Contract Research
GDOT Research Project No. 7005
Quarterly Report No. 20

SUPPLEMENT

DEVELOPING A WORK TRIP TABLE FOR THE ATLANTA REGION USING
THE 1970 CENSUS URBAN TRANSPORTATION PLANNING PACKAGE

BY
Michael S. Bronzini
Assistant Professor
School of Civil Engineering
Georgia Institute of Technology

Contract with
Department of Transportation of Georgia

In Cooperation with

U.S. Department of Transportation
Federal Highway Administration

July, 1975
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Atlanta Transportation Planning Data</td>
<td>2</td>
</tr>
<tr>
<td>Experience of Other Urban Areas</td>
<td>3</td>
</tr>
<tr>
<td>II. APPROACHES TO DEVELOPING THE TRIP TABLE</td>
<td>5</td>
</tr>
<tr>
<td>Criteria for Evaluating Alternative Approaches</td>
<td>5</td>
</tr>
<tr>
<td>Approach 1: UTPP Subset</td>
<td>7</td>
</tr>
<tr>
<td>Approach 2: Proportional Trip Reallocation</td>
<td>10</td>
</tr>
<tr>
<td>Approach 3: Separate Trip Table</td>
<td>12</td>
</tr>
<tr>
<td>Approach 4: Gravity Model</td>
<td>14</td>
</tr>
<tr>
<td>III. RECOMMENDATIONS</td>
<td>16</td>
</tr>
<tr>
<td>Summary of Recommendations</td>
<td>18</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>19</td>
</tr>
<tr>
<td>APPENDIX: STATISTICAL TESTS FOR COMPARING TRIP TABLES</td>
<td>20</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

The 1970 Census Urban Transportation Planning Package (UTPP), prepared by the Census Bureau for Standard Metropolitan Statistical Areas (SMSA's), contains work trip information tabulated at the traffic zone level. The package consists of four parts, as follows:

Part I - socioeconomic data compiled by zone of residence;
Part II - areawide characteristics of population and work trips;
Part III - employment data compiled by zone of work;
Part IV - work trips by zone of residence, zone of work, and travel mode.

The original intent of this research project was to obtain the UTPP for Atlanta and use Part IV, the work trip table, as input to a traffic assignment computer program. Relationships between peak-hour traffic volumes as observed in the field (1)* and assigned work trip volumes would then be obtained for various highway types by regression analysis.

The Atlanta UTPP was delivered to the Atlanta Area Planning Branch of the Georgia Department of Transportation (GDOT) in May of 1974, and in their judgement the work trip table is unacceptable for traffic assignment purposes. Hence alternative means of developing the required trip table were investigated. This report presents the results of that investigation. The remainder of section I discusses the Atlanta UTPP data and supplementary data available, and summarizes the experiences of other urban areas in attempting to use the UTPP. Section II presents four alternative approaches which may be used to develop the trip table. Recommendations are given in section III.

*References are given at the end of the report.
Atlanta Transportation Planning Data

The primary deficiency in the Atlanta UTPP work trip table is that only 50% of the work trip destinations are coded to traffic zones (2). The remainder are coded to "dummy" zones, assigned zone numbers in the 9000 series.

The reason for this is the manner in which work trip destinations were originally coded by the Census Bureau (3). Respondents were asked to provide the specific street address for their place of work. An address coding guide (ACG), developed by the Bureau, was then used to code the census block or tract of the work place. Unfortunately, the ACG did not cover the entire Atlanta SMSA, but was limited to the central portion of the urbanized area. Hence SMSA residents whose place of work was not within the ACG area had their trip destination coded only to zip code (ZC) and/or Universal Area Code (UAC). Respondents who did not provide a good or complete work address were treated similarly.

These latter geographic entities (ZC's and UAC's) are not useful for transportation planning purposes. Zip codes have never been mapped by the U.S. Postal Service, and one zip code will typically overlap numerous traffic zones. Universal Area Codes are very large (e.g., one UAC may correspond to an entire county), and thus contain many traffic zones, as well as areas outside the boundary of the transportation study area. Hence it is not possible to assign a ZC or a UAC to a single traffic zone. In order to avoid losing the data entirely, these areas were assigned to the fictitious or dummy traffic zones mentioned earlier.

