Effective RMS Error
03-05	0. 4	168	36. 0	9. 0	1215	34. 8
05-10	7. 5	224	30. 5	6. 5	890	29. 8
10-15	12. 5	143	27. 8	5. 5	744	27. 2
15-20	17. 5	59	18. 1	4. 5	308	17. 5
20-25	22. 5	22	13. 0	3. 8	155	12. 4
25-30	27. 5	11	7. 1	3. 5	39	6. 2
30-up	30	4	15. 1	3. 3	217	14. 7

The values shown in the last column as effective RMS error indicate that the RMS errors computed in Test 4 are of the same order of magnitude as the errors shown by Test 5, which provides a link-by-link comparison of the results. Thus, the conclusion that traffic assignments obtained in Madison are within ±15 percent of ground counts again seems to be reasonable.

Some important conclusions were also made concerning the ability of the capacity restraint process to reduce the errors in traffic assignment. These conclusions are based on the analysis made for assignment data obtained from 10 cities; they may be summarized as follows:

1. In most cases, capacity restraint reduces the overall error in traffic assignment. This may be seen by examining the results shown in Table 2 and Figure 10.
2. One application of capacity restraint seems to provide the maximum overall reduction in error on a study-wide basis, as measured by the "total weighted error" computation provided in the capacity restraint statistics; these values are also shown in Table 2 and Figure 10. However, it is more desirable to apply capacity restraint at least three times (resulting in four loads) to take advantage of the diversion effect of the process, as shown in Figure 1. This also provides more balanced network speeds, as shown by Figures 6 through 9. Finally, it results in a more desirable link-by-link comparison of assigned volumes and ground counts.
3. Reasonable assignments are obtained by using the average of four loadings. This is shown by an examination of the total weighted error in Table 2 and Figure 10, as well as the percent standard deviations shown in Tables 3, 4, and 5.
4. The "total weighted error" computation provided by the capacity restraint program does not measure the true accuracy of the traffic assignment process. This number is a relative index, showing the ability of capacity restraint to reduce the overall error in the assignment process.
5. The "percent standard deviation" computation provided by the capacity restraint program is a better indication of the accuracy obtained because statistics are provided for each volume group. However, these statistics are for average values; thus, they tend to show a larger error than indicated in a link-by-link comparison.
6. The "percent standard deviation" is approximately equal to the root-mean-square (RMS) error computed for each link (for ground counts vs assigned volumes). However, the RMS computations made for each volume group showed a slightly smaller error in assignment for each group.

7. The error in assignment is quite large for volume groups up to about 5, 000 vehicles; the error obtained for volumes greater than 10, 000 is considerably less. This may be seen by the results shown in Table 5.

8. The best way to determine the accuracy of the assignment process is a graphic presentation of the results on a link-by-link basis on a network map.

ACKNOWLEDGMENTS

The author expresses sincere thanks to those individuals in the Bureau of Public Roads regional and division offices, and the respective state highway departments and urban area transportation study staffs who provided much of the raw data needed to accomplish this research effort. Without their cooperation this work could not have been completed.

REFERENCES

1. Traffic Assignment Manual. U. S. Bureau of Public Roads, Office of Planning, Urban Planning Div., June 1964.
2. Skilton, James L. Volume Restraint and Capacity Restraint. U. S. Bureau of Public Roads, 1961.
3. Culp, C. A. Analysis of the Capacity Restraint Statistics. Unpublished paper, U. S. Bureau of Public Roads.
4. Atlanta Area Transportation Study: A Report on the Group 'A' Tables—Completeness of Vehicular Trip Information Obtained in the Internal Survey. Atlanta Area Transportation Study, Div. of Highway Planning, State Highway Dept. of Georgia.
5. Madison Area Transportation Study: Technical Report No. 8—Origin and Destination Survey. State Highway Commission of Wisconsin, Planning and Research Div.
6. Sosslau, Arthur B., and Brokke, Glenn E. Appraisal of Sample Size Based on Phoenix O-D Survey Data. HRB Bull. 253, pp. 114-127, 1960.
7. Guide for Traffic Volume Counting Manual. U. S. Bureau of Public Roads, Office of Planning.

Computer Coding of Origins and Destinations

HUBERT P. NUCCI, Urban Planning Division, U. S. Bureau of Public Roads

•THE purpose of this paper is to describe methods and procedures developed to code origins and destinations of trips by use of electronic computers. The use of computers rather than manual methods makes it feasible to code to a block level instead of the usual practice of coding to a survey zone. Coding to a block retains the original identity of the trip end in great detail, which allows almost unlimited flexibility in aggregating into any desired geographical area. The use of computing equipment was needed to speed up the process, reduce the overall cost, and maintain a high level of control.

The development of these procedures was prompted by the increased number of study areas, the ever-increasing amount of detail desired, the propensity toward human error in performing routine tasks, and the advantages inherent in the standardization of procedures. The Cleveland Seven County Transportation/Land Use Study conducted an O-D survey in 1963 and collected a 25 percent sample. The manual effort involved in coding the large amount of collected data was sufficient to justify a considerable effort in computer programming. These data were punched onto 2.05 million cards, recorded onto reels of magnetic tape, edited for obvious errors, and corrected. These reels of magnetic tape were then sent to the Bureau of Public Roads for processing.

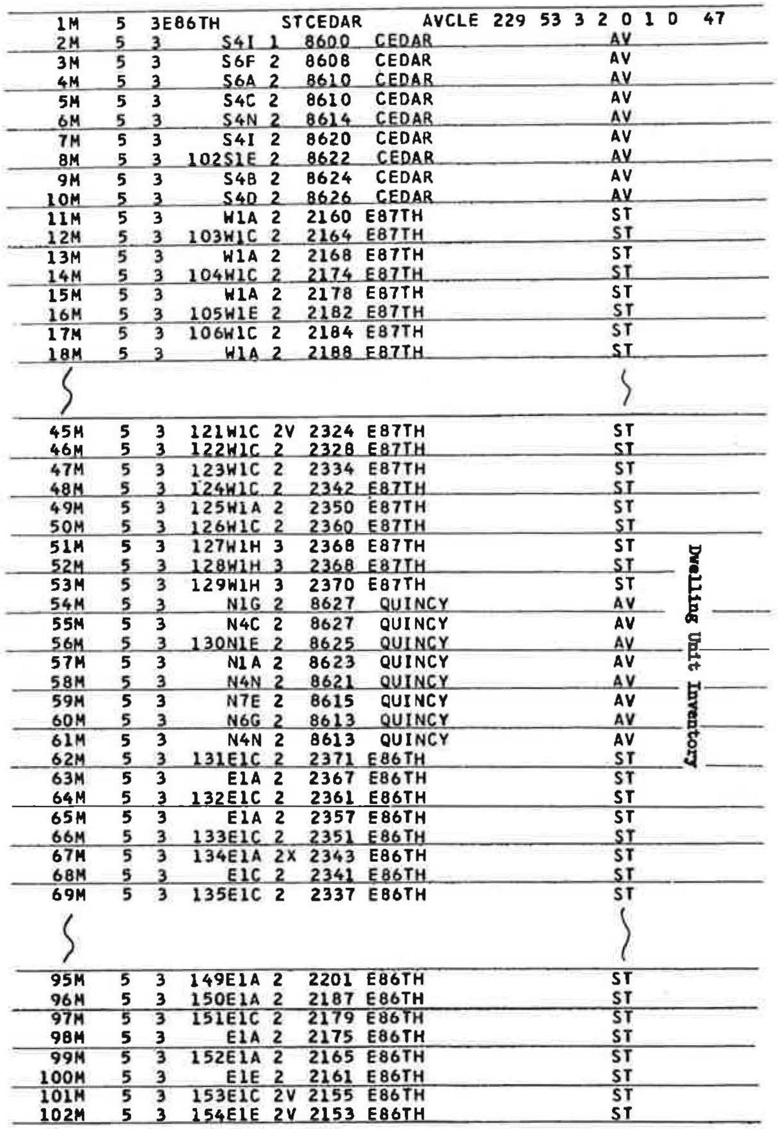
In the manual procedure, trip ends are visually referred to a zoned map or a coding index, and, when found, the corresponding zone number is manually recorded on the interview form.

Since an electronic computer has the ability to make comparisons between two items and determine if they are equal or unequal, then it is possible to insert a facsimile of a coding index into the computer and let it compare the location of trip ends to entries in the coding index, and if an equal condition occurs, let the computer enter a recode to the trip end. If this is possible, why not also retain a finer degree of identification by coding to an individual block instead of coding a trip end to a zone which has the effect of summarizing it in a zone with many other nearby trip ends? Also, if a coding index will have a location recode associated with it, why not also include a land-use recode? Since all of these are feasible, then why not let the computer search the trip ends instead of the staff? At least the computer is fast, exact, and consistent—not subject to boredom or daydreaming, and does not transpose digits. Taking advantage of the high speed of the computer, all trip ends could be compared to the coding index in a relatively short period of time and only the rejects forwarded to the editing staff for resolution. In this manner, by permitting the computer and the staff to concentrate in the area where each is most adept, an efficient, harmonious operation would result. The attributes of the computer are speed, accuracy, high volume, and repetition. These same attributes comprise the area where the staff is inefficient.

The disadvantage of the computer is that it lacks imagination and intuition. It cannot recognize items for speculation. This is the area where the staff excels. A human being can instantaneously recognize MTVIEW as Mountain View or Mount View instead of a misspelled name. Also NORTON RD would be detected by a human as being correct and not a contraction of North Orton Rd. The state of the art has not progressed to the point where we understand how a human can distinguish variations of spellings, immediately discarding certain combinations and recognizing others, or determining a misspelled name.

When the street name directory was first printed and visually inspected for alignment, ease of use, titles, format, and clarity, one entry stood out as an obvious error—

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.



```
  1M   5  3E86TH       STCEDAR      AVCLE 229 53 3 2 0 1 0  47
  2M   5  3       S4I 1  8600  CEDAR              AV
  3M   5  3       S6F 2  8608  CEDAR              AV
  4M   5  3       S6A 2  8610  CEDAR              AV
  5M   5  3       S4C 2  8610  CEDAR              AV
  6M   5  3       S4N 2  8614  CEDAR              AV
  7M   5  3       S4I 2  8620  CEDAR              AV
  8M   5  3   102S1E 2  8622  CEDAR              AV
  9M   5  3       S4B 2  8624  CEDAR              AV
 10M   5  3       S4D 2  8626  CEDAR              AV
 11M   5  3       W1A 2  2160 E87TH               ST
 12M   5  3   103W1C 2  2164 E87TH               ST
 13M   5  3       W1A 2  2168 E87TH               ST
 14M   5  3   104W1C 2  2174 E87TH               ST
 15M   5  3       W1A 2  2178 E87TH               ST
 16M   5  3   105W1E 2  2182 E87TH               ST
 17M   5  3   106W1C 2  2184 E87TH               ST
 18M   5  3       W1A 2  2188 E87TH               ST

 45M   5  3   121W1C 2V 2324 E87TH               ST
 46M   5  3   122W1C 2  2328 E87TH               ST
 47M   5  3   123W1C 2  2334 E87TH               ST
 48M   5  3   124W1C 2  2342 E87TH               ST
 49M   5  3   125W1A 2  2350 E87TH               ST
 50M   5  3   126W1C 2  2360 E87TH               ST
 51M   5  3   127W1H 3  2368 E87TH               ST
 52M   5  3   128W1H 3  2368 E87TH               ST
 53M   5  3   129W1H 3  2370 E87TH               ST
 54M   5  3       N1G 2  8627  QUINCY             AV
 55M   5  3       N4C 2  8627  QUINCY             AV
 56M   5  3   130N1E 2  8625  QUINCY             AV
 57M   5  3       N1A 2  8623  QUINCY             AV
 58M   5  3       N4N 2  8621  QUINCY             AV
 59M   5  3       N7E 2  8615  QUINCY             AV
 60M   5  3       N6G 2  8613  QUINCY             AV
 61M   5  3       N4N 2  8613  QUINCY             AV
 62M   5  3   131E1C 2  2371 E86TH               ST
 63M   5  3       E1A 2  2367 E86TH               ST
 64M   5  3   132E1C 2  2361 E86TH               ST
 65M   5  3       E1A 2  2357 E86TH               ST
 66M   5  3   133E1C 2  2351 E86TH               ST
 67M   5  3   134E1A 2X 2343 E86TH               ST
 68M   5  3       E1C 2  2341 E86TH               ST
 69M   5  3   135E1C 2  2337 E86TH               ST

 95M   5  3   149E1A 2  2201 E86TH               ST
 96M   5  3   150E1A 2  2187 E86TH               ST
 97M   5  3   151E1C 2  2179 E86TH               ST
 98M   5  3       E1A 2  2175 E86TH               ST
 99M   5  3   152E1A 2  2165 E86TH               ST
100M   5  3       E1E 2  2161 E86TH               ST
101M   5  3   153E1C 2V 2155 E86TH               ST
102M   5  3   154E1E 2V 2153 E86TH               ST
```

Figure 1. Dwelling unit inventory.

LJUBLJANA DR. The human eye had detected, presumably, the double occurrence of the LJ combination, and the brain did not accept it. Either a typographical error had occurred or the keypunch machine had malfunctioned. It was decided to attempt to find the street name in a printed index to determine the correct spelling. Searching first in the L's and ignoring the invalid J, there was found a street in the Euclid area spelled exactly as found in the street name directory. LJUBLJANA DR was a valid (correctly spelled) street name.

As a result of this speculative investigation, it was decided to assign to the staff those duties and functions best suited to the staff, and the computer would be required to perform those functions best suited to it. Therefore the computer would have to build a set of directories/dictionaries for searching purposes, would search the trip ends against these directories, and would assign the corresponding land-use code and block-number code to those trip ends found. The rejects would be separated so the

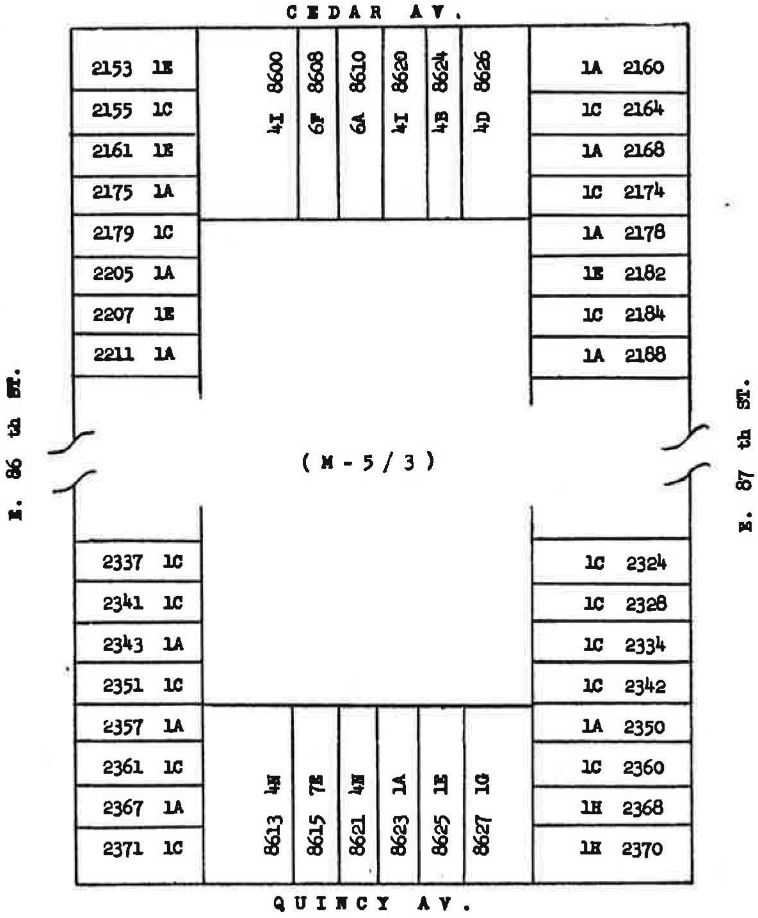


Figure 2. Illustration of Figure 1 data.

staff could resolve them, indicate the type of correction to be made, or, if possible, which land-use code and block-number code to assign to the trip end and return them to the computer for correction and final processing. Once the areas of responsibility were defined, the next step was to obtain the necessary data in a form that the computer could efficiently process. Two sets of data were required: (a) a dwelling unit inventory to identify parcels and their land-use description, and (b) the home interview data in a form that would correspond to the directories.

DWELLING UNIT INVENTORY

The dwelling unit inventory is a list of all parcels in the study area and includes such items as city (municipality), census tract, block number, house number, street name (direction, name, designation, orientation), land use, and, if applicable, one or more names associated with this address. This list was arranged in order of occurrence (Fig. 1) as though an enumerator field-listed the parcels by walking around each block. Whenever an intersection was encountered, the names of the two intersecting streets were recorded with the predominant land use. From this basic set of data, there is adequate information to build four directories/dictionaries—street name, parcel (house number), place name, and intersection.

Figure 2 is a partial illustration of the data listed in Figure 1. Both describe a block in Cleveland identified as census tract M-5 and block number 3 (M-5/3). Figure 3 illustrates the relationship of block M-5/3 with respect to adjacent blocks, each having

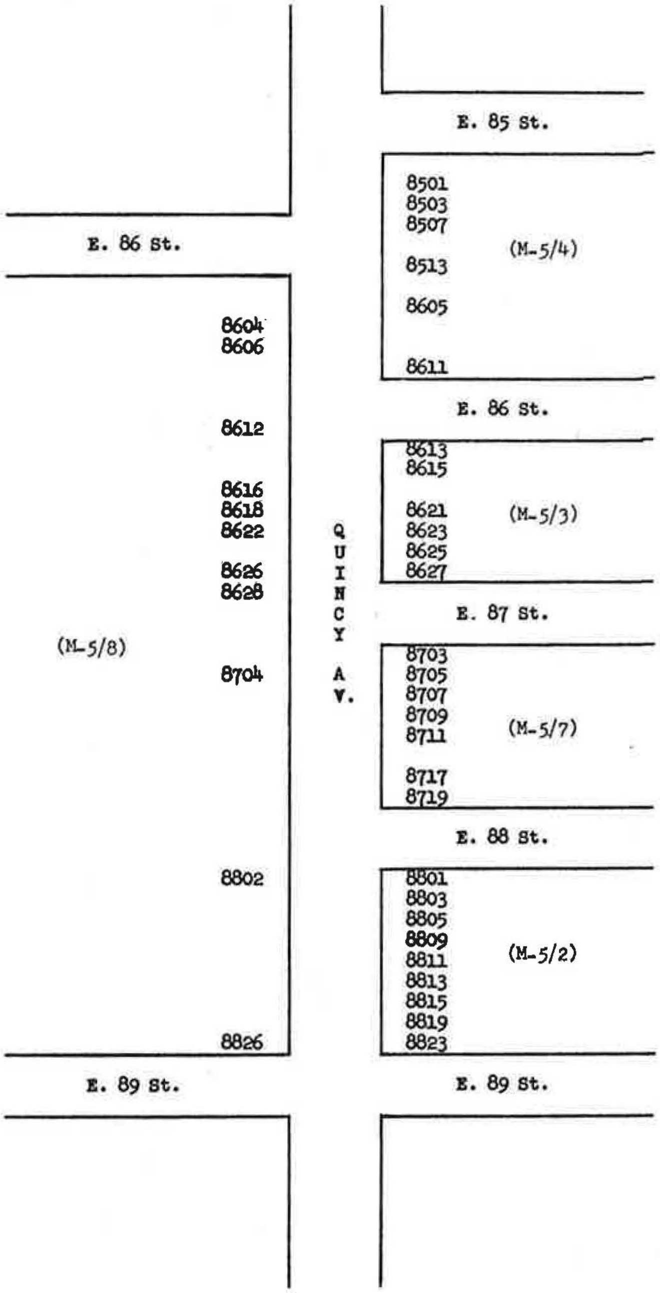
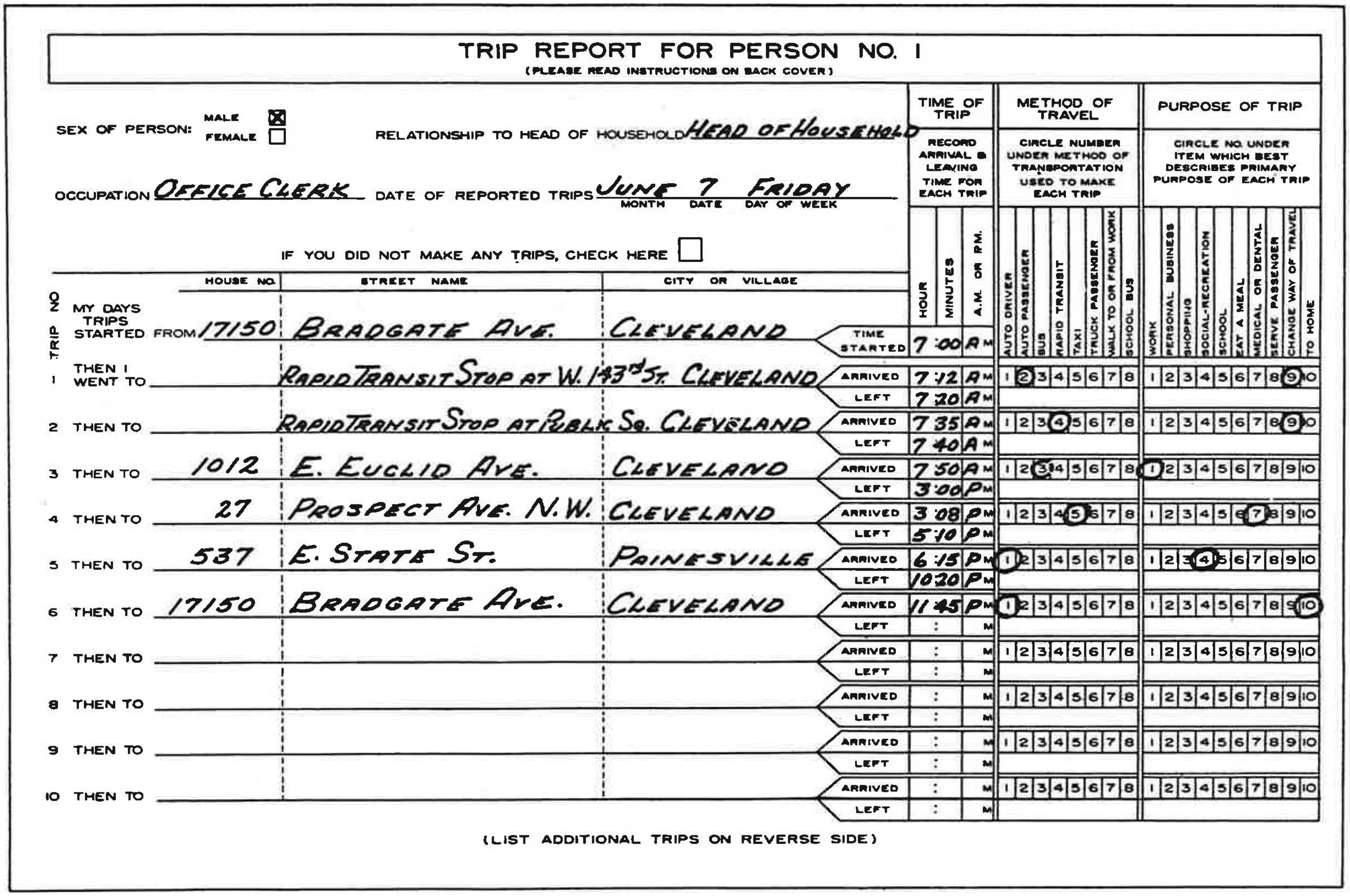


Figure 3. Relationship of Figure 2 to adjacent blocks.

a common face on Quincy Avenue. Quincy Avenue is used to illustrate the parcel (house number) dictionary (Fig. 7).

