A PARTIAL ANALYSIS OF TRIP GENERATION

by

John C. Goodknight

Engineering Research Associate

Research Report Number 60-12

Traffic Assignment

Research Study Number 2-8-63-60

sponsored by

Texas Highway Department

in cooperation with

U. S. Department of Transportation

Federal Highway Administration

Bureau of Public Roads

August, 1968

Texas Transportation Institute

Texas A & M University

College Station, Texas

TABLE OF CONTENTS

SUMMARY AND CONCLUSIONS

The objective of this phase of Research Study 2-8-63-60 was the analysis of trip generation by traffic assignment zones. Particular items of concern were the relative efficiency of different independent variables used for trip estimates, the effect of stratification of trips by purpose and/or direction, and the comparison of multiple regression with simple rates. Data collected in the 1963-64 Waco Urban Transportation Study were used in the analyses.

The analyses were primarily concerned with residential trips. Analysis of non-residential trip generation, using the 1963-64 Waco data, was conducted by the Transportation Center at Northwestern University. A report entitled "Further Comments on the Analysis of Non-Residential Trip Generation," by Thomas, Horton, and Dickey dated November 1966, discusses the findings of this research.

Regression models for non-residential trip generation productions and attractions, as well as origins and destinations, were developed and analyzed in this study.

The research conducted under Study 2-8-63-60 was directed toward the problems of trip end estimation in urban transportation studies conducted by the Texas Highway Department rather than at more elementary or basic analyses. The conclusions and recommendations reflect the findings and interpretation of the research reported herein as well as previous research and experience in trip generation.

The principal conclusions reached in the conduct of the research reported herein are:

(1) In general, basic demographic and land use activity data such as population, labor force, employment, dwelling units, and automobiles,* provide reasonable estimates of trip generation for traffic assignment zones.

Given such data for existing conditions, essentially equivalent estimates of the number of certain residential trip ends may be obtained using any one of several independent variables. For example, the use of the number of automobiles per zone instead of the number of dwelling units or the population seems to have little effect on the accuracy and/or precision of the estimate.

With respect to such "substitutable" variables, the selection of the variable to be used in estimating future trips is largely a matter of preference on the part of the person or persons making the forecasts. However, in addition to considerations of the efficiency of such variables in describing existing trips, due regard should also be given to the reliability with which such variables are forecast.

(2) The conventional or unweighted least squares models provide relatively accurate and unbiased estimates for Home Based Work Productions. Regression estimates for the Home Based Non-Work Productions also appeared to be unbiased, but were less precise. For Attractions and Non-Home Based Productions, however, the regression models consistently underestimated trips

---

* Definitions of variables are given in Appendix A.

for large zones and overestimated trips for the small ones.

(3) The relation between the variance of the number of trips and zone size indicated that the application of the weighted least squares (based on the number of interviews in each zone) would provide more precise trip estimates than the conventional least squares technique. Although weighted multiple regression estimates were not considered in this report, the use of single variable rates (a special case of weighted regression) in certain cases did provide estimates that were superior to those of the conventional multiple regression models.

(4) On the basis of the various estimates examined, the following estimators are suggested:

(a) Home Based Work Productions - Regression models are judged to be superior to rates where the characteristics of the residents are identifiable and can be projected with confidence. For undeveloped areas, however, specific variables on which these more precise regression models are based generally cannot be projected with sufficient accuracy. Therefore, rates are likely to provide estimates that are as good or better than those provided by multiple regression models.

(b) Home Based Non-Work Production - Regression estimates appeared somewhat better than rates, however, in this case it appears to make little difference whether rates or regression models are used. For undeveloped zones, rates are preferred.

(c) Home Based Attractions - For both Work and Non-Work trips, simple rates are judged to be superior to multiple regression in all cases.

(d) Non-Home Based Productions and Attractions - Rates are also considered superior to the conventional multiple regression estimates for all Non-Home Based trips.

(5) Rates calculated as "Ratio of Averages" (total number of trips in all zones divided by total number of units of the independent variable for all zones) are superior to rates calculated as the "Average of Ratios" (average of the individual rates for all zones) when estimates are based on traffic assignment zones.

(6) Stratification of trips by purpose and/or direction for estimation is useful only as long as trips of each category are best estimated by different independent variables or sets of independent variables.

(7) With the possible exception of Non-Home Based trips where Productions and Attractions might be estimated using different independent variables, stratification of trips by direction is not warranted (i.e., Ins = Outs = 1/2 Productions or Attractions).