There are other planning data available for Atlanta which may be useful for overcoming the UTPP deficiencies. These data were collected for use in the transportation and land use components of the Regional Development Plan currently being prepared by the Atlanta Regional Commission (ARC). For the most
part, the data were collected in 1972. The following data files are available (2):

1. External cordon survey;
2. Internal origin-destination (O-D) survey, 1/2% sample;
3. Employment by traffic zone;
4. Approximately 100 socioeconomic variables, also by traffic zone;
5. Trip ends, based upon the O-D survey and socioeconomic data for 1970;

Experience of Other Urban Areas

Other urban areas which have attempted to use the UTPP have encountered difficulties similar to those described above. Nationwide, the percentage of SMSA residents whose place of work responses were coded to the block level ranged from 25 to 85 percent, with an average of 65 percent (3). Those areas which have used Part IV of the package have had to modify it extensively in order to incorporate the trips coded to ZC’s and UAC’s into their work trip matrix.

Zakaria (4) has recently reported on the experience of the Delaware Valley Regional Planning Commission (DVRPC) in using the UTPP.* Only 35% of that region’s nearly 2 million work trips were coded to traffic zones; 27% were coded to zip codes, 28% to UAC’s, and 10% were uncoded. A multistep procedure was used to reduce all work trips to the zonal level. Trips coded to ZC’s were proportionally allocated to traffic zones on the basis of DVRPC’s zonal employment file, which gives employment by zip code and traffic zone. Trips coded to the UAC’s were allocated to zones following the trip patterns of the originally coded trips plus those derived from the zip code reallocation. Finally, uncoded trips were allocated to zones on the basis of all trips previously coded or allocated to zones. About 27% of the zones required further

*His report is recommended reading for anyone contemplating modification of UTPP data.
adjustment to make their destination totals agree with employment totals. Upon completion of this process, characterized by Zakaria as "a major effort," the work trip data were judged to be acceptable for planning purposes, though not as good as the socioeconomic data.

Additional experiences are reported in a recent publication of the Transportation Research Board (3). In Rhode Island, three methods were used to obtain average daily traffic volumes from assigned census work trips. The most successful method used trip end information from Parts I and III of the UTPP as input to a gravity model to produce the trip table, which was then assigned to the network, rather than using the Part IV work trip data (5). In Wilmington, Delaware, 55% of the work trips were coded to traffic zones, and 6 man-months of effort were needed to make the UTPP operational. In Albuquerque, 64% of the work trips were coded to traffic zones, while the state of California recorded an average of 57% for 14 SMSA's. The Tri-State Regional Planning Commission abandoned their attempt to use the UTPP directly, due to the high costs estimated for obtaining the package and recoding work trips, and opted instead for a specially created Worker File, not tied to any traffic zone system.

In summary, the major problems encountered in using the UTPP have been the unacceptably low percent of work trips coded to the traffic zone level and the difficulty in establishing concordance between the traffic zone system and the geographic coding system used by the Census Bureau. Considerable local data and effort are needed to make Parts III and IV of the package operational. The data in Parts I and II, on the other hand, are generally thought to be very good. The UTPP data for the Atlanta SMSA appear to follow this national pattern.
II. APPROACH TO DEVELOPING THE TRIP TABLE

In light of the findings summarized above, there are at least four basic approaches which may be used to develop a work trip table for the Atlanta Region:

1. Use the completely coded subset of the UTPP trip table, without trip reallocation;
2. Use proportional trip reallocation to create a complete zonal level trip table;
3. Use a separate trip table based on local data;
4. Use UTPP trip ends as input to a gravity model. Before describing and analyzing each of these approaches, however, it is necessary to consider briefly some criteria for evaluating them.