HOME INTERVIEW DATA

The home interview data (trip reports) were collected, using a reporting form similar in design to the ones normally used in O-D studies (Fig. 4). The difference is noted in that the first origin is recorded for a person and from then on only the succeeding



TRIP REPORT FOR PERSON NO. 1

(PLEASE READ INSTRUCTIONS ON BACK COVER)

SEX OF PERSON: MALE ☒ FEMALE ☐

RELATIONSHIP TO HEAD OF HOUSEHOLD HEAD OF HOUSEHOLD

OCCUPATION OFFICE CLERK

DATE OF REPORTED TRIPS JUNE (MONTH) 7 (DATE) FRIDAY (DAY OF WEEK)

IF YOU DID NOT MAKE ANY TRIPS, CHECK HERE ☐

TIME OF TRIP: RECORD ARRIVAL & LEAVING TIME FOR EACH TRIP

METHOD OF TRAVEL: CIRCLE NUMBER UNDER METHOD OF TRANSPORTATION USED TO MAKE EACH TRIP

PURPOSE OF TRIP: CIRCLE NO. UNDER ITEM WHICH BEST DESCRIBES PRIMARY PURPOSE OF EACH TRIP

TRIP NO.		HOUSE NO.	STREET NAME	CITY OR VILLAGE		TIME OF TRIP			METHOD OF TRAVEL								PURPOSE OF TRIP									
						HOUR	MINUTES	A.M. OR P.M.	AUTO DRIVER	AUTO PASSENGER	BUS	RAPID TRANSIT	TAXI	TRUCK PASSENGER	WALK TO OR FROM WORK	SCHOOL BUS	WORK	PERSONAL BUSINESS	SHOPPING	SOCIAL-RECREATION	SCHOOL	EAT A MEAL	MEDICAL OR DENTAL	SERVE PASSENGER	CHANGE WAY OF TRAVEL	TO HOME
	MY DAYS TRIPS STARTED FROM	17150	BRADGATE AVE.	CLEVELAND	TIME STARTED	7	00	AM																		
1	THEN I WENT TO		Rapid Transit Stop at W. 143rd St.	CLEVELAND	ARRIVED	7	12	AM	1	[✓ 2]	3	4	5	6	7	8	1	2	3	4	5	6	7	8	[✓ 9]	10
					LEFT	7	20	AM																		
2	THEN TO		Rapid Transit Stop at Public Sq.	CLEVELAND	ARRIVED	7	35	AM	1	2	3	[✓ 4]	5	6	7	8	1	2	3	4	5	6	7	8	[✓ 9]	10
					LEFT	7	40	AM																		
3	THEN TO	1012	E. EUCLID AVE.	CLEVELAND	ARRIVED	7	50	AM	1	2	[✓ 3]	4	5	6	7	8	[✓ 1]	2	3	4	5	6	7	8	9	10
					LEFT	3	00	PM																		
4	THEN TO	27	PROSPECT AVE. N.W.	CLEVELAND	ARRIVED	3	08	PM	1	2	3	4	[✓ 5]	6	7	8	1	2	3	4	5	6	[✓ 7]	8	9	10
					LEFT	5	10	PM																		
5	THEN TO	537	E. STATE ST.	PAINESVILLE	ARRIVED	6	15	PM	[✓ 1]	2	3	4	5	6	7	8	1	2	3	[✓ 4]	5	6	7	8	9	10
					LEFT	10	20	PM																		
6	THEN TO	17150	BRADGATE AVE.	CLEVELAND	ARRIVED	11	45	PM	[✓ 1]	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	[✓ 10]
					LEFT		:	M																		
7	THEN TO				ARRIVED		:	M	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
					LEFT		:	M																		
8	THEN TO				ARRIVED		:	M	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
					LEFT		:	M																		
9	THEN TO				ARRIVED		:	M	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
					LEFT		:	M																		
10	THEN TO				ARRIVED		:	M	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10
					LEFT		:	M																		

(LIST ADDITIONAL TRIPS ON REVERSE SIDE)

Figure 4. Home interview form.

destinations until all trips are accounted for. The trip reports were examined by the editing staff prior to keypunching. At this time, the editing staff classified the trip ends by type, in the following manner:

Any addresses which were the same as the residence were coded as type H (approximately 41 percent in Cleveland); place names were coded as type P; intersections were coded as type I; city names external to the study area were coded as type E; and all other addresses (house number and street name) were not coded, which resulted in a code of blank.

Each line entry was punched onto a single card and included basic information regarding the residence (city, census tract, block number, sample number, and serial number), the type of address (H, P, E, I, or blank), the address of the trip end, and other pertinent data (mode, purpose, time, etc.). The address field was used as a common field to accommodate any of the five types of addresses (see Fig. 5). Those trips classified as type H (home-based trip end) were punched by leaving the address field blank and only identifying the type as H. This reduced considerably the amount of keypunching. Also, any person who did not make any trips had only the origin card punched containing person data (sex, occupation, industry, etc.) and a code indicating that no trips were made by this particular person. The remainder of this trip was left blank. A summary card containing basic information pertaining to the residence was punched and included city, census tract, block number, serial number (always 1), land use, number of call-backs, type of address, address of residence, number of persons, persons over 5 years of age, roomers, number of cars, number employed, and number of persons making no trips. Control punches were used for identification of three types of cards—A identified the dwelling unit summary, B identified a person and his trip origin, and C identified each of the trip destinations for this person. Each card was serialized to establish a sequential order. As can be seen from the listing of the home interview trip report (Fig. 5), a considerable amount of data was not keypunched. In the Cleveland O-D study, three-quarters of a million cards (approximately 54.0 percent) were classified as home-based (H) or no trip, which resulted in a considerable reduction in the amount of keypunching required.

After keypunching, the cards were recorded onto magnetic tape, edited, corrected, and recoded. Then they were separated into groups by type of address so that the external (type E) could be searched against the external directory; the intersections (type I) could be searched against the intersection directory; the place names (type P) could be searched against the place name directory; and the addresses (type blank) could be searched first against the street name directory and second against the parcel (house number) directory. The home-based (type H) and no-trip cards were combined and held aside. No searching was necessary for these cards since they related back to the residence which was recorded in the trip summary card. This resulted in 54 percent of the cards not having to be searched in any of the directories.

BUILDING THE DIRECTORIES

Prior to punching the dwelling unit inventory onto cards, it became apparent that there was sufficient duplication of land use from one parcel to the next to warrant a deletion of duplication. Actually what was necessary was the first (or last) address of a string of parcels with the same land use in any block. To reduce the volume of data, duplication was eliminated and an additional condition was stipulated—that the first and last parcel on a block face be the minimum requirement (see Fig. 11). This reduced volume of data was punched onto 384,000 cards.

A summary of each block was punched onto a card (see card formats, Appendix A) containing identification and hand-summarized quantitative information such as city (municipality), census tract, block number, serial number, the names of the intersecting streets where the enumeration began, total number of parcels in the block, number of parcels in the sample, number of vacancies, and a distribution of types of refusals (in Cleveland the home interview was resident-respondent and was delivered and collected by the enumerator). The summary card was followed by detail cards, serially numbered, which recorded the addresses, land use, and associated place names of the parcels

```
 1CC 43   4 1281A4               6200  HEMINGWAY           RD  MAY 2 20220        A Trip Summary Record
 2CC 43   4 1281M4 HHJL 9TU H                                  MAY       555A 0   B Home Based
 3CC 43   4 1281            23555  EUCLID              AV  EUC 615A 330P11    C
 4CC 43   4 1281             5112  EASTOVER            RD  LYN 400P 430P12    C
 5CC 43   4 1281            15816  GROVEWOOD           AV  CLE 600P 800P12    C          Person-1
 6CC 43   4 1281             6295  WILSONMILLS         RD  MAY 810P 820P22    C
 7CC 43   4 1281          H                                MAY 825P     20    C Home Based
 8CC 43   4 1282F0 WIJL 9TU H                                  MAY       710A 0   B Home Based
 9CC 43   4 1282             6295  WILSONMILLS         RD  MAY 715A 720A12    C
10CC 43   4 1282             7828  WIRE                AV  CLE 800A 230P21    C          Person-2
11CC 43   4 1282             6700  GRANT               AV  CLE 235P 250P11    C
12CC 43   4 1282          E        PITTSBURG               PEN 300P     21    C External
```

```
 1CC 43   8 2201A1               6823  GLENVIEW            RD  MAY 4 40212        A Trip Summary Record
 2CC 43   8 2201M1 HHSE17TU H                                  MAY       800A 0   B Home Based
 3CC 43   8 2201             6635  WILSONMILLS         RD  MAY 804A 925A11    C
 4CC 43   8 2201          H                                MAY 930A1030A10    C Home Based
 5CC 43   8 2201             6635  WILSONMILLS         RD  MAY1035A1100A11    C
 6CC 43   8 2201             1345  MAYFIELD            RD  GAM1110A1205P11    C           Person-1
 7CC 43   8 2201          H                                MAY1215P1250P10    C Home Based
 8CC 43   8 2201          I        CHAGRIN   RDMAYFIELD   RDGAM 115P 150P11    C Intersection
 9CC 43   8 2201          E        RUSSELL                 OHI 205P 535P11    C External
10CC 43   8 2201          H                                MAY 555P 645P10    C Home Based
11CC 43   8 2201             6635  WILSONMILLS         RD  MAY 650P 730P11    C
12CC 43   8 2201          H                                MAY 735P     10    C Home Based
13CC 43   8 2202F9 WISE17TU1                                                  B No-Trip          P-2
14CC 43   8 2203M9 SOSE17TU H                                  MAY       730A 0   B Home Based
15CC 43   8 2203             1123  SOMCENTER           RD  MDH 800A 300P85    C                  P-3
16CC 43   8 2203          H                                MAY 330P     80    C Home Based
17CC 43   8 2204M9 SOSE17TU1                                                  B No - Trip        P-4
```

Figure 5. Home interview trip report.

on each face of the block. Intersections were punched onto cards (see card formats) depicting the predominant land use and the two street names at the intersection. All of the cards describing a block were serially numbered to correspond to the sequence of occurrence. Maintaining this sequence, the cards were entered into the computer which was able to distinguish one block from another. Each block was assigned a unique number (internally generated) starting with one and continuing in increments of one until the last one was assigned. This number is referred to as the CTB number—a five-digit number which actually represents the eight-character census tract and block number identification. The computer then scanned all of the data for a block, examined street names, intersections, house number sequences (ascending or descending), even and/or odd house number sequences in order to identify each block face. Once these determinations were made, the records were written out in a more compact form with the majority of identification recoded numerically for compactness. Each individual record had a land-use recode and a CTB number attached to it. The detail records had an additional bit of information attached to them (an A or a D) which signified whether the inventory of each block face was recorded in an ascending or descending order. This is essential in determining the appropriate land-use code to use in searching.

Statistics were accumulated for the various land uses on a block, census tract, city, county, and total level. From the summary records, enough data were available to summarize (by blocks) to the census tract level and compute an expansion factor. This was written out onto a separate reel of magnetic tape in the form of a table containing identification, expansion factor, and a correspondence table of CTB numbers and their equivalent census tract and block number.

The census tract expansion factor was calculated to two decimal places in the following manner:

$$F = \frac{A - (C \cdot A/B)}{B - (C + D_1 + D_2 + D_3 + D_4)}$$

where

F = factor for census tract (xx.xx);
A = number of parcels in the block;
B = number of parcels in the sample;
C = number of parcels vacant;
D_1 = number of refusals;
D_2 = number of returns with no information;
D_3 = number of no contact, no return; and
D_4 = number of contacts with no return.

At this point each record has a land-use recode and a CTB number attached to it, which is the identification needed to identify a trip end to the block level and the appropriate land use. These records were further processed in order to extract the information for four directories. As the data are passed through the computer, each record is examined to determine what has to be extracted. The detail records are separated for use in the street name and parcel directories. If a place name appears on a detail record, it is written out for use in the place name directory. All intersections are written out separately for use in the intersection directory.

The detail records are arranged into sequence by street name and house number (house numbers in ascending order—all even numbers preceding the odd) grouped by city (municipality). As the file passes through the computer, two directories are created—a street name directory and a detailed parcel directory.

Street Name Directory

The street name directory is a list of each street name, grouped by city and arranged in alphabetic order. Associated with each one is a unique street sequence number. This number (5 digits) is actually a recode (for brevity) and represents up to 25 characters allowed for the maximum street name length (Fig. 6).

Parcel (House Number) Dictionary

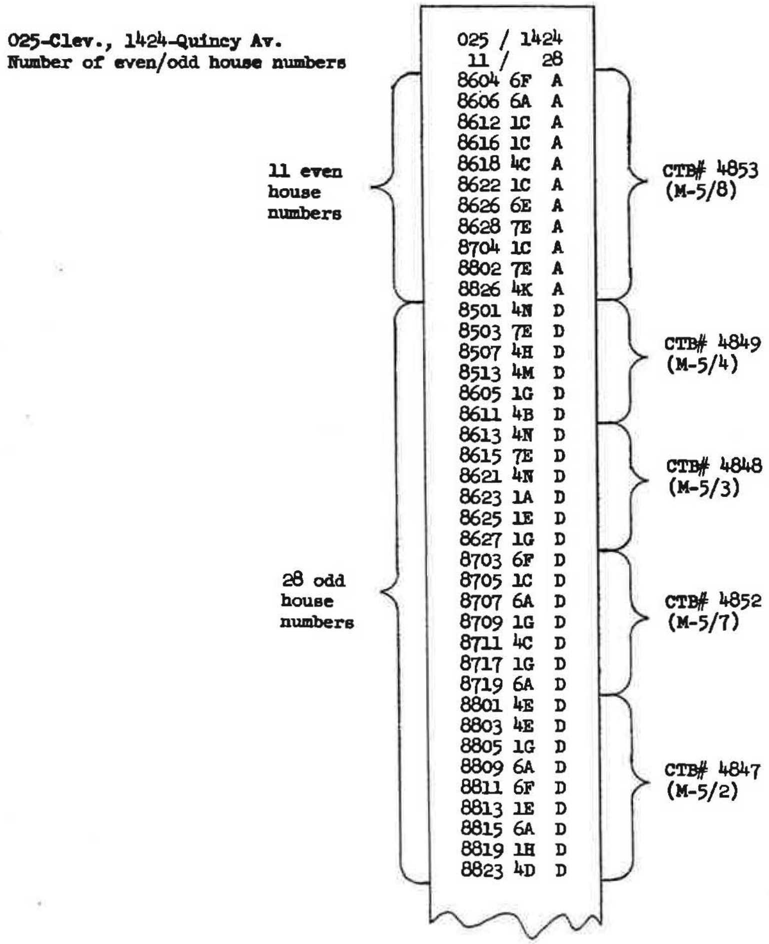
The parcel dictionary (Fig. 7) is a list of all house numbers for each individual street within a city (municipality). It is identified by the street sequence number instead of the actual street name. The house numbers are arranged in ascending order with the even house numbers preceding the odd numbers. This actually represents the even side of a street for the complete length of its existence within a city followed by the odd side. Each house number has associated with it a land-use recode, a CTB number, and the A or D code (ascending or descending).

Place Name Dictionary

A place name dictionary (Fig. 8) is created by extracting any and all names associated with an address from the dwelling unit inventory records. The corresponding land-use code and the CTB number is also picked up and associated with the place name. The place names are arranged in alphabetic sequence and grouped by city (municipality).

CITY RECODE NUMBER			025
CITY ABBREVIATION			CLE
	(STREET NAME - - - -	DESGN)	(SEQ#)
	ABBEY	AV	1
	ABELL	AV	2
	ABERDEEN	AV	3
	ABINGTON	RD	4
	ABLEWHITE	AV	5
	ACKLEY	RD	6
	ADAMS	AV	7
	≀	≀	≀
	CEDAR	AV	301
	CENTER	ST	302
	≀	≀	≀
	QUINCY	AV	1424
	QUINN	CT	1425
	≀	≀	≀
	WASHINGTON	AV	1800
	WASHINGTON	BV	1801
	≀	≀	≀
	E 86TH	ST	2078
	E 87TH	ST	2079
	≀	≀	≀
	W LLOYD	RD	2308
	W SCHAAF	RD	2309

Figure 6. Street name dictionary (quick reference).



```
025 / 1424
 11 /   28
8604 6F  A
8606 6A  A
8612 1C  A
8616 1C  A
8618 4C  A
8622 1C  A
8626 6E  A
8628 7E  A
8704 1C  A
8802 7E  A
8826 4K  A
8501 4N  D
8503 7E  D
8507 4H  D
8513 4M  D
8605 1G  D
8611 4B  D
8613 4N  D
8615 7E  D
8621 4N  D
8623 1A  D
8625 1E  D
8627 1G  D
8703 6F  D
8705 1C  D
8707 6A  D
8709 1G  D
8711 4C  D
8717 1G  D
8719 6A  D
8801 4E  D
8803 4E  D
8805 1G  D
8809 6A  D
8811 6F  D
8813 1E  D
8815 6A  D
8819 1H  D
8823 4D  D
```

Figure 7. Parcel (house number) dictionary.

CITY RECODE NUMBER		025
CITY ABBREVIATION		CLE
(PLACE NAME)	(L.U.)	(CTB#)
ABHARTSCHOOL	7A	5193
AMERICANSTEELWIRE	2E	4266
ARTMUSEUM	7D	6176
BURKEAIRPORT	3F	3761
CLEVELANDHOPKINSAIRPORT	3F	7394
CLEVELANDZOO	8G	3533
HOPKINSAIRPORT	3F	7394
LAKEFRONTAIRPORT	3F	3761
MUNICIPALSTADIUM	8E	3712
NASA	2I	7394
REIDSTATION	3C	5221
ROSEBUILDING	6F	3882
STADIUM	8E	3712
STLUKEHOSPITAL	7B	6309
UNIVERSITY HOSPITAL	7B	6239
VETERANSHOSPITAL	7B	6166
ZOO	8G	3533

Figure 8. Place name dictionary.

CITY RECODE NUMBER				025
CITY ABBREVIATION				CLE
	1ST ST.SEQ#	2D ST.SEQ#	(LU)	(CTB#)
(CEDAR AV./E.86TH ST.)	301	2078	1A	4848
(CEDAR AV./E.87TH ST.)	301	2079	1A	4848
(QUINCY AV./E.86TH ST.)	2078	1424	1A	4848
(QUINCY AV./E.87TH ST.)	2079	1424	1A	4848
CUL DE SAC	501	501	1A	2222
DEAD END STREET	603	603	1A	2222

Figure 9. Intersection dictionary.

Intersection Dictionary

An intersection dictionary (Fig. 9) is created from all intersection records encountered in traveling around eack block. Each of the two street names has its corresponding street sequence number associated with it. Therefore, the dictionary has two street sequence numbers arranged low to high instead of two street names and the predominant land use at the intersection with the CTB number.

STATE ABBREV. AND CODE	ALA - 01		
	(CITY)	(COUNTY)	(CITY)
	ABBEVILLE	067	0010
	ALBERTVILLE	095	0020
	ALEXANDER CITY	123	0030
	ALICEVILLE	107	0040
	ARZ - 02		
	AJO	019	0010
	AVONDALE	013	0020
	WYO - 49		
	BUFFALO	019	0010

Figure 10. External dictionary.

External Dictionary

The external dictionary is a list of all cities in the United States with a population of 2,500 or more arranged in alphabetic sequence by state (Fig. 10). Associated with each city is a nine-digit identification code (2-digit State code, 3-digit county code, and a 4-digit city code). This code structure is contained in the IBM publication "Numerical Code for States, Counties and Cities of the United States." The last digit of all city codes is a zero, which allows an additional nine entries to be added to the list. Since many external trip ends were to places within Ohio other than cities with a population of 2,500, an additional table was composed of over 5,000 entries such as resort areas, camps, parks and lakes. These entries were assigned last digit codes other than zero which readily distinguish them from the original IBM list.

