Evaluating Trip Forecasting Methods with an Electronic Computer

G.E. BROKKE and W.L. MERTZ,
Highway Research Engineers, U.S. Bureau of Public Roads

In forecasting future trip distribution from the existing pattern, the average factor method, the Detroit method, and the Fratar method are equally accurate if each method is carried through a sufficient number of successive iterations. In all cases tested, the second approximation of the Fratar method was of maximum accuracy while four or more approximations were usually required with the other two methods.

The results of the test emphasized the fact that the majority of the trips within a metropolitan area consist of a very large number of small-volume zone-to-zone movements, where the zones are of normal size. With sampling rates used these individual small-volume movements are not accurately determined. The accumulation of the small-volume movements into volumes associated with ramps, streets, and expressways, should result in acceptable accuracy as calculated by statistical formulas, but this will have to be definitely established by additional research.

The advantage of using an electronic computer on research projects of this type can hardly be overestimated, notwithstanding the difficulty and time consumed in preparing the program. In this series of tests, the speed and accuracy of the computer permitted the attainment of results in hours instead of years after the program had been completed, without a single error attributable to the computer.

● HOME-INTERVIEW origin and destination studies were made in the Washington, D.C. area in 1948 and in 1955. In the earlier study a 5-percent sample was obtained by interviewing the residents of one of every 20 dwelling units. In the 1955 study, the sample rate was 1 in 30 in the District of Columbia and 1 in 10 elsewhere within the area.

These two surveys offered, for the first time, an opportunity to study the changes occurring over a period of several years in a metropolitan area in the pattern of trips, that is, the differences in the numbers of trips between the same origins and destinations. They also provided data that could be used to evaluate methods of forecasting future trip volumes.

TRIP FORECASTING ELEMENTS

Two basic elements are involved in the forecasting of trips. One is the increase in the number of trip origins and destinations in a particular part of the city such as a zone. For brevity the number of trip origins and destinations combined have been labeled trip ends. Thus, for example, 2 trips originating in a zone and 3 trips destined to it would be counted as 5 trip ends. Therefore, if only the trips made wholly within an area are considered, the total number of trip ends in the area is exactly twice the number of trips.

The ratio of the future trip ends expected in a particular zone to the present trip ends in the zone is called the growth factor for that zone. Much work has been and is being done in this field to determine the best method of arriving at the proper growth factor. Up to now, however, forecasts, so far as total trip ends are concerned, are dependent to some extent on personal judgment. In order to eliminate this variable, and isolate the elements being studied, the growth factors were calculated for each zone by taking the ratio of the reported trip ends in each zone in 1955 to the reported trip ends in each zone in 1948. Thus any variability in predicting growth factors will not affect this study of forecasting methods.

The other basic element involved in forecasting zone-to-zone movements is the application of the growth factors of the two terminal zones in predicting the number of future trips between them. Various mathematical formulas have been developed with that end in view. It is the purpose of this study to evaluate the accuracy of these methods and the formulas used therein. Certain other methods of trip forecasting, based on population distribution, trip-attraction distribution, and distance or travel time, directly predict the number of zone-to-zone trips, but as these methods are still in the process of development they will not be further discussed.

CHARACTERISTICS OF THE AREA

One problem that had to be resolved in beginning the test was that the area covered in the 1948 Washington survey was somewhat smaller than that of the 1955 survey, and the extent and identifying numbers of many of the zones had been changed. The first step was to reconcile these differences by rezoning the metropolitan area into 254 zones and to determine both the 1948 and 1955 volumes of trips into and out of these zones.

The 254 zones covered in area which contained 96 percent of the population that lived within the 1948 cordon and 93 percent of that living within the 1955 cordon.

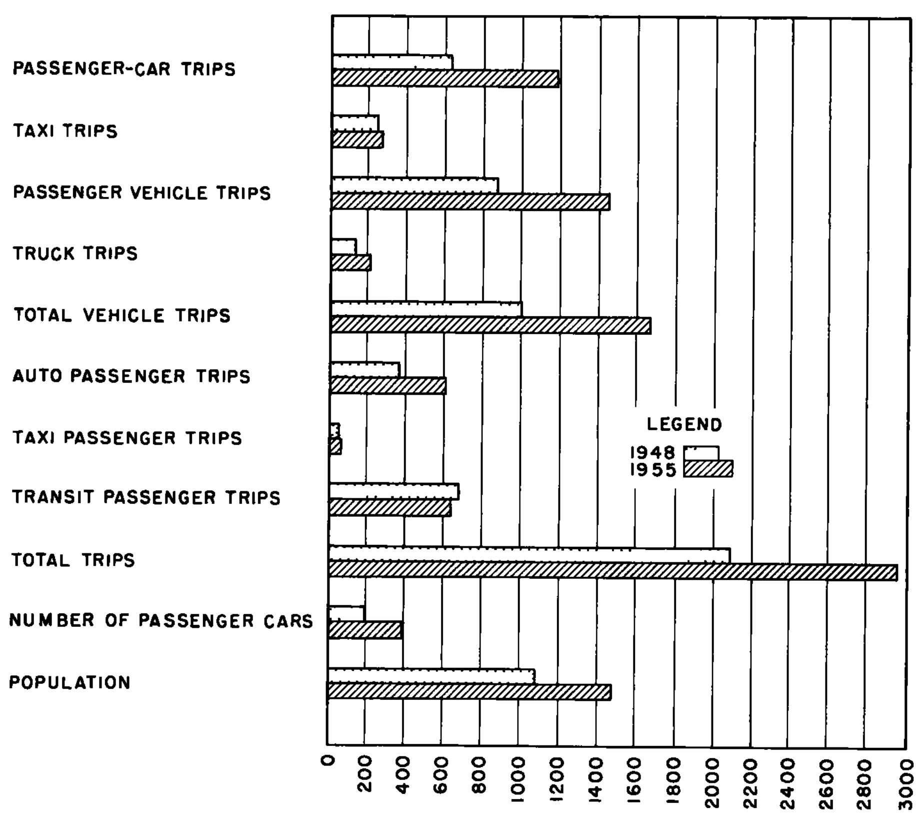

Figure 1. Area within the 1948 cordon, Washington, D. C. Number of trips, cars and transportation 1948-1955.

Most of the external cordon stations in 1955 were placed at different locations from those in 1948, and therefore the trips crossing the cordon, called "external" trips, are omitted from the study, the two surveys not being comparable in regard to this class of trip.

Within the 254 zones the population increased 38 percent during the 7-year interval while the number of trips by persons increased 42 percent. (In this paper "trips by persons" includes trips by drivers of automobiles, taxis and trucks and by passengers in automobiles, taxis and mass transit vehicles. Walking trips and the small number of trips by passengers in trucks are not included.) This represents a small increase in trips per person from 1.95 to 2.00. During the same interval the number of passenger cars owned by residents almost doubled, increasing 96 percent, and the number of

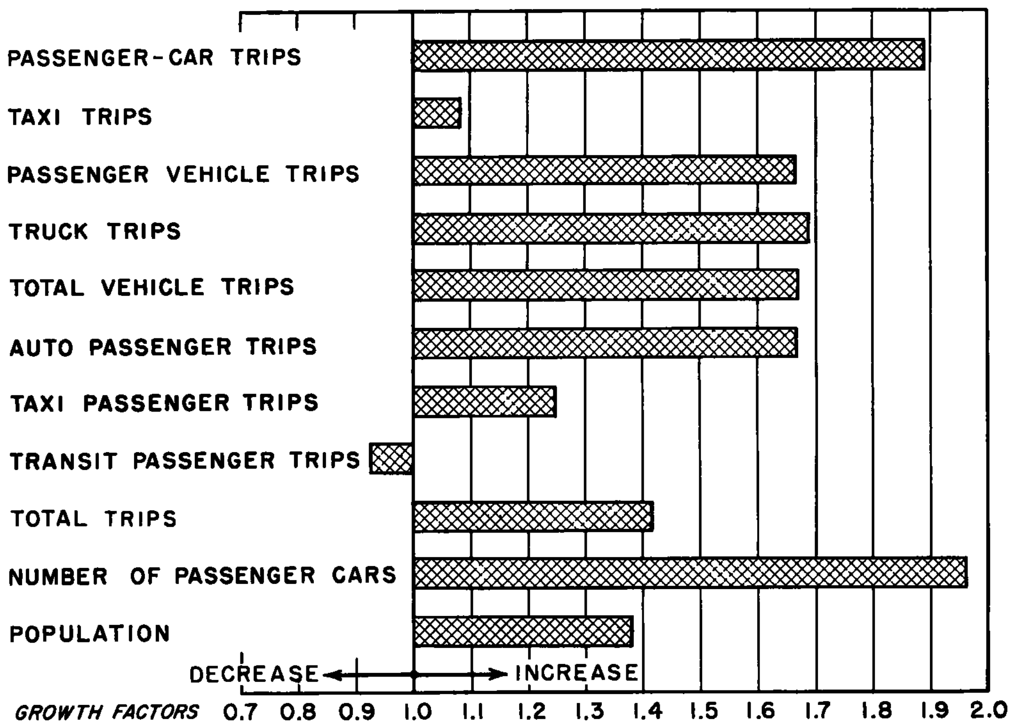

Figure 2. Area within the 1948 cordon, Washington, D. C. Change from 1948 to 1955.

trips made by these passenger cars went up 89 percent. This represents a small decrease in the numer of passenger-car trips per passenger car from 3.15 to 3.05. These figures seem to indicate that the number of car trips increases roughly in proportion to the increase in the number of cars, and total person trips increase about as population does in the Washington area.