Criteria for Evaluating Alternative Approaches

A prime consideration in selecting a method for developing the trip table is the impact of the method on the future usefulness of the peak-hour traffic model. The main objective of this research, it will be recalled, is to develop statistical relationships between peak-hour traffic and assigned work trips. In support of this objective, peak-hour volumes were recorded at selected counting stations throughout the Atlanta area in Phase I of the study. The traffic model, to be derived on the basis of the work trip network assignment, may then be used in the following types of applications:

1. To estimate current peak-hour volumes for links not included in the Phase I count program;
2. To estimate future peak-hour volumes, based upon future assigned work trips.

The method of trip table development will not have a substantial impact on the first type of application, provided that estimates are made only for links within the general area covered by the count stations. For the second type of application to be successful, however, the future assigned volumes must represent
the same collection of urban travel patterns as are represented in the assigned volumes used to develop the model.

To restate this important point in another way, the peak-hour traffic model to be developed in this research may be expressed mathematically as follows:

\[ V_{ik} = A_k + B_k Q_{ik} \]  

where

- \( V_{ik} \) = peak-hour traffic volume on link \( i \) in functional class \( k \)
- \( Q_{ik} \) = work trips assigned to link \( i \) in functional class \( k \)
- \( A_k, B_k \) = constants determined by regression analysis.

For purposes of estimating the constants \( A_k \) and \( B_k \), the ground counts compiled during Phase I are used as the values of variables \( V_{ik} \), and the corresponding link volumes resulting from assigning the work trips to the network are used for the \( Q_{ik} \). For estimating future \( V_{ik} \) values, future assigned volumes \( Q_{ik}' \) obtained by assigning estimated future work trips to the network, are substituted into equation 1, along with the constants \( A_k \) and \( B_k \) estimated as indicated above. In order for these future \( V_{ik} \) to be valid, the future \( Q_{ik} \) must be consistent with the \( Q_{ik} \) used to estimate \( A_k \) and \( B_k \). That is, the present and future work trips used to derive the \( Q_{ik} \) must have the same definition.

Assignment of a work trip table capturing all work travel within the Atlanta region would seem to be the approach which would yield the most useful peak-hour traffic model. This would allow any future work trip matrix covering the entire region to be used as input to the model, irrespective of how that matrix was obtained. But it would also be possible to develop the model on the basis of some nonexhaustive set of work trips. For example, only those work trips which are destined to locations within the railroad cordon surrounding Atlanta might be used. If this were to be done, future peak hour volumes could
be validly estimated with the model only if the future work trips used to obtain the future $Q_{ik}$ are also only those work trips with destinations inside the cordon.

Additional criteria for comparing alternative approaches to developing the trip table are more obvious than that discussed above. These include such things as prospective resource requirements, availability of needed data, and ease of implementation.

A final point to consider concerns methods which might be used to assess the adequacy of the trip table. An indirect but simple way to make this judgement is to use the trip table to develop the peak-hour traffic model (i.e., use the $Q_{ik}$ resulting from assigning the work trips to the 1970 Atlanta network to estimate $A_k$ and $B_k$ in equation 1). If a statistically valid model is obtained, it may be inferred that the trip table is adequate.

A second method hinges upon the availability of an independent trip table, such as that given in the 1972 O-D survey, or the gravity model work trips based upon that survey. The two trip tables may be compared by using the standard procedures for examining the accuracy of a trip distribution model, i.e., by looking at average trip length, trip length frequency distributions, and spider network assignments. Several statistical tests for comparing the two trip tables directly are also available. As these are less well known to transportation planners and researchers than the standard procedures, these statistical tests are described in the Appendix to this report.

The approaches to developing a trip table listed at the beginning of this section are taken up successively below.

**Approach 1: UTPP Subset**

This approach consists of proceeding with the original study design, deleting those census work trips not coded to traffic zones. That is, the 50% of the work trips for which both the origin and destination zones are
known would be assigned to the 1970 Atlanta network, and the remaining work trips would be ignored.