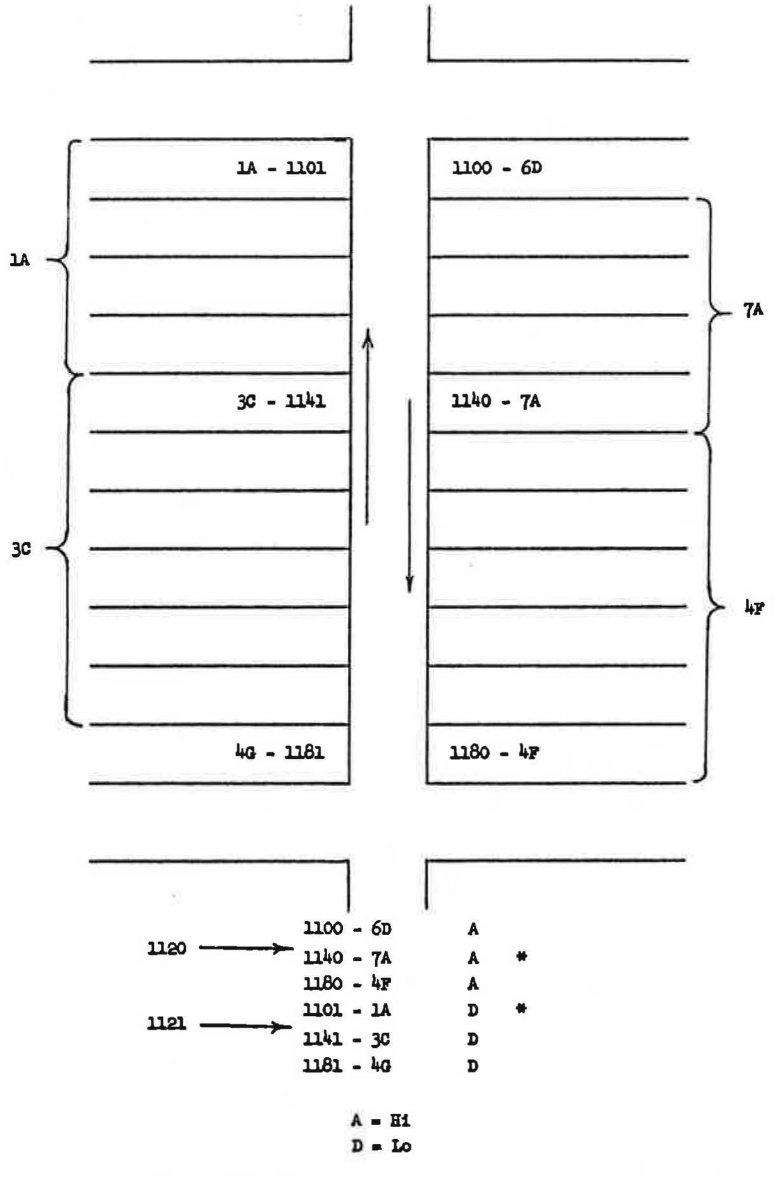


Figure 11. Illustration of directory searching.

DICTIONARY SEARCHES

Those trip ends classified as addresses (type = blank) are searched against the street name directory and, if successful, are then searched against the parcel directory.

The trip ends are grouped by city and searched against the street names for that city. If a match occurs, the street sequence number is inserted into the record. If a match does not occur, the trip end is written out onto a separate error tape and is sent to the staff for resolution. Those trip ends that matched are then arranged by street sequence number within each city and searched against the parcel directory. When a match occurs on city and street sequence number, the house number is examined to determine whether it is even or odd. If it is even, it is searched in the front of the directory. If it is odd, searching begins beyond the last even entry (Fig. 7). If an equal match occurs on house number, the land-use recode and CTB number are attached to the trip end record. If the house number falls between two house numbers in the dictionary, further examination is necessary. The CTB numbers of the two house numbers are checked to see if they are the same; if so, it signifies that both are in the same block. If this is so, then the A/D is examined to determine which land-use recode to use. If it is A, then the land-use recode of the lower house number is used. Conversely if it were D, then the higher house number is used (Fig. 11).

If the two CTB numbers were different, it would signify that the house number is between two blocks (out in the intersection) which should be rejected as a not-found. Before rejecting this record, another attempt is made to accept it. The house number hundreds range is examined to see if it is in the same range. If so, it uses the recodes of the house number with the same hundreds range. If both house numbers are in the same hundreds range, preference is shown to the low house number. Detailed counts are kept of the number of trip ends found and not found, and also for all of the variations as explained above. Those that are rejected are transmitted to the staff for resolution.

Figure 11 illustrates a typical street enumerated in two different directions. The even side was recorded in ascending house number sequence and the odd in descending sequence. The brackets in both margins show the range of parcels with common land uses and the house number representing each particular group. The bottom of the figure illustrates the order of house numbers as they would appear in the parcel directory. Two sample addresses (1120 and 1121) are used as examples for searching, and if reference is made to the physical arrangement of the blocks above, it can be seen that 1120 would be assigned the land-use code of the group represented by 1140 (the high address) and flagged by the asterisk (*). In the case of 1121, the low address is used and again flagged by an asterisk. To summarize, if an address falls between two addresses in the directory and they are coded A (ascending), use the high address. The converse is true for D (descending).

Place Name Search

Those trip ends to place names (type = P) are grouped by city and searched against the place name directory. Those that match exactly have the corresponding land-use recode and CTB number attached to the record. Those that are not matched are written out separately for transmittal to the staff for resolution.

External Search

Those trip ends to external cities (type = E) are grouped by state, city, and station of interview, and searched against the external directory. Those that are matched are recoded to the corresponding state, county, and city code and have a key code of 7 inserted in an error position. Those that are not found are merely bypassed and noted by printing a copy of the error. This error list is transmitted to the staff for resolution. The external directory has the added feature of allowing additional cities or variations of spellings of cities to be included. Multiple entries are allowed such as: WASHDC, WASHINGTONDC, etc. This enables a high rate of matching on subsequent searches.

Intersection Search

Those trip ends to intersections (type = I) have to be grouped by city and searched against the street name directory in order to recode each of the two street names to their street sequence number. If this is successful then a search is made against the intersection directory to match the combination of two street sequence numbers to select the land-use recode and CTB number. Rejects from either search are transmitted to the staff for resolution.

Visual Reference Material

After the directories are built, each one is formatted and printed so the staff will have access to the same information the computer has been working with. They are supplied with the street name directory, the parcel directory, the place name directory, the intersection directory and a copy of the street name directory rearranged to show all of the cities a street name occurs in. A copy of "Numerical Code for States, Counties and Cities of the United States" and a listing of all popular places in the state are supplied for resolving trips to external places.

ERRORS AND CORRECTIONS

When processing large volumes of data (more than a million), a moderate error rate of 10 percent becomes a staggering figure. If it is approached in a logical manner, advantage may be taken of certain conditions that occur. The errors were scanned to determine the most efficient procedure to rectify them. It was noticed that there were a significant number of cases where the reason for not matching entries in the directory was a difference in spelling of one or two characters. In other instances there was clustering—a recurrence of the same error due to a common misspelling or a missing entry of a popular place in the directory (e.g., Ford Motor Co. Assembly Plant).

This prompted two things: (a) a general correction routine program and (b) an orderly arrangement of errors to ease the task of the staff.

General Correction Routine

When a minor error occurs, it is usually necessary to repunch the data in their entirety and also reflect the change (usually one or two characters). In punching a correction, it quite often occurs that an additional error is introduced in duplicating the original correct portion of the record. It was decided to develop a flexible correction program which minimized keypunching, reflected only the change to be made without disturbing the correct portion of the record and would make a common change to more than one record from one correction card. To avoid lengthy identification, it was decided to identify a record by number (relative position in a file). In order to conform with this condition, each record rejected in any of the searches was printed with its record number. As an example, assume that the only error in the 127th record was CENTRAL RD instead of CENTRAL AV and if the street designation (RD or AV) appeared in positions 57-58 of the record, then the staff should be able to represent this in a punched card as C127, 57*AV*. This involves keypunching only 11 columns of the correction card instead of repunching the entire record. Translated it reads: change (C) the 127th record (127) of this file and place in position 57 whatever is found between the two asterisks (*). The comma separates the record number from the starting position number. Using the same example, this time let us assume that the same error exists in 75 consecutive records. This same correction can be laced forward by the following correction card: C127/201,57*AV*. This involves keypunching only 15 columns of the card and translated reads: change (C) starting with the 127th record (127) through and including (/) the 201st record (201) by inserting, starting in the 57th position (57) of the record whatever is included between the two asterisks (*). This program was found to be quite simple to use, flexible, and extremely easy to comprehend by the staff.

Orderly Arrangement of Errors

A very simple technique was employed in many instances to sequence the errors (depending on type of search) to group duplication. The staff had only to resolve the first case and then carry the same correction forward for as many times as it occurred, resulting in the keypunching of only one card for each group of duplicates. Although the errors encountered in processing the Cleveland data were numerous, it was surprising how fast and efficiently the staff resolved them.

Additional techniques also evolved. In the case of addresses where the street name could not be found, the staff had to find the street name in order to assign the street sequence number, but as time went on they were able to assign directly to the land-use code and CTB number. In these cases a code of 2 or 3 was recorded in a key position of each record which the search programs interrogated. Code 2 informs the street name search to bypass this record—it is already coded to street sequence number. Code 3 informs both the street name search and the parcel search to bypass this record. A code 7 was used similarly in the external search to indicate that the particular record had been found.

Post-Searching

After all of the trip ends are searched and found and/or resolved, the summary records have the expansion factor inserted into each record, and the original file is reestablished by putting it back into sequence, representing each individual interview form. The next step is to create standard trip cards. A table is supplied by the staff indicating the zone number each CTB number is to be assigned to. This table, with the file, is passed through the computer which creates a number 1 card (dwelling unit summary) from the trip summary record and inserts the appropriate zone number for the CTB number in the record. Both are retained. From then on, each person in the interview has his trips developed into number 2 trip cards with the expansion factor carried forward from the summary record and CTB numbers converted to zone. Again, both (CTB numbers and zones) are retained. Since each record contains only one trip end, each pair of records is held in order to create one trip card. That is, the first and second trip end records will create the first trip card. Then, the second and third trip end records create the second and so on. Had the trip end been a type H (home-based), the land-use recode, CTB number, zone and the expansion factor would be picked up from the summary record.

Comparability/Flexibility

If it is desired to compare this survey to a preceding one, all that is necessary is to repeat the trip card creation procedure and substitute a table that indicates to which zones (previous) the CTB numbers are to be assigned. The same applies if it is necessary to evaluate more than one zone configuration of the study area.

These procedures are quite flexible in that the external, truck, and taxi surveys can be combined with the internal survey after each is split by type of address. These combined groupings can be searched and corrected in one pass through the procedures instead of four individual ones. They can then be separated into their original survey groupings.

Subsequent Use of Directories

The initial cost and effort to build a set of directories is a one-time occurrence. For subsequent use, all that is necessary is to reflect the changes that have occurred during the elapsed time. Procedures for updating the directories are available.

This battery of computer programs was written for the IBM 1401 and the IBM 7090/94 under the IBSYS monitor. This monitor afforded two distinct advantages in the input/output areas (IOCS). Most of the data files were multireel, the largest being contained on 10 reels of magnetic tape even at a density of 800 characters per inch. Each file was labeled for identification, and each reel within a file had a reel sequence number automatically recorded on it. Internally, computer memory is buffered (by the monitor) so that once a reel of tape was set into motion (reading or writing) it did not stop, since there was adequate room to receive the data. As each piece of data was processed, its buffer area was released and made available again to incoming data. As a result of this buffering, the majority of the programs operate at tape speed (6 to 8 minutes per full reel of tape).

REMOTE OPERATION

The majority of the large computer facilities are concentrated in the major metropolitan areas. This has the tendency of creating a hardship on many potential users since there is usually a breakdown in long distance communications, or an additional burden of expense and time is incurred in sending a staff to a distant installation to oversee processing.

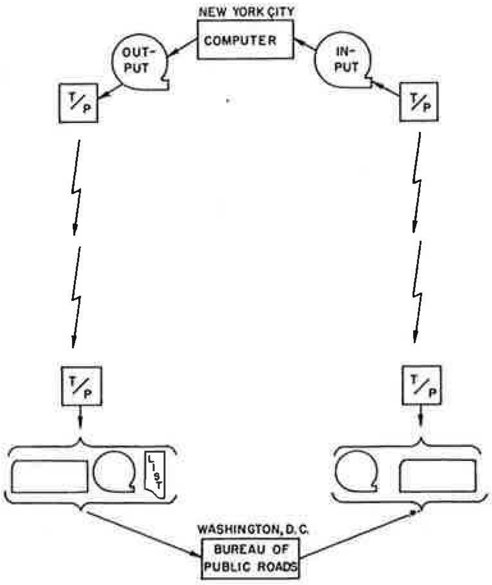


Figure 12. Remote operation utilizing teleprocessing equipment.

During the development of these procedures, teleprocessing equipment became commercially available between the Washington, D. C., area and a computer facility in New York City (Fig. 12). With this equipment, it was possible to transmit over leased telephone lines information (programs and data) between the two cities in a relatively short period of time. One reel of magnetic tape can be transmitted in 10-45 minutes depending on the length of tape records and the quantity.

It was decided to investigate the use of this equipment to determine if it would be feasible for isolated users. Many technical problems were encountered in the beginning, but as time progressed the transmission of data between the two cities evolved into a rather smooth operation. The most sensitive area of difficulty was narrowed down to the highly critical alignment of the tape unit read/write heads between installations. The best insurance was found to be in maintaining adequate backup of data transmitted in the form of a duplicate copy of the data on another reel of magnetic tape. If data were properly transmitted, received, and processed, then the backup copy could be released. If not, then lost time was minimized since an additional copy of the data was readily available. The end result was a saving in man-hours on the part of the staff, several operations were proceeding simultaneously, and the staff was better able to utilize its time for other functions.

ACCURACY OF SEARCHING

The accuracy of the searches was verified by randomly spot checking records that were found to insure that the correct land-use code and CTB number were recorded, and also the errors were similarly verified by checking that there was no match for them in the directories. An exact match is necessary for a record to be considered as found. A difference of one character in spelling is enough for a rejection (VERMILLION vs VERMILION). The results of searching are dependent upon the accuracy and completeness of the directories.

In the Cleveland survey, which was resident respondent, the home interview forms were delivered to the dwelling units by the enumerator at the time the field listing was recorded. One of the ground rules was that the address of the dwelling unit had to be recorded on the field listing and identified by a sample number. When the home interview trip reports were received, punched onto cards, and separated by types of address, the dwelling unit summaries were held aside and searched separately as the first search operation. This was done for two reasons. First, the volume was low (approxi-

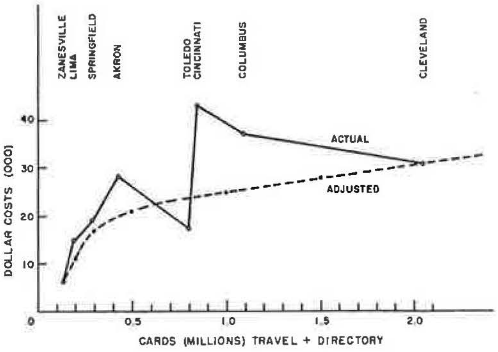


Figure 13. Relationship of dollar costs to card volume.

mately 10 percent), and second, the results of the search would indicate the level of accuracy of the street name and parcel directories at the earliest stage of overall processing.

ACCEPTANCE

As of this writing, these procedures have been successfully used to process the O-D survey data for Cleveland, Akron, Columbus, and Cincinnati in Ohio and Louisville, Kentucky. Four additional Ohio cities (Toledo, Lima, Springfield, and Zanesville) are in process. The State of Ohio has also collected data for Steubenville and Canton to be processed in the same manner.

PERFORMANCE OF SEARCHING

By examining the bar charts depicting the results of searching (Appendix B), it can be concluded that, in general, the larger the volume of records, the better the match rate. The lowest match rate occurred in searching intersections. This was due to the fact that the intersection directory was insufficient to be representative of the study area.

Intersections also constitute the major problem area of these procedures. If the search finds an intersection, there are up to four entries in the directory, representing the four corners, each having a different CTB number. Which of the four should be assigned to the trip end? If you select randomly, you have one chance in four of being correct, but how representative is the land use at that corner to the actual land use at the terminal end of the trip? Isn't it true that, if a trip ends at an intersection, the actual terminal end is more probably in the middle of the block with some other land use? Possibly, one alternative would be to assign a miscellaneous land-use code for trip ends to intersections. In this set of procedures, the first intersection encountered in the directory was used. It was realized (too late during the development of this battery of programs) that the intersections were not required in the field listing of the inventory data. Since each block could be isolated, then intersections could be generated (in the computer) at the recognition of change in street names as the block periphery was scanned. It is true that this would generate four times the number of actual intersections, but duplication could be removed. The total number of actual intersections can be estimated by the equation

$$\Sigma I = C\left[1.99 + (\Sigma\ B)^{1/2}\right]^2$$

where

I = total number of intersections;
C = squareness (shape) of the study area, ranging from 1.000 (for square) to 1.969 (for maximum elongation); and
B = total number of blocks.

COST COMPARISONS

The cost to manually code trips is $1.17 per home interview or 20.6 cents per trip if a trip rate of 5.7 trips per household, as experienced in the Cleveland survey, is used. The figures of $1.17 and 20.6 cents are averages based on data for 10 studies in the Missouri, Texas, and Arizona areas ranging over the years 1958-1964. To manually code the 143,496 home interviews in the Cleveland survey would cost approximately

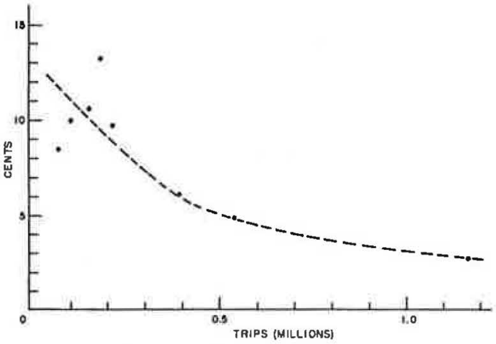


Figure 14. Cost per trip.

$167,000. The computer method costs less than $35,000 at a rate of 3.1 cents per trip. This figure can be misleading if other factors are not taken into consideration. A ratio of travel cards to directory cards of four to one seems to be ideal, as experienced in processing eight cities in Ohio. One city, Cincinnati, had a one-to-two ratio, and the cost was 7.2 cents per trip. The average cost for eight cities in Ohio was 7.0 cents per trip. At this time, it appears that the cost of producing the directories is a relatively expensive item.

In order to compensate for this expense, a proportional number of trips must be searched to reduce the unit cost. Personnel costs average to about 23 percent of total cost with a range of 18-31 percent for the eight cities. The condition (accuracy) of the input data is by far the most influential factor affecting total cost. Processing time is currently 15-16 weeks for a study area.

Considerations

In processing the data for the Cleveland survey, there were 1,668,199 trip end records searched for all four surveys (internal, external, truck and taxi). The total error rate (for all four surveys) was 10.7 percent or 178,619 trip ends not found. In resolving these errors, it only required approximately 15,000 cards to make these corrections, which is 1 percent of the total trip ends searched.

Figure 13 shows the relationship of total dollar costs to total card volume (travel and directory) for eight Ohio cities, indicated by the solid line. Three cases are atypical—Akron, Cincinnati, and Columbus. These costs were unusually high because they were the first cities to use the method, and the procedures required a learning process and some minor modifications. The dashed line is a smoothed fit to adjust to a more uniform representation. Toledo, which lies well below the line, illustrates a case where the data were exceptionally clean, and no problems were encountered with the programs and operations. Using the adjusted cost curve, the cost per trip is shown in Figure 14 and ranges from 2.7 cents to 13.3 cents. The points represent the same eight cities from Figure 13.

CONCLUSIONS

It is the opinion of the author that these procedures are a step in the right direction of utilizing the capabilities of electronic computers to assist traffic engineers and urban planners in coping with an ever-increasing task. The logic of the procedures is quite sound, but there is definitely room for improvement. With the advent of newer, faster, larger, and more economical computers, it is planned to convert these procedures to the new generation of machines, incorporating many improvements and efficiencies that were realized during and after the initial development.

The restraining bonds which have hampered all concerned in the past seem to be loosening. The computer will allow much larger volumes of data to be analyzed than before, which should reflect a larger sample, larger study area, or both. The staffing requirements are less although they will have to be of a higher caliber. Remote operation is now feasible with a minimum of delay and encumbrance and allows several studies to be conducted by the same staff concurrently.

The retention of microscopic identification (at the block level) adds a tremendous amount of flexibility and utility to the data. A definite breakthrough has been experienced in the area of resolving and correcting errors. The point has been reached where we are now able to process accurately larger volumes of data, faster, at a lower cost, and still maintain a tight control.

In many areas, assessors' records are rapidly becoming mechanized. These records contain a wealth of information, the least of which is sufficient to provide the basic elements of the search directories. If these records were combined with utility billing records which contain telephone numbers, names, addresses, and zip codes, the result would be a master file from which the search directories could be constructed, and also a sample could be selected and established as a separate file. The sample file could be used to print mailing labels, and the interview forms could be delivered and returned by the post office. Returns could be matched against the sample file to account for reporting and nonreporting. Periodically, the nonreporters could be listed from the sample file for follow-up. If a phone number is available in a nonreported record, the follow-up could be handled by telephone. If a telephone number is not available, then the follow-up could be handled first by mailing another interview form and second by a personal visit. All returns could be checked against the sample file to maintain control and account for all samples selected.