(8) For the three purposes considered, stratification by purpose improved the regression estimates. For the Home Based Productions estimated using rates, stratification into Work and Non-Work trips did not improve the precision of the estimates. However, these estimates were based on only a limited number of independent variables, and it appears that by further stratifying trips by purpose, increased precision might be obtained using rates.

(9) While over-stratification of trips may result in unstable and highly unsatisfactory estimating equations when the conventional least squares approach is used, instability is not a problem with rates and trips can thus be stratified to a much higher degree.

INTRODUCTION

It is generally accepted that traffic patterns within an urban area are influenced by such factors as the pattern of land use, the socio-economic characteristics of the population, and the nature of the transportation system. The manner in which these factors influence traffic patterns and the generation of trips, however, is not widely agreed upon.

Specific questions which have not yet been fully answered involve: the particular variables to be used as the basis for the estimates; the appropriate form of the "trip estimator," the degree to which trips should be stratified by purpose and/or directions; and the optimum "observation unit" (i.e., traffic zones vs. individual dwelling unit).

In considering trip generation, it is essential that a distinction be drawn between analysis and forecast. In the case of analysis it is of interest to:

(1) Explain as much of the variation of the dependent variable (number of trips) as possible;

(2) Determine the relationship between the dependent variable (i.e., number of trips) and various pertinent independent variables such as population, employment, labor force, automobiles, etc.;

(3) And of lesser importance, determine the relationships between the several independent variables.

Multiple correlation and regression techniques can be used to advantage in such an analysis; however, the use of correlation and regression as an analysis tool has led to some confusion. The fact that regression techniques provide mathematical models for the estimation and/or forecast of the dependent variables has led to the false conclusion that such models are appropriate and useful for the forecast or estimation of trip ends for some future point in time.

Such rationalization fails to take into account the fact that while certain independent variables can be measured very accurately for the survey year, it is difficult or perhaps impossible to project these variables by anything other than a sheer guess. It is inappropriate to include such independent variables in any regression model that is to be used for forecasting. A frequent error, however, has been to include independent variables which show a high correlation with the dependent variable in the forecasting model, even though the technology or understanding of this variable is not sufficient to permit its forecast value to have any real validity.

It has been observed with the Waco data, as well as with data from other cities, that certain independent variables are "substitutable." In other words, certain trip ends can be estimated just as well with one of these "substitutable" variables as the other. Thus, the variable which can be most easily and precisely forecast is the appropriate one to use in a forecast model.

A recent study by the Transportation Center at Northwestern University has dealt with trip generation for non-residential land uses. This study was based on data collected in the 1963-64 Waco Urban Transportation Study. Specific items considered in this analysis of non-residential trip generation were the effects of land use activity grouping and areal aggregation, the influence of locational variables on the attraction of trips to commercial parcels, and the linkages between land use activities or multi-purpose trips. (1)*

The study reported herein was concerned principally with the analysis of residential trip generation, although consideration was also given to the estimation of non-residential trip ends. Trip generation estimates considered here were also based on data from the 1963-64 Waco Urban Transportation Study. For this investigation, the traffic assignment zone was taken as the observation unit.

---

* Numbers in parentheses refer to reference with corresponding number.

PURPOSE AND SCOPE

The general objectives of this investigation were to:

(1) Identify the independent variables which provide the most efficient (statistically as well as logically) estimates of residential trip generation;

(2) Compare the use of rates calculated on the basis of single independent variables with the use of conventional multiple regression models for trip generation estimates, and to suggest conditions under which each of these methods might be most appropriate;

(3) Determine the conditions under which the stratification of trips by purpose and/or direction can be used to improve the precision of trip generation estimates.

For this study, all observations of both trips and the independent variables were aggregated at the traffic assignment zone level so that only zonal estimates were considered. Attention was given to Home Based as well as Non-Home Based trips estimated from the home interview survey. Independent variables used in this analysis included descriptors of the land use (Commercial Area, Net Residential Area etc.) and population characteristics (Total Population, Persons in Labor Force, etc.) for each zone.

Two techniques for estimating trip generation were studied. In the first case, the use of rates based on a single variable describing land use of population characteristics was considered. In the second case, several alternate models, each including a number of independent variables, were examined using multiple regression analysis.

In regard to the stratification of trips, two purposes, Home Based Work and Home Based Non-Work, were considered in detail. Separate estimates obtained for these purposes were combined and compared with the non-stratified estimate. Likewise, directional estimates were compared with the non-directional estimates for each purpose.