Note that the trip table used in this approach consists of those work trips made by SMSA residents to locations within the ACG area. Hence adopting this approach is equivalent to hypothesizing that peak-hour traffic volume is a function of centrally destined work trips. (The initial hypothesis, of course, was that peak-hour volume is a function of primary work trips, regardless of their travel orientation.) This assumption may not be as severe as it appears, since the Phase I volume counts "placed greater emphasis on the ... area within which the Address Coding Guide was developed." (1, p. 7) Hence it might be expected that central area work trips would dominate these peak-hour volumes. Nonetheless, it is true that a significant number of volume counts were taken in suburban and fringe suburban areas. Thus a careful selection of count stations to be used in developing the traffic model would have to be made.

This approach also assumes essentially complete and accurate reporting of work trip information by census respondents working within the ACG area. That is, it assumes that all ZC and UAC coded destinations are outside of the ACG area. This is not entirely true, since some respondents did not provide accurate work address information. The seriousness of such errors can be determined by examining the numbers of trips coded to ZC's and UAC's within the ACG area. If significant numbers of trips are involved, they should simply be allocated proportionally (according to the pattern of completely coded trips) to the appropriate zones.

The primary disadvantage of this approach is the limitation which it places upon the future applicability of the peak-hour traffic model. Future traffic estimates could reliably be made only for streets and highways within the ACG area. Although there is currently a large amount of redevelopment occurring within this area, and the area continues to serve as a strong commercial
and cultural focus for the region, it is also true that there has been and continues to be extensive development at suburban and fringe suburban locations. Further, street and highway facilities within the ACG area are well developed and stabilized, while the outlying areas provide the largest number of design and traffic operations projects requiring peak-hour volume estimates. Hence the utility of a model which works only for the central portion of the region is highly questionable.

A second hazard is the possibility that this approach will not yield a statistically acceptable traffic model. Atlanta is characterized by a significant amount of reverse commuting and crosstown-circumferential work trip activity. (The 50% figure for work trips destined to the ACG area supports this observation.) Thus a traffic model which ignores these trips may well turn out to have poor statistical properties, precluding use of the model.

The nature of the future trip matrix required for input to the model has both positive and negative aspects. On the plus side, only future work travel to zones within the ACG area is required. This could be obtained by surveying trip makers at their place of work, rather than by the costly home interview procedure. On the minus side, future travel surveys (including the 1980 census) will likely cover the entire region. Hence the work trips needed for use in conjunction with the traffic model would have to be extracted from a larger matrix, and care would have to be exercised to insure that a proper match of trip pattern coverage was obtained.

The major advantage of this approach is its ease of implementation. No significant additional processing of the UTPP trip table would be required, and no additional data are needed. Resource requirements are also very modest. It would likely take about one to three man-months to implement this approach, including the time required to perform the statistical analyses.
Approach 2: Proportional Trip Reallocation

In contrast with the first approach, the present approach calls for using all of the work trip information in the UTPP, rather than discarding half of it. To do so, it will be necessary to recode those trips with destinations known only at the ZC and UAC level to valid traffic zones. This approach conforms with that adopted in other urban areas. The specific procedures included represent a tailoring of the DVRPC procedures (4) to fit Atlanta.

First, those trips coded to zip codes should be allocated to the traffic zones covered by each zip code area. This will necessitate mapping of the Atlanta region's zip codes. This can be done in short order by GDOT and ARC personnel with the aid of postal clerks employed on a part time basis. Overlaying the zip code map on a map of traffic zones will quickly reveal which traffic zones (or portions thereof) are located within each zip code area. Using the employment-by-zone data file referred to in section one of this report, the total employment within each zip code area and the percentage of employment for each zone within each zip code can be calculated. Based upon those percentages, trips from a given zone to a zip code can be allocated to the various zones within the zip code area. Special handling will be required for those zip codes which represent post office boxes or mailing addresses, rather than actual work locations. Trips to such zip codes for which the actual work place can be determined should be manually posted to the proper zone. The remainder should be added to the uncoded category.

Next, trips coded to UAC's should be allocated to zones, on the basis of all previously determined zonal trip interchanges (including the zip code reallocations), in conjunction with ARC's 34 superdistricts for the Atlanta region. Trips previously coded from all zones within a given superdistrict of origin should be aggregated, and the percentage of these trips destined
to each zone within each UAC should be calculated.* Using these percentages, trips from each zone within the superdistrict coded to a UAC should be allocated to zones within the UAC.