The number of trips by various vehicle types and modes of transportation together with the population and passenger-car ownership are shown in Figure 1. The growth factors resulting from the changes between 1948 and 1955 are shown in Figure 2.

The increase of 89 percent in the number of passenger-car trips during the 7-year interval is rather high. It represents an average increase of almost 13 percent annually on a straight-line basis or about 9.5 percent if compounded annually.

This high rate of increase in the Washington area, however, has the advantage of providing growth factors that are somewhat similar to the growth factors that have been forecast for about 25 years in some of the larger cities. For example, the growth factors for total vehicle trips as measured in Washington and those predicted for Detroit and Cleveland are shown in the following table:

Item	Washington	Detroit	Cleveland
Period covered	1948 to 1955	1953 to 1980	1952 to 1975
Over-all growth	1.66	1.67	1.79
Percent of zones with growth factors:			
Less than 1.00	5	2	1
1.00 - 1.49	38	64	54
1.50 - 1.99	25	9	17
2.00 - 2.99	18	9	8
3.00 - 4.99	9	8	15
5.00 - 9.99	3	6	5
Over 10.00	2	2	0

Thus, although the test of the forecasting methods is confined to the growth of Washington from 1948 to 1955, the actual growth factors are not entirely dissimilar to forecasts into the future for Detroit and Cleveland although the latter two are for longer periods of time.

In this study, the passenger-car trips and taxi trips were combined into one category of passenger-vehicle trips The over-all growth factor for these trips was 1.67. As for individual zones, about 6 percent had fewer trip ends in 1955 than in 1948, and 50 percent had a growth factor smaller than 1.55. A more detailed distribution of the individual zone growth factors is as follows:

Growth Factor	Percent of Zones
Less than 1.00	6
1.00 to 1.50	40
1.50 to 2.00	23
2.00 to 3.00	19
3.00 to 5.00	7
5.00 to 10.00	2
Over 10.00	3

METHODS OF FORECASTING TRIPS

The 1948 zone-to-zone trips were expanded to 1955 by various formulas and the predicted values compared with those obtained in the 1955 sample. The actual methods are described as follows:

T_{ij} = Observed 1955 trips between zone i and j

$T_{ij'}$ = Calculated 1955 trips between zone i and zone j

T_{i-j}' = Calculated 1955 trips from zone i to zone j

T_{j-i}' = Calculated 1955 trips from zone j to zone i

t_{ij} = Observed 1948 trips between zone i and zone j

T_i = Summation of observed 1955 trip ends in zone i

t_i = Summation of observed 1948 trip ends in zone i

F_i = Growth factor for zone i = $\frac{T_i}{t_i}$

T = Summation of 1955 trip ends in entire area

t = Summation of 1948 trip ends in entire area

F = Growth factor for entire area = $\frac{T}{t}$

t_{ix} = 1948 trips between zone i and each of all other zones designated as zone x

F_x = Growth factor for zone x

Uniform Factor Method

The most simple method of expanding trips is to compute a single factor for the entire area and use it to multiply all zone-to-zone trips. This particular method is seldom used now, but because of its wide use in the past it was evaluated. Mathematically the expansion formula is as follows:

$$T_{ij}' = t_{ij} F$$

There is no possibility of successive approximations with this method, such as those used in the methods subsequently described.

Average Factor Method

In this method each of the 1948 zone-to-zone movements is multiplied by the average of the growth factors for the two zones involved as follows:

$$T_{ij}' = t_{ij} \frac{(F_i + F_j)}{2}$$

After the trips from one zone (i) to all other zones have been computed by this method, the sum of all trip ends in that zone as determined from this calculation (T_i') will probably not equal the actual 1955 trip ends in that zone (T_i). This discrepancy can be eliminated by a series of iterations producing successively closer approximations, as follows:

Let F_i' equal the factor needed to bring the calculated number of trip ends (T_i') to actual number (T_i) or $F_i' = T_i/T_i'$, and similarly $F_j' = T_j/T_j'$

Then for the second approximation, $$T_{ij}'' = T_{ij}' \frac{(F_i' + F_j')}{2}$$

Similarly for a third approximation, $$T_{ij}''' = T_{ij}'' \frac{(F_i'' + F_j'')}{2}$$

The process can be repeated until the F factors for a new iteration equal the limiting value of 1.00.

One of the inherent disadvantages of the average factor method is that the calculated trips into zones with higher-than-average growth factors generally total less than the predicted number of trips. Conversely the calculated trips into zones with lower-than-average growth factors total more than the predicted total of trips. This systematic bias of the predicted values could result in an inordinate number of approximations and may affect the accuracy of the method.

Detroit Method

A method to alleviate this difficulty was developed by Carroll's staff for the Detroit study. In this method they assumed that the trips from zone i will increase as predicted by F_i and will be attracted to zone j in the proportion $\frac{F_j}{F}$. The predicted trips from zone i to zone j can then be calculated as follows:

$$T_{i-j}' = t_{i-j} \frac{(F_i \cdot F_j)}{F}$$

Similarly the trips from zone j can be considered as increasing as predicted by F_j and will be attracted to zone i in the proportion $\frac{F_i}{F}$.

The predicted trips from zone j to zone i can then be calculated as follows:

$$T_{j-i}' = t_{j-i} \frac{(F_j \cdot F_i)}{F}$$

Therefore the number of trips between zone i and zone j is equal to the sum of the trips from i to j and from j to i or

$$T_{ij}' = T_{i-j}' + T_{j-i}'$$

or

$$T_{ij}' = t_{i-j} \frac{(F_i \cdot F_j)}{F} + t_{j-i} \frac{(F_j \cdot F_i)}{F}$$

$$= (t_{i-j} + t_{j-i}) \frac{(F_i \cdot F_j)}{F}$$

$$= t_{ij} \frac{(F_i \cdot F_j)}{F}$$

As in the case of the average factor method the calculated trip ends in a particular zone will probably not equal the predicted trip ends in that zone. Therefore new F factors can be determined as follows:

$$F_i' = \frac{T_i}{T_i'}$$

$$F_j' = \frac{T_j}{T_j'}$$

and a second approximation can be calculated as follows:

$$T_{ij}'' = T_{ij}' \frac{(F_i' \cdot F_j')}{F'}$$

This same procedure can be used to calculate a third and subsequent approximations until the new F factors equal the limiting value of 1.00.

Fratar Method

The first method in which the iterative process was used in predicting future trips was developed by Thomas J. Fratar in connection with the forecast for Cleveland, Ohio. Fratar considers that the distribution of the trips from any zone i is proportional to the present movements out of zone i modified by the growth factor of the zone to which these trips are attracted. The volume of the trips, however, is determined by the expansion factor of zone i.

If the trips between zones i and j, as calculated by considering all trips from zone i are represented by the symbol $T_{ij(i)}'$ and those as calculated by considering all of the trips from zone j by the symbol $T_{ij(j)}'$, then

$$T_{ij(i)}' = t_{ij} \cdot F_j \cdot \frac{\Sigma (t_{ix} \cdot F_i)}{\Sigma (t_{ix} \cdot F_x)} \qquad (1)$$

Noting that $\Sigma t_{ix} \cdot F_i$ can also be written as $F_i \cdot \Sigma t_{ix}$ then Eq. 1 can be written as

$$T_{ij(i)}' = t_{ij} \cdot F_j \cdot F_i \cdot \frac{\Sigma t_{ix}}{\Sigma (t_{ix} \cdot F_x)} \qquad (2)$$

The last term in Eq. 2 basically represents the reciprocal of the average attracting pull of all other zones on i. It has been labeled the "Location" or "L" factor since it is somewhat dependent on the location of the zone with respect to all other zones. Thus since

$$\frac{\Sigma t_{ix}}{\Sigma(t_{ix} \cdot F_x)} = L_i \qquad (3)$$

Eq. 2 can be rewritten as

$$T_{ij(i)}' = t_{ij} \cdot F_j \cdot F_i \cdot L_i \qquad (4)$$

Then for all trips from zone j, it can similarly be shown that

$$T_{ij(j)}' = t_{ij} \cdot F_i \cdot F_j \cdot L_j \qquad (5)$$

Thus the trips between zone i and zone j have been computed twice—once for all trips out of zone i and once for all trips out of zone j. The most probable value is an average of the two computations or

$$T_{ij}' = \frac{T_{ij(i)}' + T_{ij(j)}'}{2} \qquad (6)$$

Substituting the identities from Eqs. 4 and 5 into Eq. 6 and factoring out the common terms, the final equation is developed

$$T_{ij}' = t_{ij} \cdot F_i \cdot F_j \cdot \frac{(L_i + L_j)}{2} \qquad (7)$$

After all the zone-to-zone trips have been computed by this formula, the calculated trip ends in a particular zone will probably not agree with the predicted trip ends in that zone. Therefore new factors can be calculated as follows:

$$F_i' = \frac{T_i}{T_i'}$$

$$F_j' = \frac{T_j}{T_j'}$$

$$L_i' = \frac{\Sigma T_{ix}'}{\Sigma T_{ix}' \cdot F_x'}$$

$$L_j' = \frac{\Sigma T_{jx}'}{\Sigma T_{jx}' \cdot F_x'}$$

A second approximation can then be calculated as follows:

$$T_{ij}'' = T_{ij}' \cdot F_i' \cdot F_j' \cdot \frac{(L_i' + L_j')}{2}$$

The same procedure can be used for subsequent approximations until the new F factors equal the limiting value of 1.00.

