Trip Production and Attraction Characteristics in Small Cities

M. D. HARMELINK, Department of Highways, Ontario; G. C. HARPER, N. D. Lea and Associates, Vancouver, B. C.; and H. M. EDWARDS, Queen's University, Kingston, Ontario

Several factors affecting the estimation of trips produced by and attracted to various areas of a small city were examined for various trip purposes. The trip production study was conducted for one city only, Kingston, Ontario, while the trip attraction study treated two cities, Kingston and Barrie, Ontario. Planning and assessment departments and home-interview O-D surveys provided all the basic data. For the most part, 24-hour person trips by vehicle were studied.

In the production study, attempts were made to relate trip production to car ownership, population, distance from CBD, residential density, proportion of single-family dwelling units, and percentage of the population less than five years of age. The most reliable predictor of trip production was found to be car ownership. Results indicated decreased accuracy of trip estimation with increased segregation of trip purpose, and showed little improvement in the trip estimates with increased sample sizes.

In the attraction study, relationships were found between attracted trips and the assessed value of land and buildings and area of land in each of four land-use categories. The results of the study have expected application in both conventional O-D surveys and mathematical traffic models.

For proper evaluation of the transportation problems in an urban area, a knowledge of traffic movements within the area is essential. A quantitative understanding of the factors which influence traffic movement is, therefore, of major importance.

As an outgrowth of comprehensive origin-destination studies and the relation of area traffic movements to land use and economic and population characteristics of the area, mathematical models (formulae) describing traffic movements have been developed and used for predicting future traffic movements in the area. Home-interview O-D surveys are still required for model formulation since they provide the most reliable input data necessary for the model (1). It is claimed, however, that fewer interviews are required for model formulation than for conventional O-D surveys.

This paper describes trip production and attraction characteristics in small cities. Certainly, previous attempts have been made to relate the number of trips produced by residents of an area to certain population and location characteristics. Kudlick, Fisher, and Vance (2) used car ownership and population as predictors of trips from a zone. Others have used labor force as a predictor of work trips, where such data were available (3). Mertz and Hamner (4) used car ownership, population density, income per household, and distance from the CBD as predictors of total trips, but found that car ownership was the most reliable predictor.

Attempts have also been made to relate the number of trips attracted to an area to land-use characteristics of the area. Voorhees (3, 5), Barnes (6), and Kudlick, Fisher,
and Vance (2) all used employment and population as predictors of trips attracted to a zone. Harper and Edwards (7) found a strong relationship between the total number of daily person trips attracted to the CBD and the floor-space area in three use categories—retail, service-office, and manufacturing-warehousing.

However, few investigations of the factors affecting trip production and attraction appear to have been made for small cities (less than 75,000 population), where good land-use data are less generally available. Moreover, the effects of sample size on accuracy of trips estimates and on the optimum trip purpose groupings to be used have seldom been examined. These matters were felt to warrant investigation. A further innovation was relating trip attraction to assessment values (readily available in small cities) rather than to employment in each of several categories, estimates which may not be so readily available in small cities. Besides reflecting intensity of land use and competitiveness between different land uses, assessment will also reflect multiple usage of land or buildings. It is recognized that assessment values will probably not be equal to real market values, but they should reflect relative values within the same land-use categories. Moreover, since the assessment process is generally treated in a quasi-judicial manner (by the method governing appeal), assessed values of land and buildings have little tendency to reflect speculation in land and building values and thus supply stable, consistent measures of land-use intensity as it exists—not as the market thinks it should exist.

The city selected for the trip production study was Kingston, Ontario (study area population of 63,000). The cities selected for the attraction study were Kingston and Barrie, Ontario (study area population of 22,000). The data used in the studies were obtained from the planning and assessment departments in each city and from the home-interview origin-destination study conducted in each city in 1961 according to Bureau of Public Roads recommendations.

TRIP PRODUCTION

Research Procedure

The trip production study was concerned only with daily (24-hour) person trips by vehicle made by residents of the study area. Data from separate truck, taxi, and external roadside surveys were excluded, and there was no modal split. Fifty-five traffic zones were used in the study, of which 49 were Group A zones (predominantly single-family dwelling units) and 6 were Group B zones (predominantly multiple-family dwelling units).

Three sample sizes (based on the total number of dwelling units) were selected to cover a reasonable range of sizes and to bracket the number of 1,000 interviews suggested as adequate for gravity model formulations (8). The emphasis in this study was placed on systematic sampling rather than cluster sampling. Cluster sampling has been used in several model formulations; however, Hansen, Hurwitz, and Madow (9), Deming (10), and Lieder (11) all suggest reasons for preferring systematic sampling to cluster sampling, the primary reason being smaller variance. The samples were selected by systematically sampling the original 12 1/4 percent dwelling unit sample identification number; the results of the sample selection are shown in Table 1. Although no definite conclusions could be drawn regarding variability among samples of the same size, the two 5 percent samples were tested to observe qualitatively the kind of differences that might arise.

To examine the effect of various trip purpose categories on accuracy of prediction, trip production studies were made for seven trip purposes, as follows:

1. Total trips,
2. Home-origin work trips,
3. Non-work home-origin trips,
4. Non-home-based trips (neither origin nor destination at home),
5. Home-origin shopping trips,
6. Home-origin social-recreational trips, and
7. Miscellaneous home-origin trips.
The category "total trips" for a given zone refers to all person trips by vehicle made by the residents of that zone, regardless of origin or destination of trip, and excludes all trips made to or from the zone by non-residents of the zone. The "miscellaneous" category includes trips made for the purposes of personal business, medical-dental service, school, eating a meal, changing mode of travel, and serving passengers.