An operation of this type, including the printing of labels, is quite feasible with the computer equipment available today. It would minimize the time and expense involved in obtaining the home interview and a consistent field listing of the parcels and their land uses.

Appendix A

CARD FORMATS

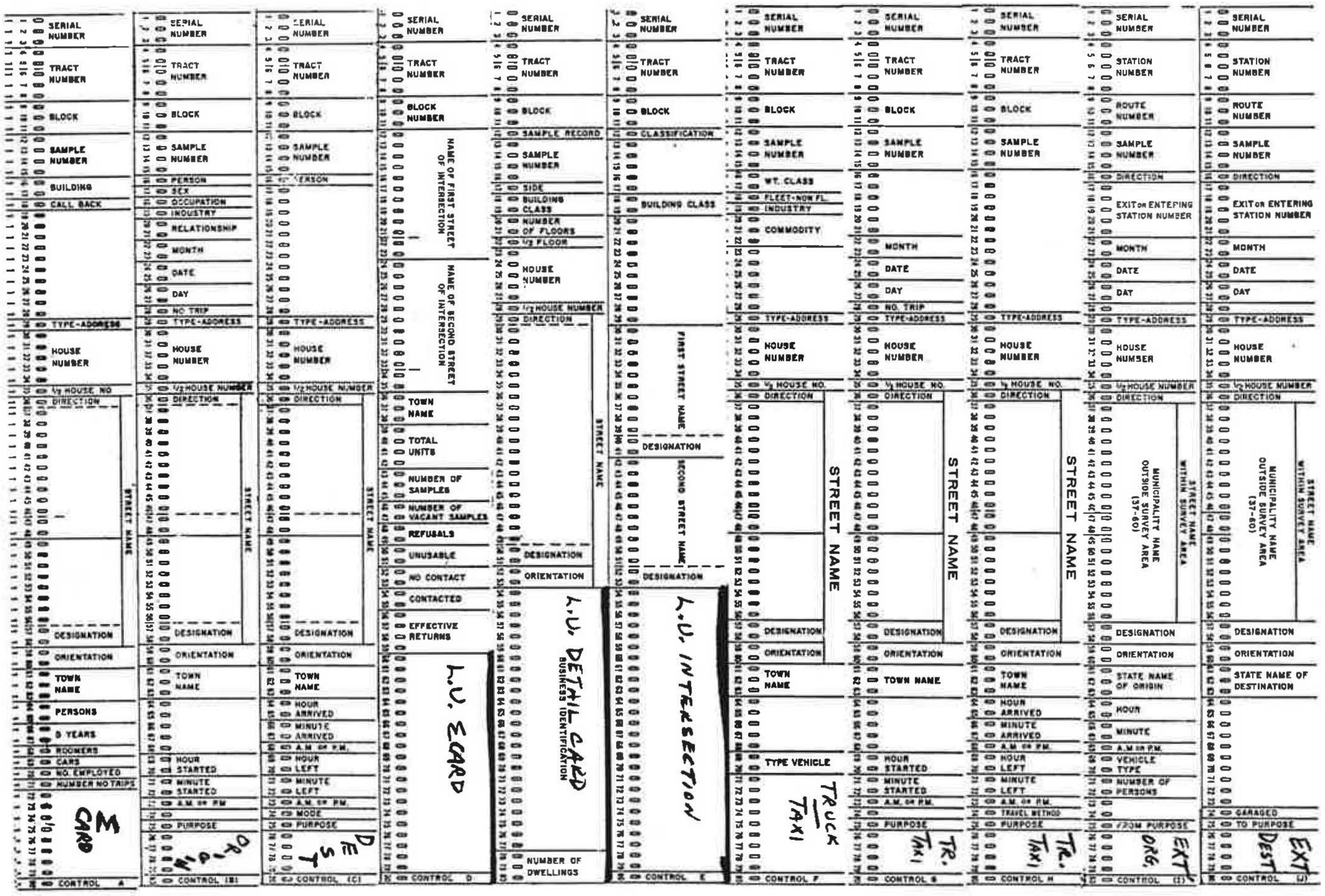


Figure 15.

Appendix B

SEARCH PERFORMANCE

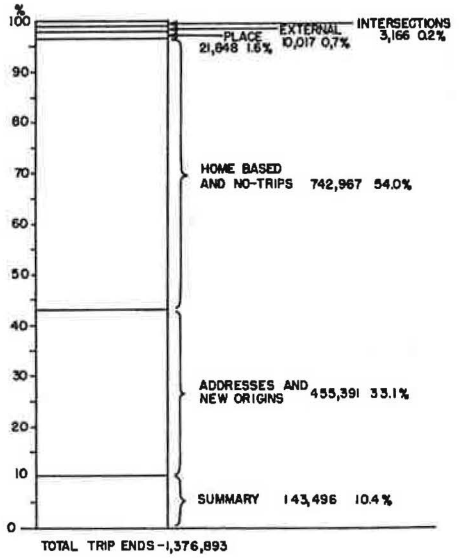


Figure 16. Distribution of trip ends by type, internal survey.

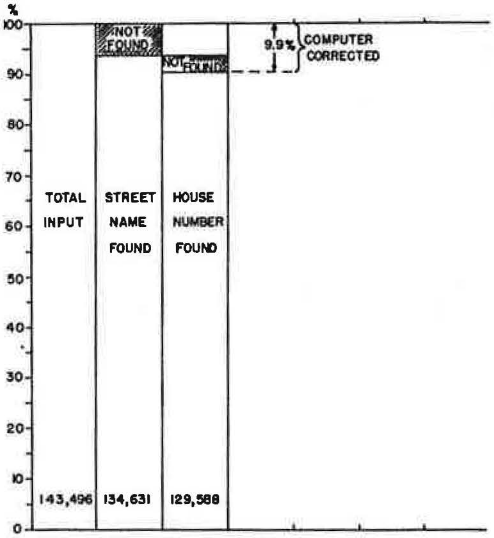


Figure 17. Trip summary records (house number and street name), internal survey.

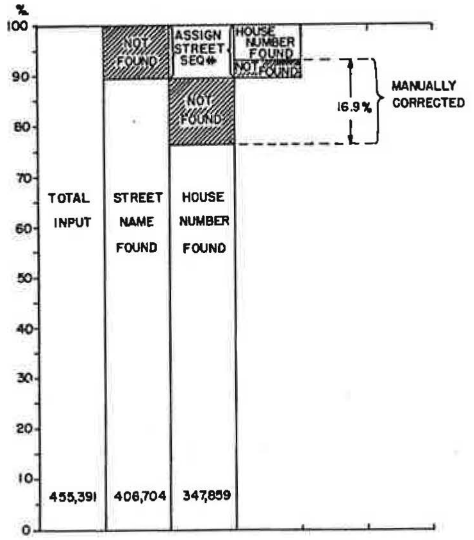


Figure 18. Trip ends to addresses (house number and street name), internal survey.

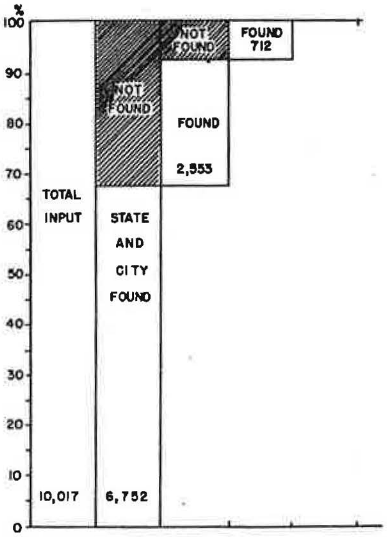


Figure 19. Trip end to external cities, internal survey.

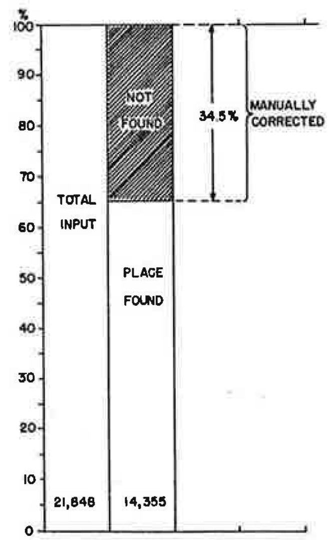


Figure 20. Trip end to place name, internal survey.

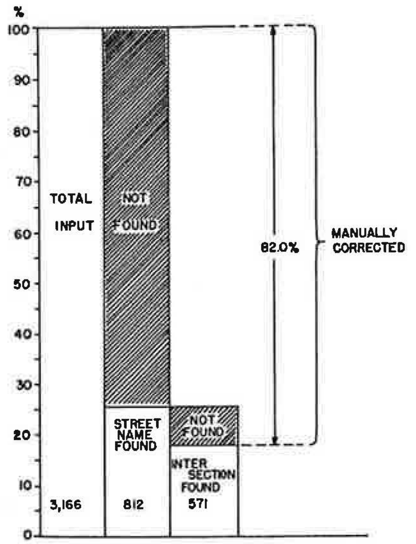


Figure 21. Trip end to intersection, internal survey.

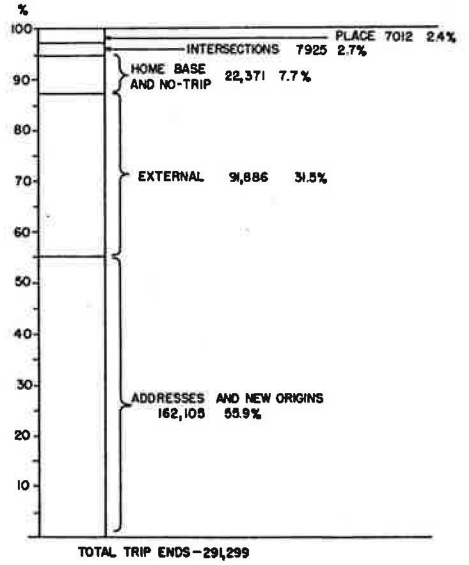


Figure 22. Distribution of trip ends by type, external, truck and taxi surveys combined.

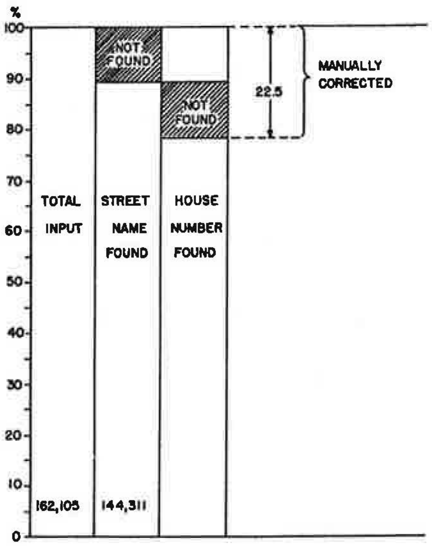


Figure 23. Trip ends to addresses (house number and street name), external, truck and taxi surveys combined.

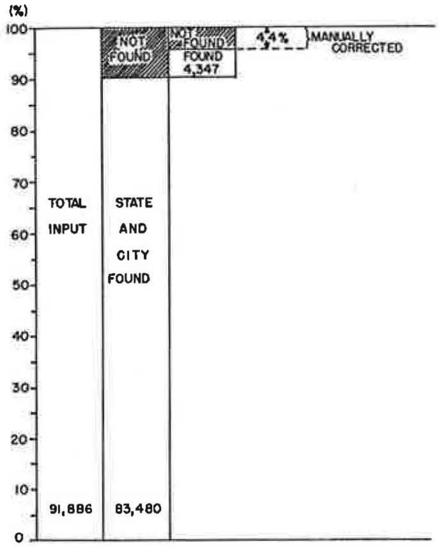


Figure 24. Trip ends to external cities (state and city), external, truck and taxi surveys combined.

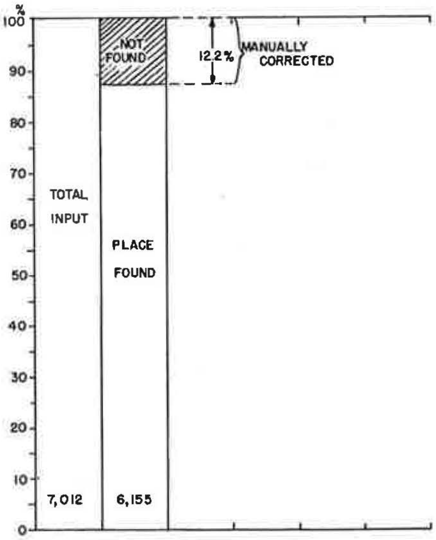


Figure 25. Trip end to place name, external, truck and taxi surveys combined.

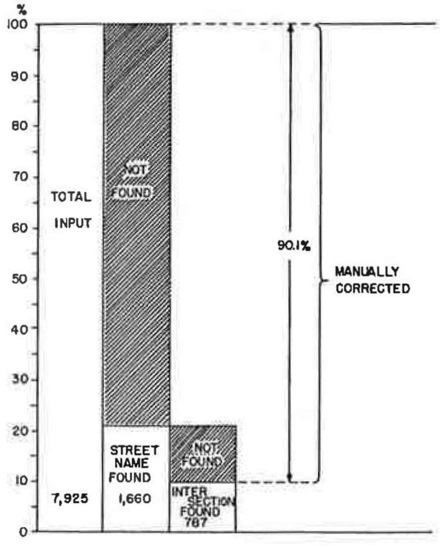


Figure 26. Trip end to intersection, external, truck and taxi surveys combined.

Appendix C

FLOW CHARTS

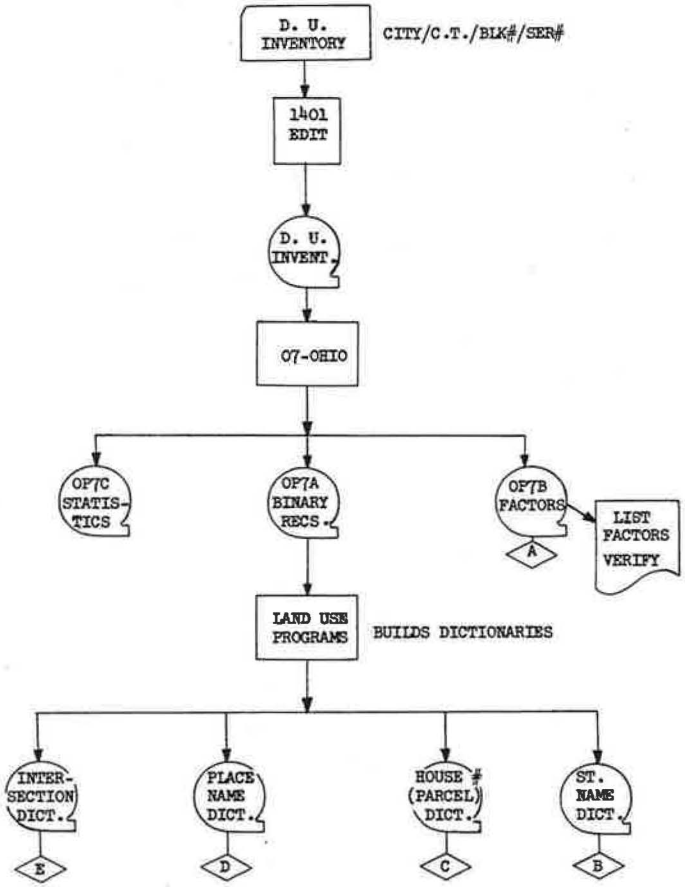


Figure 27.

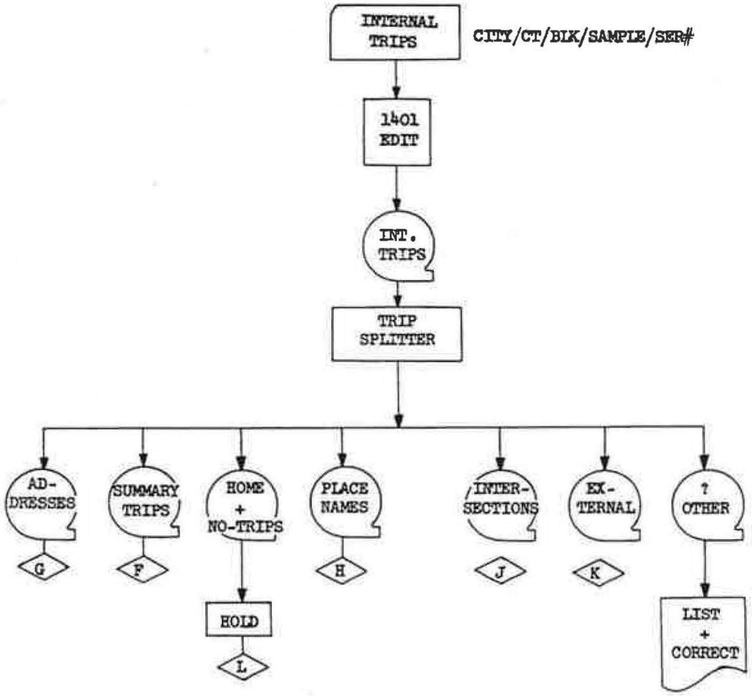


Figure 28.

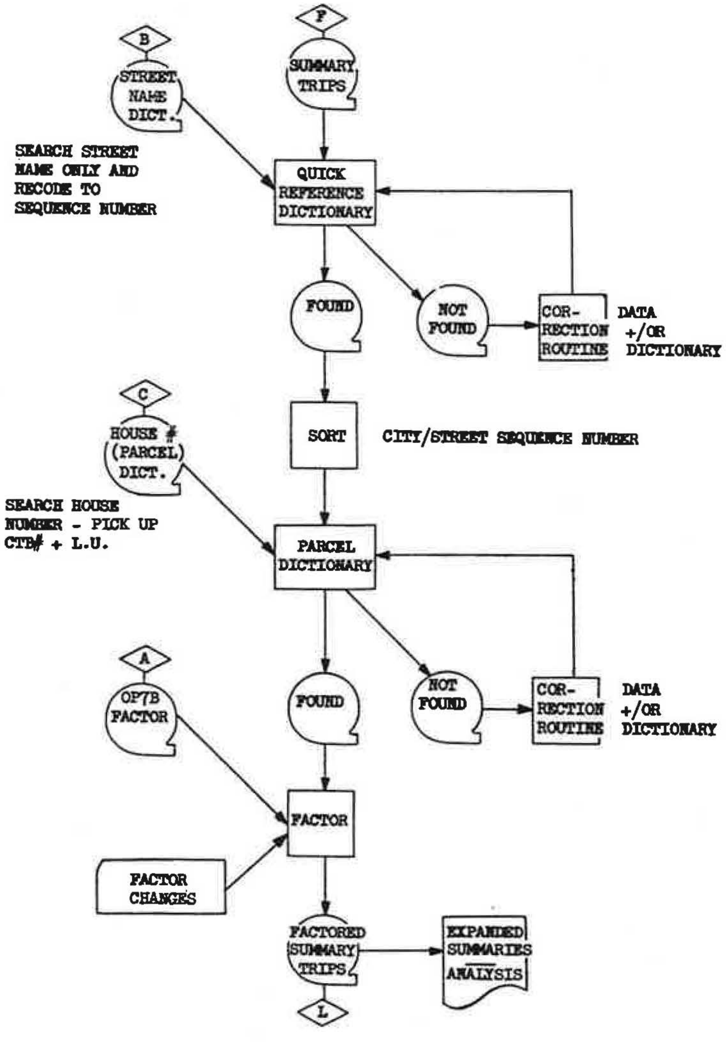


Figure 29.

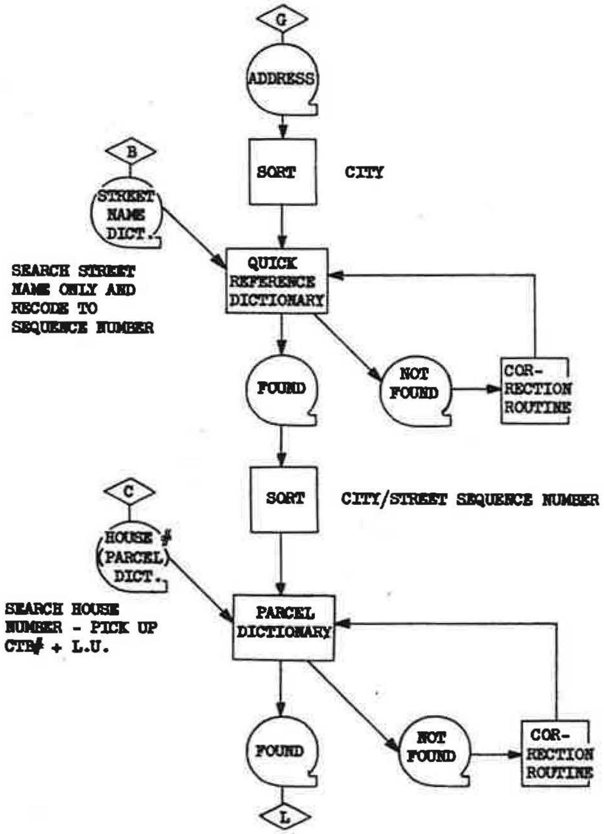


Figure 30.