DESCRIPTION OF DATA

Study Area

Data used for this study were obtained from the 1963-64 Waco Urban Transportation Study. The study area had a population of approximately 132,000. The CBD was the largest trip generator, accounting for 26 percent of all daily trip ends. Of the 47,000 household units in the area surveyed (excluding Baylor University, Connally Air Force Base, and the other public institutions) approximately 90 percent were single-family dwelling units.

Classification of Trips

Trip data were available for each of the 206 traffic assignment zones of the study. Internal auto-driver trips only were used for this analysis. External trips and trips by other modes were excluded along with intrazonal trips (those trips whose origins and destinations were within the same zone).

The three trip purposes considered were: Home Based Work, Home Based Non-Work, and Non-Home Based. For each purpose, trip ends in each zone were classified as Productions and Attractions. Trips were further divided into Ins (trips originating within the zone of interest with destinations outside that zone) and Outs (trips originating outside the zone of interest with destinations inside that zone). These are summarized as follows:

Home Based Work	Productions	Attractions
Home to Work	Outs	Ins
Work to Home	Ins	Outs
Home Based Non-Work	Productions	Attractions
Home to Non-Work	Outs	Ins
Non-Work to Home	Ins	Outs
Non-Home Based	Productions	Attractions

Independent Variables

A total of 22 independent variables describing the population and land use characteristics available from the survey data were used in the trip generation analysis. These variables are listed in Appendix A.

Zones Eliminated from Consideration

Although 206 traffic assignment zones were included in the survey data, 11 of these were excluded from consideration because of the following irregular land uses:

Urban Renewal
Sewage Treatment Plant
Baylor University
Waco State Home
Veterans Administration Hospital
Airport
James Connally Air Force Base

In considering the residential trip generation rates, additional inconsistencies caused unreasonable or undefined rates for certain individual traffic assignment zones. Reasons for these discrepancies for some of the zones included:

1. There was no residential activity within the zone at the time of the survey (for this discussion, no residential activity is taken as synonomous with a total population of zero).

2. There were errors in the original data (this is evidenced by the fact that for certain zones with no residential activity some Home Based trips were reported).

3. Home interviews were too few to adequately represent the statistical population (one interview per 10 dwelling units was judged to be the minimum adequate sample rate).

The analysis here was based on the entire set of data excluding only those 11 zones with the irregular land uses listed above. However, to determine the effects of the latter inconsistencies on the generation rates and regression models, the analysis of residential trip generation was repeated after excluding those zones with no residential activity.

GENERATION RATES

Three items of interest with respect to the "rates" portion of the study were: (1) the relative efficiency of rates as opposed to regression models as trip estimators; (2) the adequacy of various independent variables as the basis for these estimates; (3) the appropriateness of various methods of calculating trip generation rates. The efficiency of rate estimates is discussed in a later section in which rates and regression estimates are compared. In this section, the independent variables and calculation procedures were compared on the basis of the estimated errors and biases.

Only the home based trips at the zone of production were included in the analysis of rates. Estimates for each of these trips were based on the four following independent variables:

Total Population,
Population 5 Years and Older,
Total Dwelling Units,
Net Residential Area.

Procedure

For each combination of trip classification, independent variable, and calculation procedure, a rate was calculated for the entire study area. Using this rate, the number of trips in each individual zone was estimated and compared with the number from the survey data. The residuals (deviations between observed and estimated number of trips) were then calculated and the standard deviations of these residuals (analogous to standard error of estimate for regression) were taken as indicators of the relative efficiencies of the various estimates.

Three methods were used to calculate trip generation rates for each combination of trip type and independent variable. These are designated here as "Average of Ratios," "Ratio of Averages," and "Quartile Averages."

Average of Ratios

The "Average of Ratios" was determined by first calculating the trip generation rate for each individual zone (excluding all zones where the value of the independent variable was zero because the zone rate is undefined in such cases) and then averaging these individual rates. Effectively, small zones were weighted equally with large zones regardless of the zone size. Mathematically this rate or average can be expressed as follows:

$$b_{AR} = \frac{\sum_{i=1}^{n} \frac{Y_i}{X_i}}{n}$$

Where Y_i = Number of trips in the ith zone

X_i = Number of units of the independent variable in the ith zone

n = Number of zones (exclusive of those zones where the value of the independent variable was zero)

Ratio of Averages

For the "Ratio of Averages" the rate was taken simply as the ratio of the total number of trips for all zones to the total number of units of the independent variable for all zones. The relationship can be expressed as:

$$b_{RA} = \frac{\sum_{i-1}^{n} Y_i}{\sum_{i-1}^{n} X_i}$$

where the variables are defined as on the preceding page except that n includes all zones.