If it is assumed that all home-origin trips return home, estimates of total trips may be made in three ways, using different combinations of the given trip-purpose categories, as follows:

Total Trip Estimate 1: Trip Purpose Category 1
Total Trip Estimate 2: Trip Purpose Categories 2, 3, 4
Total Trip Estimate 3: Trip Purpose Categories 2, 4, 5, 6, 7

For the sample sizes tested, attempts were made by means of simple and multiple regression analysis to relate sampled trips per D. U. for various purposes to sampled values of the following land-use parameters and characteristics of the zonal population:

\[ x_1 \text{ Cars/dwelling unit} \]
\[ x_2 (\text{Cars/dwelling unit})^2 = x_1^2 \]
\[ x_3 \text{ Persons/dwelling unit} \]
\[ x_4 (\text{Persons/dwelling unit})^2 = x_3^2 \]
\[ x_5 \text{ Airline distance from CBD (tenths of miles)} \]
\[ x_6 \log_{10} (\text{airline distance from CBD}) = \log_{10} x_5 \]
\[ x_7 \text{ Residential density (dwelling units/net residential acre)} \]
\[ x_8 \log_{10} (\text{residential density}) = \log_{10} x_7 \]
\[ x_9 \text{ Proportion of single-family dwelling units} \]
\[ x_{10} \text{ Percentage of population less than 5 years of age} \]

In the regression analysis, the observations were weighted in accordance with the number of dwelling units in the zone.

The estimation equations of best fit are fully documented (12). The equations for total trips, home-origin work trips, and home-origin shopping trips are given in Table 2.

The trip estimation equations showed that in Kingston, car ownership was the most reliable single predictor of the trips produced by residents of a zone. For almost all estimates, car ownership was a variable in the estimation equations, and only for home-origin shopping or social-recreational trips was the multiple regression equation an improvement over the simple regression equation using car ownership alone as the independent variable or predictor.

An important observation was that, for a given trip purpose category, although the regression equations derived from the larger sample sizes appear to be better estimating equations by virtue of their larger coefficients of correlation and smaller standard errors, the actual regression equations for different sample sizes show a distinct similarity. This similarity was evident for all trip purpose categories except home-origin shopping and social-recreational trips, where greater equation differences occur.
**TABLE 2**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total Trips</th>
<th>Home-Orig Work Trips</th>
<th>Home-Origin Shopping Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>S - .025 - K</td>
<td>$y' = -2.327 + 8.757x_1$</td>
<td>$y' = 0.149 + 1.002x_1$</td>
<td>$y' = -0.017 + 0.448x_1 - 0.008x_{10}$</td>
</tr>
<tr>
<td></td>
<td>$r = 0.87$</td>
<td>$r = 0.71$</td>
<td>$r = 0.47$</td>
</tr>
<tr>
<td></td>
<td>$S_1 = 1.39$</td>
<td>$S_1 = 0.29$</td>
<td>$S_1 = 0.24$</td>
</tr>
<tr>
<td>S - .05 - 1 - K</td>
<td>$y' = -2.684 + 8.289x_1$</td>
<td>$y' = -0.025 + 1.095x_1$</td>
<td>$y' = 0.902 + 1.155x_1 + 0.859x_2 + 0.328x_9$</td>
</tr>
<tr>
<td></td>
<td>$r = 0.90$</td>
<td>$r = 0.75$</td>
<td>$r = 0.85$</td>
</tr>
<tr>
<td></td>
<td>$S_1 = 1.05$</td>
<td>$S_1 = 0.25$</td>
<td>$S_1 = 0.12$</td>
</tr>
<tr>
<td>S - .05 - 2 - K</td>
<td>$y' = -2.710 + 8.751x_1$</td>
<td>$y' = -0.012 + 1.145x_1$</td>
<td>$y' = 0.038 + 0.399x_1 - 0.012x_{10}$</td>
</tr>
<tr>
<td></td>
<td>$r = 0.91$</td>
<td>$r = 0.79$</td>
<td>$r = 0.86$</td>
</tr>
<tr>
<td></td>
<td>$S_1 = 1.04$</td>
<td>$S_1 = 0.23$</td>
<td>$S_1 = 0.14$</td>
</tr>
<tr>
<td>S - .10 - K</td>
<td>$y' = -3.041 + 6.983x_1$</td>
<td>$y' = -0.024 + 1.124x_1$</td>
<td>$y' = -0.356 + 0.155x_1 + 0.203x_6$</td>
</tr>
<tr>
<td></td>
<td>$r = 0.94$</td>
<td>$r = 0.81$</td>
<td>$r = 0.81$</td>
</tr>
<tr>
<td></td>
<td>$S_1 = 0.70$</td>
<td>$S_1 = 0.19$</td>
<td>$S_1 = 0.10$</td>
</tr>
</tbody>
</table>

Note: $y' = \text{Trips/D.U.}$

**Trip Estimates and Discussion of Errors**

From the regression equations, the number of zonal trips for each trip purpose category was calculated from the actual zonal values of the basic parameters listed previously and the total number of zonal dwelling units. The actual zonal values of the basic parameters (except for distance from the CBD and residential density, which were measured directly) were obtained from an expansion of the original 12½ percent sample to 100 percent, due to lack of better data. However, where sample sizes smaller than 12½ percent were tested, it was assumed that base data comparable to those obtained from the expansion of the 12½ percent sample to 100 percent would be available from planning sources, and the comparable base data could be substituted in the estimating equations.