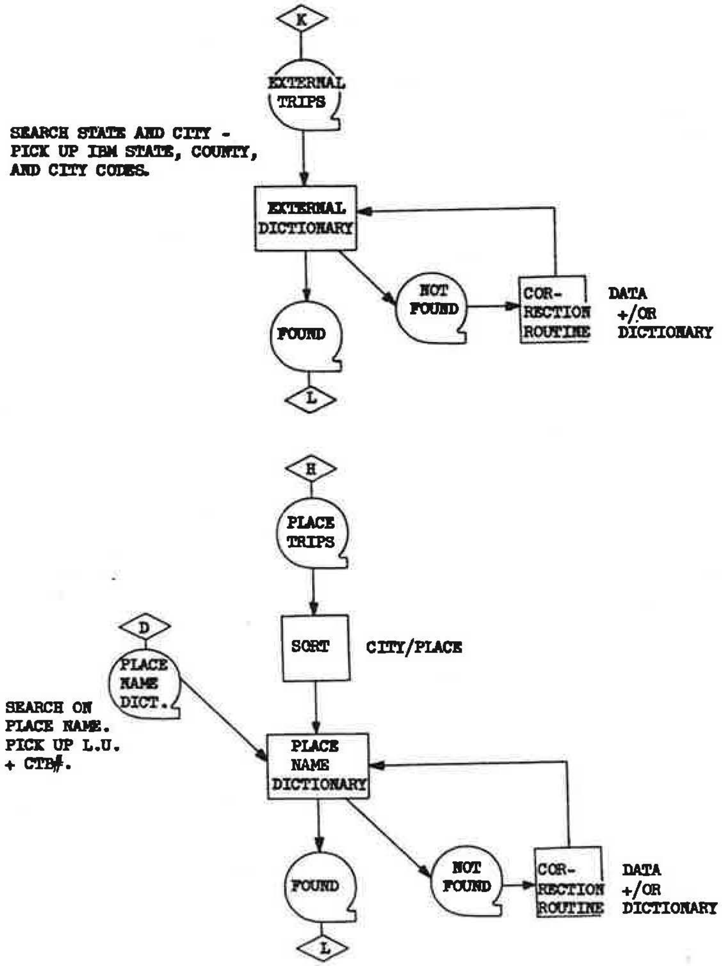


Figure 31.

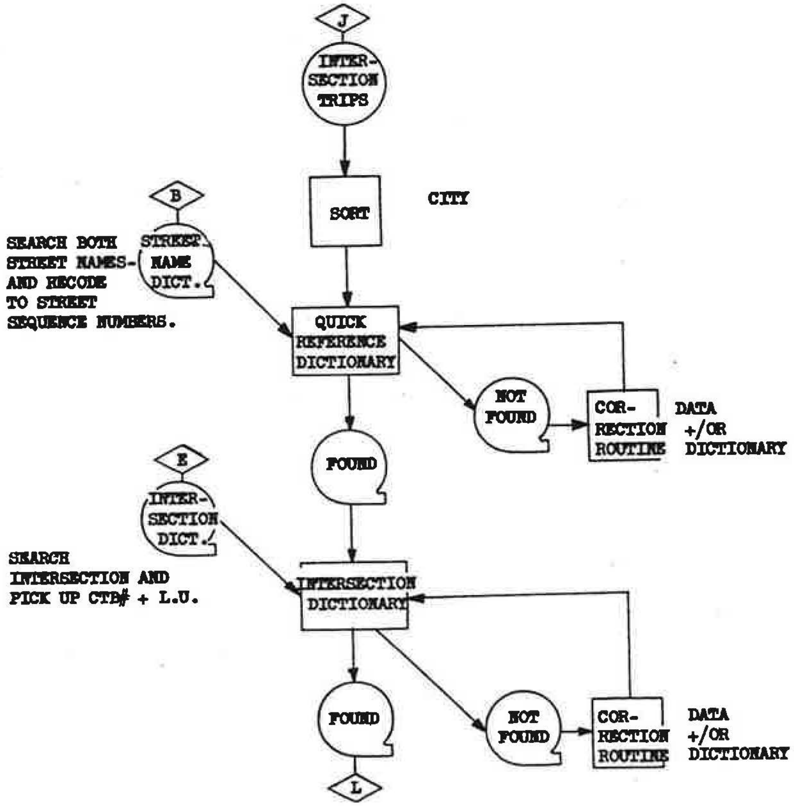


Figure 32.

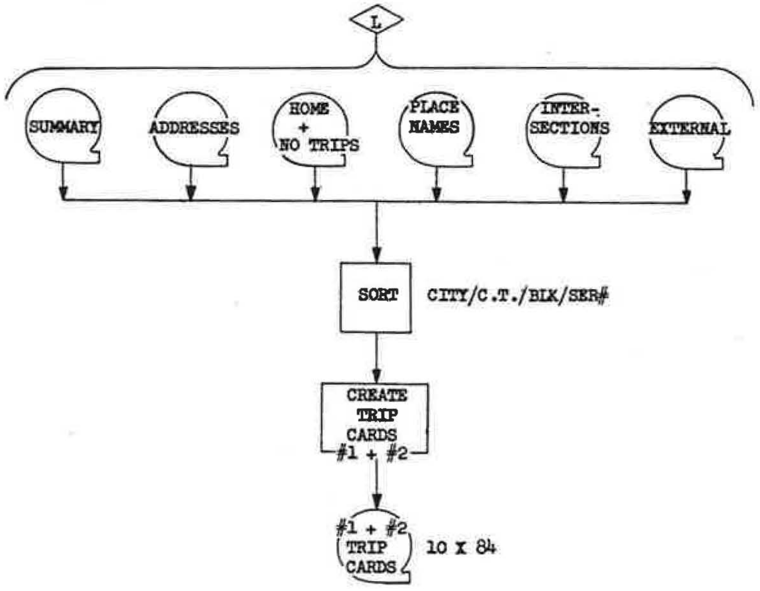


Figure 33.

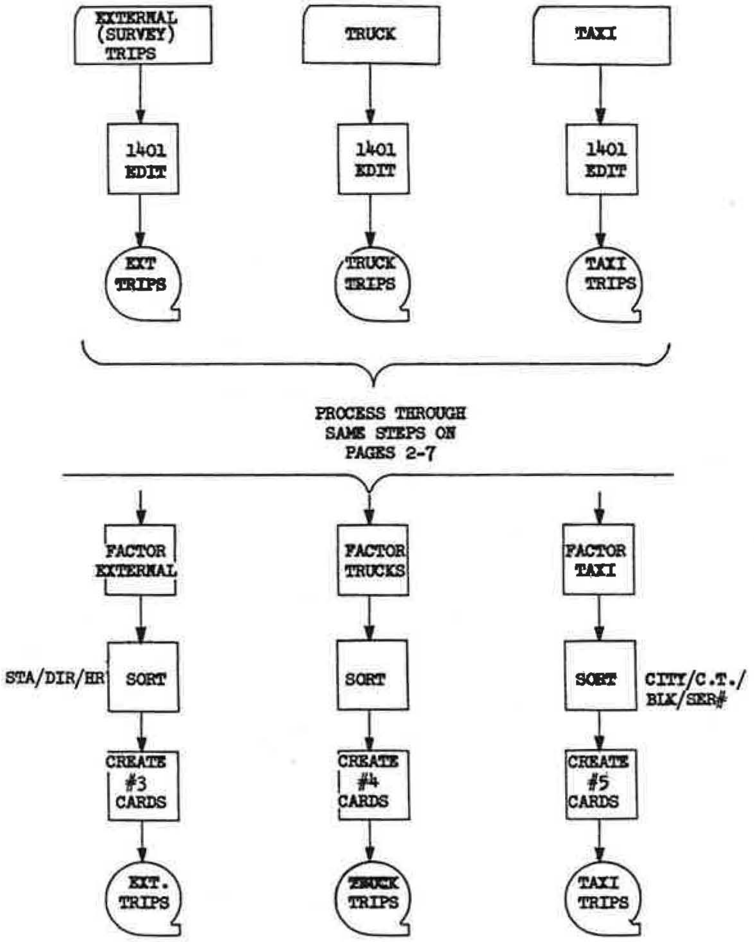


Figure 34.

Discussion

F. E. COLEMAN, Highway Associate Engineer, Connecticut Highway Department—Before discussing Mr. Nucci's paper, I would like to comment in general about this concept. We in Connecticut developed and successfully used a form of this type of automation in 1964. At that time, we were not aware that this was to be a forerunner of a far more detailed program involving the use of the basic methodology by the U. S. Bureau of the Census. I am certain those involved in urban transportation studies at least are aware of the methodology to be employed by the Bureau but for those that aren't I would like to mention a few highlights.

The Census of 1970 is going to be carried out through the use of the mail, in all urbanized areas that have city mail delivery. This does not preclude having a mail-out in some rural areas. In fact, I understand they are considering, for instance, doing the entire State of Connecticut as a mail-out. With this type of survey they will code the statistical information to the block face, which means that census data can be obtained at a much finer detail than in the past. What does this have to do with computer zone coding? Just this: the Census is going to establish master files which will identify each individual block in urban areas and will then match the individual census return to its appropriate block. There are considerably more "goodies" to be obtained than I have alluded to; however, I first wanted to demonstrate the emergence of this type of approach as acceptable and reasonable as well as saving time and costs.

Since the Connecticut Highway Department has used this method extensively for all of its recent surveys for transportation purposes, it is obvious that we feel this is a highly recommendable approach. I think that the work that the Bureau has done, and I assume the Ohio Highway Department, is a further step in hardening and extending the process. However, a few things were not clear to me from the paper and this could be my fault. With this in mind, I felt that to automatically account for some coding, spelling, and interview errors would have added that much more to the process and would eliminate some of the tedious hand-coding. We developed this type of approach for our use.

Again, I am not sure of the reasoning behind the use of the different dictionaries which I assume are on different tapes. It would seem that if they were made one, a time savings would result. Possibly the extensiveness of these dictionaries necessitated this type of approach.

We experienced quite a bit of trouble with the substitution of "street" for "avenue," etc., in the recording of the address. Extensive tests were made to find a way of eliminating the common inaccuracies in recording interviewed trips. We found by reducing the trip end address and the dictionary from 12 positions to 9 positions and then running the rejects from the 12-position match, that 25 percent of the rejects would be zone coded. The computer program was written to eliminate any zone coding errors, for example, if a town had Washington Street and Washington Avenue in two different zones, a 9-position match would not assign a zone number to Washington at all. This then would have to be hand-coded.

I disagree wholeheartedly with the author's conclusion that this type of approach can reflect a larger sample. For instance, in the home interview the coding is not the great cost, it is the field collection, and I feel, especially in urbanized studies, that the data collection can get out of hand and the money spent is not commensurate with the use of the additional data.

All in all, I felt that the author has made a definite contribution.

JOSEPH M. MANNING, Planning Project Director, Massachusetts Department of Public Works—Many have watched with interest this computer coding of O-D trip ends for the Cleveland Study. The results in accuracy, speed and cost reported here clearly show the computer coding approach is superior to manual coding in this case. However, there are significant aspects of this approach which are buried in the paper, due to the multitude of objectives it ambitiously attempts (and achieves on the whole), and an important happening which the paper ignores.

This discussion is not meant to be critical of Mr. Nucci's paper; rather it is an attempt to select a few elements in the paper which I consider especially significant and to discuss them. The first aspect I consider significant is the amount of use or number of trips to be coded that decrees the approach, i.e., computer or manual coding. For example, in the Tri-State (New York) area, where a 1 percent (vs 25 percent in Cleveland) home interview sample was used, a significant majority of the block sides in a computer coding directory would never be used. So the Tri-State Study did not develop area-wide coding guides; rather the coding was done by manual methods.

This resulted in a unit cost of coding a trip end that was higher than Cleveland. But how much higher would it have been if computer coding directories were prepared for about 10 times as many addresses as Cleveland and with only about one-fourth of the trip ends to code? Maybe 40 times the unit cost of the Cleveland Study! There are, then, certain primary questions which have to be asked:

1. How much will it cost to prepare coding directories (a) for manual coding (these may be maps), and (b) for computer coding?
2. How much will it cost to code the O-D data from these directories (a) for manual coding, and (b) for computer coding?
3. Which of these approaches is more economical?

The second significant aspect I would like to discuss is the preparation of the O-D trip end for computer coding. Here is the tie-in between the coding directory and the trip end. The computer demands an exact match between a trip end address and a directory entry. If no match occurs because of address miscoding, the trip has to be manually coded or corrected. If too many mismatches occur, it might have been more economical to manually code the whole O-D survey. This means that the trip end coding of addresses has to be exactly the same format and of very high quality.

Another significant aspect, I feel, is the use of the IBM 7090 computer to perform this coding task. Since the actual coding operation is an item-by-item comparison of O-D addresses vs a coding directory (both in the same sort), then the amount of computer working space needed is quite small.

This being the case, a smaller computer could have been used, even if it might be slightly more expensive on an information-unit-processed basis. Employing a smaller, more popular computer, such as an IBM 1401, would mean that the programs could be much more widely used. So, for the rewriting of the programs, consideration might be given to utilizing a smaller, more popular computer.

The happening, which the paper ignores, is the 1970 U. S. Census. It is to be a mail census with the geographic information coded to individual block faces for urban areas. This block face coding will be done using computer coding directories, almost exactly like those used in the Cleveland Study. This means that coding guides for each urban area will be available by 1970. Since the major cost in computer coding is the compilation of coding directories, this is a significant happening. Of course, it means that any O-D study will only need to be designed with these coding guides in mind and computer coding could well become the rule, rather than the exception.

HUBERT P. NUCCI, Closure—Mr. Coleman mentioned the automatic accounting of errors and comparing on a field reduced from 12 positions to 9. Anything other than an exact match is in the realm of speculation. It would still be subject to human review. This is the area where human judgment is far superior to the computer. There are techniques, such as soundex, which the Bureau of the Census is investigating, but much more research and investigation are required before any conclusions can be drawn as to their utility. This obviously is the next area of intensive investigation.

Mr. Coleman disagreed with my remark that this approach could reflect a larger sample. In many instances we accept a compromise sample size due to cost, time, volume, and processing limitations. If this is the case, the economies from this phase could be reflected elsewhere, such as a larger sample size if it applies.

Mr. Manning mentioned that the majority of block sides were never used. This is true, and I agree with him. But to my knowledge, there is no way of determining beforehand which blocks to include and which ones not to include in the directories.

Mr. Manning then questions the cost for directories ten times as large as those in the Cleveland area and the cost to code O-D trips. The cost to build the directories in Cleveland was of the order of $10,000 for a quarter of a million entries, and it was shown in Figure 14 that as the volume of trips increased, the unit cost declined. For a large area such as New York City, there are techniques to compress enormous files to a manageable size. One method is the removal of duplication, such as in the suburbs, resulting in house number ranges instead of every item of data being present.

Mr. Manning said that if too many mismatches occur, it might be more economical to manually code the whole O-D survey. It was just shown that in Cleveland the error rate was of the order of 10 percent. In the other study areas, the error rate was even lower. The controlling factor is the quality of the search directories. If a high error rate should occur, say as high as 25 percent, remember that the other 75 percent was coded at one-tenth the manual cost.

Mr. Manning questions the choice of computers—the IBM 7090 vs the IBM 1401. True, the 1401 is more popular and readily available, but two factors alone rule out the small computer—speed and cost. Both would be prohibitive on the 1401. Commercially, the 7090 is 7 times more expensive, but internally, it is more than 30 times faster. The greater capacity of the 7090 allowed a very efficient organization of the directories which lent themselves to binary searching.

To sort on the 1401 is inefficient, slow, and costly. Recovery is another factor. Whenever you have a bad run, be it a sort, bad tape, or some other situation, you can reestablish yourself in minimum time on the 7090. The 1401 is an interlocked, unbuffered machine, which means you can only do one operation at a time. The 7090 is buffered, which allows input/output operations of data to be performed while, internally, searching is going on.

When we rewrite this battery of programs for the new generation of computers, it will function on a wide range of models, thereby eliminating this old argument of large computers vs small computers.

Both Mr. Coleman and Mr. Manning mentioned that the 1970 Census will use techniques similar to these to code their census returns. When they become available, they should further enhance these procedures. We have worked very closely with the Bureau of the Census for over three years on details of these problems, and we still have some items to resolve. For example, the primary search argument of the Census directories is postal zip code, which we feel would be difficult to obtain in a survey. Without the zip code in the trip end, it might evolve into a less efficient operation.

Evaluation of Trip Distribution and Calibration Procedures

FRANK E. JAREMA, U. S. Bureau of Public Roads; and
CLYDE E. PYERS and HARRY A. REED, Cleveland Seven County Land Use Transportation Study

This paper discusses a research project designed to evaluate the calibration and testing techniques to two different trip distribution models currently used in urban transportation planning—the intervening opportunities model and the gravity model.

In analyzing the application of calibration procedures for the intervening opportunities model, two approaches which differed basically in the treatment of the L value or probability factor were investigated. The first method entailed the use of single area-wide L values by trip purpose or category for the entire study area. The second method involved the application of L values which varied by analysis area for each trip category.

The calibration of the gravity model is described with primary emphasis on the analysis used to identify areas of bias in the distribution and on the measures applied to effect an accurate calibration.

The results of a comprehensive series of analytical and statistical tests which were applied to each model are reported. Included are tests traditionally used on each of the models over the past several years, thus making possible an evaluation of these tests on a common basis.

•TWO papers concerning the calibrating, testing, and performance of trip distribution procedures were presented at the 1965 meeting of the Highway Research Board (1, 2). The first paper dealt with the calibration and forecasting capability of the intervening opportunities model, while the other was a comparative evaluation of this model with other trip distribution procedures. As a result of this previous research, several questions were raised concerning the methods of calibrating the intervening opportunities model and the testing utilized to check the reasonableness of the calibration. In an attempt to answer these questions, additional research was undertaken in the Urban Planning Division of the Bureau of Public Roads.

This paper expands on the calibration process of the intervening opportunities model by attempting to evaluate the various procedures in use. It also investigates those parameters which are important in terms of calibration and testing. In addition, the various refinements in calibrating and testing a gravity model, utilizing similar data, are described. The results of both models are compared on the basis of a comprehensive series of tests applied to each technique.

STUDY AREA

The data used in this project were obtained in the 1960 Honolulu Metropolitan Area Traffic Survey (3). Honolulu is located on the island of Oahu, known as the Standard Metropolitan Statistical Area of Honolulu. Oahu is one of the eight major islands (Fig. 1) which comprise the State of Hawaii. Although its 595 square miles embrace only 10 percent of Hawaii's land area, Oahu's 1960 population of over 500,000 comprised 80 percent of Hawaii's population.

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.

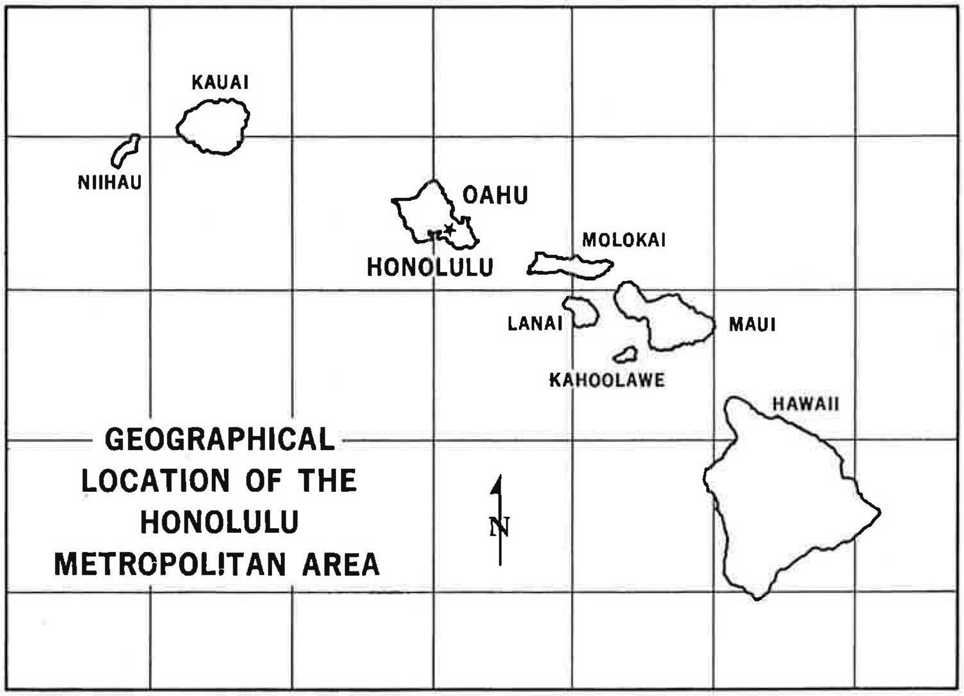


Figure 1. Location of Honolulu study area.

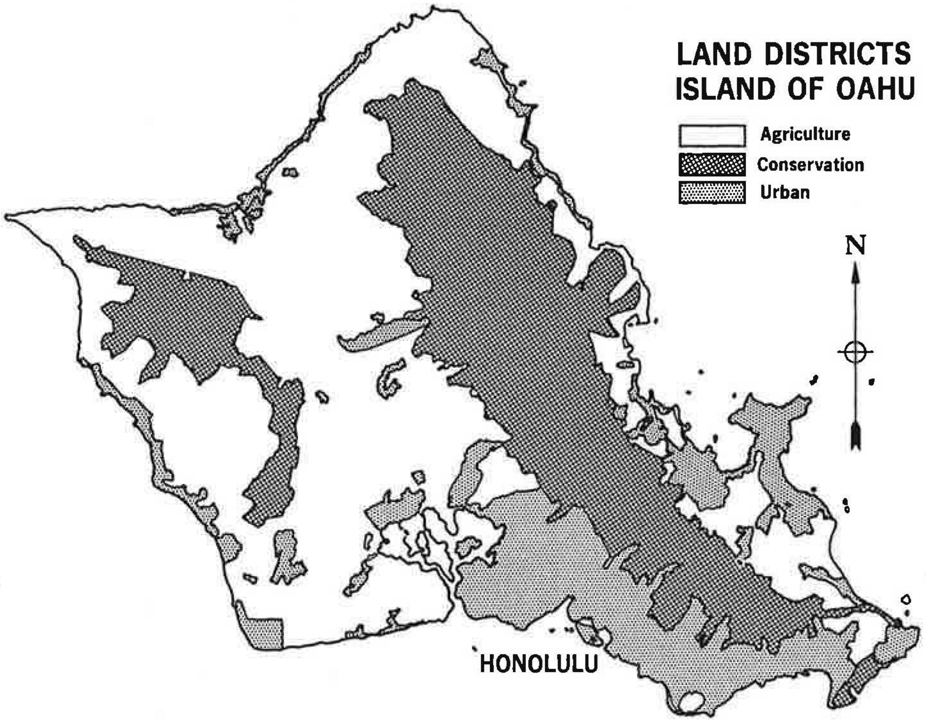


Figure 2. Oahu land districts.