Finally, trips in the uncoded category should be allocated to zones on the basis of all trips previously coded or allocated to traffic zones.

External trips, made by persons residing outside of the SMSA to locations within the region, pose an additional problem. Data on such trips are available at the UAC level in the 1970 Census Journey to Work Report. These trips, which should constitute only a small percentage of total work trips, should be allocated to zones on the basis of the trip patterns previously established for all trips and zones in the region. Alternatively, the 1972 external survey, appropriately deflated to represent 1970 travel, can be used to supplement the census data.

Completion of the steps outlined above will produce a table of 1970 interzonal work trips covering the entire study area. This table should be carefully examined for reasonableness, and any indicated changes made manually. Further checks can be obtained by comparing the trip distribution pattern in this table with that revealed in the 1972 travel surveys or gravity model, making allowances for any changes in development patterns between 1970 and 1972. Techniques for making such comparisons have been described previously.

The main advantage of this approach is that the peak-hour traffic model derived from it will be applicable to the entire Atlanta region, and will be directly compatible with later areawide work trip surveys. For example, the model could be used in conjunction with the 1972 travel data to provide updated peak-hour volume estimates. Also, since the UTPP is based on a 15% sample, the completed work trip data should provide area planners with a better planning base than the 1972 travel survey, which used a 1/2% sample.

*If numerous small percentages are obtained, counties can be used instead of superdistricts, as was done by DVRPC.
A less tangible advantage of this approach is that it will enhance the GDOT's ability to use the 1980 census work trip information effectively. Presumably the quality of the 1980 work trip data will be much better than that of the 1970 data, and will require much less local processing. After going through the procedures given above, GDOT personnel will be very familiar with methods for updating and correcting census data, and thus will be able to accomplish any local processing expeditiously. The peak-hour traffic model itself may be usable directly with the 1980 data (pending verification of model coefficients), and thus may provide a very quick means of estimating the peak-hour volumes needed for system monitoring and plan revision.

Obviously, the major disadvantage of this approach is the large amount of effort required to implement it. Using national experience as a guideline, approximately six man-months will be needed.

This also means that the peak-hour traffic model would not be available until mid-1976 at the earliest, which is a rather late date to be using 1970 or 1972 peak-hour volumes for decision-making. Fortunately, estimated work trips for 1980 and 2000 are already available from ARC, so the model would be quite useful for planning and design purposes.

Approach 3: Separate Trip Table

A third possible approach to obtaining a work trip table is to use the locally developed 1972 work trips, which are currently being used as the basis for the ARC Regional Development Plan. Since the Phase I traffic count data are for an average weekday in April of 1970, the 1972 trip data would have to be correspondingly adjusted. The UTPF trip data could be used to make this adjustment.

A simple and straightforward way to deflate the 1972 trips would be to multiply 1972 work trips from zone i to zone j by the ratio of the UTPF trip origins in zone i to the 1972 trip origins in zone i. Symbolically,

\[ T'_{ij} = \frac{T_{ij}}{t_{ij}} \]  

(2)
where

\[ T_{ij} = 1970 \text{ work trips from } i \text{ to } j \]
\[ t_{ij} = 1972 \text{ work trips from } i \text{ to } j \]
\[ T_i = \text{ UTPP work trips originating in } i \]
\[ t_i = 1972 \text{ work trips originating in } i. \]

Equation (2) may be recognized as a simple origin zone growth factor model.

Use of this often criticized technique is acceptable here because a time span of only two years is involved, and because data for developing reliable destination zone growth factors (e.g., for use in the Fratar model) are not available.* Destination zones which experienced abnormally high growth between 1970 and 1972 can be accounted for by manually reallocating trips to other zones on a proportional basis. Some searching of county building permits or tax records may be required to identify such zones.