For a comparison of actual and estimated trips, the actual number of trips was obtained in each case from an expansion of the original 12½ percent home-interview survey data to 100 percent. Although it was recognized that these actual trip values are probably in error, they were used as the standard of comparison in this analysis because they are the generally accepted figures, because they have been used as a standard of comparison in other studies (7, 13, 14), and because there was no available substitute.

The estimated trip volumes were compared with the corresponding actual trip volumes and the errors or discrepancies were calculated. These results are documented elsewhere (12), and are shown graphically for total trips, home-origin work trips, and home-origin shopping trips in Figures 1, 2, and 3 respectively.

So that Figures 1-3 could be plotted, actual zonal volumes of trips were grouped into arbitrary ranges of trip volumes. For each range of trip volumes, the root-mean-square (RMS) error was calculated and plotted as a percent at the arithmetic mean of the trip volumes in the range:

$$\text{RMS error} = \sqrt{\frac{\sum_{i=1}^{n} (V_{est} - V_{act})^2}{n}}$$

where

- $V_{est} = \text{estimated volume}$,
- $V_{act} = \text{corresponding actual volume}$, and
- $n = \text{number of zones in range}$.

$$\text{Percent RMS error} = \frac{\text{RMS error}}{V} \times 100$$
Figure 1. Errors in production estimates of total trips for various sample sizes.

Figure 2. Errors in production estimates of home-origin work trips for various sample sizes.
Figure 3. Errors in production estimates of home-origin shopping trips for various sample sizes.

where $\overline{V}$ = mean actual trip volume in range. The percent RMS error was plotted against $\overline{V}$.

Now, similar estimating equations, though derived from different sample sizes, would be expected to produce trip estimates of comparable accuracy. Generally this was found to be true, as shown by Figures 1-3. For total trips, home-origin work trips, non-work home-origin trips, non-home-based trips, home-origin social-recreational trips, and miscellaneous home-origin trips, the 5 percent samples appear to produce results as good as or better than the 10 percent sample, with some overlap of the curves. The 2½ percent sample is generally slightly poorer than the 5 percent or 10 percent samples, but the differences are not great. For home-origin shopping trips, larger sample sizes appear to improve the estimates slightly, but again the differences are not great. This seems to indicate that sample size (within the range of sample sizes tested) has little effect on accuracy of trip production estimates, and that a considerable cost saving could be made by using a relatively small systematic sample.

The use of a small sample is, of course, contingent upon the adequacy of a small sample

<table>
<thead>
<tr>
<th>Trip Purpose Category</th>
<th>Average RMS Error at Zonal Trip Volume of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
</tr>
<tr>
<td>Home-Origin Work</td>
<td>20%</td>
</tr>
<tr>
<td>Non-Work Home-Origin</td>
<td>110%</td>
</tr>
<tr>
<td>Non-Home-Based</td>
<td>-</td>
</tr>
<tr>
<td>Home-Origin Shopping</td>
<td>194%</td>
</tr>
<tr>
<td>Home-Origin Soc.-Rec.</td>
<td>116%</td>
</tr>
<tr>
<td>Miscellaneous Home-Origin</td>
<td>44%</td>
</tr>
</tbody>
</table>
for trip attraction and distribution as well as for trip production. The effect of sample size on trip attraction estimates is described in a later section of this paper.

In all cases, the percent RMS errors decrease with increasing trip volume. With increased segregation of trip purposes, the trip volumes considered are smaller, and the percent RMS errors are generally larger over the range of trip volumes. However, Table 3 shows that at a given zonal trip volume, the errors are about the same size for all trip purpose categories except for total trips and non-home-based trips, whose errors are somewhat larger and approximately equal. Although the percent RMS error at 10 zonal trips shows considerable variability, the absolute error is quite small and insignificant.

No definite conclusions can be made at present regarding the number of trip purpose categories that should be used in a small city. Table 3 indicates that for a given volume, average percent RMS errors will be approximately equal for all trip purpose categories except total trip and non-home-based trips. It would appear then that the smaller trip purpose categories may be used for production estimates, but the larger trip purpose categories should be used as a check on them. Even then, however, the check may not provide much of an improvement over the individual purpose categories used separately. For example, in a given zone assume:

<table>
<thead>
<tr>
<th></th>
<th>Approx. RMS Error</th>
<th>Approx. Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Home-Origin Work Trips = 100</td>
<td>42%</td>
<td>42</td>
</tr>
<tr>
<td>No. of Non-Work Home-Origin Trips = 100</td>
<td>41%</td>
<td>41</td>
</tr>
<tr>
<td>No. of Non-Home-Based Trips = 30</td>
<td>120%</td>
<td>36</td>
</tr>
</tbody>
</table>

Assuming that the home-origin trips return home,

<table>
<thead>
<tr>
<th></th>
<th>Approx. RMS Error</th>
<th>Approx. Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Home-Oriented Work Trips = 200</td>
<td>42%</td>
<td>84</td>
</tr>
<tr>
<td>No. of Home-Oriented Non-Work Trips = 200</td>
<td>41%</td>
<td>82</td>
</tr>
<tr>
<td>No. of Non-Home-Based Trips = 30</td>
<td>120%</td>
<td>36</td>
</tr>
</tbody>
</table>

Total No. of Zonal Trips = 430  Total Error = 202

Thus using the individual purpose categories, the RMS error $\approx 202/430 = 47$ percent. Using Figure 1, for total trips, at 430 trips, the RMS error is about 37 percent, an improvement of about 10 percent. Using an even further segregation of purpose categories, and assuming the 200 home-oriented non-work trips to be made up of 70 home-oriented shopping trips, 70 home-origin social-recreational trips, and 60 miscellaneous home-oriented trips, the RMS error on total trips is about 62 percent, compared with 47 percent and 37 percent as determined earlier.