THE PROBLEM OF EVALUATION

With the Washington area divided into 254 zones, the number of possible zone-to-zone movements is $\frac{N(N+1)}{2} = 32,385$. (Zone-to-zone movements as used in this article also include intrazone movements.) It was expected that some of the zone-to-zone movements would be zero in 1948 and 1955 and would not need to be computed. However, it was estimated conservatively that perhaps 30,000 of the 1948 zone-to-zone movements would require expansion to 1955.

To determine whether the accuracy of the prediction was influenced by vehicle type or mode of transportation, the trips were separated by mode of travel into six categories: passenger-vehicle trips, truck trips, total vehicle trips, transit-passenger trips, auto-passenger trips, and total trips by persons; and each group was expanded separately.

Expanding each of the 30,000 zone-to-zone trips made in 1948 would be almost meaningless unless some method of summarizing the comparison to the 1955 survey movements had been determined. The most obvious answer to this problem was to subtract the computed number of trips from the reported number of 1955 trips, square the difference and accumulate the result. The sum of the differences squared could then be used to calculate the root-mean-square error of the number of trips as expanded from the 1948 data.

This summary, however, had a serious disadvantage in that the actual volume of zone-to-zone trips varies from zero to several thousand. A root-mean-square error could be inordinately affected by the relatively few large movements. Similarly, if the difference were converted to a percentage of the 1955 actual movement and the root-mean-square of the percentage computed, the result could be as greatly affected by the small movements that probably lack sufficient stability to provide meaningful information. It was therefore decided to stratify the 1948 movements by volume classes, thus: by volumes of tens to 100, by volumes of hundreds to 1,000, and all volumes over 1,000. The numerical root-mean-square error was then computed for each volume class and the percentage error for the class was obtained from the ratio of the numerical root-mean-square to the average 1955 volume. The proportion of all 1955 volumes in each volume class was then determined and each of the percentage errors was weighted by the proper proportion to obtain the over-all percentage error.

The over-all percentage error as described above was regarded as the proper measure to evaluate the various predictive formulas. In addition, the accuracy of larger movements could be measured by an extrapolative process in which the number of average zone-to-zone movements required to make up a larger volume was determined and the basic error was divided by the square root of the number of movements required.

In addition to the computation described above, it was desirable to know the root-mean-square error for each of the zones so that the error can be related to the growth factor of the individual zones. Therefore, the difference between the expanded 1948 and the 1955 movements squared was accumulated for each of the zones. It was considered likely that from this computation any inordinate error found in a particular zone could be recognized.

As has been previously explained, it is possible to carry the average factor method, the Detroit method, and the Fratar method through a number of iterations to produce successively closer approximations. To be reasonably certain that this process was continued a sufficient number of times, it was decided to calculate 10 successive approximations by each method.

NEED FOR AN ELECTRONIC COMPUTER

It has been estimated that roughly 25 million computations would be required for this test. On the very optimistic basis that one computation can be completed in 10 seconds, using ordinary desk calculators, the project would require some 30 man-years for completion. Clearly this is not feasible. However, with electronic computers, the time required can be reduced enormously.

The three methods that use iterations are similar in that the input for the first calculation is made up of the original data, the output of this calculation and each successive iteration becomes the input of the next iteration and so on until 10 calculations have been made if necessary for satisfactory closure. In addition each iteration of the Fratar method requires two passes of the input—one to determine the L factor and one to make the required expansion. Thus, the 30,000 zone-to-zone movements have to be processed 42 times, including the one pass of the data required to obtain the original growth factors.

In deciding upon the particular type of computer to be used, the first problem was to decide whether to use one with card input and output, or one with tape. The card type would require about 200 hours of computer time provided sufficient memory were available. With tape, only about 10 hours of computer time, or less would be required, again assuming sufficient memory, so obviously the tape-using type was preferable. In the actual test, 30 hours of computer time were used principally due to additional tests on larger zone groupings.

COMPUTER CHARACTERISTICS

Computer problems in general fall into two categories—data-processing problems and computation problems. For instance, a problem of testing forecasting methods, requires a great deal of input and output but rather simple internal operations and is properly classified as a data-processing problem. On the other hand computing such things as log tables or trigonometric functions requires much computation but very little input and output.

The problem then was to select a machine designed to process data with good "read and write" characteristics and with a large memory. Part-time use of an IBM 705 machine which met all these requirements very well was arranged. The machine had a core memory of 40,000 characters. The memory capacity of computers is sometimes reported in "characters" and sometimes in "words." In computer terminology, a "character" may be a digit, a letter or a symbol, while a "word" consists of a group of characters. In some computers the work is of a constant length and in other computers the word length may be varied at the option of the programer. The computer used was a variable-word-length machine and the core-memory capacity is therefore given in characters rather than words.

The machine was equipped with two tape-record coordinators, more commonly referred to as buffers. The purpose of the buffers is to shorten the reading and writing time. Each buffer has a core storage capacity of 1,024 characters. The buffers are loaded from the tape units, and when the computer requires more data it obtains it at electronic speed from the buffer. As soon as the buffer is called upon for data, the tape unit feeding it begins to accelerate, so that by the time the buffer is empty the tape unit has begun to refill it with the next record to be processed. The same process works essentially in reverse for output. Thus the machine can go on with other work while records are being fed into and out of the buffers.

The 1,024-character capacity of the buffers also allows a number of card records to be grouped so that when the buffer calls on the tape units for data, one tape record can receive information from a number of cards without exceeding the buffer capacity. The program for this study was designed to have the machine read or write the equivalent of 24 cards of information per reading or writing cycle. As it turned out in production, the computer took a longer time to process the data on the 24 cards than was required to fill or empty the buffers so that the machine never had to wait for data, and all reading or writing time was essentially "free."

The 705 is a decimal machine, meaning that the entire content of memory is in condition always to be printed out as alpha-numeric information directly without conversion from binary to decimal. The machine performs internal operations at an average rate of 8,300 per second.

PREPARATION OF PROGRAM

This was the first large scale electronic computer project done by the Bureau of Public Roads, therefore, the services of the Bureau of Standards Computer Laboratory were retained to aid in the selection of the machine and in the preparation of the program. Their wide experience in this field has proven invaluable in the completion of this work.

In programing the problem the first difficulty was one of memory space, in spite of the large amount available. It was necessary to overlap the program and resort to external tape storage. Even so it was necessary to split the problem into two parts. All of the vehicle types (passenger cars, trucks, and total vehicles) were handled in the first run. The second run processed the person trips: auto passengers, transit passengers, and total persons. The same basic program was used for both runs, however.

It was necessary to prepare a preliminary program in order to group the single card records into groups of 24, which is the maximum number of cards within buffer capacity, and to separate them into vehicle or person categories. Included in this preliminary program was an editing routine, a volume-classification routine, and a check-sum routine. The editing routine rejected any cards with alphabetic information, for example, over-punches and inconsistent zone numbers. The volume classification routine classified the 1948 volumes of trips into 20 volume classes. The check-sum routine summed all the volumes by modes and zones and compared the totals with a 254-card summary deck prepared independently. The preliminary program also permitted the preparation of the main program while the input data were being compiled. Any changes in the arrangement of the data taken from the cards could be taken care of in the first program without affecting the main program.

The program was designed to have all the final output written on tape for subsequent printing. It was found that the machine would be slowed down a great deal if a printer were connected "on line." As originally designed, there was one line of printed information for each of 20 volume classes times 6 modes of travel, one line for each of 254 zones times 6 modes, and one line for each over-all citywide volume for each of the 6 modes. For all of the methods tested, with their iteration, this amounted to 52,800 printed lines, or almost 1,760 pages—a real data-processing problem.

TEST RESULTS

The initial run of the computer was made for vehicle trips by passenger cars, trucks, and total vehicles. The 1948 zone-to-zone movements were projected to 1955 by a uniform factor, by the average factor method, by the Detroit method, and by the Fratar method. These projected or "forecasted" results were compared with the measured 1955 volumes and the differences were squared, accumulated, and used to compute a root-mean-square error for the average movement, for the various volume classes and for the individual zones.

These results are shown in Table 1 for the three types of vehicle trips. So far as number of trips is concerned, the errors are not large, considering that the sample was as small as 1 in 30 for an important part of the data, and in no case larger than 1 in 10. On a percentage basis, however, the errors are very large.