Quartile Rates

For the third calculation method, all zones were ranked in order of increasing value of the individual zone rates. This ranked set of zones was then divided into quartiles and the "Ratio of Averages" for each quartile was computed. Trips for each zone were estimated using the appropriate "Quartile Rate" calculated in this manner.

Results

Rates and standard errors for the various combinations of calculation procedure, independent variable, and trip classification are included in the following tables. Table 1 applies for the set of data which included zones with no residential activity; these zones were excluded from the data presented in Table 2.

"Quartile Rates" are included only for certain trip categories shown in Table 1. The mean rates shown are actually the averages of the four individual "Quartile Rates."

TABLE 1

SUMMARY OF TRIP GENERATION RATES

(Including Zones With No Residential Activity)

Trip Classification	Independent Variable	Average of Ratios Mean Rate	Average of Ratios Standard Error	Ratio of Averages Mean Rate	Ratio of Averages Standard Error	Quartile Rates Mean Rate	Quartile Rates Standard Error
H-W(OUT)	TOT POP	.18	52	.18	52	.16	26
	POP 5+	.20	51	.21	51	.17	27
	TOT DU	.48	59	.53	56	.45	26
	NET RES	.27	66	.24	63	.21	22
H-NW(OUT)	TOT POP	.42	148	.39	141		
	POP 5+	.46	138	.44	135		
	TOT DU	1.03	144	1.12	143		
	NET RES	.60	161	.51	144		
W-H(IN)	TOT POP	.16	46	.17	46	.15	23
	POP 5+	.18	44	.19	44	.17	22
	TOT DU	.43	54	.48	50	.42	20
	NET RES	.25	61	.22	57	.20	20
NW-H(IN)	TOT POP	.42	147	.39	141	.36	56
	POP 5+	.46	138	.44	134	.40	54
	TOT DU	1.07	145	1.12	145	1.00	67
	NET RES	.60	161	.51	144	.47	62

TABLE 1 (cont'd)

Trip Classification	Independent Variable	Average of Ratios		Ratio of Averages	
		Mean Rate	Standard Error	Mean Rate	Standard Error
HBW-P	TOT POP	.34	96	.35	96
	POP 5+	.38	93	.39	92
	TOT DU	.91	110	1.01	104
	NET RES	.52	126	.46	118
HBNW-P	TOT POP	.84	293	.78	281
	POP 5+	.91	274	.88	268
	TOT DU	2.10	288	2.24	287
	NET RES	1.21	322	1.03	287
W-H(IN) + NW-H(IN)	TOT POP	.58	177	.56	172
	POP 5+	.64	165	.63	164
	TOT DU	1.50	183	1.60	181
	NET RES	.85	208	.74	187
H-W(OUT) + H-NW(OUT)	TOT POP	.59	180	.57	176
	POP 5+	.65	169	.64	167
	TOT DU	1.50	187	1.64	183
	NET RES	.87	211	.75	190

TABLE 2

SUMMARY OF TRIP GENERATION RATES

(Excluding Zones With No Residential Activity)

Trip Classification	Independent Variable	Average of Ratios		Ratio Of Averages	
		Mean Rate	Standard Error	Mean Rate	Standard Error
HBW-P	TOT POP	.34	100	.35	100
	POP 5+	.38	97	.40	96
	TOT DU	.94	111	1.01	108
	NET RES	.53	135	.46	123
HBNW-P	TOT POP	.84	305	.78	292
	POP 5+	.91	285	.87	279
	TOT DU	2.14	299	2.24	299
	NET RES	1.23	345	1.03	299
W-H(IN) + NW-H(IN)	TOT POP	.58	184	.56	179
	POP 5+	.64	172	.63	170
	TOT DU	1.51	190	1.61	188
	NET RES	.88	225	.74	194
H-W(OUT) + N-NW(OUT)	TOT POP	.59	188	.57	184
	POP 5+	.65	175	.64	174
	TOT DU	1.56	191	1.65	190
	NET RES	.89	228	.75	197

Comparison of Calculation Procedures

Since the "Average of Ratios" method involves calculation of rates for each individual zone, it was necessary to exclude all zones having zero values for units of the independent variable in order to avoid undefined rates. The effects of omitting these zones were not considered significant for the independent variables and home based trip ends considered here; however, in certain cases this may result in biased estimates of trip generation rates. It can also be shown that the "Average of Ratios" indeed produces a biased estimate of the true trip generation rate and that the magnitude of this bias is directly proportional to the square root of the sample size.