Grouping zones into districts also has the general effect of reducing percentage errors. When the trip estimates for individual zones were summed to district totals, the percentage errors for the district totals were generally found to be quite small in relation to many of the percentage errors in the individual zonal estimates. As with trip purpose categories, as greater definition and segregation are sought (in this case, of land areas), the accuracy of trip production estimates is decreased.

In traffic model analysis, trip production values have generally been assumed correct, and where discrepancies arise in screenline checks or trip length distributions, adjustments have been made to either the attraction figures or the time function used in distribution. This study indicates, however, that errors in trip production estimates may be sizable. Thus, it would seem appropriate to examine the effect that higher production rates might have on trip distribution and assignment, especially in critical areas.
Theoretical Standard Error Determination About the True Zonal Production Values

In the preceding discussion of errors, the errors were determined using the actual, or expanded home-interview, numbers of trips produced by each zone as the standard of comparison. It is recognized that these actual values may themselves be in error, but they were accepted as being the best available standard of comparison.

While it is not possible to determine the errors measured about the true or correct, but unknown, zonal values (of which the actual zonal values are estimates), it is possible to determine statistically the range in which the true zonal value is expected to lie, and this expected error in the estimate of the true zonal value may be combined with the standard error of the regression line to obtain a statistical expected standard error of the zonal trip estimate.

The derivation of this expected standard error is as follows:

Sampling is carried out in a number of zones in a specified area, and $Z_T$ is the total number of sampled zones. For each zone $i$, the sample provides an estimate of the mean number of trips per D.U.,

$$
y_i = \frac{\text{Sample trips, zone } i}{\text{No. of sampled D.U.'s in zone } i} = \frac{\sum y_{ij}}{n_i} = \text{Sample No. of trips per D.U. produced by zone } i
$$

where $y_{ij} =$ sample number of trips from D.U. $j$ in zone $i$. For zone $i$, the variance of the number of trips per D.U. is (9):

$$
\sigma^2_{y_i} = \frac{\sum (y_{ij} - \bar{y})^2}{n_i - 1} \left(1 - \frac{n_i}{N_i}\right)
$$

where $N_i =$ total number of D.U.'s in zone $i$, and

$$
\left(1 - \frac{n_i}{N_i}\right) = \text{finite population correction, zone } i, = fpc_i
$$

These $y_i$ values, with their corresponding $\bar{x}_i$ values, were combined in a regression analysis. This produced the regression line $y' = a + bx$. The estimate of the number of trips produced by zone $i$ is then:

$$
Y_i' = N_i y_i'
$$

where $y_i'$ = regression estimate of trips per D.U. for zone $i$. The variance of this estimate, from which the standard error of the estimated $Y_i'$ may be determined, is:

$$
\text{Var} (Y_i') = \text{Var} (N_i y_i') = N_i^2 \text{Var} (y_i')
$$

The true variance of $y_i'$, the regression estimate of trips per D.U., is:

$$
\sigma^2_{y_i'} = \frac{\sum (y_i' - \bar{y}_i)^2}{v_i}
$$

where

$\bar{y}_i =$ true, or correct, value of trips per D.U. for zone $i$, and

$v_i =$ degrees of freedom, zone $i$. 

This true variance \( S^2_{Y_1} \) is made up of two parts:

1. The variance between the regression estimates and the sample estimates of trips per D. U.—this variance is the square of the standard error of the regression line.

2. The variance between the sample estimates of trips per D. U. and the true values of trips per D. U.—this part is normally not considered, because it is usually assumed that the observed values used in regression analyses are true, not sampled, values.

Considering the sums of squares only,

\[
\Sigma (y_i - \bar{y}_1)^2 = \Sigma [(y_i' - \hat{y}_1) + (y_i - \bar{y}_1)]^2 = \Sigma (y_i' - \hat{y}_1)^2 + 2\Sigma (y_i' - \hat{y}_1)(\bar{y}_1 - \bar{y}_1) + \Sigma (y_i - \bar{y}_1)^2
\]

The two parts of the variance are independent. Therefore, the cross-product term becomes zero, and

\[
\Sigma (y_i - \bar{y}_1)^2 = \Sigma (y_i' - \hat{y}_1)^2 + \Sigma (y_i - \bar{y}_1)^2
\]  

(6)

or

\[
S^2_{Y_1} = \frac{\Sigma (y_i' - \hat{y}_1)^2}{\nu_1} + \frac{\Sigma (y_i - \bar{y}_1)^2}{\nu_1}
\]

i.e.,

\[
S^2_{Y_1} = S^2_e + \frac{\Sigma (y_i - \bar{y}_1)^2}{\nu_1}
\]  

(7)

The true value of the zone average, \( \bar{y}_1 \), is unknown. However, in general terminology, the variance of sample means about the true mean may be determined from

\[
S^2_Y = \frac{S^2_Y}{n} = \frac{\Sigma (y - \bar{y})^2}{n(n-1)} \text{ (fpc)}
\]

Now, in a given zone \( i \), a sample of \( n_i \) D. U.'s provides \( \hat{y}_1 \) and

\[
S^2_{\hat{y}_1} = \frac{\Sigma (y_{ij} - \hat{y}_1)^2}{n_i - 1} \text{ (fpc)}
\]

and the variance of \( \hat{y}_1 \) about \( \bar{y}_1 \), the true mean, will be

\[
S^2_{\hat{y}_1} = \frac{S^2_{\hat{y}_1}}{n_i}
\]  

(8)

and for a given zone,

\[
S^2_{\hat{y}_1} = S^2_e + S^2_{\hat{y}_1}
\]  