The procedure described thus far will not work for external trips, since \( T_i \) is zero for external zones. After completion of the above calculations for all internal trips, 1970 external trips may be estimated as follows:

\[ T_{ij} = \frac{T_i}{t_i} t_{ij} \]  \hspace{1cm} (3)

where

\[ T_{ij} = 1970 \text{ work trips from } i \text{ to } j \]
\[ t_{ij} = 1972 \text{ work trips from } i \text{ to } j \]
\[ T_j = \text{ total 1970 internal work trips destined to } j \]
\[ t_j = \text{ total 1972 internal work trips destined to } j \]
\[ i = \text{ an external zone.} \]

Equation (3) may be recognized as a simple destination zone growth factor model, which must be used in this case because origin zone growth factors are not available.

---

*The UTPP Part III data are plagued with the same difficulties as the Part IV data.*
This approach is conceptually very appealing, due to its simplicity, ease of implementation, and potentially high accuracy. The UTPP data compiled by zone of residence have generally been found to be quite good, and this is the type of data which would be used to adjust the 1972 O-D matrix. Furthermore, this approach amounts to using trip information from a 15% sample to modify that based on a 1/2% sample, which would seem to be a valid procedure. Also, the only additional assumption required (over and above those normally embodied in trip tables) is that trip distribution patterns were similar in 1970 and 1972, which again appears to be very reasonable.

This approach is akin in spirit to approach two, since both of them attempt to construct a complete areawide work trip matrix. Hence the peak-hour traffic models developed with either method would have the same range of applicability. The present approach does fail to exploit the UTPP data to its fullest. This is irrelevant, however, and should be viewed in much the same light as an economist views a sunk cost.

From a resource perspective, approach three falls in between approaches one and two. Approximately three man months of effort should be sufficient to implement it.

Approach 4: Gravity Model

This approach calls for using the trip production and attraction information contained in the UTPP as input to a gravity model, which would then provide the desired interzonal trip table. Travel time factors derived for the 1972 work trip gravity model could be used, thus obviating the need for an extensive model calibration process. Trip productions can be obtained from Part I of the UTPP, which provides number of employed persons by zone of residence. Attraction factors can be obtained from Part III, which provides employment by work zone. These latter data will require modification similar to that needed for the trip destinations in approach two.

The advantages and disadvantages of this approach are virtually identical
to those of approach two. Anticipated resource requirements are also the same as those of approach two (i.e., approximately six man-months).

Approach four has been included here primarily because of the success achieved with it in the Rhode Island study (5).
III. RECOMMENDATIONS

Four alternative methods for developing a work trip table for Atlanta, for the year 1970, have been presented in the preceding section. Each of these methods is workable, and should provide a trip table of acceptable accuracy. Based upon this observation, the principle recommendation resulting from this study is that one or more of these methods should be implemented. In other words, the project should not be abandoned due to the apparent flaws in the UTPP data. There are sufficient ancillary data and methods at hand to overcome these difficulties.

Selection of a particular method or methods is dependent upon the intended use of the peak-hour traffic model and the level of resources which GDOT is willing to commit to the effort. Approach one is the least expensive of the four, but also places the greatest limitations upon the usefulness of the model. Approaches two and four are the most costly, but have the advantages of relying on a relatively large (15% sample) data base and preparing the Department to capitalize on the 1980 census. Approach three is intermediate in cost, and is probably the simplest and most accurate way to fulfill the initial study objectives. The latter three approaches all lead to a peak-hour traffic model with areawide applicability, and requiring a complete work trip table for input.

Considering the points summarized above, approach three stands out as the best candidate for implementation. It would be the easiest to use, would lead to speedy completion of the project, and would base the peak-hour traffic model upon a trip matrix with which local planners are intimately familiar. It also avoids much of the difficulty associated with external trips. It is therefore recommended that approach three be selected.
The availability of the 1972 travel surveys, and the fact that they have already been used to update the Atlanta region transportation models, mitigated against selection of approaches two and four. Implementation of either of these approaches would have to be based on a careful assessment by GDOT of their needs relative to the use of 1980 census data. However, it is recommended that if either approach two or four is chosen, the other of the pair should also be implemented. The reason for this recommendation is that both approaches require virtually identical adjustment procedures, and hence the incremental cost of including both is relatively small.