The standard errors for the "Quartile Rates" were appreciably lower than for the other calculation methods in the cases considered. However, because grouping zones on the basis of trip generation rates requires that both the number of trips as well as the number of units of the independent variable be known, this method is of no value in forecasting future trips. Furthermore, considerable computational effort is required to rank the zones according to trip generation rates.

Of the three methods considered, the "Ratio of Averages" provided the best estimates of the trip generation rates, at least for estimates made at the zone level. Since the "Ratio of Averages" is simply the ratio of total number of trips to total number of units of the independent variable for all zones considered, it can be calculated with a minimum of effort. In all cases, the standard error for the "Ratio of Averages" was as low or lower than that for the "Average of Ratios" and the resulting estimates were not unduly sensitive to variations for the small zones.

Independent Variables

Again, using the computed standard error as the figure of merit, results contained herein indicated that, of the variables considered, population measurements (i.e., Total Population or Population 5 Years and Older) provided the best trip generation estimates (on the basis of a single independent variable) with Total Dwelling Units exhibiting the next best fit. The poorest fit was obtained using Net Residential Area. With respect to the population parameters, estimates based on Population 5 Years and Older consistently exhibited standard errors lower than those for estimates based on Total Population. However, these differences were not great and may be of no practical significance.

It is emphasized that these judgments are made in the context of fitting relationships to measured data rather than in forecasting future trips. It is quite possible that other variables, such as Residential Area, can be projected with sufficiently greater reliability than Total Dwelling Units. Likewise, Total Population or some other variable may, in the long run, provide better estimates of future trips than can be obtained using Population 5 Years and Older.

Effect of Zones with No Residential Activity

The influence of deleting zones with no residential activity is illustrated by comparing aggregated trip rates in Tables 1 and 2. For the "Average of Ratios" method, the mean rates calculated with those zones excluded were essentially the same as the values previously obtained, with a slight increase in the standard error. That this change in standard error has any practical significance is doubtful since the zones deleted contributed little to the residual variation but still accounted for one degree of freedom each. Furthermore, since zones having zero values for units of the independent variable were removed during the original calculations, the first and second sets of data were essentially the same with respect to the "Average of Ratios" calculation method.

For the "Ratio of Averages" method, differences in the calculated rates were negligible when zones with no residential activity were excluded. This can be attributed to the relative insensitivity of this method to errors in the small zones. As stated previously, the standard error when these zones were deleted was increased, but again this can partly be attributed to a reduction in the degrees of freedom without a corresponding reduction in the residual sums of squares.

MULTIPLE REGRESSION

In considering the application of multiple regression for trip generation estimates, models were developed for Home Based Productions and Attractions as well as Non-Home Based trips. These trips were also stratified by direction for separate consideration. For each trip classification, an attempt was made to develop the most efficient and reasonable model based on selected measures of the land use and population characteristics of each zone.

Procedure

The form of the model obtained for estimating the number of trips of each classification is:

$$Y_j = B_o + B_1X_{1j} + B_2X_{2j} + ... + B_mX_{mj}$$

where:

Y_j = Number of trips estimated for the jth zone

X_{ij} = Observed value of the ith variable in the jth zone

B_i = Partial regression coefficient associated with the ith independent variable

m = Total number of independent variables included in the model

Assumptions made in determining the confidence with which trips are estimated using this model include the following:

1. The regression coefficients (B_i) are constant.
2. The independent variables are measured without error.
3. Errors or deviations of estimated values from actual observations are independent of each other and the X's.

4. The variance of the errors is everywhere constant (homoscedasticity).

Details of the multiple regression analysis technique are given in most references dealing with statistical methods. A discussion of the technique with specific reference to trip generation analysis has been given by Schuldiner, et al. (2).