(9)

Over all or selected zones, a pooled or weighted estimate of this variance may be determined from

\[
S^2_Y = \frac{\nu_1 S^2_{y_1} + \nu_2 S^2_{y_2} + \nu_3 S^2_{y_3} + \ldots + \nu_p S^2_{y_p}}{\nu_1 + \nu_2 + \nu_3 + \ldots + \nu_p}
\]

where \( p \) represents the \( p \)th and final zone to be included in the pooled variance calculation.
Now, since in each zone,

\[ S_{y_1}^2 = \frac{S_{y_1}^2}{n_i} = \frac{SSD_i}{n_i \nu_i} \quad (fp_c) \]

where SSD = sum of squares of deviations, then,

\[ \nu_i S_{y_1}^2 = SSD_i \quad (fp_c) \]

Therefore:

\[ S_y^2 = \frac{SSD_1}{n_1} \frac{(fp_c)}{\nu_1} + \frac{SSD_2}{n_2} \frac{(fp_c)}{\nu_2} + \ldots + \frac{SSD_p}{n_p} \frac{(fp_c)}{\nu_p} \]  

(10)

Also, \( \nu_i = n_i - 1 \). Hence,

\[ \Sigma \nu_i = \sum_{i=1}^{p} (n_i - 1) = \Sigma n_i - p \]

so that

\[ S_y^2 = \frac{\sum_{i=1}^{p} \frac{SSD_i}{n_i} \frac{(fp_c)}{\nu_i}}{\sum_{i=1}^{p} n_i - p} \]

Since \( S_{y'}^2 = S_e^2 + S_y^2 \), therefore

\[ S_{y'}^2 = S_e^2 + \sum_{i=1}^{p} \sum_{j=1}^{n_i} \frac{(y_{ij} - \hat{y}_i)^2}{n_i} \left( 1 - \frac{n_i}{N_i} \right) \]  

(11)

Two assumptions have been made in the derivation of Eq. 11:

1. Within a given zone, the frequency distribution of the number of trips per D. U. is approximately normal; and
2. Among all zones, the variances of the number of trips per D. U., \( S_{y_1}^2 \), are approximately equal.

Both assumptions were found to be false. The frequency distribution in the first assumption appeared to be a truncated normal distribution rather than normal, and the \( S_{y_1}^2 \) varied considerably from zone to zone, often because of an occasional very high value of trips per D. U. of 20 or more.

For this analysis (total trips only), the weighted average zonal variance was determined for each sample size by summing the individual sums of squares of deviations about the zonal means and dividing by the total number of sampled D. U.'s less the number of zones. Thus, for zone i,
\[ s_{y_i}^2 = \frac{SSD_i}{n_i - 1} \text{(fpc)} \]

Over all zones, and assuming fpc is the same for all zones, the weighted average zonal variance becomes

\[ \overline{s_y^2} = \frac{\sum SSD_i}{\sum (n_i - 1)} \text{(fpc)} = \frac{\sum SSD_i}{\sum n_i - Z_T} \text{(fpc)} \]

where \( Z_T \) = total number of sampled zones. For any zone having \( n_i \) sampled D. U.'s,

\[ s_{y_i}'^2 = s_e^2 + \frac{\sum SSD_i}{n_i} \]

The values of \( s_e^2, \frac{\sum SSD_i}{\sum n_i - Z_T} \), and fpc for each sample size (the two 5 percent samples have been averaged) are given in Table 4. On the basis of Table 4, Figure 4 was plotted for comparison with Figure 1. In a sense Figure 4 is somewhat artificial, for the \( s_{y_i}'^2 \) values, from which the RMS errors in Figure 4 are derived, are related to the number of dwelling units rather than the number of trips. Each point plotted in Figure 4 represents a certain range of trip volumes; in order to provide a comparison with Figure 1, the same zones grouped for each range in Figure 1 were grouped for each point in Figure 4.

The RMS error and \( s_{y_i}' \) are analogous measures of dispersion, whose squared values are equal to the average value of the squares of the deviations from a central value. In the case of the RMS error in Figure 1, the central value is the expanded or actual number of trips; in the case of the RMS error in Figure 4 derived from \( s_{y_i}' \), the central value is the true number of trips, and this RMS error expresses the statistical expectation of the variation from this value. The values plotted in Figure 4 were obtained by finding \( \bar{x} \), the average number of zonal total trips in each range, and \( \bar{N} \), the average number of total zonal dwelling units in each range. The average RMS error for each range is then, from Eq. 4,

\[ \text{RMS error} = \frac{\bar{N} \cdot s_{y_i}'}{\bar{x}} \]

Expressed as a percentage of the average zonal trip volume in the range,

\[ \text{Percent RMS error} = \frac{\bar{N} \cdot s_{y_i}'}{\bar{x}} \times 100 \quad (12) \]