In 1960, nearly 300,000 of the 500,000 residents lived in Honolulu proper, which is built on a narrow strip of land located between the ocean and a series of steeply rising foothills. This topography and related limited accessibility have prevented the city from assuming the growth normal to most American communities and have forced a ribbon development along the southern coast of Oahu. With the exception of scattered developments around the perimeter of Oahu, the remainder of the land is devoted largely to agricultural uses and forest reserves. Figure 2 shows boundaries of zoning districts established by the State Land Use Commission.

The study area included the entire island of Oahu and was considered an "internal area" within which travel data by residents were obtained. The dwelling unit survey was based on a sample of 6.67 percent or one in every 15 dwelling units. To complete the inventory of both people and vehicles, samples of 20 percent of commercial vehicles and 100 percent of taxi cabs were taken. For the research reported in this paper, only the information from the dwelling unit survey was utilized.

Data from the 1960 survey were analyzed on the basis of 159 internal zones. For summary and general analysis purposes, these 159 zones were combined into 13 districts or analysis areas.

INTERVENING OPPORTUNITIES MODEL CALIBRATION

Theory and Formulation

The Intervening Opportunities Trip Distribution Theory was developed in the late 1950's by Morton Schneider for the Chicago Area Transportation Study (CATS). This theory is based on the premise that in urban travel, total travel time from a point is minimized, subject to the condition that every destination has a stated probability of being acceptable if considered. The model states that the probability of a trip originating in one zone finding a destination in another zone is proportional to the possible trip destinations in the other zone and to the number of trip destinations previously considered:

$$T_{ij} = O_i \left[e^{-LD} - e^{-L\,(D\,+\,D_j)} \right]$$

where

T_{ij} = trips originating in zone i and destined for zone j,
O_i = trip origins in zone i,
D = trip destinations considered prior to zone j,
D_j = trip destinations in zone j, and
L = measure of probability that a random destination will satisfy the needs of a particular trip.

Four parameters must be known before T_{ij} can be computed. First is the number of trips originating in a zone (O_i) and second is the number of trips ending in a zone (D_j). The third parameter, travel time, is a measure of the zonal spatial separation. It is used as a means of ranking all zones in descending order from any given zone. The fourth parameter is the L value or probability factor, which is empirically derived and describes the rate of trip decay with increasing trip destinations and increasing trip length.

Trip Stratification

Trips were stratified into categories similar to those utilized in previous research and current operational studies. The trip purposes, shown in Table 1, are defined as follows:

1. Long residential—All home-to-work trips and trips from home outside of the central business district (CBD) to areas in the CBD for any other purpose.

TABLE 1

INTERVENING OPPORTUNITIES MODEL TRIP STRATIFICATION

Trip Purpose	Trip Type			
	Interzonal	Intrazonal	Total	Percent of Total
Long residential	92,498	7,363	99,861	17
Long nonresidential	83,919	6,933	90,852	15
Short	338,108	60,102	398,210	68
Total	514,525	74,398	588,923	100

Note: Total number of trips shown above differs slightly from the gravity model due to a reduction resulting from the use of factoring programs to obtain the three-category intervening opportunities model trip stratification.

2. Long nonresidential—All work-to-home trips and trips for any other purpose which originate in the CBD and are destined to homes outside of the CBD.
3. Short—All other trips not considered as long.

Difference in BPR and Chicago Computer Program Package

At this point, a distinction should be made in the version of the intervening opportunities model distribution developed for CATS (currently being used by the Upstate New York Transportation Studies[1]) and the version used by the Bureau of Public Roads.

The procedure developed by CATS incorporates distribution theory, traffic assignment, and capacity restraint into one computer program. The system as designed selects at random an origin zone, builds a tree for that zone, calculates the trip interchange between that zone and all other zones, and assigns the calculated interchanges to the minimum time paths previously determined. The process is repeated for all zones in the study area. The system also has the option of adjusting the link speeds based on a volume-to-capacity ratio after each origin zone is processed. Therefore, the speed adjustment may have an effect on the distribution of the trips from the succeeding zones. However, research by Saltman (4) showed there was very little effect on the zonal trip distribution due to the application of capacity restraint.

The Bureau of Public Roads system is considerably different. It consists of a stepwise approach wherein the trip distribution, traffic assignment, and capacity restraint are completely independent and do not interact. In this way, the spatial separation between zones does not change during the distribution phase of the process as it does in the CATS system.

Calibration Procedures

Basically, the intervening opportunities model is calibrated by varying the L values until a satisfactory simulation of existing travel patterns is obtained. Two different calibration procedures were investigated. The first method employed the use of a single area-wide L value for each of the three trip purposes. This procedure was similar to that employed in prior production and research (1, 5). The second method of calibrating the intervening opportunities model involved the application of L values which varied by analysis area for each trip purpose. This procedure follows quite closely the methods developed and utilized by the Upstate New York Studies (6).

In contrast to the first method, which involved only one set of L's by trip category for each calibration, the second method utilized 13 sets of L values—one set for each of the analysis areas of the study.

---

[1] Upstate New York Transportation Studies of the Subdivision of Transportation Planning and Programming, New York State Department of Public Works.

The accuracy of the calibrations was determined on the basis of the results of a series of tests applied to the distribution by each method.

Single L Method—Two approaches were examined in developing a single area-wide L value for each purpose. The first approach involved the selection of an area-wide L value by purpose which would result in a reasonable agreement between the actual and estimated average trip lengths in minutes and trip length frequency curves by purpose. The second approach was based on the selection of area-wide L values by purpose resulting in a distribution in which only the total purpose average trip lengths and trip length frequency curves would agree, in addition to estimating the correct number of total intrazonal trips.

While the second approach provided significantly more overall accuracy than the first, neither method proved satisfactory in duplicating travel patterns in Oahu. These results are somewhat in contrast to the findings determined in the Washington, D. C., research project (1), although the calibration procedures utilizing single L values in both the Washington project and this project were similar. However, this finding apparently reflects the thinking by the various production studies using the intervening opportunities model who have essentially discontinued the use of a single set of area-wide L values and have adopted the variable L technique.

Variable L Method—In contrast to the use of single area-wide L values by trip category in the calibration of the intervening opportunities model, the alternate approach analyzed was the use of variable L values. Both the Chicago and Upstate New York studies have recognized the problems associated with the use of single area-wide L values and have since applied the use of variable L's. In addition, both groups have indicated that the variation in L is related to trip-end density.

With increased opportunities or trip density, a trip can afford to be more selective. Thus, a trip originating in a zone with many opportunities has a high probability of being satisfied, whereas a trip originating in a low-density area has few opportunities available and a corresponding lower probability of being satisfied.

An analysis of travel patterns for Oahu indicated, for example, that for the short-trip category, densities ranged from an average of 3,500 trips per square mile within a 100-square-mile area of Honolulu to 500 trips per square mile for the remainder of the island. Therefore, with the use of one set of area-wide L values, trips from the high-density areas were distributed short of their actual trip lengths, while trips originating in low-density areas were sent too far.

This situation is also illustrated by a formula applied in previous research (1) to estimate single area-wide L values prior to the development of the variable L approach:

$$\bar{r} = K \sqrt{\frac{1}{PL}}$$

where

$\bar{r}$ = average trip length in minutes,
K = proportionality constant approximately equal to 2π,
L = probability factor, and
P = density of area expressed in trip-ends per square mile.

For an average trip length of 10 minutes, the L value increases as the trip-end density of a particular zone decreases. On the other hand, the L value decreases as the density increases. Thus, a relatively higher L value is needed in a low-density area to insure that a trip finds a suitable destination, whereas in a high density area, a lower L value is utilized to lower the probability of stopping too soon.

Based on the preceding analysis and unsatisfactory results with the single L approach, variable L values were applied in the calibration of the intervening opportunities model.

Determination of Variable L Values—Using data obtained in the O-D survey, a curve as shown in Figure 3 may be hand-plotted by zone for each of the three trip categories. This curve indicates that as opportunities or destinations are encountered, trip origins

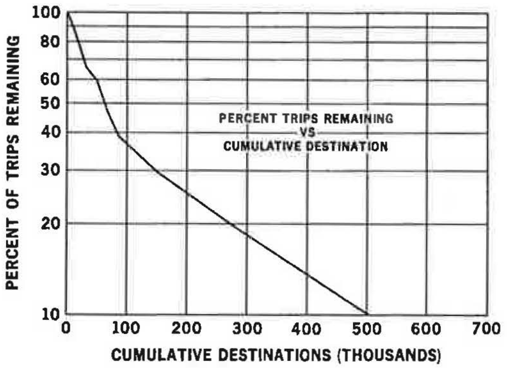


Figure 3. Decay curve.

are satisfied. However, a certain number of origins are selective and consider many destinations before being satisfied. The curve takes the form of a decay-type function with the slope being the L value. To obtain a function which best represents the data, it is necessary to do a least-squares fit of a function with the form e^{-bx}.

A computer program (7) written and utilized by the Upstate New York Studies accomplishes the above manipulation and calculates a least-squares fit for each trip category by zone. The resultant output of this computer program is zonal L values for each trip category. Input data to the program include (a) "skim" trees or minimum time paths for all zones; (b) trip destinations by purpose for each zone; and (c) origin-destination movements for each zone to all other zones by purpose.

Basically, the computer program operates in the following manner. The accumulated destinations ("subtended volume") are calculated for a zone by summing the trip-end input data by destination zone order. In turn, the destination zone order is obtained from the tree records. The cumulative number of trips delivered is calculated by summing the O-D data by destination zone. The percent of trips remaining is calculated by dividing the total number of delivered trips into the number of trips remaining for that particular zone.

Next, the resultant zonal L values by trip category are weighted to obtain district L values by trip category for each of the 13 districts or analysis areas. The following procedure was used in computing district L values:

$$L_{ds} = \frac{\Sigma\ (L_{zs} \times D_{zs})}{D_{ds}}$$

where

L_{ds} = district short L,
L_{zs} = zonal short L,
D_{zs} = zonal short trip destinations, and
D_{ds} = total district short trip destinations.

$$L_{dl} = \frac{\Sigma\ [(L_{znr} \times D_{zr}) + (L_{zr} \times D_{znr})]}{D_{dr} + D_{dnr}}$$

where

L_{dl} = district long L,
L_{znr} = zonal nonresidential L,
L_{zr} = zonal residential L,
D_{dr} = total district long residential trip destinations,
D_{dnr} = total district nonresidential trip destinations, and
D_{zr}, D_{znr} = zonal residential and nonresidential destinations.

Calibration Technique—The calibration technique utilized in this project is the one employed by the Upstate New York Studies. Essentially, the accuracy of the calibration is determined on the basis of a series of vehicle-miles of travel comparisons between the O-D and model data. Vehicle-miles of travel resulting from an assignment of the estimated interchanges are summarized by district and highway facility type. A similar summary is determined for the VMT resulting from an assignment of actual or O-D

TABLE 2

LEVELS OF VEHICLE-MILES OF TRAVEL COMPARISONS

District	Facility Type			
	Primary	Secondary	Freeway	Total
A	4	4	4	3
B	4	4	4	3
C	4	4	4	3
Z	4	4	4	3
Total	2	2	2	1

Numbers refer to levels of vehicle-miles of travel comparisons made for both the survey and model total purpose trips as indicated:

1—Total overall vehicle-miles of travel
2—Total vehicle-miles of travel by facility type
3—Total vehicle-miles of travel by district
4—Vehicle-miles of travel by facility type within each district.

TABLE 3

SUMMARY COMPARISON, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Trip Purpose	Number of Trips	Average Trip Length (min)		Vehicle Hours of Travel		Intrazonal Trips	
		O-D	Model	O-D	Model	O-D	Model
Long residential	99, 861	19. 46	20. 96	32, 390	34, 883	7, 363	2, 698
Long nonresidential	90, 852	19. 12	18. 45	28, 957	27, 930	6, 933	2, 626
Short	398, 210	13. 34	12. 90	88, 503	85, 629	60, 102	48, 115
Total	588, 923	15. 27	15. 12	149, 851	148, 443	74, 398	53, 439

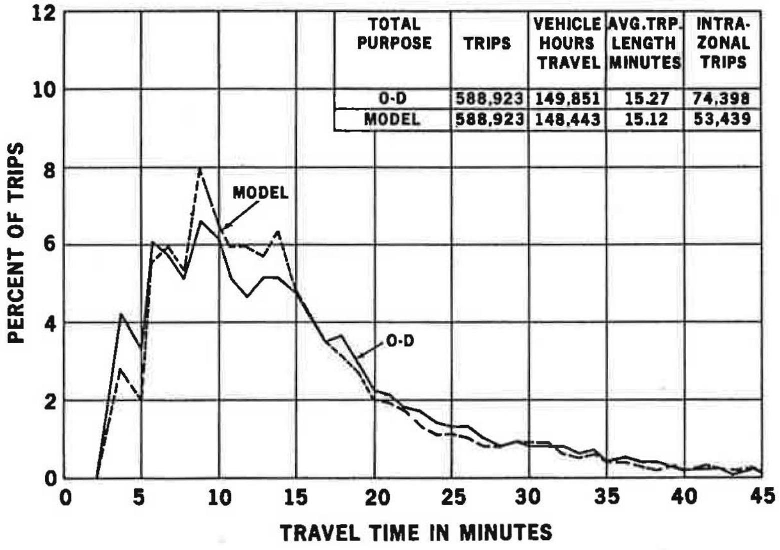


Figure 4. Comparison of trip length distribution, total purpose trips, O-D vs intervening opportunities model (final calibration).

TABLE 4

FREQUENCY DISTRIBUTION OF TOTAL PURPOSE DISTRICT MOVEMENTS, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	65	240	199	224	93
500-999	17	737	840	668	91
1, 000-1, 999	17	1, 615	2, 113	1, 259	78
2, 000-2, 999	17	2, 453	2, 068	1, 056	43
3, 000-3, 999	7	3, 463	3, 440	1, 883	54
4, 000-4, 999	10	4, 400	4, 174	3, 076	70
5, 000-5, 999	10	5, 499	5, 772	2, 271	41
6, 000-7, 999	9	6, 534	6, 011	2, 277	35
8, 000-9, 999	5	9, 018	9, 272	1, 930	21
10, 000-14, 999	7	11, 715	13, 854	4, 539	39
15, 000-73, 999	5	36, 507	33, 937	3, 950	11

TABLE 5

COMPARISON OF TOTAL TRIPS CROSSING SELECTED SCREENLINES, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Screenline	O-D	Model	Percent Difference
Kalihi	78, 532	78, 696	0
Kapolomo	98, 940	85, 776	-13
Nuuahu	122, 456	96, 310	-21
Mondo Pololo	122, 292	124, 520	+2
Total	422, 220	385, 302	-9

TABLE 6

COMPARISON OF TOTAL TRIPS CROSSING KOOLAU MOUNTAIN RANGE, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Crossing	O-D	Model	Percent Difference
Southern	2, 924	3, 492	+19
Lower central	11, 140	13, 026	+17
Upper central	14, 864	12, 084	-19
Northern	1, 012	708	-30
Total	29, 940	29, 320	-2

TABLE 7

FREQUENCY DISTRIBUTION OF TOTAL PURPOSE TRIPS ASSIGNED TO SPIDER NETWORK, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	275	201	235	143	71
500-999	165	726	806	299	41
1, 000-1, 999	185	1, 508	1, 638	601	40
2, 000-2, 999	77	2, 404	2, 531	593	25
3, 000-3, 999	60	3, 460	3, 782	1, 204	35
4, 000-4, 999	34	4, 379	4, 573	883	20
5, 000-5, 999	36	5, 426	5, 934	1, 316	24
6, 000-6, 999	24	6, 581	6, 482	1, 269	19
7, 000-7, 999	30	7, 519	7, 937	1, 599	21
8, 000-8, 999	25	8, 523	7, 651	1, 754	21
9, 000-10, 999	31	9, 927	8, 843	2, 532	26
11, 000-12, 999	18	11, 850	10, 305	2, 918	25
13, 000-14, 999	22	13, 962	12, 622	2, 617	19
15, 000-19, 999	26	17, 559	15, 045	4, 500	26
20, 000-29, 999	23	24, 517	23, 267	4, 404	18
30, 000 and over	20	35, 534	30, 663	8, 253	23

TABLE 8

FREQUENCY DISTRIBUTION OF TOTAL PURPOSE TRIPS ASSIGNED TO HIGHWAY NETWORK, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	450	251	304	211	84
500-999	365	745	840	272	37
1,000-1,999	521	1,455	1,610	913	63
2,000-2,999	372	2,420	2,510	667	28
3,000-3,999	203	3,520	3,510	870	25
4,000-4,999	125	4,450	4,530	886	20
5,000-5,999	93	5,430	5,680	1,163	21
6,000-6,999	101	6,450	6,250	1,285	20
7,000-7,999	69	7,500	7,630	1,517	20
8,000-8,999	51	8,480	7,750	1,412	17
9,000-10,999	79	9,850	9,630	1,917	20
11,000-12,999	74	11,750	10,950	1,982	17
13,000-14,999	43	13,650	12,450	2,136	16
15,000-19,999	64	17,450	15,600	3,127	18
20,000-29,999	54	24,800	19,000	8,882	36
30,000 and over	3	31,000	28,100	5,814	19

interchanges. Using these summaries, various levels of comparison between O-D and model VMT are made as shown in Table 2.

Both categories of L values are adjusted by district until the resultant VMT comparisons indicate a predetermined level of accuracy. The procedure shown below is applied to obtain adjusted L values for each additional distribution by the model. However, after each distribution using the adjusted district L values, the various VMT comparisons are repeated. For the short L,

$$L_2 = L_1 \left(\frac{\text{VMT model}}{\text{VMT O-D}} \right)^2$$

TABLE 9

COMPARISON OF VEHICLE-MILES OF TRAVEL BY DISTRICT, O-D VS VARIABLE L INTERVENING OPPORTUNITIES MODEL

District	Assignment of Total Purpose O-D Trips	Assignment of Total Purpose Model Trips	Percent Difference
1	123,890	97,990	-21
2	178,431	150,363	-19
3	119,136	118,940	0
4	205,154	205,238	0
5	269,564	261,394	-3
6	407,250	409,305	+1
7	228,460	250,492	+10
8	51,291	55,295	+8
9	96,237	113,661	+18
10	53,019	58,633	+11
11	271,625	300,737	+11
12	486,053	495,910	+2
13	193,709	177,307	-9
Total	2,683,819	2,695,265	+0.4

where

L_2 = calculated district short L for second calibration,
L_1 = district short L used in calibration 1,
VMT model = district VMT resulting from assignment of estimated short trips, and
VMT O-D = district VMT resulting from assignment of actual short trips.

For the long L,

$$L_2 = L_1\left(\frac{\text{VMT model}}{\text{VMT O-D}}\right)^2$$

where

L_2 = calculated district long L for second calibration,
L_1 = district long L used in calibration 1,
VMT model = district VMT resulting from assignment of estimated long trips, and
VMT O-D = district VMT resulting from assignment of actual long trips.

Utilizing the initial set of weighted district L values, the estimated trip interchanges resulting from the first distribution were assigned to the network. Based on data from the assignment, a computer program (8) prepared by the Oahu Transportation Study was used to summarize VMT for short and long trips by facility type within each of the 13 districts. This estimated summary of VMT resulting from the model was compared with a similar tabulation actual VMT. Based on this comparison, an adjustment of L values was made by substituting the appropriate values of estimated and desired VMT into the equation described in the previous section. Thus, for those districts in which the model VMT was overestimated, the L value was adjusted upward. This increased the probability of a trip orgin finding a satisfactory destination and, in effect, resulted in a lower estimated VMT for that district. On the other hand, for those districts where VMT was underestimated, the adjustment resulted in a lower L value and a corresponding higher VMT in the next calibration.

For the Oahu area, two adjustments of L values were made as described. It was also found necessary to balance destinations similar to prior research (1) to insure that approximately the correct number of trips were received by zone. An additional distribution using the second set of adjusted L values and balanced destinations was run. Analysis of this final distribution indicated that a satisfactory estimate of travel patterns was simulated. The various tests used in this analysis and the respective results are described in the next section.

Results—One of the primary tests applied to check the accuracy of the calibration of the model by its users is the VMT comparison—that is, how good the estimated VMT is when compared to the VMT from an O-D assignment.

Additional tests utilized in the BPR package were also applied to further analyze the accuracy of the model calibration. These tests included the following:

1. Comparison of average trip lengths and vehicle hours of travel (Table 3).
2. Comparison of trip length frequency distribution (Fig. 4).
3. Statistical comparison of district-to-district movements (Table 4).
4. Comparison of trips crossing selected screenlines (Table 5).
5. Comparison of trips crossing Koolau Mountain Range (Table 6).
6. Statistical comparison of loadings on a spider network (Table 7).
7. Statistical comparison of loadings on the highway network (Table 8).
8. Actual and estimated vehicle-miles travel comparison (Table 9).

GRAVITY MODEL CALIBRATION

Summary of Calibration Procedure

In addition to the intervening opportunities model, a second distribution technique, the gravity model, was also calibrated for Oahu. This was a six-purpose model which was calibrated using standard procedures (9).