When the results of the first computer run became available, the question immediately arose whether the errors were primarily attributable to the forecasting methods being tested or to the preponderance of low-volume zone-to-zone movements, which are known to lack accuracy or stability at the sample rates used.

This problem was attacked by two methods. One was by a systematic enlargement of the zones to increase the volume of the zone-to-zone movements and then testing these larger volumes through the computer program. The other method was to determine the percentage distribution of the zone-to-zone trip volumes within the city and by statistical techniques to determine the accuracy that might be expected in the original trip expansion. If the 1948 zone-to-zone trip volumes as expanded from the sample were unreliable, the error would be carried on into the forecast data, and if the 1955 expanded volumes were also unreliable, the result could be to compound the effect of the errors due to sample variability in comparing the forecasts with the 1955 data.

Enlarging Zones

Inasmuch as zone boundaries are chosen from land-use and geographic features, the number of trip ends in each zone is not uniform. In this study the variability was intensified by the fact that the area had to be rezoned so that it would be identical in both years, and the trip ends in the individual zones vary over wide limits. For example the number of 1948 passenger-vehicle trip ends averaged 6,900 per zone but varied from as little as 193 to as much as 59,870. As the initial step, therefore, adjacent zones were combined until each zone group had a minimum of 10,000 passenger-vehicle trip ends in 1948. To minimize the effect of sample variability on the errors, this procedure was repeated to accumulate a minimum of 20,000 trip ends per zone group, and again to accumulate a minimum of 30,000 trip ends per group, then to divide the entire area into 7 groups and finally into 2 groups.

TABLE 1

ROOT-MEAN-SQUARE ERROR IN THE NUMBER OF ZONE-TO-ZONE TRIPS FORECASTED FOR 1955 FROM 1948 DATA, COMPARED TO 1955 SURVEY RESULTS, EXPRESSED IN NUMBER OF TRIPS AND PERCENTAGE

Approximation Number	Numerical RMS Error				Percent RMS Error[1]			
	Uniform Factor	Average Factor	Detroit	Fratar	Uniform Factor	Average Factor	Detroit	Fratar
				Passenger Cars[2]				
1	165	133	234	140	151	136	192	140
2		132	129	131		136	133	134
3		133	148	132		136	143	134
4		134	129	132		137	133	135
5		134	136	132		137	137	135
6		135	131	132		138	134	135
7		135	133	132		138	135	135
				Trucks[2]				
1	18	57	59	55	163	160	172	162
2		55	58	55		160	161	161
3		55	55	55		161	163	161
4		55	56	55		162	161	161
5		55	55	55		162	162	161
6		55	55	55		162	161	161
7		56	55	55		162	162	161
				Total Vehicles[2]				
1	174	137	229	138	141	124	175	125
2		133	131	130		122	120	120
3		133	144	131		121	128	121
4		133	129	131		122	119	121
5		133	134	131		122	122	121
6		133	130	131		122	120	121
7		133	132	131		122	121	121

[1] Calculated by determining the error in the various volume groups and weighting the error in each group in proportion to the percentage of all trips in that group.
[2] Calculated on the basis of number of zone-to-zone movements that had more than 0 trips in either 1948 or 1955.
Passenger cars average zone-to-zone volume = 84,
Trucks average zone-to-zone volume = 28;
Total vehicles average zone-to-zone volume = 90.

The number of zones in these successive groupings, the number of 1948 passenger-car trip ends in the average zone, and the average number of area-to-area trips are as follows:

Number of Areas	Number of Trip Ends Per Area: Minimum	Average	Number of Area-to-Area Possibilities	Average Number of Area-to-Area Trips
254 zones	193	6,900	32,385	27
122 groups	10,000	14,400	7,503	116
66 groups	20,000	26,600	2,211	394
49 groups	30,000	35,800	1,225	711
7 groups	214,305	250,000	28	31,092
2 groups	731,000	870,000	3	290,192

Test Results of Enlarged Zones

The results of tests of each forecasting procedure for the various zone groupings are shown in Table 2. The average-factor method, the Detroit method, and the Fratar method each reach essentially the same minimum error although the Fratar method reaches this minimum in the second approximation, whereas more iterations are generally required for the other methods.

The minimum percentage error, by any of the methods tested after any number of iterations, for the various zone groups and the average number of area-to-area passenger-car trips were as follows:

Number of Areas	1948 Average Number of Area-to-Area Trips	Minimum Percent Error
254 zones	27	133
122 groups	116	70
66 groups	394	41
49 groups	711	34
7 groups	31,092	14
2 groups	290,192	11

In the case of the 2-group division a second test was made by dividing the area with a line roughly at right angles to the first. The minimum percent error for this second grouping was the same as that for the first, to the nearest percent (11 percent).

TABLE 2

ROOT-MEAN-SQUARE ERROR OF THE FORECASTED 1955 PASSENGER CAR TRIPS FOR THE 254 ZONES AND FOR DIFFERENT ZONE GROUPINGS

Approximation Number	254 Zones Error in Trips Number	Percent[1]	122 Groups Error in Trips Number	Percent[1]	66 Groups Error in Trips Number	Percent[1]	49 Groups Error in Trips Number	Percent[1]	7 Groups Error in Trips Number	Percent[1]
					Uniform Factor Method					
	165	151	415	98	499	57	747	49	19,641	38
					Average Factor Method					
1	133	136	244	77	458	49	655	42	10,180	20
2	132	136	204	72	364	44	542	37	8,300	16
3	133	136	196	71	344	43	517	36	7,810	15
4	134	137	194	71	338	42	509	35	7,690	15
5	134	137	193	71	337	42	507	35	7,540[2]	15
6	135	138	193	71	336	42	506	35		
7	135	138	193	71	336	42	506	35		
					Detroit Method					
1	234	192	388	89	720	64	871	51	10,300	20
2	129	133	228	74	396	46	543	37	8,960	17
3	148	143	299	74	344	42	504	35	7,700	15
4	129	133	194	70	337	42	484	34	7,510	15
5	136	137	196	70	320	41	478	34	7,500	15
6	131	134	189	70	326	41	480	34	7,570	15
7	133	135	190	70	319	41	476	34	7,480	14
					Fratar Method					
1	140	140	205	71	339	42	498	35	7,360	14
2	131	134	188	70	322	41	480	34	7,460[2]	14
3	132	134	188	70	322	41	478	34		
4	132	135	188	70	321	41	479	34		
5	132	135	188	70	322	41	478	34		

[1] A weighted percent error obtained by determining the percent error in each volume class and weighting this error by the proportion of trips in that volume class. Not applicable for the 7 zone group because all volumes were in the largest volume class.
[2] No further iterations required because new F factor for all zones was 1.00.

The relationship between average area-to-area trip volume and the minimum percent error is shown in Figure 3. As the minimum error for the three iterative methods is about the same, the chart can be considered as applicable to any one of them. This chart is difficult to interpret because part of the error is due to sample variability and part is due to the projection method being tested. Differences in sampling rate introduce a further complication. In the 1948 survey the sampling ratio was 1:20 throughout the area, while in the 1955 survey it was 1:30 for the District of Columbia and 1:10 for the Maryland and Virginia suburbs. However, the curve should give some indication of the error to be expected in using any of the three iterative methods where the sampling rate is about the same as the average for the two Washington surveys, that is, about 5 percent.

The shape of the curve suggests that it will level off at about 10 percent. In other words an error of about 10 percent seems to be inherent in the methods tested, regardless of size of sample or areas.

Rate of Closure. A measure of the efficiency of the various forecasting methods is the rapidity with which the individual zone growth factors converge toward the limiting F factor of 1.00 in successive iterations. The difference between the computed F factor at the end of an iteration and 1.00 is the factor residual error that remains in the individual zones.

The factor residual error for the 254 zones is shown in the following tabulation for the various iterations of the three methods. The first column indicates the method and the second column indicates the approximation number. The third column shows the percent of the zones that have no residual error (new F factor = 1.00) at the end of the approximation shown in the second column. The fourth column indicates the percent of zones with a residual error less than 0.01 (new F factor between 0.99 and 1.01). The next four columns similarly show the percent of zones with residual errors less than 0.02, 0.03, 0.05, and 0.10. The last column shows the percent of zones with residual errors greater than 0.10 (new F factor less than 0.90 or more than 1.10).

Method	Approximation No.	Percent of zones with a factor residual error of—						
		0.00	Less than 0.01	Less than 0.02	Less than 0.03	Less than 0.05	Less than 0.10	0.10 and over
Average factor	1	1	8	11	17	26	47	53
	2	6	15	24	35	50	75	25
	3	10	30	44	55	77	93	7
	4	19	47	71	84	98	99	1
	5	33	70	88	93	98	99	1
	6	49	84	92	94	98	100	0
	7	64	92	97	99	100	100	0
Detroit	1	2	4	9	14	23	44	56
	2	1	3	11	15	22	57	43
	3	4	13	25	37	63	95	5
	4	5	18	38	58	95	98	2
	5	10	36	68	93	98	99	1
	6	16	62	95	97	98	100	0
	7	28	85	98	99	100	100	0
	8	37	96	99	100	100	100	0
Fratar	1	6	20	33	54	68	79	21
	2	60	97	100	100	100	100	0
	3	98	100	100	100	100	100	0
	4	99	100	100	100	100	100	0

As can be seen from this table the Fratar method is extremely efficient in its rate of closure. Since the F factor must be obtained for each new iteration and since these new F factors may be easily summarized, it is suggested that they be used to indicate the desirability for additional iterations.