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>( s^2_e )</th>
<th>( \frac{\sum SSD_i}{\sum n_i - Z_T} )</th>
<th>fpc</th>
<th>( s_{y_i}'^2 )</th>
<th>( s_{y_i}' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%</td>
<td>1.93</td>
<td>28.05</td>
<td>0.975</td>
<td>( 1.93 \times 28.05 \times 0.975 )</td>
<td>( s_{y_i}' = 1.93 \times 28.05 \times 0.975 )</td>
</tr>
<tr>
<td>5%</td>
<td>1.09</td>
<td>22.40</td>
<td>0.950</td>
<td>( 1.09 \times 22.40 \times 0.950 )</td>
<td>( s_{y_i}' = 1.09 \times 22.40 \times 0.950 )</td>
</tr>
<tr>
<td>10%</td>
<td>0.49</td>
<td>22.58</td>
<td>0.900</td>
<td>( 0.49 \times 22.58 \times 0.900 )</td>
<td>( s_{y_i}' = 0.49 \times 22.58 \times 0.900 )</td>
</tr>
</tbody>
</table>
Comparison of Figures 1 and 4 indicates that larger errors may be expected statistically than are apparently obtained when the actual number of trips is used as the standard of comparison. It is to be expected that some of the actual trip volumes used as standard in Figure 1 are in fact in error, so that the true errors in the trip estimates cannot be determined. On the other hand, if the trip estimation equations are very similar among sample sizes, and if they produce very similar zonal trip volume estimates, it follows that the error curves must be very similar, as they are in Figure 1. Figure 4 therefore shows not the true RMS error, but rather what we might consider the limit of the expected RMS error and the dependence of that limit on sample size.

TRIP ATTRACTION

Research Procedure

The trip attraction study was conducted for both Kingston and Barrie, and it too was concerned only with daily person trips by vehicle made by residents of the study area.
Fifty zones were used in the Kingston study and 31 zones in the Barrie study. For Kingston, the same sample sizes were used as for the production study: 2 1/2 percent, 5 percent, 10 percent, and in addition, the full 12 1/2 percent sample. For Barrie, the systematic sample sizes tested were 4 percent, 8 percent, 16 percent, and the full 20 percent sample, with the 16 percent sample size approximating 1,000 interviews, as shown in Table 5.

The trip purpose categories examined in the attraction study were the same for both Kingston and Barrie, and were designated as follows:

- **ALL**: The total of all purpose categories
  - 1: Trips to work
  - 2: Trips made to conduct personal business
  - 8: Trips made to shop
  - 0: Trips made to home

The modal breakdown was also examined to some extent in that examinations were made for:

1. All modes combined (excluding walking trips), and
2. Auto driver and passenger combined.

Since the investigation was based on the assumed existence of a relationship between the trips attracted to an area and certain assessment-area variables, it seemed reasonable to assume that if a relationship did exist, it would exist not only for the total trips attracted to any area for various trip purposes but also for different sized samples of these trips. In an attempt to determine a possible relationship, multiple regression analyses were performed on trips obtained from the various samples selected and the following assessment-area variables:

\[
\begin{align*}
    x_1 & \text{ Assessed value of residential land (thousands of dollars)} \\
    x_2 & \text{ Assessed value of residential buildings (thousands of dollars)} \\
    x_3 & \text{ Assessed value of commercial land (thousands of dollars)} \\
    x_4 & \text{ Assessed value of commercial buildings (thousands of dollars)} \\
    x_5 & \text{ Assessed value of industrial land (thousands of dollars)} \\
    x_6 & \text{ Assessed value of industrial buildings (thousands of dollars)} \\
    x_7 & \text{ Assessed value of public land (thousands of dollars)} \\
    x_8 & \text{ Assessed value of public buildings (thousands of dollars)} \\
    x_9 & \text{ Area of residential land (acres)} \\
    x_{10} & \text{ Area of commercial land (acres)} \\
    x_{11} & \text{ Area of industrial land (acres)} \\
    x_{12} & \text{ Area of public land (acres)}
\end{align*}
\]

The assessment information was obtained by examining the assessment rolls of each city and recording the assessed value of each piece of land and building under its particular land-use category (residential, commercial, industrial, and public) for the traffic zone in which it was located. The land areas in each land-use category for each traffic zone were obtained by planimetry from the land-use maps prepared by each city.

The estimation equations of best fit are documented in a separate report (12). The equations for all purposes, work trips, and shopping trips, for all modes combined are given in Tables 6 and 7 for Kingston and Barrie respectively.

Again, the equations obtained for each travel mode category and trip purpose appeared to be similar in form, for a given city, although there was little or no similarity between cities. This does not necessarily imply that a common relationship is impossible. Rather, it is believed that the differences exhibited in this study may reflect variations in assessment philosophies and methods in the two cities.

The relationships developed also appear to be realistic for each city, in that the assessment-area variables tending to have the most pronounced effect in the estimating equations are those which are most logically associated with the trip purpose under consideration; that is, shopping trips are very dependent upon the commercial assessment-area variables, work trips are dependent upon commercial, industrial, and public assessment-area variables, etc.
### TABLE 6