TABLE 10

GRAVITY MODEL TRIP STRATIFICATION

Trip Purpose	Trip Type			
	Interzonal	Intrazonal	Total	Percent of Total
Military work	38,704	9,605	48,309	8
Civilian work	114,373	4,694	119,067	20
Shop	74,147	14,741	88,888	15
Social-recreation	93,382	11,633	105,015	18
Miscellaneous	78,933	13,658	92,591	16
Nonhome-based	116,312	20,068	136,380	23
Total	515,851	74,399	590,250	100

The development of trip tables for the gravity model basically involved two types of trips, home-based and nonhome-based trips. These were stratified into six purposes as shown in Table 10. In Oahu, circumstances dictated the use of two separated models in the distribution of home-based work trips. This decision is based primarily on two factors:

1. Military work trips (those trips between a person's place of residence and his place of employment for the purpose of military work) accounted for 29 percent of total work trips; and

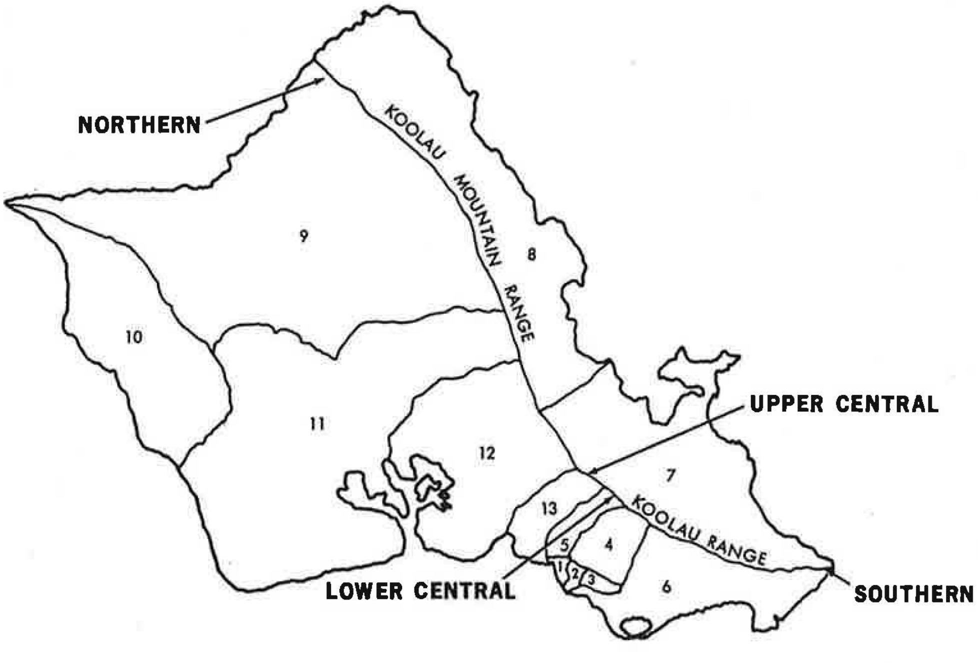


Figure 5. Location of Koolau Mountain Range crossings.

TABLE 11

SUMMARY INFORMATION, GRAVITY MODEL (CALIBRATION 8)

Trip Purpose	Trips	Average Trip Length (min)			Vehicle-Hours of Travel		
		O-D	Model	Percent Difference	O-D	Model	Percent Difference
Military	48, 316	18. 96	18. 96	0. 0	15, 266	15, 268	0. 0
Civilian work	119, 141	19. 73	19. 82	0. 4	39, 179	39, 353	0. 4
Shop	88, 940	11. 35	11. 48	1. 1	16, 819	17, 022	1. 2
Social-recreation	105, 146	15. 58	15. 98	2. 5	27, 301	28, 010	2. 6
Miscellaneous	92, 656	14. 18	14. 00	1. 2	21, 894	21, 627	1. 2
Nonhome-based	136, 537	13. 36	12. 97	2. 9	30, 344	29, 522	2. 7

2. Gravity model procedures do not differentiate between types of work installations except when treated as separate-purpose models. Because of the size and location of military installations with respect to other employment areas, a single work purpose model would result in an inadequate distribution.

Upon completion of the preparation of basis input data, calibration of the model was initiated. Travel time factors were adjusted and the gravity model trip distribution was rerun until the average trip lengths by purpose were within an acceptable level of the O-D trip lengths and the resulting trip length frequency curves compared favorably with the O-D curves. For Oahu, four calibrations were necessary to achieve this criterion.

Tests applied to the output at this stage of the calibration process indicated that a bias existed and that the model was overestimating movements across the mountain range and movements to the CBD. An effort was made to improve this situation by introducing time penalties to the facilities crossing the Koolau Range (Fig. 5).

Two additional runs were made using time penalties of 4-4-4-0 and 5-6-6-2 minutes on the southern, lower central, upper central and northern crossings, respectively. Extensive checking of the latter calibration indicated the model was estimating movements across the mountains satisfactorily. However, several district movements estimated by the gravity model indicated an excessive error.

TABLE 12

FREQUENCY DISTRIBUTION OF TOTAL PURPOSE DISTRICT MOVEMENTS, O-D VS GRAVITY MODEL (CALIBRATION 8)

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	64	182	215	105	51
500-999	19	720	640	227	34
1, 000-1, 999	19	1, 428	1, 427	320	37
2, 000-2, 999	15	2, 551	2, 473	743	28
3, 000-3, 999	9	3, 586	3, 522	905	21
4, 000-4, 999	12	4, 564	4, 471	679	30
5, 000-5, 999	6	5, 597	5, 411	1, 003	16
6, 000-7, 999	5	6, 889	5, 934	1, 614	18
8, 000-9, 999	5	8, 813	8, 788	925	14
10, 000-14, 999	9	11, 165	11, 260	1, 759	14
15, 000 and over	6	31, 217	32, 467	3, 372	10

TABLE 13

COMPARISON OF TOTAL TRIPS CROSSING SELECTED SCREENLINES, O-D VS GRAVITY MODEL (CALIBRATION 8)

Screenline	O-D	Model	Percent Difference
Kalihi	78,522	85,188	+8.5
Kapolomo	99,102	95,824	-3.3
Nuuahu	123,600	118,012	-4.5
Mondo Pololo	124,828	125,540	+0.7
Total	426,052	424,564	-0.3

Examination of the causes of this error led to a general belief that the use of the basic network, i.e., before any capacity restraint was applied, would have improved the situation. This determination was made upon analyzing the changes in network speeds brought about by capacity restraints in the outlying portions of the island vs those changes in the portion around Honolulu. The speed adjustments resulting from capacity restraint appeared reasonable on an area-wide basis. However, in some corridors having low counts but high capacity values coded on links (4 to 1 ratio), an 8 to 10 percent increase of speeds occurred.

In addition, the number of intrazonal trips estimated by the model was compared to the actual intrazonal trips. This analysis indicated that the district movements in question would have been substantially improved by the correct estimation of intrazonal trips.

To test the effect of the unrestrained network with the distribution model, another calibration was run. The results indicated that there was an overall improvement in the distribution patterns with the use of the unrestrained network. It was also evident that some form of time penalty would still be needed on the Koolau Range crossings.

Based on the experience of previous calibration runs, it was anticipated that all of the significant adjustments could be incorporated into a final model. These included (a) zonal time separations based on the unrestrained network; (b) time barriers across the Koolau Range of 5-6-6-6 minutes, respectively, on the crossings from north to south; (c) adjusted travel time factors based on the initial distribution using the unrestrained network; and (d) adjustments in 33 intrazonal travel times. Since the model was not simulating intrazonal trips satisfactorily, selected intrazonal travel times (intrazonal driving time plus terminal time) were adjusted to provide a reasonable estimate of these trip movements.

Results

An extensive series of tests applied to the output of this calibration indicated that a simulation of travel patterns was accomplished within an acceptable degree of accuracy. The final calibration was checked by the following analysis:

1. The average trip length in all six models was within 3 percent of the related O-D average trip lengths.
2. Vehicle hours of travel in all six models was within 3 percent of the O-D hours of travel.

TABLE 14

COMPARISON OF TOTAL TRIPS CROSSING KOOLAU MOUNTAIN RANGE, O-D VS GRAVITY MODEL (CALIBRATION 8)

Crossing	O-D	Model	Percent Difference	Time Barrier (min)
Southern	2,936	2,864	-2.4	6
Lower central	10,956	10,916	-0.3	6
Upper central	14,972	17,884	+19.4	6
Northern	1,024	528	-48.4	5
Total	29,888	32,192	+7.7	

TABLE 15

FREQUENCY OF TOTAL PURPOSE TRIPS ASSIGNED TO SPIDER NETWORK, O-D VS GRAVITY MODEL (CALIBRATION 8)

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	317	218	241	130	60
500-999	159	741	751	216	31
1,000-1,999	156	1,441	1,592	409	29
2,000-2,999	94	2,429	2,548	479	20
3,000-3,999	50	3,388	3,738	696	21
4,000-4,999	31	4,448	4,509	1,110	24
5,000-5,999	27	5,455	5,519	671	13
6,000-6,999	26	6,498	6,735	915	14
7,000-7,999	20	7,378	7,719	1,161	17
8,000-8,999	24	8,481	8,763	1,234	16
9,000-10,999	27	9,855	9,612	1,141	11
11,000-12,999	23	11,908	12,345	1,416	13
13,000-14,999	14	13,984	13,886	1,562	10
15,000-19,999	21	16,445	16,321	1,423	10
20,000-29,999	23	23,301	22,075	1,905	8
30,000 and over	26	39,998	37,478	3,176	8

3. Estimated zonal interchanges were compressed into district-to-district movements and compared with similar information from the O-D data. A statistical analysis of this comparison indicated a reasonable model distribution.

4. Four screenline comparisons of assigned O-D and gravity model trips showed total assigned gravity model trips to be 0.3 percent less than the related assigned O-D trips. The individual screenline comparisons ranged from +8.5 percent to -4.5 percent.

5. The gravity model distribution across the Koolau Mountain Range for all four crossings was within +7.7 percent of the related O-D trips.

TABLE 16

FREQUENCY OF TOTAL PURPOSE TRIPS ASSIGNED TO HIGHWAY NETWORK, O-D VS GRAVITY MODEL (CALIBRATION 8)

Volume Group	Frequency	Mean Volume		Root-Mean-Square Error	Percent RMSE
		O-D	Model		
0-499	488	236	274	128	54
500-999	429	741	762	222	30
1,000-1,999	505	1,466	1,490	304	21
2,000-2,999	306	2,442	2,442	308	16
3,000-3,999	176	3,452	3,416	503	15
4,000-4,999	128	4,462	4,513	561	18
5,000-5,999	118	5,526	5,647	835	15
6,000-6,999	81	6,413	6,571	927	14
7,000-7,999	55	7,563	7,683	1,022	14
8,000-8,999	41	8,464	8,396	935	11
9,000-10,999	61	10,041	9,865	1,193	12
11,000-12,999	47	12,108	11,923	1,454	12
13,000-14,999	44	13,989	13,989	1,734	12
15,000-19,999	84	17,102	16,304	2,533	12
20,000-29,999	48	24,482	24,066	1,757	7
30,000 and over	12	34,524	32,532	2,364	7

TABLE 17

COMPARISON OF VEHICLE-MILES OF TRAVEL BY DISTRICT, O-D VS GRAVITY MODEL (Calibration 8)

District	Assignment of Total Purpose O-D Trips	Assignment of Total Purpose Model Trips	Percent Difference
1	123,434	114,098	-7.5
2	173,482	159,267	-8.1
3	120,750	117,763	-2.4
4	216,688	210,770	-2.7
5	269,791	240,168	-10.9
6	386,390	382,804	-0.9
7	227,892	242,257	+6.3
8	51,891	45,874	-11.5
9	95,901	91,875	-4.1
10	53,569	49,044	-8.4
11	270,866	388,322	+6.6
12	483,361	523,533	+8.3
13	192,722	215,680	+11.9
Total	2,666,737	2,681,455	+0.5

6. Both final model and O-D total purpose trips were assigned to a spider network and the results compared statistically. This test also indicated a reasonable distribution of trips by the model.

7. In addition, final model and O-D total purpose trips were assigned to the entire traffic assignment network. A statistical analysis of link loadings by volume groups appeared to be satisfactory.

8. A final check was the comparison of VMT resulting from the assignment of total purpose model and O-D trips. This check was made for total VMT, total VMT by highway classification, total VMT by district and VMT within district by highway classification. This check also indicated reasonable results.

The results of the final model are shown in detail in the following references: (a) comparison of average trip lengths and vehicle hours of travel, Table 11; (b) trip length frequency distribution by purposes, Figures 6-11; (c) statistical comparison of district-to-district movements, Table 12; (d) comparison of trips crossing selected screenlines, Table 13; (e) comparison of trips crossing the Koolau Mountain Range, Table 14; (f) statistical comparison of loadings on a spider network, Table 15; (g) statistical comparison of loadings on the highway network, Table 16; and (h) actual and estimated vehicle-miles of travel comparison, Table 17.

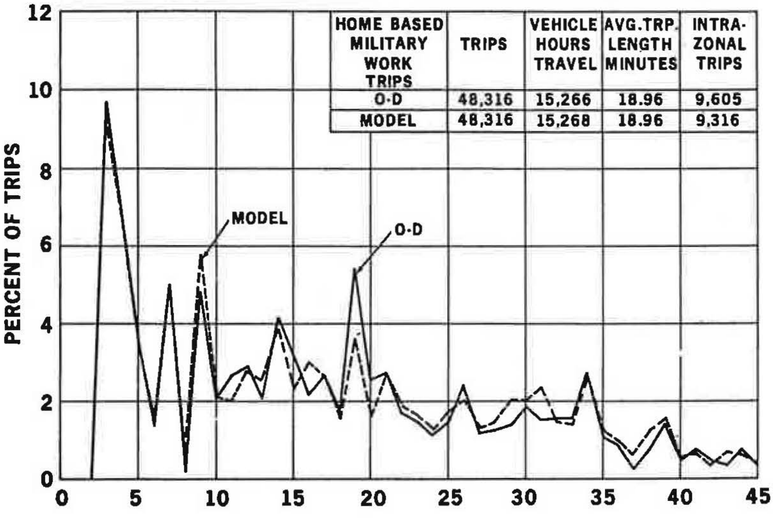
COMPARISON OF RESULTS

A comparison of the results of several analytical and statistical tests applied to the final calibration of the intervening opportunities model and the gravity model is shown

TABLE 18

SUMMARY COMPARISONS, TOTAL PURPOSE TRIPS, SURVEY AND MODELS (FINAL CALIBRATION)

Parameter	Gravity Model			Intervening Opportunities Model		
	Survey	Model	Percent Difference	Survey	Model	Percent Difference
Average trip length (min)	15.32	15.32	0	15.27	15.12	-1
Vehicle hours of travel	150,805	150,804	0	149,851	148,443	-1
Intrazonal trips	74,398	68,491	-8	74,398	53,439	-28



HOME BASED MILITARY WORK TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	48,316	15,266	18.96	9,605
MODEL	48,316	15,268	18.96	9,316

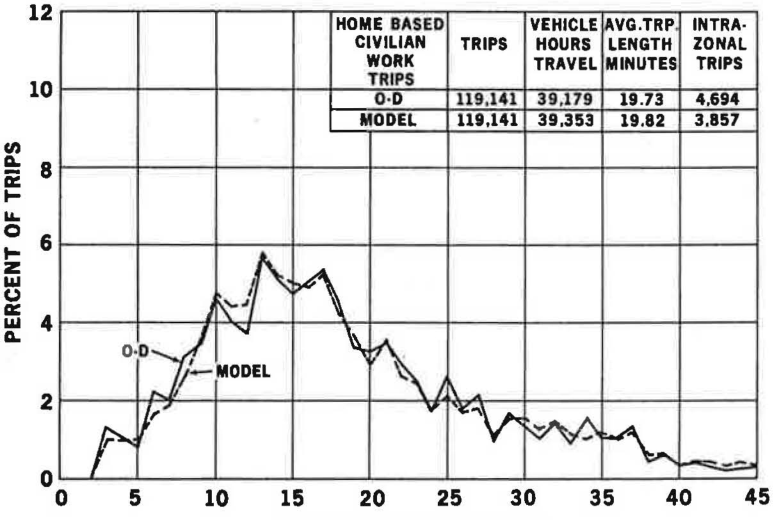
Figure 6. Comparison of trip length distribution, home-based military work trips, O-D vs gravity model (calibration 8).

in Table 18 and Figures 12-17. Although detailed results relating to each of the two models have been shown earlier, the summary comparisons are presented mainly to indicate the relative performance of each model.

EVALUATION OF PROCEDURES

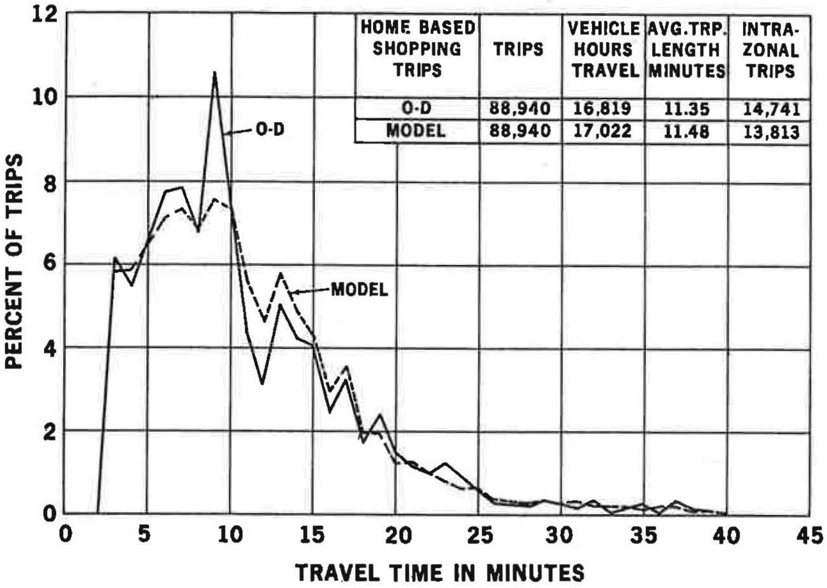
Trip Distribution Techniques

Probably two of the most significant questions about trip distribution techniques are "How accurately do they perform?" and "How stable are the parameters of the model?"



HOME BASED CIVILIAN WORK TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	119,141	39,179	19.73	4,694
MODEL	119,141	39,353	19.82	3,857

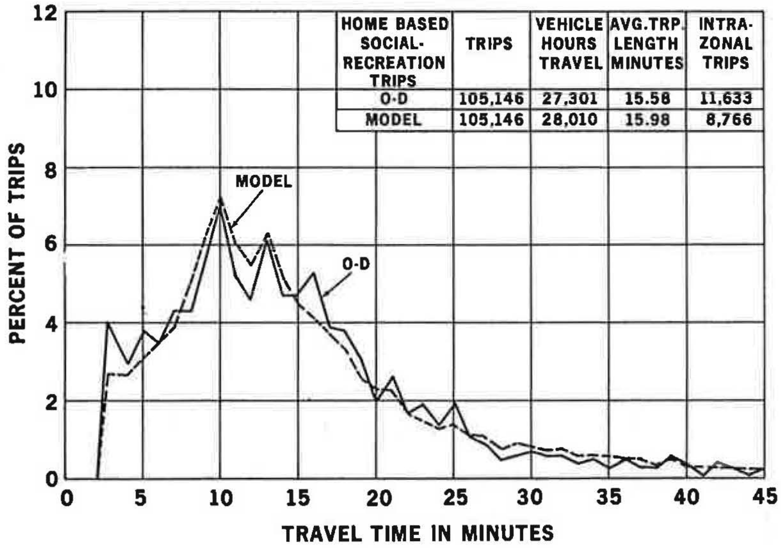
Figure 7. Comparison of trip length distribution, home-based civilian work trips, O-D vs gravity model (calibration 8).



HOME BASED SHOPPING TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	88,940	16,819	11.35	14,741
MODEL	88,940	17,022	11.48	13,813

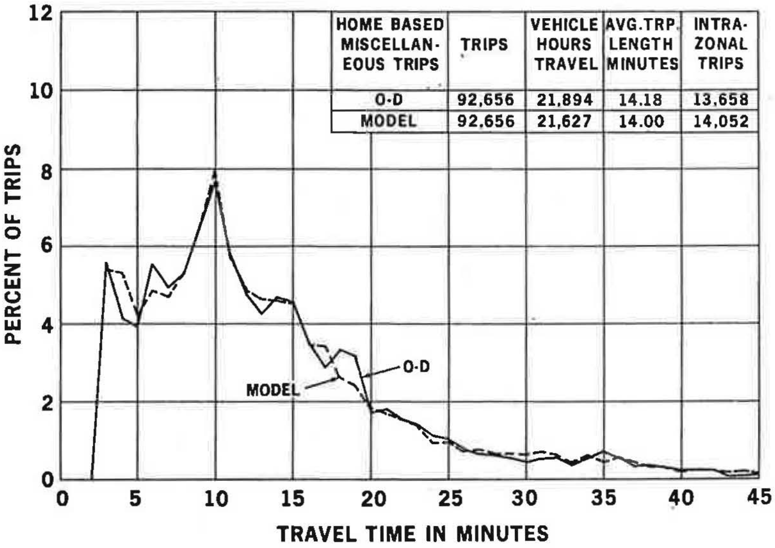
Figure 8. Comparison of trip length distribution, home-based shopping work trips, O-D vs gravity model (calibration 8).

First, it is important to specify the overall requirements of the model. Significant differences could exist, depending on whether the requirement was corridor planning or developing design volumes. Presumably the transportation planning process, and therefore trip distribtuion techniques, must be oriented to provide answers for both of these requirements, even though one comes later than the other in the process. Since the data used in this research covered only one point in time, it is difficult to obtain precise quantitative answers to the questions raised. Given some of the problems and



HOME BASED SOCIAL-RECREATION TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	105,146	27,301	15.58	11,633
MODEL	105,146	28,010	15.98	8,766

Figure 9. Comparison of trip length distribution, home-based recreation work trips, O-D vs gravity model (calibration 8).