For this analysis, multiple regression models were developed for each of the independent variables using the stepwise multiple regression technique. Initially, regression models were "constructed" by selecting the order of variables to be included in the model as that giving the highest R^2 at each successive step. At later stages of the analysis, alternative combinations of independent variables were examined. The criteria used for evaluation of these alternatives and for the selection of the final models are given below:

(1) <u>Coefficient of determination (R^2)</u>

The test for statistical significance of an increase in R^2 with the addition of each independent variable to the model is included in the regression program. The test statistic is calculated as:

$$F_{(k-m),(n-k-1)} = \frac{R_k^2 - R_m^2}{k - m} \bigg/ \frac{1 - R_k^2}{n - k - 1}$$

where:

(k-m) and (n-k-1) are the degrees of freedom for the 'F' distribution,

R_K^2 = coefficient of determination for the model with

R_m^2 = coefficient of determination for the subset of the above model with "m" independent bariables;

n = number of observations (i. e. zones).

If the value of 'F' calculated above is larger than the theoretical value of 'F', then the increase in the value of R^2 by the inclusion of the independent variables, m + 1, m + 2, ..., k, is declared statistically significant.

(2) Standard Error of Estimate of the model (SE)

The reduction in the SE, together with the rate of change of this reduction as additional independent variables are included in the model, provide alternative measures of the efficiency of these variables in explaining the variations of the dependent variable. Models were generally truncated at the point where reduction in SE became relatively minor and the rate of the decrease diminished.

(3) Reasonableness of the combination or mix of independent variables

"Explanatory" or independent variables should be logically related to the dependent variable (in this case, number of trips). Most desirable, of course, is the case where the independent variables provide direct measures of the factors "causing" the generation of trips. Furthermore, independent variables should, as much as possible, measure separate characteristics of the causative factors, and should not merely be redundant measurements of the same characteristic.

(4) Availability of data with respect to the independent variables

Variables selected should be those which can be measured with relative ease, kept current or updated without excessive effort, and projected to a future planning date with reasonable accuracy.

(5) Number of variables to be measured

The model containing the smallest number of measured variables to be projected is generally preferred, other things being equal.

(6) Sensitivity of model to "errors" in estimates in input data

The calculated value of "y" should not fluctuate wildly when a small change is made in the value of an independent variable that is subject to considerable variation and/or is difficult to project with precision.

(7) Plots of Residuals

Plots of Residuals (deviations between observed and estimated values) versus the dependent variable are frequently useful in detecting biased models when the degree of bias is a function of zone size. An attempt was made here to select models for which the systematic deviations between observed and estimated values were least significant.

Regression Models

The models selected for each trip classification are given on the following pages. Included are R^2 and Standard Error of Estimate(SE) for each of the models shown. Numbers shown in parentheses under each coefficient represent the SE for that particular coefficient. (When multiplied by 1.96, these values give the 95% confidence intervals for the coefficient.) Plots of the residuals vs. the independent variables for Production and Attraction models are included in Appendix B.

HOME BASED WORK (Productions)

$$Y = 3.6 + \underset{(0.048)}{0.67}\ X_4 - \underset{(0.025)}{0.099}\ X_5 + \underset{(0.089)}{0.064}\ X_{22}$$

$R^2 = .961$

SE = 52

X_4 = Number of Cars

X_5 = Total Population

X_{22} = Total Labor Force (home end)

Substitution of Civilian Labor (X_8) for Total Labor in the above model resulted in a slight decrease in R^2 (to .950) and an increase in SE (to 59). Inclusion of additional variables did not appreciably change the R^2 or SE.

HOME TO WORK (Out)

$$Y = 2.3 + 0.36\ X_4 - 0.069\ X_5 + 0.37\ X_{22}$$
$$(0.026) \quad (0.013) \quad (0.047)$$

$R^2 = .960$

SE = 27

X_4 - Number of Cars

X_5 - Total Population

X_{22}- Total Labor Force (home end)

Alternate:

$$Y = 0.4 - 0.11\ X_5 + 0.19\ X_{12} + 0.73\ X_{22}$$
$$(0.024) \quad (0.047) \quad (0.056)$$

$R^2 = .925$

SE = 38

X_{12}= Total Dwelling Units

In this model, Cars and Total Dwelling Units were considered essentially substitutable, with no additional gain when both were included in the model. Although in the model with Cars, the R^2 was slightly higher and SE lower, estimation on the basis of Dwelling Units rather than Cars might be preferred as a more logical alternative.