**REGRESSION EQUATIONS FOR TRIP ATTRACTION ESTIMATES, ALL MODES, KINGSTON**

<table>
<thead>
<tr>
<th>Sample</th>
<th>All Purposes</th>
<th>Work Trips</th>
<th>Shopping Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>S - 025 - K</td>
<td>$Y' = 163 - 4.058x_1 + 1.432x_2 - 0.468x_3 + 5.430x_4 + 0.226x_5 + 23.64x_6 - 19.37x_7 + 3.76x_{11}$</td>
<td>$Y' = 100 - 0.976x_1 + 1.352x_2 + 3.746x_3 + 1.658x_4 + 0.054x_5 + 6.88x_6 - 13.41x_7 + 2.05x_{11}$</td>
<td>$Y' = 15 - 0.790x_1 + 1.687x_2 - 0.98x_3 + 27.09x_{10}$</td>
</tr>
<tr>
<td>$r = 0.98$</td>
<td>$r = 0.97$</td>
<td>$r = 0.95$</td>
<td></td>
</tr>
<tr>
<td>S - 05 - 1 - K</td>
<td>$Y' = 183 + 0.900x_2 - 0.715x_3 + 4.236x_4 - 1.785x_5 + 0.292x_6 + 10.57x_7 - 26.04x_{10} + 2.87x_{11}$</td>
<td>$Y' = -64 - 0.471x_1 + 0.838x_2 + 1.767x_3 + 1.514x_4 + 0.074x_5 + 4.17x_6 - 12.30x_{10} + 1.90x_{11}$</td>
<td>$Y' = 2 - 0.804x_1 + 1.619x_2 + 1.10x_3 + 25.95x_{10}$</td>
</tr>
<tr>
<td>$r = 0.98$</td>
<td>$r = 0.96$</td>
<td>$r = 0.93$</td>
<td></td>
</tr>
<tr>
<td>S - 05 - 2 - K</td>
<td>$Y' = 301 + 0.944x_2 + 3.244x_3 + 0.247x_6 + 2.75x_{11}$</td>
<td>$Y' = 33 + 0.352x_1 + 1.275x_2 + 1.350x_3 + 0.305x_4 + 0.071x_5 + 0.999x_6 + 3.00x_7 - 19.41x_{10} + 0.73x_{11} - 1.21x_{12}$</td>
<td>$Y' = 4 - 0.142x_1 + 0.604x_2 + 1.516x_3 + 18.93x_{10}$</td>
</tr>
<tr>
<td>$r = 0.96$</td>
<td>$r = 0.98$</td>
<td>$r = 0.96$</td>
<td></td>
</tr>
<tr>
<td>S - 10 - K</td>
<td>$Y' = 162 + 3.097x_1 + 1.115x_2 + 0.503x_3 + 4.424x_4 + 0.209x_6 + 20.47x_7 + 24.16x_{10} + 2.93x_{11}$</td>
<td>$Y' = -61 - 0.425x_1 + 1.031x_2 + 1.424x_3 + 0.195x_4 + 1.250x_5 + 0.692x_6 + 3.50x_7 - 10.31x_{10} + 1.26x_{11} - 1.18x_{12}$</td>
<td>$Y' = 9 - 0.679x_1 + 1.519x_2 + 1.06x_3 + 21.50x_{10}$</td>
</tr>
<tr>
<td>$r = 0.98$</td>
<td>$r = 0.98$</td>
<td>$r = 0.95$</td>
<td></td>
</tr>
<tr>
<td>S - Total - K</td>
<td>$Y' = 103 - 2.850x_1 + 1.278x_2 - 0.377x_3 + 4.180x_4 + 0.189x_5 + 17.23x_6 - 21.76x_{10} + 3.16x_{11}$</td>
<td>$Y' = -2 + 0.811x_1 + 0.459x_2 + 0.731x_3 + 0.126x_4 + 1.19x_5 + 5.88x_{10} + 0.70x_{11} - 2.37x_{12}$</td>
<td>$Y' = 1 - 0.138x_1 + 0.702x_2 + 1.540x_{10} + 18.58x_{10}$</td>
</tr>
<tr>
<td>$r = 0.98$</td>
<td>$r = 0.98$</td>
<td>$r = 0.96$</td>
<td></td>
</tr>
</tbody>
</table>

Note: $Y' = \text{No. of trips attracted to zone}$

### TABLE 7

**REGRESSION EQUATIONS FOR TRIP ATTRACTION ESTIMATES, ALL MODES, BARRIE**

<table>
<thead>
<tr>
<th>Sample</th>
<th>All Purposes</th>
<th>Work Trips</th>
<th>Shopping Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>S - 04 - B</td>
<td>$Y' = 79 + 1.478x_1 + 6.022x_2 + 11.148x_3 + 9.29x_{11} + 28.39x_{12}$</td>
<td>$Y' = -16 + 0.849x_2 + 1.028x_3 + 0.216x_4 + 1.07x_5 + 4.20x_{11} + 2.14x_{12}$</td>
<td>$Y' = 32 + 1.058x_1 - 0.276x_2 + 1.02x_3 + 4.90x_{11} + 2.67x_{12}$</td>
</tr>
<tr>
<td>$r = 0.92$</td>
<td>$r = 0.97$</td>
<td>$r = 0.92$</td>
<td></td>
</tr>
<tr>
<td>S - 08 - 1 - B</td>
<td>$Y' = -159 + 4.719x_1 + 0.617x_2 + 6.137x_3 + 1.449x_4 + 17.76x_{10} + 0.33x_{11}$</td>
<td>$Y' = -23 + 1.113x_2 + 2.127x_3 + 0.628x_4 - 1.078x_5 + 6.696x_{10} + 0.277x_{11} + 1.61x_{12} + 0.28x_{11}$</td>
<td>$Y' = -57 + 2.630x_1 - 0.100x_2 + 1.348x_3 + 6.079x_{10} - 1.15x_{11}$</td>
</tr>
<tr>
<td>$r = 0.96$</td>
<td>$r = 0.98$</td>
<td>$r = 0.91$</td>
<td></td>
</tr>
<tr>
<td>S - 08 - 2 - B</td>
<td>$Y' = -53 + 1.342x_1 + 5.751x_2 + 1.429x_3 - 1.175x_4 + 0.735x_{10} + 9.18x_{11} + 0.33x_{12} + 1.25x_{12}$</td>
<td>$Y' = 16 + 3.326x_1 - 0.325x_2 + 0.47x_3 - 1.239x_4 + 6.173x_5 + 1.2x_{11} + 1.72x_{12}$</td>
<td>$Y' = -10 + 1.161x_1 + 0.425x_2 - 0.647x_3 + 0.458x_4 - 0.87x_{11}$</td>
</tr>
<tr>
<td>$r = 0.95$</td>
<td>$r = 0.99$</td>
<td>$r = 0.98$</td>
<td></td>
</tr>
<tr>
<td>S - 16 - B</td>
<td>$Y' = -139 + 2.866x_1 + 0.926x_2 + 2.062x_3 + 6.940x_4 + 0.724x_5 + 15.30x_{10} + 7.38x_{11} + 8.83x_{12}$</td>
<td>$Y' = -5 + 0.631x_1 + 2.478x_2 + 0.385x_3 - 0.65x_4 + 0.574x_5 + 0.195x_6 + 1.00x_7 + 2.41x_{11} + 1.49x_{12}$</td>
<td>$Y' = -75 + 2.505x_1 + 0.335x_2 + 0.196x_3 + 1.80x_4 + 7.15x_{10} + 3.90x_{11}$</td>
</tr>
<tr>
<td>$r = 0.97$</td>
<td>$r = 0.99$</td>
<td>$r = 0.91$</td>
<td></td>
</tr>
<tr>
<td>S - Total - B</td>
<td>$Y' = -96 + 2.118x_1 + 1.309x_2 + 6.230x_3 + 1.164x_4 + 3.18x_5 + 0.939x_6 + 11.42x_{10} + 9.45x_{11} + 11.22x_{12}$</td>
<td>$Y' = -14 + 0.953x_1 + 6.825x_2 - 0.738x_3 - 1.616x_4 + 1.083x_5 + 1.06x_6 + 1.66x_{12}$</td>
<td>$Y' = 89 + 1.046x_1 + 0.570x_2 + 0.223x_3 + 2.21x_4 + 0.02x_{10} + 4.48x_{11}$</td>
</tr>
<tr>
<td>$r = 0.98$</td>
<td>$r = 0.99$</td>
<td>$r = 0.90$</td>
<td></td>
</tr>
</tbody>
</table>