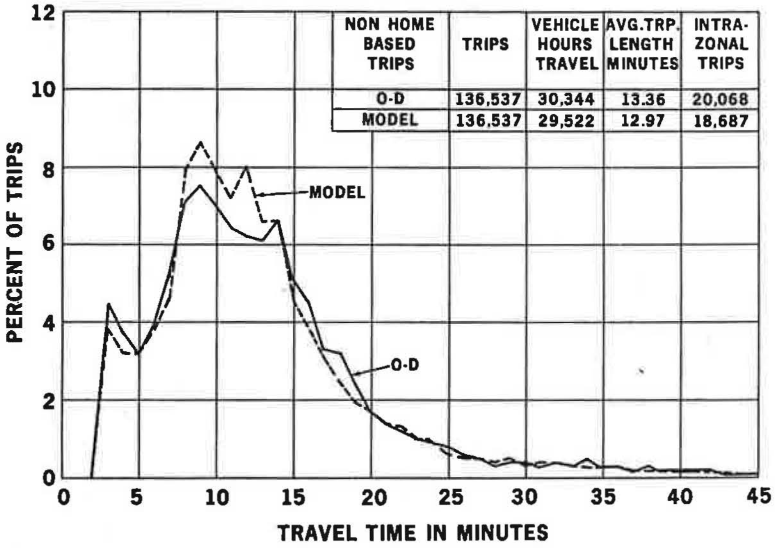


HOME BASED MISCELLANEOUS TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	92,656	21,894	14.18	13,658
MODEL	92,656	21,627	14.00	14,052

Figure 10. Comparison of trip length distribution, home-based miscellaneous work trips, O-D vs gravity model (calibration 8).

parameters of a mathematical model, however, it is possible to obtain some helpful insights into applying trip distribution techniques.

With this background and the examination of test results using Oahu data, how does each of the models perform as a simulation technique for trip distribution patterns? One of this paper's most significant values in answering this question is that for the first time all tests used previously with one model have been applied to the results of both models. Each test was evaluated to determine its significance. Special emphasis was given to evaluating these tests and the data used as the major calibration control and as a basis for forecasting model parameters.



NON HOME BASED TRIPS	TRIPS	VEHICLE HOURS TRAVEL	AVG.TRP. LENGTH MINUTES	INTRA-ZONAL TRIPS
O-D	136,537	30,344	13.36	20,068
MODEL	136,537	29,522	12.97	18,687

Figure 11. Comparison of trip length distribution, nonhome-based work trips, O-D vs gravity model (calibration 8).

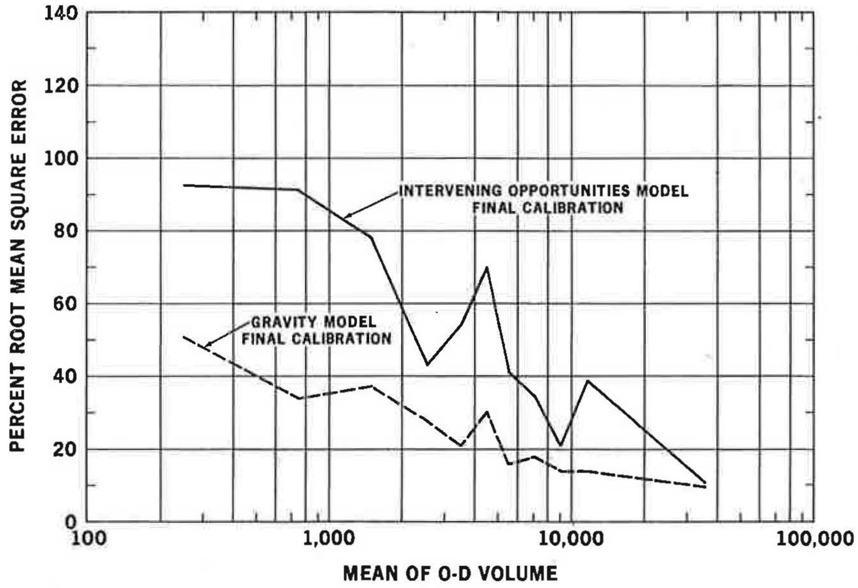


Figure 12. Statistical comparison of district-to-district movements, total purpose trips, intervening opportunities and gravity models.

The major controls on each model are outlined both in calibration and in forecasting. Also illustrated are those variables or data which are to be used for additional testing purposes. An attempt is made to weigh the performance of each model on each test.

Model Calibration

Each of the models tested—the gravity model and the intervening opportunities model—has unique methods of calibration. The gravity model, as calibrated for Oahu,

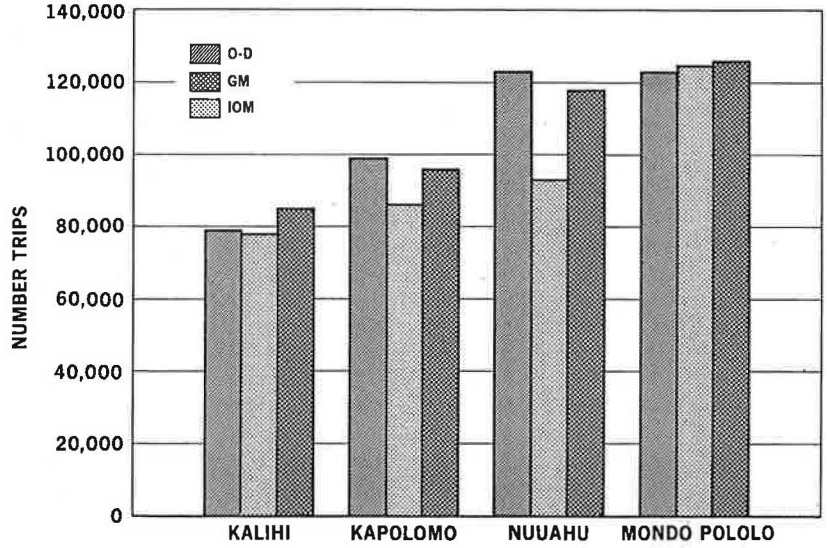
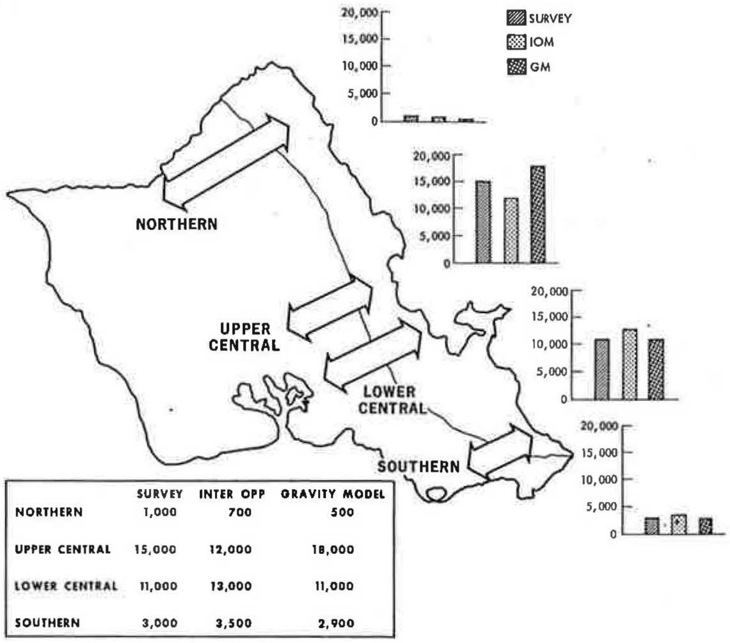


Figure 13. Comparison of total purpose trips crossing selected screenlines, survey and models (final calibration).



	SURVEY	INTER OPP	GRAVITY MODEL
NORTHERN	1,000	700	500
UPPER CENTRAL	15,000	12,000	18,000
LOWER CENTRAL	11,000	13,000	11,000
SOUTHERN	3,000	3,500	2,900

Figure 14. Comparison of total purpose trips crossing Koolau Mountain Range, survey and models (final calibration).

simulates the actual trip length frequency through the use of friction factor curves and also incorporates topographical barriers for the major mountain range on the island. To accomplish this calibration, it was necessary to obtain a set of friction factors and explore the need for topographical barriers, based on data from observed O-D movements.

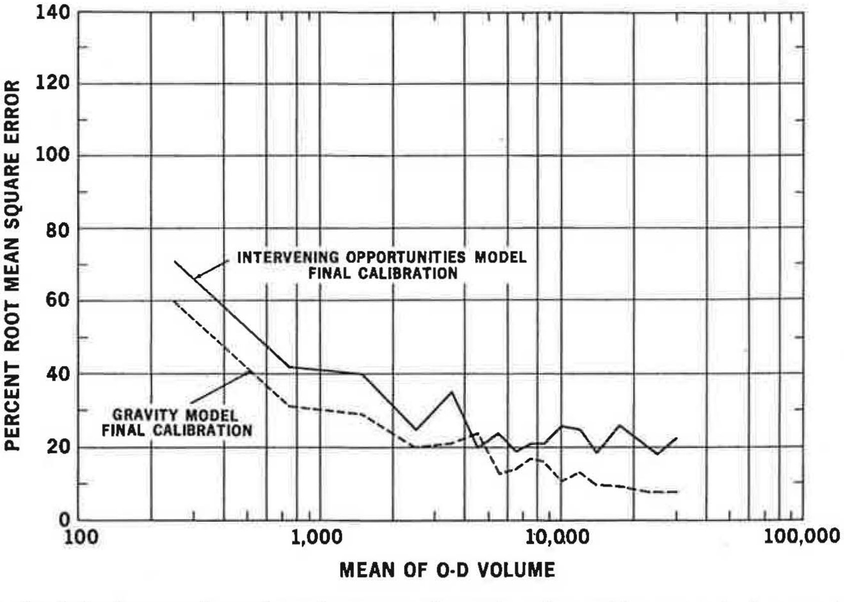


Figure 15. Statistical comparison of total purpose trips assigned to spider network, intervening opportunities and gravity models.

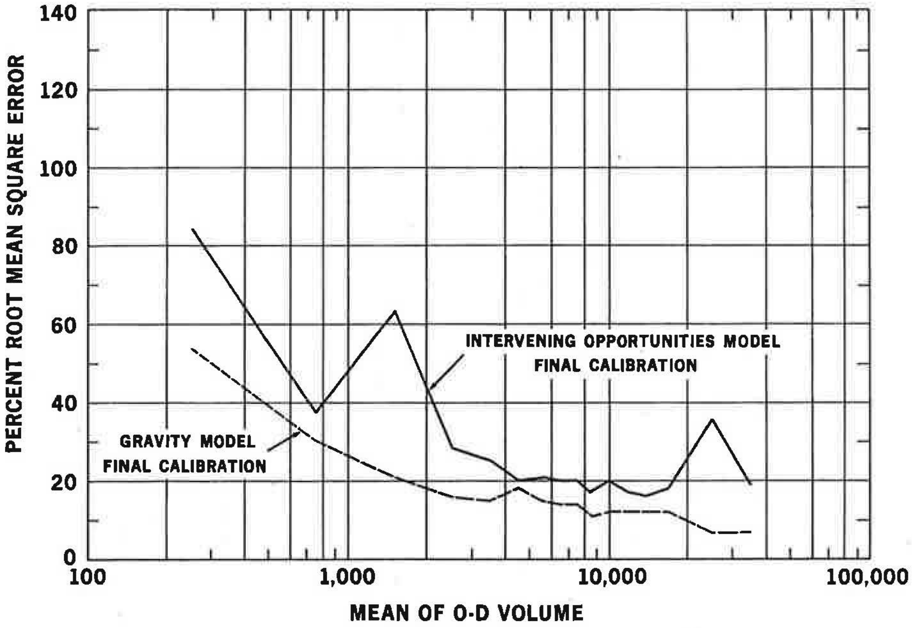


Figure 16. Statistical comparison of total purpose trips assigned to highway network, intervening opportunities and gravity models.

In some instances where the gravity model has been used previously, other model parameters have been required, such as K factors. These have been applied in relatively few cases; exploration in the calibration process for Oahu indicated they were not needed.

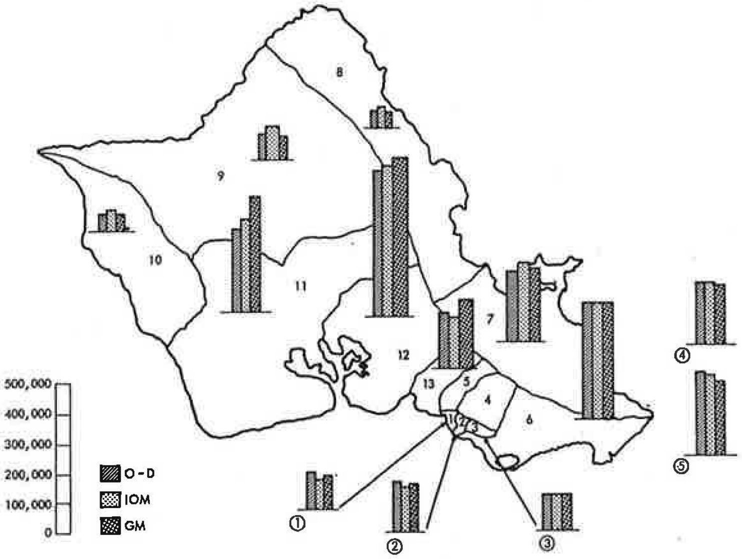


Figure 17. Comparison of vehicle-miles of travel by district, survey and models (final calibration).

When using the gravity model as a forecasting tool, the same variables—friction factor curves—are assumed to remain the same, and forecasts of the level of future travel barriers are required.

The intervening opportunities model essentially utilizes the observed O-D data to develop a decay rate, described as an L value. This indicates the rate at which origins are attracted to or utilize the opportunities in the form of destinations available to them in other zones. This is similar to the friction factor curves used in the gravity model.

The use of the decay rate tends to treat changes in the network somewhat less precisely because it depends only on the relative position of destinations and not on absolute value of time separation.

After the initial set of L values is determined, vehicle-miles of travel, both estimated and actual, are then used to adjust L values by district until a properly calibrated model is obtained. No other parameters entered into the calibration process with the intervening opportunities model.

When using the intervening opportunities model as a forecasting tool, the calibrated L values are related to trip-end densities and future L values are calculated based on projected trip-end densities. In most cases, L values have been predicted to change on the basis of changes in trip-end densities.

Model Tests

As indicated earlier, the transportation planning process, and therefore trip distribution techniques, should be designed to provide answers on a range of transportation planning problems. One of the first problems is to adequately estimate corridor demand. The planning process must also provide design volumes on particular links in the network. This is significant in looking at tests of model performance on a variety of needs.

It could be argued that as long as one is able to predict volumes on the network at acceptable accuracies, other measures of trip distribution, such as trip length, frequency curves, or vehicle-miles of travel by district, would be relatively unimportant. However, the question remains as to what testing and calibration techniques are the most important for indication of model performance.

Several tests were applied to the results of each model. These include (a) estimated link volumes on actual and spider networks compared to assigned O-D data; (b) district-to-district movements compared to actual O-D movements; (c) screenline checks comparing model volume accumulations to survey volume accumulations; (d) the exploration of the relative accuracy of predicting movements across topographical barriers, such as mountain crossings, by comparing the estimated link volumes to survey volumes; (e) actual mean trip lengths and trip length frequency curves from the observed O-D data compared to model results; and (f) vehicle-miles of travel by district and facility type.

Items (e) and (f) are the controls governing the calibration of the models. Item (e) is applied primarily to the gravity model and item (f)—vehicle-miles of travel—is used primarily to calibrate the intervening opportunities model. Items (a) through (d) are independent tests used to determine the accuracy of various elements of each model.

In almost every instance, the gravity model results were as good, if not somewhat better, than the intervening opportunities model results. (This was particularly true when the trips were assigned to the highway network, as recorded in Figure 16.)

Results of specific movements—district-to-district movements as compared to the orgin-destination data—were also somewhat better in the gravity model results. Screenline checks over the four screenlines were approximately equal in overall accuracy. On the mountain crossing checks, the intervening opportunities model was better, even with the time barriers used in the gravity model calibration. Without barriers, this difference was more critical.

Average trip length and trip length frequency were better with the gravity model. This is to be expected since the friction factor curves are developed on the basis of trip length frequency, and this is an independent check for the intervening opportunities model and is not used in the calibration process.

Vehicle-miles of travel was checked in four different ways—overall total, total by facility type, total by district, and total by facility type within each district. This particular parameter is used as a major calibration element for the intervening opportunities model adjustments. An analysis of the test VMT comparison indicates accuracy for the gravity model and the intervening opportunities model as approximately equal.

Additional Considerations

One additonal item, balancing destinations, which has not been mentioned previously, was determined to be necessary in both models as an additional calibration step. This insures that the correct number of trips are attracted or sited at available destinations.

Tests outlined in previous sections indicate the improved accuracy of balancing destinations, and certainly it seems not only reasonable but extremely important that this control be applied to model results in the present or forecast year. Whether trip destinations are determined from an observed study or in a trip-end forecast, it would seem that these trip destinations should adhere to and the distribution be forced to meet these conditions.

Special mention should be made of topographical features. Observations here and in other research sometimes indicate the necessity of incorporating travel time barriers in the trip distribution model. In most cases, these barriers to free movement of travel are extremely important in the transportation planning process. Facilities crossing barriers such as mountains or rivers are among the most costly and critical transportation planning decisions. The expense of providing facilities crossing these barriers undoubtedly has caused the historical development of restrictions on free movement. In the Oahu application, the intervening opportunities model did not require the use of a topographical barrier, while the gravity model did. This raises significant problems in determining future travel demand across these barriers when given certain levels of service and in determining adequate facilities. It also merits considerable exploration in utilization of trip distribution techniques.

CONCLUSIONS

There has been much discussion about the best controls and tests on trip distribution procedures to insure the best possible answers to transportation planning questions. Obviously, of the various tests applied in this rearch, link volumes—the demand that will occur on links in the network— is the most significant, both for corridor planning and design purposes.

A specific aspect of this is determining what volumes will occur on particular links or across a particular line, such as a topographical barrier. These particular items are probably the most important of the tests. If one can do well on these, then district-to-district movements, trip length frequencies and vehicle-miles of travel by district are of lesser consideration.

The importance of this phase of the transportation planning process certainly demands that investigation be given to as many significant tests as possible. While three tests have been indicated as being less significant, it is still felt that their investigation is important and proper. The models can be used improperly or properly, but cannot be expected to provide adequate answers if used improperly.

It is important to note that while the gravity model calibration was fairly straightforward with no particular innovations being made, some different procedures were evaluated in terms of calibrating the intervening opportunities model. It is evident that the most appropriate method of calibrating an intervening opportunities model was with the use of variable short and long L values by district, based on the decay rate and adjusted on the basis of vehicle-miles of travel. The use of single area-wide L values by trip category did not provide a satisfactory simulation of Oahu travel patterns, as pointed out earlier. The use and documentation of the variable L values by district is relatively new. The authors feel that this is a satisfactory and meaningful way to improve the performance of the model without introducing undue manipulation of the data, based primarily on the fact that the L values are related to a sound parameter in trip-end density. Trip-end densities are used as a control, both at the present time and in the forecast procedure.

The following comments relate to the significance of the test results by considering the relationships between the model parameters and the urban phenomena, and how this relationship may meet the test of time.

As a review, the forecast using the intervening opportunities model in Oahu would be based on future L values derived from future trip-end densities by district. It appears that this is a very sound approach, based on the relative confidence which can be placed in forecasts of trip-end density by district. Unfortunately, the relationship between L values and trip-end density has not been fully tested over a forecast period. This factor raises some questions regarding this particular assumption.

With the gravity model, the F factors, which are developed giving relative weight to trip propensity at various time levels, are assumed to remain constant. The other variable used in the gravity model calibration—topographical travel barriers—would require forecasting. Some research has been done on this variable, but it remains a serious problem requiring further study. The assumption made regarding the standard, or continuation, of the friction factor curve into time has also met some serious criticism and questions when significant changes in a region's level of service occur. Based on the tests in Oahu and some investigation and exploration of parameters used, the gravity model appears to perform better in most present situations and test applications. However, there are significant questions regarding the assumptions and forecasts required to predict future trip distribution patterns. With the intervening opportunities model, as indicated, the trip-end density control is a sound procedure. The L values's relationship with trip-end density is more of a problem.

In conclusion, the authors feel that significant information has been developed regarding the testing and calibration parameters used in trip distribution models. Based on the tests here and investigation of the parameters, they also feel that either model evaluated in this paper can be used, if applied properly, to simulate present and future travel patterns.

REFERENCES

1. Pyers, Clyde E. Evaluation of Intervening Opportunities Trip Distribution Model. Highway Research Record 114, pp. 71-98, 1965.
2. Heanue, Kevin E., and Pyers, Clyde E. | A Comparative Evaluation of Trip Distribution Procedures. Highway Research Record 114, pp. 20-50, 1965.
3. 1960 Honolulu Metropolitan Area Traffic Survey. Volume I, Historical Data and Design of the Study, Sept. 1962.
4. Saltman, Theodore J. The Effects of Alternate Loading Sequences on Results From the Chicago Trip Distribution Model. Pittsburgh Area Transportation Study, June 1964.
5. Final Report, Chicago Area Transportation Study, Volume 2.
6. Niagara Frontier Transportation Study. Volume 2, Final Report, January 1966.
7. Marshall, Ralph J. Program to Calculate a Fitted "L" to Empirical Data. Program No. MO-505, Upstate New York Transportation Studies.
8. Essene, Karen. Program to Summarize Vehicle-Miles of Travel. Program No. 4004, Oahu Transportation Study.
9. Calibrating and Testing a Gravity Model for Any Size Urban Area. Urban Planning Division, Office of Planning, U. S. Bureau of Public Roads, October 1965.