Number of Iterations Required. From the tests that have been run, the minimum root-mean-square error has always been reached in the second approximation by the Fratar method. By the other methods, however, this minimum error may not be reached until the fourth or fifth approximation. There is also a possibility that an unusual set of growth factors will develop that will not close as rapidly as those occurring in the test data.

Considering the division of the Washington, D.C., area into 49 zone groups with a minimum of 30,000 passenger-car trip ends per group, the factor residual error was accumulated for all groups at the end of each iteration. The accumulated residual error was then divided by the number of groups to obtain the average residual error per group. This average residual error was then related to the RMS error already computed for each approximation. The following table indicates the results:

Approximation No.	Average Factor		Detroit		Fratar	
	Average residual error	RMS error	Average residual error	RMS error	Average residual error	RMS error
		(percent)		(percent)		(percent)
1	0.084	42	0.123	51	0.035	35
2	0.034	37	0.070	37	0.003	34
3	0.014	36	0.032	35	0.001	34
4	0.007	35	0.022	34	[1]	34
5	0.003	35	0.014	34	[1]	34

[1] Less than 0.001

To be reasonably certain that greater accuracy cannot be obtained with additional iterations, it is suggested that iterations be continued until the average residual error per zone is less than 0.01.

The computer time required for each method, however, is not uniform but is approximately related to the complexity of the method. During the test, the computer time as recorded for each iteration of each method and adjusted proportionately to a common base of 10,000 zone-to-zone movements is as follows:

TABLE 3

PERCENT ROOT-MEAN-SQUARE ERROR BY VOLUME CLASS OF ZONE-TO-ZONE MOVEMENTS—PASSENGER VEHICLES

1948 Volume Class	254 Zones	122 Zone Groups	66 Groups	49 Zone Groups
20-29[1]	195.0	121.0	95.6	106.1
30-39	222.1	101.4	86.7	83.4
40-49	279.7	118.1	81.5	73.4
50-59	141.0	96.3	82.3	89.4
60-69	139.9	107.5	92.6	89.7
70-79	108.4	74.2	78.2	76.2
80-89	154.3	95.2	94.9	54.4
90-99	128.2	83.0	54.8	73.2
100-199	133.4	93.3	64.8	64.8
200-299	80.5	69.7	58.6	52.1
300-399	60.7	64.0	56.0	43.8
400-499	54.9	56.6	50.9	38.4
500-599	42.6	42.3	45.9	44.1
600-699	51.2	40.6	46.7	48.6
700-799	88.8	41.1	31.1	26.5
800-899	38.9	26.5	57.2	31.6
900-999	18.4	21.4	33.1	38.4
Over-1,000	28.6	27.5	21.2	25.3

[1] Because of home interview sampling ratio of 1.20 in 1948, errors for volume classes below 20 cannot be accurately appraised.

Computer time per iteration of 10,000 area-to-area movements:

Average factor method - 6 minutes
Detroit method - 9½ minutes
Fratar method - 12 minutes

From this, the computer time required for the average factor method is half that required for the Fratar method. These times, however, include the computer time needed to develop and store the various statistical measures. In an ordinary forecasting procedure these measures would not be required and the above times would be reduced by a constant but indeterminate amount. The average factor method should therefore require something less than half the time per iteration required by the Fratar method.

Since the RMS error for the average factor method at the end of four approximations is about equal to the RMS error for the Fratar method at the end of two approximations, the over-all computer time required for equal RMS accuracy is about the same. However, the rate of closure of the Fratar method is more than twice as rapid as the average factor method and it would therefore appear to be the preferred method.

Percent RMS Error for Accumulated Volumes

The data presented thus far have to do only with the error for the area-to-area volumes. As has been shown, the average volume between zones of the size ordinarily used is relatively small and the percent root-mean-square error is, roughly speaking, correspondingly large.

In actual practice, the individual zone-to-zone volumes are assigned to the highway network and, therefore, each portion of the highway network represents an accumulation of zone-to-zone volumes. The volumes assigned to the highway network are our primary concern. The errors to be expected in such accumulated volumes can only be determined from actual tests and these have not yet been made. However, some indication of the magnitude of the errors to be expected can be obtained from purely theoretical considerations.

From a statistical standpoint, if the percent error of an average zone-to-zone volume is X, the percent error of a group of average zone-to-zone volumes is $\frac{X}{\sqrt{N}}$ where N is the number of individual zone-to-zone movements in the group. By dividing 10,000 by the average zone-to-zone volume for each of the zone groupings, the number of zone-to-zone movements (N) required to accumulate to a volume of 10,000 can be determined and the percent RMS error of the group can therefore be calculated.

The above relationship holds true only if the mean error of the group is zero. If the movements, however, are heavily weighted by trips from an individual zone, as they would be in the case of ramp volumes, the factor residual error (as previously explained) may be appreciable in the early iterations.

For example, if the trips on a ramp are essentially from two zones and these two zones have an average F factor for the next iteration of 1.30, the summation of trips into and out of the zones at the end of the present iteration are too low. The total number of trips for a group of zone-to-zone movements from these zones, therefore, will have a tendency to approach a volume which should be increased by 30 percent. To take this error into account, the root-mean-square of the residual errors was determined for each iteration of each method. The number of zones required to provide a volume of 10,000 was determined and the root-mean-square residual error was added to the RMS error for individual zone-to-zone movements by taking the square root of the sum of the squares of the two errors to determine the total error. The results of this test are shown in Figure 4.

Assuming that this chart has some validity with reference to the problem, it indicates that the error for a volume of 10,000 trips is within acceptable limits and that it does not make too much difference what size zone group is used although a minimum error was obtained for the zone grouping which had 10,000 trip ends per group. Manifestly, however, this conclusion is dependent, to a substantial degree, on statistical inference and should be subjected to an actual test before it can be fully accepted.

Distribution of Zone-to-Zone Volumes

Even though the accumulated volumes of 10,000 or more apparently will have errors of rather modest proportions, it is desirable to inquire into the reasons for the inaccuracies of the movements between zones as they were originally planned and subsequently enlarged.

The test program was set up to count the number of zone-to-zone movements in each of several 1948 volume classes as previously described. This procedure was followed for the original 254 zones and for the subsequent groupings to 122 areas, 66 areas, and 49 areas.

The results of this test are shown in Figure 5 for passenger cars (including taxis). About 93 percent of the movements between the 254 zones have a volume of less than 100. When the number of areas was reduced to 122, about two-thirds of the area-to-area movements were less than 100; with 66 areas about one-fourth of the movements were less than 100; and with 49 areas about 10 percent were less than 100. Also, the number of area-to-area movements which are less than the mean exceeds the number that are greater than the mean showing that the distribution is skewed. This is true for each of the zone groups although slightly less pronounced as the number of areas is successively decreased.

The test program also permitted the determination of the percent of the 1955 trips that were accumulated in each of the 1948 volume classes (Fig. 6). Fifty percent of

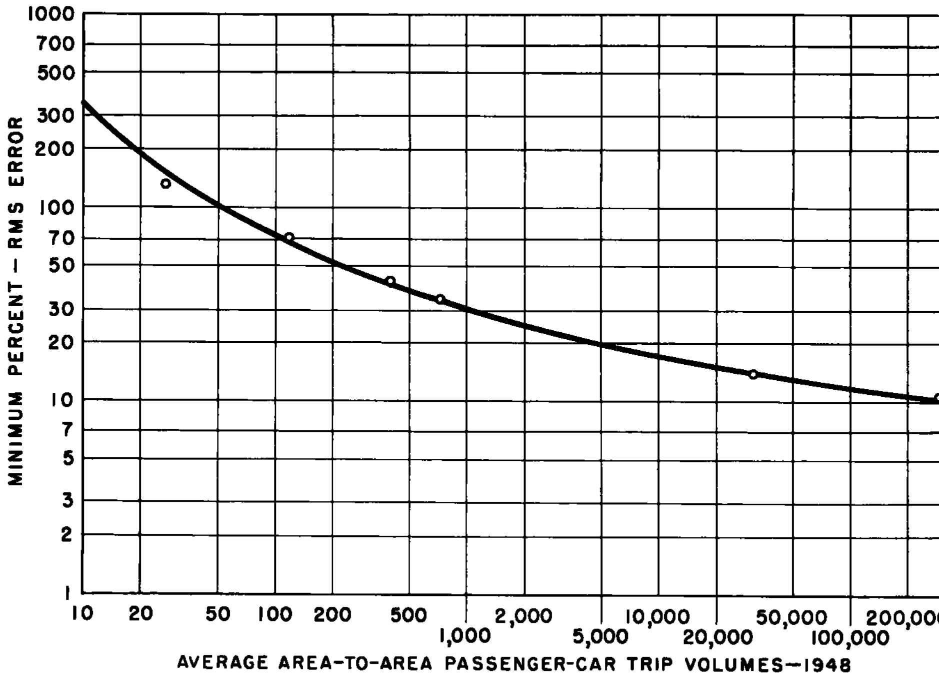

Figure 3. Relationship between average 1948 passenger car trip volume and the minimum percent RMS error.

the 1955 passenger-car trips were made between zone pairs that, in 1948, had a volume of less than 100 passenger-vehicle trips per day. Values for other zone groups and other 1948 volumes can be read from the chart.