Note: $Y' = \text{No. of trips attracted to zone}$
Trip Estimates and Discussion of Errors

Substitution of the zonal assessment-area data in the estimation equations provided estimates of the number of trips attracted to each zone within the city by travel mode category and trip purpose for each sample size. These trip estimates were compared with the expanded home-interview O-D trip data (actual) and the differences were calculated. The results of these comparisons (12) are shown for all purposes, work trips, and shopping trips, for all modes combined, in both Kingston and Barrie, in Figures 5 through 10. So that these figures could be plotted, actual zonal volumes of attracted trips were grouped into arbitrary ranges of trip volumes, and the percent RMS errors were calculated in the same way as for the trip production estimates.

As in the trip production study, one would expect similar estimation equations to produce trip estimates approximately comparable in accuracy. Examination of the Kingston results shows that this expectation generally holds true. For work trips, personal business trips, shopping trips, and "to home" trips, the 5 percent samples appear to produce results as good as the 10 percent sample with some overlap of the curves. The 2½ percent sample is generally somewhat poorer than the 5 percent or 10 percent samples. For trip purpose ALL, the errors appear to decrease with increase in sample size. Regression equations based on the total O-D sample do not appreciably improve the accuracy of trip estimates.

The error results for Barrie are more variable, although in all cases but shopping trips, the regression equation based on the total O-D sample does appear to improve considerably the accuracy of trip estimates. Only for shopping trips does the 4 percent sample give consistently poorer results than the 8 percent or 16 percent samples. For
Figure 6. Errors in trip attraction estimates, Kingston.

Figure 7. Errors in trip attraction estimates, Kingston.
Figure 8. Errors in trip attraction estimates, Barrie.

Figure 9. Errors in trip attraction estimates, Barrie.
the other four trip purposes, the error curves overlap in such a way that no conclusions can be drawn as to the effect of sample size on estimates of attracted trips.

It may be seen, therefore, that the sample size (within the range of sample sizes tested) has little or no definite and consistent effect on the accuracy of trip attraction estimates. Consequently, it appears that a considerable saving in cost could be achieved by using a relatively small O-D sample.

For the smaller attracted volumes, it may be noted that the percent RMS error is so great as to cause one to question the value of small trip volumes. This implies that perhaps zone boundary selection should be governed to some extent by the number of trips attracted to the zone.

The attraction study shows that trip attraction estimates, as well as production estimates, may exhibit sizable errors. Again, however, it is suggested that the estimates are useful, and that it would seem appropriate to examine the effects of higher production and attraction rates on trip distribution and assignment in critical areas.

Although no calculations were made, the method of determining theoretical standard errors about the true zonal values, as outlined in the production study, should also be applicable, with minor modifications, to trip attraction.

**CONCLUSIONS**

1. Estimation equations may be developed for small cities which relate trip production to car ownership primarily as well as to other population and land-use parameters,
and which relate trip attraction to assessment and land-use parameters. For trips having the same trip purpose and mode of travel, the equations developed for a given city show similarity and also appear to be logical in that the expected significant variables play a predominant role in the estimating equations.

2. Comparison of the estimating equations for trip attraction for the two cities indicates some dissimilarities. It is suggested that variations of this kind might be reduced if equalization of assessment between cities could be accomplished.

3. The size of the origin-destination sample used to develop the estimating equations does not appear to affect significantly the accuracy of trip estimates, although theoretically it would be expected to. (Accuracy here means agreement with expanded O-D survey results.)

4. The size of the percent RMS error in trip estimation appears to increase with more extensive breakdown or segregation of trip purpose categories and land areas. On the one hand, variation of estimating equations among trip purposes is quite evident (especially for trip attraction), indicating that separate equations would appear to be necessary for estimating these different kinds of trips. On the other hand, it would appear that attempts to stratify trips too extensively will result in rather sizable errors.

REFERENCES