WORK TO HOME (In)

$$Y = 1.6 + .31\,X_4 - 0.035\,X_5 + 0.28\,X_{22}$$
$$(0.025) \quad (0.013) \quad (0.047)$$

$R^2 = .953$

SE = 27

X_4 - Number of Cars

X_5 - Total Population

X_{22}- Total Labor Force (home end)

Alternative:

$$Y = 0.5 - 0.075\,X_5 + 0.18\,X_{12} + 0.58\,X_{22}$$
$$(0.023) \quad (0.044) \quad (0.052)$$

$R^2 = .923$

SE= 27

X_{12} - Total Dwelling Units

As in the case of HW (Out), Total Dwelling Units and Cars appeared to be essentially substitutable, with the Cars model giving only a slightly better R^2 and SE Residuals were relatively small with no marked deviation of estimated from observed values throughout the range of values considered.

HOME BASED WORK (Attractions)

$$Y = 22.6 + 0.79\ X_{11} + .80\ X_{13} + 1.7\ X_{15}$$
$$\quad (0.043) \quad (0.05) \quad (0.42)$$

R^2 = .799

SE = 117

X_{11}- Employment (work end)

X_{13}- Area in Commercial Use

X_{15}- Area in Office Use

Much of the variation in HBW-A's can be accounted for by Employment (R^2 = .728). With Employment and Commercial Area alone, the R^2 was .782 and the SE, 121, but a slight skewness was noted for the residual plots, indicating a slight underestimate in the trips for the larger zones. The addition of Office Area improved the statistical properties of the model, but further improvement with each additional variable using the stepwise procedure was slight.

HOME TO WORK (In)

$$Y = 10.2 + 0.43\ X_{11} + 0.38\ X_{13} + 0.79\ X_{15}$$
$$\quad (0.022) \quad (0.082) \quad (0.22)$$

R^2 = 0.797

SE = 62

X_{11}- Employment (work end)

X_{13}- Area in Commercial Use

X_{15}- Area in Office Use

For this model, deleting Office Area reduced the R^2 only to .783 and increased the SE to 64. Improvements in statistical properties with addition of terms to the model shown using the stepwise regression procedure were negligible.

WORK TO HOME (Out)

$$Y = 12.2 + 0.36\ X_{11} + 0.41\ X_{13} + 0.88\ X_{15}$$
$$(0.021) \quad (0.074) \quad (0.20)$$

$R^2 = 0.791$

SE = 56

X_{11} - Employment (work end)

X_{13} - Area in Commercial Use

X_{15} - Area in Office Use

Exclusion of Office Area from the above model resulted in an R^2 and SE of 0.770 and 59. Residuals plotted for this model were slightly skewed with underestimates for the large zones and over-estimates for the smaller zones, although the magnitude of the residuals was relatively small.

HOME BASED NON-WORK (Productions)

$$Y = 7.4 + 2.4\ X_4 - 0.16\ X_5$$
$$(0.14) \quad (0.051)$$

$R^2 = .914$

SE = 170

X_4 - Number of Cars

X_5 - Total Population

Using only one variable at a time, 90.9% of the variation in trips can be explained by variations in Cars or 77.4% by variations in Total Population. Taken together, these two variables still account for only 91.4% of the variation, and the standard error is decreased from 174 for Cars only to 170 for Cars and Population. Additional variables added to the model changed the statistical properties only slightly, with a negligible change attributable to each additional variable included.

HOME TO NON-WORK (out)

$$Y = 4.4 + 1.2\ X_4 - 0.08\ X_5$$
$$\qquad (0.069) \quad (0.026)$$

$R^2 = .910$

SE = 87

X_4 - Number of Cars

X_5 - Total Population

In this case, substitution of Total Dwelling Units for Cars changed the R^2 and SE to .771 and 138. Residuals for the model shown indicated only a slight underestimation of the number of trips for the very large zones.

NON-WORK TO HOME (In)

$$Y = 3.2 + 1.2\ X_4 - 0.084\ X_5$$
$$\qquad (0.067) \quad (0.025)$$

$R^2 = .915$

SE = 84

X_4 - Number of Cars

X_5 - Total Population

Inclusion of Total Dwelling Units in the above model did not change either the R^2 or SE Substitution of Dwelling Units for Cars gave an R^2 of .773 and SE of 138 indicating a poorer fit than for the model shown.

Residuals for the model given here indicated a reasonably good fit of the regression equation to the observed data.