In summation then, the preponderance of zone-to-zone movements within a metropolitan area is exceedingly small but because of the large number of such movements, they do, in the aggregate, account for a substantial portion of the present or predicted trips.

Prediction Error as Related to Zone-to-Zone Volume

There is, of course, no a priori reason why small zone-to-zone volumes cannot be forecast with accuracy equal to large zone-to-zone volumes. The converse would likely seem true, that the forecast error should be independent of the volume.

However, there is a priori probability that the error in the expanded number of trips from a "sample" survey is inversely proportional to the number of "sample" trips interviewed as shown below.

Of all trips into and out of zone i there is a certain proportion (p) that will be from or to zone j. Therefore, (p) is the proportion of trip ends in zone i with the other end of the trip in zone j and i-p (equals q) is the proportion of trip ends in zone i that do not have the other end of the trip in zone j.

If (s) trips with an end in zone i are reported by interview, the probable number ($\bar{X}$) with the other end in zone j is (s.p). Mathematically: $\bar{X} = sp$

From any standard statistical text it can also be shown that the standard deviation (σ) of the number of trips reported between zone i and zone j from sample interviews made of trips with one end in zone i is $\sqrt{spq}$. Mathematically: $\sigma = \sqrt{spq}$

The standard deviation (σ) as a proportion of the expected number ($\bar{X}$) is:

$$\frac{\sigma}{X} = \frac{\sqrt{spq}}{sp} = \sqrt{\frac{q}{sp}}$$

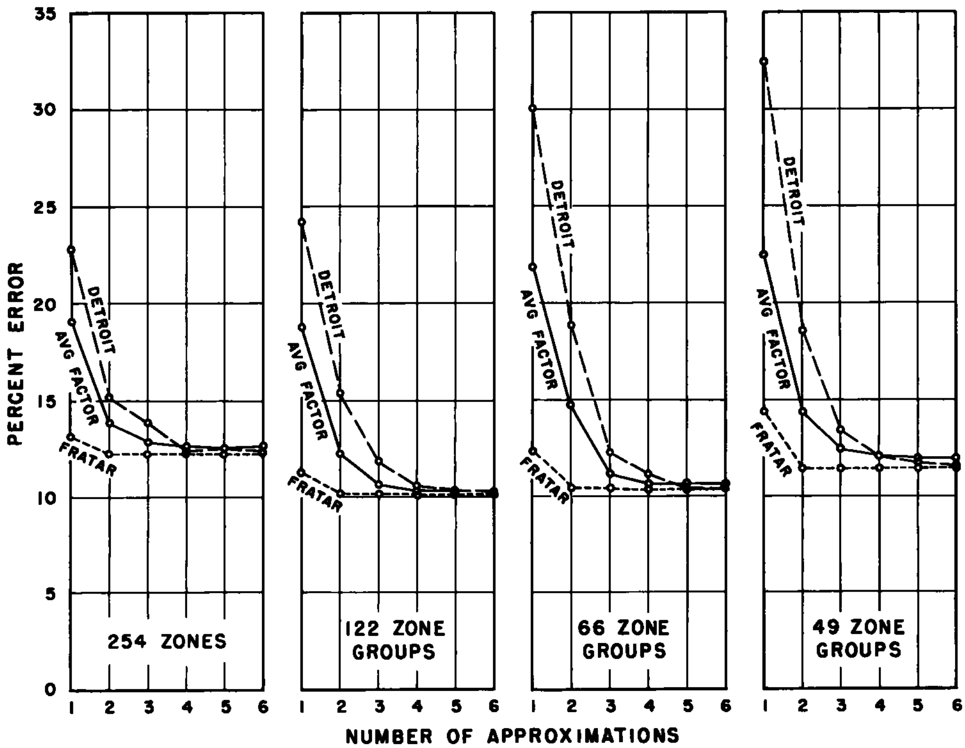

Figure 4. Percent error to be expected for 10,000 trip volumes based on theoretical conditions.

Since p and q are constants for any pair of zones, the percent error varies inversely with the square root of the number of trips between zone i and zone j obtained by interview.

For 254 zones p will have an average value of $\frac{1}{254}$ and q of $\frac{253}{254}$. Thus,

$$\frac{\sigma}{\bar{X}} = \sqrt{\frac{\frac{253}{254}}{\frac{s}{254}}} = \sqrt{\frac{253}{s}}$$

Again the average zone had 6,900 trip ends in 1948, of which about 345 were obtained by interview, so on the average s = 345

$$\frac{\sigma}{\bar{X}} = \sqrt{\frac{253}{345}} = 0.86$$

Thus, for an average movement in 1948 a standard deviation percent error of 86 percent could be expected.

Since the test of forecasting procedures uses the reported 1948 trips as the base data for calculating the 1955 trips and then compares the result with the reported 1955 trips, it would be expected that the error would be increasingly large for the smaller grip volumes, not because of errors in the forecasting procedure but due to sample variability.

The RMS error in the various 1948 volume classes for passenger vehicles in the seventh approximation by the Fratar method behaves in this manner as is shown in Table 3 for each of the zone groups.

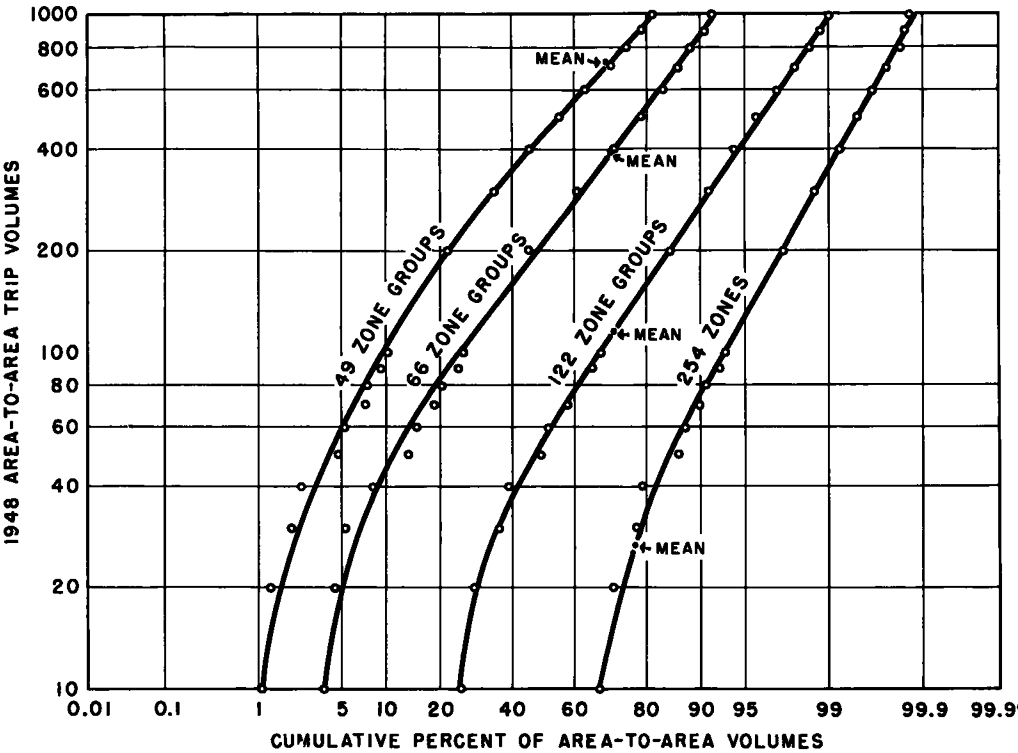

Figure 5. Frequency of 1948 area-to-area trip volumes with various zone groupings. (passenger cars)

RECOMMENDED PROCEDURES

The primary purpose of forecasting zone-to-zone volumes is for the selection and assignment of trips to a transportation network. To do this with reasonable accuracy particularly at each ramp on an expressway network, it is imperative that the zones be of a size consistent with the distance between ramps. Thus, while increasing the size of the zone increases the accuracy of predictions of the zone-to-zone movements, this procedure adversely affects the primary purpose of forecasting. Pending further studies as outlined at the end of this article, it is recommended that zones be established in accordance with present practice and that the zone-to-zone movements be forecast by the Fratar method.

Figure 7 is a flow chart for accomplishing this forecast on an electronic computer. This chart is made up for an input of zone-to-zone volumes as file A and a combination input of present trip ends, future trip ends and growth factors by zones as alternate inputs for file B. The program includes appropriate editing routines for checking the present trip ends and the growth factors if these are available independently from the zone-to-zone volumes in file A. The switches shown on the flow chart are programmed transfer points resulting from decisions made at a remote point in the program. They are not switches on the console of the computer.

A program written from this flow chart includes the computation of the frequency of the new F factors, to be used as a guide for determining the need for continuing additional iterations.