The above recommendations are all based upon operational considerations. From a research viewpoint, it would be desirable to implement and compare all four approaches. This would enable the Department to select the most accurate peak-hour traffic model derivable from this project for operational use. This would also put the Department in a good position to easily update the model using the 1980 census, since the best techniques to use will already be known.

Whichever approach is selected, it is also recommended that GDOT should undertake a comparison of the UTPP work trip data and the 1972 data, using the methods described in section one of this report. (In fact, this should be done even if the Department decides to abandon the project). This will allow inferences to be made as to whether or not these two independent surveys measured the same underlying travel behavior, and should provide valuable insights into the compatibility of local data and census data. Of course, without modification of either data set, only corresponding subsets should be compared.

The Department should also begin planning now for the 1980 census. Forms for the 1980 census will be closed out in 1977, so GDOT recommendations concerning collection of travel data need to be forwarded to the Census
Bureau immediately.* Once the precise contents of and procedures for the 1980 census are finalized, the Department should develop plans for updating the peak-hour traffic model and for other uses of the census data.

A final recommendation concerns use of the traffic volume information compiled in Phase I of this project. This constitutes a valuable data source which has not been effectively exploited by the Department. These data could be used to study such things as variations in peak-hour percent of daily traffic, and directional split percentage, by facility type and location. Traffic volumes observed at locations corresponding to regular GDOT count stations can be used to examine stability of these relationships over time and to develop short range growth factor forecasting models. Availability of these data should also be made known more widely.

**Summary of Recommendations**

1. GDOT should proceed with development of the peak-hour traffic model, using one or more of the approaches described in this report.

2. If a single approach is to be chosen, approach three should be selected.

3. If either of approaches two or four are selected, the other should also be implemented.

4. From a research viewpoint, it is recommended that all four approaches be tested and compared.

5. A comparison of the UTPP work trip data and the locally developed 1972 trip data should be undertaken.

6. The Department should begin planning immediately for the 1980 census.

7. The Department should make more effective use of the excellent traffic volume information collected in Phase I of the project.

*It is suggested that these recommendations be channeled through Committee A1C03, Information Systems and Data Requirements, of the Transportation Research Board, since the Census Bureau has requested their assistance.
REFERENCES


2. Meeting with G. Booth, Atlanta Area Planning Branch, Georgia Department of Transportation, April 29, 1975.


5. Fertal, M. Informal presentation to Committee AlC03, Information Systems and Data Requirements, 54th Annual Meeting, Transportation Research Board, Washington, D. C., January 1975
APPENDIX: STATISTICAL TESTS FOR COMPARING TRIP TABLES

Several statistical tests are available for directly comparing one trip table with another, assuming that both tables are intended to represent a common underlying set of travel patterns. Two of these deemed appropriate for use here are the Chi-Square test and the Smirnov Maximum Deviation test.

The following notation is used to describe these tests:

\[ X_{ij} = \text{trips from } i \text{ to } j \text{ in the base or presumably more accurate matrix} \]
\[ Y_{ij} = \text{trips from } i \text{ to } j \text{ in the other matrix} \]
\[ N = \text{total number of zones} \]
\[ T_X = \text{total trips in the } X \text{ matrix} \]
\[ T_Y = \text{total trips in the } Y \text{ matrix} \]
\[ x_{ij} = \text{empirical probability that a random trip in the region is from } i \text{ to } j = \frac{X_{ij}}{T_X} \]
\[ y_{ij} = \text{the } Y \text{ matrix counterpart of } x_{ij} = \frac{Y_{ij}}{T_Y} \]

Chi-Square Test

This test is used to compare two sets of frequency data, or to test the goodness of fit of one distribution to another. The null hypothesis tested is

\[ H_0 : X = Y \]

against the general alternative

\[ H_1 : X \neq Y \]

For this application, if \( T_X \neq T_Y \), then the \( Y \) values should be multiplied by the ratio \( \frac{T_X}{T_Y} \) prior to conducting the test. The test statistic is then calculated as
\[
\chi^2 = \sum_{i} \sum_{j} \frac{(X_{ij} - \bar{Y}_{ij})^2}{\bar{X}_{ij}}
\]

If the calculated value exceeds the tabulated value of \(\chi^2\) with \(N^2 - 1\) degrees of freedom, then \(H_0\) is rejected at the \(\alpha\) significance level.

Values \(X_{ij} = 0\) cannot be used in the test, as they make the denominator of the corresponding term in equation 4 zero. In fact, small values of \(X_{ij}\) are notoriously unstable in trip distribution analysis, and should be excluded. Specifically, it is suggested that all \(X_{ij} < 10\) be excluded from the test. This reduces the degrees of freedom of the test statistic to \(n - 1\), where \(n\) is the number of cells included in the test.

Values of \(\chi^2\) are typically tabulated only for \(n - 1 \leq 30\). Since \(n\) will be very large in this case, it will be necessary to use a Normal approximation. Specifically, the critical value of the test statistic should be calculated as

\[
\chi^2_{\alpha} = \frac{1}{2}(I_{\alpha} + \sqrt{2n-3})^2
\]

where \(I_{\alpha}\) is the \(\alpha\) percentage point of the cumulative Normal distribution with zero mean and unit variance.* Values of \(I_{\alpha}\) for several commonly used significance levels are given below.

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_{\alpha})</td>
<td>1.282</td>
<td>1.645</td>
<td>2.327</td>
<td>2.575</td>
</tr>
</tbody>
</table>

In view of the limited power of this test, a value of \(\alpha = 0.05\) is recommended.

*That is, \(\int_{-\infty}^{I_{\alpha}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1-\alpha\)
Smirnov Maximum Deviation Test

This test is a non-parametric test which can be used to determine whether or not two samples have been drawn from the same population. The test statistic used is the maximum deviation between the cumulative probability distributions of the two samples. The hypothesis of equality is rejected if this deviation is excessively large.

For this application, let

\[ F(IJ) = \sum_{i} \sum_{j} x_{ij} \]
\[ G(IJ) = \sum_{i} \sum_{j} y_{ij} \]

That is, \( F \) and \( G \) are the cumulative interzonal trip interchange probability distributions. Then the null hypothesis is

\[ H_0 : F(IJ) = G(IJ) \quad I = 1, \ldots, N \]
\[ J = 1, \ldots, N \]

against any general form of nonidentity

\[ H_1 : F(IJ) \neq G(IJ) \]

Note that these hypotheses are essentially the same as those for the Chi-Square test.

The test assumes that sampling is random and that the distributions are continuous. The latter assumption is violated here, but the degree of violation is slight, since graphs of \( F \) and \( G \) would closely resemble continuous curves. In any case, using the test for discrete distributions is conservative.

The test statistic is calculated as

\[ D = \max_{I, J} \left| F(IJ) - G(IJ) \right| \]

That is, it is merely necessary to go through the \( x \) and \( y \) matrices cell by cell, calculating the cumulative probabilities and keeping track of the largest absolute value of the differences between the two.

For large sample sizes, as is the case here, the critical value of \( D \) may be approximated as
\[ D_{\alpha} = K_{\alpha} \sqrt{\frac{2}{N}} \]  

(7)

\( H_0 \) is rejected at the \( \alpha \) significance level if \( D > D_{\alpha} \). Values of \( K_{\alpha} \) for commonly used significance levels are given below. Further details about this test are given by Bradley.*

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
<th>0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_{\alpha} )</td>
<td>1.224</td>
<td>1.358</td>
<td>1.628</td>
<td>1.858</td>
</tr>
</tbody>
</table>

Discussion

The Chi-Square and Smirnov tests are both tests of the same properties of the X and Y trip tables, and either one should provide a good indication of their similarity. The latter test, however, makes fewer assumptions than the former, and does not require discarding of cells for which \( X_{ij} < 10 \). Also, many statisticians believe that the Chi-Square test is of dubious validity, while the Smirnov test is theoretically sounder. For these reasons, the Smirnov test is the preferred choice. Many researchers, however, are not familiar with this test, but are well versed in the Chi-Square technique. Hence both have been included in this Appendix.