FUTURE RESEARCH

The work to date indicates that the three iterative methods are equally accurate in computing the future trips but the Fratar method arrives at the minimum error in fewer

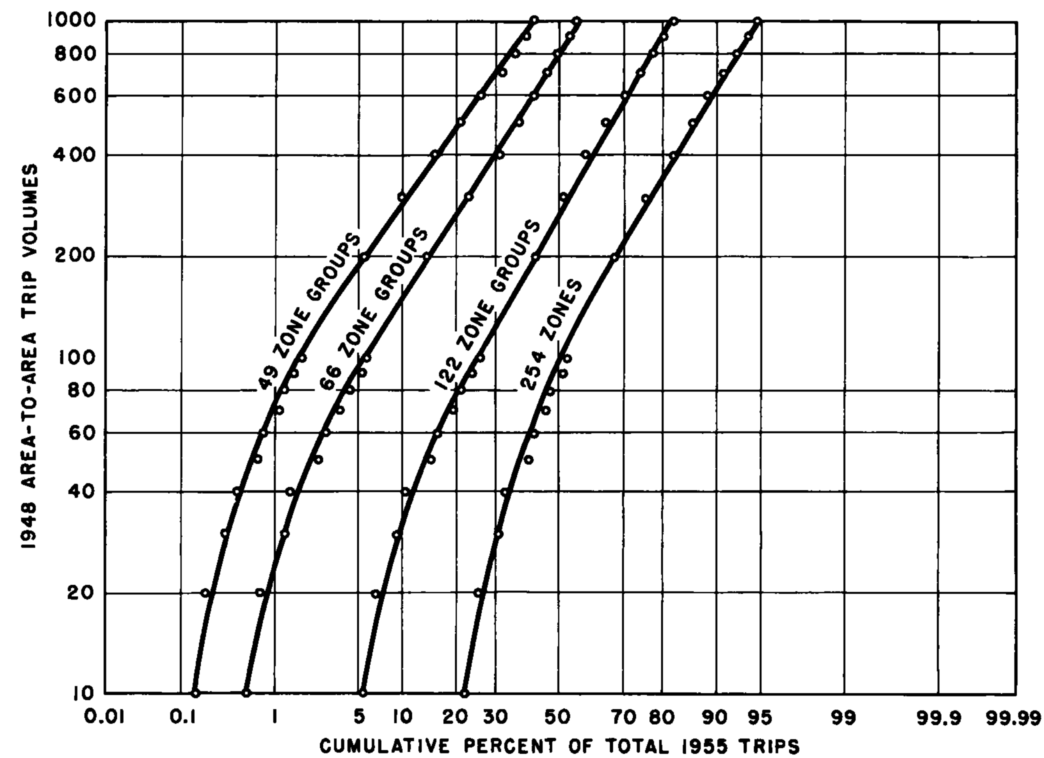

Figure 6. Cumulative percent of 1955 trips by 1948 area-to-area trip volumes with various zone groupings.

approximations. It is also more efficient in its rate of closure.

In addition it has been found that the preponderant small volume movements are individually affected by an inherent sample variability. The summation of these movements accounts for a substantial portion of the total trips. It is also known that the small volume zone-to-zone movements are, on the average, the longer trips within the city that account for proportionately more vehicle-miles of travel and are also the trips most likely assignable to high-type highway facilities.

Accumulating Trips Across A Grid

However, it is not necessary that the individual zone-to-zone movements be accurate if a summation of these volumes crossing the city is reasonably representative of the actual travel. To determine this relationship it is proposed that each zone-to-zone movement be traced from the X and Y coordinates of one zone to the X and Y coordinates of the other. Each time this trace intersects a predetermined section of a grid line the number of 1955 trips for this zone-to-zone movement as projected from the 1948 data and the number as determined from the 1955 survey are "remembered." When all zone-to-zone movements have been similarly traced, the number of "remembered" trips over an appropriate interval of each grid line is totaled. The comparison of the total number of trips projected from the 1948 data with the total number determined from the 1955 survey will give a measure of the accuracy of the projection.

The use of the trace principle automatically "weights" the longer trips properly in that longer trips will cross many grid lines whereas the short trips may cross none or only a few. Further, by choosing an appropriate length along each grid line, the summation of trips to various volumes can be accomplished.

While the trace method appears as a difficult manual task it is comparatively simple to program for an electronic computer and would require about the same computer time as a single iteration of the Fratar method if the number of grid lines is held to a reasonable minimum.

Stabilizing the Small-Volume Movements

Another research project which should be undertaken is the testing of methods for stabilizing the small-volume zone-to-zone movements.

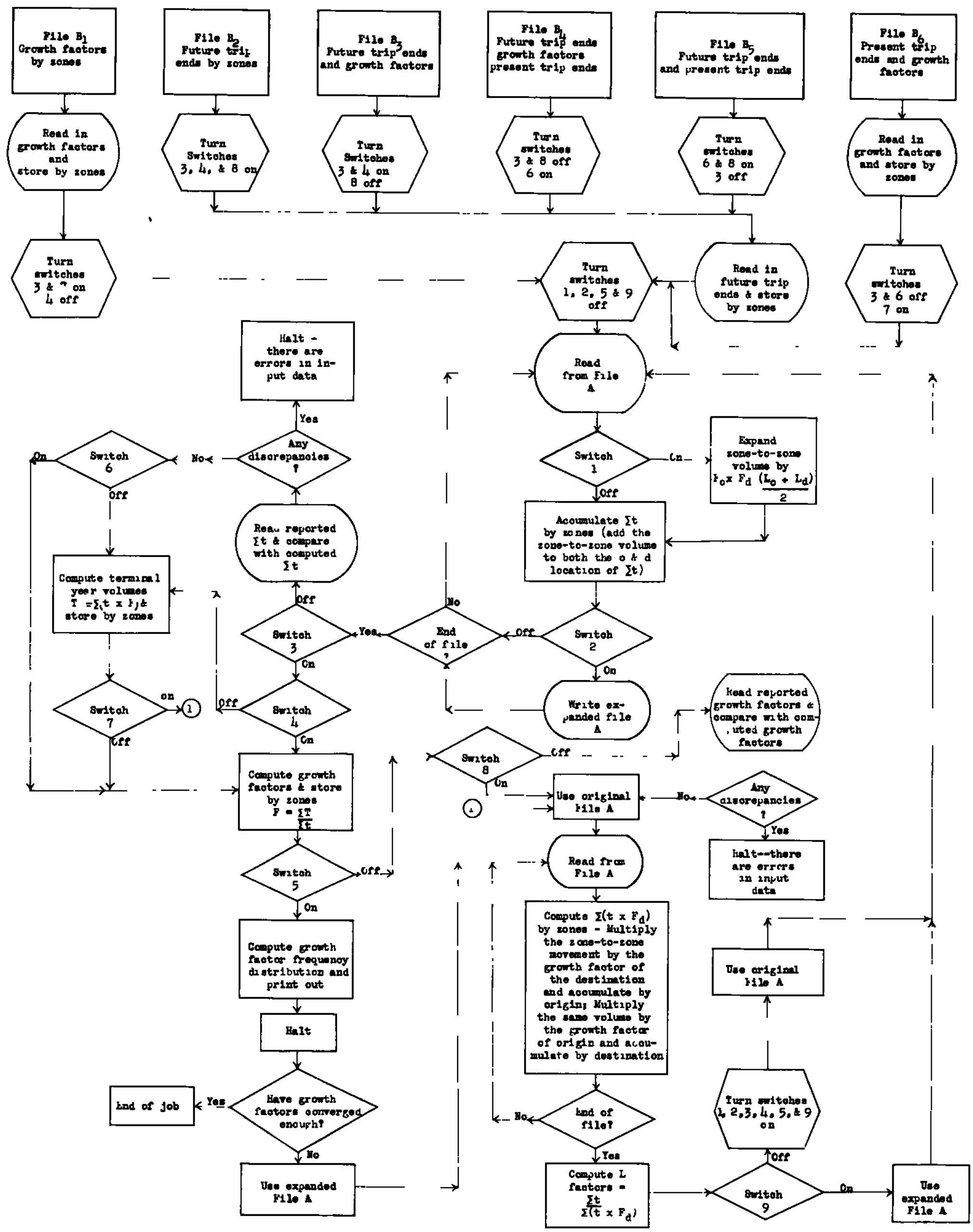

Figure 7. Electronic computer flow chart for forecasting future zone-to-zone trips fratar method.

One of the more difficult problems in forecasting traffic is that of predicting the future number of trips for the zone-to-zone movements that are zero at the present time since all methods tested require the multiplication of existing trips by various factors. The magnitude of this problem can be most easily visualized by remembering that about 67 percent of all possible 1948 zone-to-zone movements in the Washington, D.C., survey were zero and these same zone-to-zone movements account for 22 percent of all 1955 trips.

In addition to the zero volumes, other low volumes are also inherently inaccurate. For an understanding of this problem it is necessary to again resort to statistical formulas. The proportional error that may be expected in random sampling of all of the trips from one zone to a particular other zone is given by the following expression which has been previously defined:

$$\frac{\sigma}{\bar{X}} = \sqrt{\frac{q}{sp}}$$

For example, suppose that zone i has a total of 6,000 trip ends. Of these 6,000 suppose that 666 have their other ends distributed as follows: 600 in zone j, 60 in zone k, and 6 in zone 1. With a 1 in 20 sample the 6,000 trip ends in zone i will represent 300 trip ends obtained by interview; therefore s = 300. Considering all of the trip ends in zone i, the probability of any trip end having the other end in zone j is 600 out of 6,000 or 0.1. Therefore p = 0.1 and q = 1 - p = 0.9. Thus for the trips from zone i to zone j the proportional error

$$= \sqrt{\frac{q}{sp}} = \sqrt{\frac{0.9}{300 \times 0.1}} = 0.17 .$$

Thus with a one-twentieth sample, a standard deviation accuracy of 17 percent for the trips between zone i and zone j could be expected. Similar values for zones k and 1 are shown in the following table:

Zone-to-Zone	Volume	s	p	q	Proportional error	Percent error
i to j	600	300	0.1	0.1	0.17	17
i to k	60	300	0.01	0.99	0.57	57
i to 1	6	300	0.001	0.999	1.83	183

If the sample rate is 1 in 20, it is manifestly impossible that the 6 trips between zone i and zone 1 will ever be reported as 6 trips. However, from the expansion of the binomial $(p + q)^s$, the frequency distribution of the reported trips between zone i and zone k and between zone i and zone 1, is estimated as follows:

	Probability of Reported Value—	
Reported Number of Trips	Zone i - zone k	Zone i - zone 1
0	0.05	0.74
20	0.15	0.22
40	0.22	0.03
60	0.23	0.01
80	0.17	-
100	0.10	-
120	0.05	-
140	0.02	-
160	0.01	-

Thus the 6 trips between zone i and zone 1 are never correctly reported; 74 percent of the time they are reported as zero; and the remaining 26 percent of the time they are reported as 20 or more trips. Similarly the 60 trips between zone i and zone k are reported as 60 only 23 percent of the time, while the remaining 77 percent of the time the reported trips are in error by more than 33 percent.

Adjacent to zone i there will be other zones i_2 and i_3. To illustrate, it can be assumed that zone i_2 has 8,000 trip ends and zone i_3 has 6,000 trip ends; therefore since zone i has 6,000 trip ends, in zones i, i_2, and i_3 there would be a total of 20,000 trip ends which would represent 1,000 interviews.

Since zones i_2 and i_3 are adjacent to zone i, it is probably true that their movements to zone 1 are similar in volume to the movements from zone i to 1. This assumption is justified by the high correlation of distance and trip volume as established by Carroll and others. If this be true it could be assumed that $\frac{1}{1,000}$ of all trip ends in zone i_2, and i_3 are to or from zone 1 as was the case with zone i, i.e., the 20 trips between zone 1 and the three i zones are divided 30 percent to zone i, 40 percent to zone i_2 and 30 percent to zone i_3. If the expansion process $(p + q)^s$ is again used to obtain the probability of the zone-to-zone movements, the results are shown in the following table:

Reported Number of Trips to Zone 1—

Total	From i	From i_2	From i_3	Probability
0	0	0	0	0.37
20	6	8	6	0.37
40	12	16	12	0.18
60	18	24	18	0.06
80	24	32	24	0.02

From a comparison of the above table with the previous one it appears that grouping of zones will improve the accuracy of the low-volume movements. This process can be continued for other groupings. However, instead of using this rather time-consuming computation, the proportional error can be obtained by the simpler equation $\frac{\sigma}{\bar{X}} = \sqrt{\frac{q}{sp}}$ which does not indicate the frequency of the various movements but does, in one operation, compute the resulting error.

The trips between zones i, i_2, and i_3 combined and zone 1 will have a proportional error of

$$\frac{\sigma}{\bar{X}} = \sqrt{\frac{999}{1000 \times .001}} = \sqrt{.999} = 1.00 \text{ or } 100 \text{ percent.}$$

Thus by grouping three zone-to-zone movements the standard-deviation error has been reduced from 186 percent to 100 percent assuming that the distribution of trips between zone 1 and zones i, i_2, and i_3 is proportional to the distribution of trip ends in the three i zones. It is true, of course, that this assumption will not be exactly correct, but it seems likely that the error of this assumption will be less than the error added by the individual zone sample variability.

The problem of how large a grouping is desirable can be approximated from the equation: $\frac{\sigma}{\bar{X}} = \sqrt{\frac{q}{sp}}$ The limiting value of q is 1.00. It can never be as large as 1.00 and in almost all applications it is not smaller than 0.9. At the same time sp is equal to the number of trips obtained by interview between a pair of zones with perfect sampling. The value of sp can range from zero up to the value of s. The relation between the error of a zone-to-zone movement and the number of interviewed trips making this zone-to-zone movement can be approximated as follows:

Number of Interviewed Trips	Percent Error
0	α
1	100
2	71
5	45
10	32
15	26
20	22
30	18

Considering the undesirability of combining too many zones, it appears that zone-to-zone movements might well be grouped until the accumulated movement represents about 10 interviews.

A method to accomplish this purpose has been worked out but not tested except by hand computation from a small sample. The sample computation indicated that the error is reduced by approximately one-third.

The method requires the use of the binary system in coding a group of zones. For illustrative purposes, a group of 4 zones can be considered although in actual practice probably 16 would be required.

To illustrate, suppose that in region A the area is divided in half with 2 zones in each half. One-half of the zones would be designated AO and the other half A1. These pairs of zones would again be divided in half or into single zones designated A00, A01, and A11. A separate region B would be similarly separated into B00, B01, B10, and B11.

If only trip volumes that represent 10 interviews (or 200 trips with a $\frac{1}{20}$ sample) are considered sufficiently stable so as not to warrant readjustment, a method of combining zones that is amenable to computer operation is required. A suitable method is as follows:

The number of trips from A00 to B00 is examined. If it is less than 200 (10 interviewed trips), combine zone B00 and B01 and find the number of trips between A00 and B00 + B01. If it is still less than 200, combine zone A00 and A01 and find the number of trips between A00 + A01 and B00 + B01. If the number of trips is still less than 200, again double the B area to include 4 zones, and if necessary, double the A area to include 4 zones, and so on until the figure 200 is reached.

The advantage of the binary coding is that it provides the desired grouping through a simple arithmetic operation. The arithmetic operation simply combines the binary portion of the region code by alternate digits. For example if A00 is written as AA_1A_2, and B00 is written as BB_1B_2 the combination is written $A_1B_1A_2B_2$. Starting with 0000, the digit 1 is added successively until the sum 1111 is obtained. The subtotals are then decoded by the $A_1B_1A_2B_2$ pattern and the zone-to-zone movements are then in the proper order for combining as follows:

Combination	A zone	B zone	Combination— 1	2	3	4
0000	00	00				
0001	00	01	x			
0010	01	00				
0011	01	01	x	x		
0100	00	10				
0101	00	11	x			
0110	01	10				
0111	01	11	x	x	x	
1000	10	00				
1001	10	01	x			
1010	11	00				
1011	11	01	x	x		
1100	10	10				
1101	10	11	x			
1110	11	10				
1111	11	11	x	x	x	x

From the original zone-to-zone volumes and the combination totals, the preselected volume can be determined.

The combination volume is then reassigned to individual zone-to-zone movements as follows: Suppose that in the previous example all 4 of the A zones were combined and all 4 of the B zones were combined to provide a preselected volume. The total trip ends in the individual zones in the A group are added and the proportion of the total in each A zone is determined (P_{A1}, P_{A2}, etc.). Similarly, each zone of the B group is a certain proportion of the B total trip ends (P_{B_1}, P_{B_2}, etc.). Then the total volume (V) between the group is reassigned to individual zone-to-zone volumes ($V_{A_1-B_1}$) by the equation:

$$V_{A_1-B_1} = V \cdot P_{A_1} \cdot P_{B_1}$$

$$V_{A_1-B_2} = \cdot P_{A_1} \cdot P_{B_2}$$

etc.

From sample tests made to date this reassignment procedure appears to improve the accuracy of predicting the future zone-to-zone trip movements. Whether it improves the accuracy of predicting accumulated volumes such as would occur on road sections or ramps should be tested by the process of accumulating the trips across a grid, as described in the preceding section.

Discussion

THOMAS J. FRATAR, Partner, Tippetts-Abbett-McCarthy-Stratton—When the successive approximation, or iterative, method for the distribution of future zonal and intra-zonal traffic was developed it was not assumed that the attractiveness factor selected for the initial application to Cleveland area traffic was the only one that could be used, or necessarily the best. In fact it was recognized that further experimentation would be needed.

Those familiar with the Hardy Cross moment distribution procedure used in structural analyses will recognize that the attractiveness factor used in the successive approximation method for traffic distribution is comparable to the stiffness factor used in moment distribution. Messrs. Brokke and Mertz have referred to different attractiveness factor bases as different methods.

It is gratifying to learn from the research of the authors and their associates with the Bureau of Public Roads that the attractiveness factor basis originally selected when the writer initiated the successive approximation technique for traffic distribution remains the favored one.