HOME BASED NON-WORK (Attractions)

$$Y = 94.6 + \underset{(0.13)}{0.64\ X_{12}} + \underset{(0.45)}{5.1\ X_{13}} + \underset{(0.52)}{2.4\ X_{19}}$$

$R^2 = .500$

SE = 414

X_{12} - Total Dwelling Units

X_{13} - Area in Commercial Use

X_{19} - Area in Public and Institutional Use

For this model the skewness of the residuals was very pronounced with underestimates on the order of 1,500 trips for several of the large zones. Although significant increases in R^2 and decreases in SE were possible, these were obtained only with the addition of a large number of terms, and using the stepwise procedure, the increase with each additional term was small.

HOME NON-WORK (In)

$$Y = 48.2 + \underset{(0.021)}{0.099\ X_5} + \underset{(0.22)}{2.6\ X_{13}} + \underset{(0.26)}{1.4\ X_{19}}$$

$R^2 = .517$

SE = 203

X_5 = Total Population

X_{13} = Area in Commercial Use

X_{19} - Area in Public and Institutional Use

As indicated by the residual plots, the above model tends to underestimate trips for larger zones and overestimate for the smaller ones. In general, these deviations were large, but most of the zones fell within the range for which the deviations were least pronounced.

Inclusion of additional terms in the model did not result in substantial improvements in the statistical properties.

NON-WORK TO HOME (Out)

$$Y = 53.7 + 0.11\ X_5 + 2.6\ X_{13} + 1.0\ X_{19}$$

$$\qquad\quad (0.022) \quad (0.24) \quad (0.27)$$

$R^2 = .471$

SE = 216

X_5 - Total Population

X_{13} - Area in Commercial Use

X_{19} - Area in Public and Institutional Use

In this case, significant skewness of the residuals plotted against observed trips was noted, indicating consistent underestimation for larger zones. Inclusion of Dwelling Units in the model did not result in any further improvement.

NON-HOME BASED (Productions)

$$Y = 57.9 + .087\ X_5 + 0.20\ X_{11} + 3.0\ X_{13} + 1.7\ X_{19}$$
$$(0.021) \quad (0.077) \quad (0.25) \quad (0.27)$$

$R^2 = .639$

SE = 200

X_{11} - Employment (work end)

X_{13} - Area in Commercial Use

X_{19} - Area in Public and Instititutional Use

X_5 - Total Population

Alternate:

$$Y = 39.4 + 0.30\ X_4 + 0.22\ X_{11} + 3.0\ X_{13} + 1.6\ X_{19}$$
$$(0.053) \quad (0.075) \quad (0.24) \quad (0.26)$$

$R^2 = .663$

SE = 193

X_4 - Number of Cars

Choice of models for NHB-P is largely a matter of preference in using Cars or Total Population as the fourth independent variable. Although the increase in R^2 with either one was statistically significant, no improvement resulted when both were included.

The plot of residuals against the observed number of trips for both of the above models indicated a definite skewness with an underestimation of trips for the larger zones. Using the stepwise procedure, however, a substantial increase in R^2 was obtained only after a large number of terms were added to the model, resulting in a rather unwieldy equation.

NON-HOME BASED (Attractions)

$$Y = 70.6 + 0.33 X_{12} + 3.2 X_{13} + 0.73 X_{15} + 1.3 X_{19}$$
$$\quad\quad (0.07) \quad\quad (0.27) \quad\quad (0.80) \quad\quad (0.29)$$

R^2 = .581

SE = 217

X_{12}- Total Dwelling Units

X_{13}- Area in Commercial Use

X_{15}- Area in Office Use

X_{19}- Area in Public and Institutional Use

Substitution of Total Population for Total Dwelling Units in this model changed the R^2 and SE by a negligible amount (.576 and 218). Residuals plotted for this model indicated a general underestimation of trips for larger zones and a slight overestimate for the smaller ones, regardless of whether Total Population or Total Dwelling Units was used. The change in R^2 and SE with additional terms using the stepwise procedure was small.

Discussion of Models

On the basis of R^2, SE, and the plots of the residuals against the dependent variables, the multiple regression equations appeared to describe home based trips recorded at the production end reasonably well. For trips observed at the attraction end and for non-home based trips, the regression equations tended to underestimate trips for larger zones and overestimate trips for smaller zones. For home based trips, variations of data about the fitted regression line were greater for non-work than work trips at both the attraction and production ends.

At least three possible causes for the failure of certain models developed, particularly for the Non-Home Based and the Non-Work Trips, to reasonably describe the data are suggested:

(1) Errors in the measurement of data;

(2) Failure to include the appropriate variables or functional relationships in the model;

(3) Violation of the assumption of homoscedasticity.

Errors in Measurement

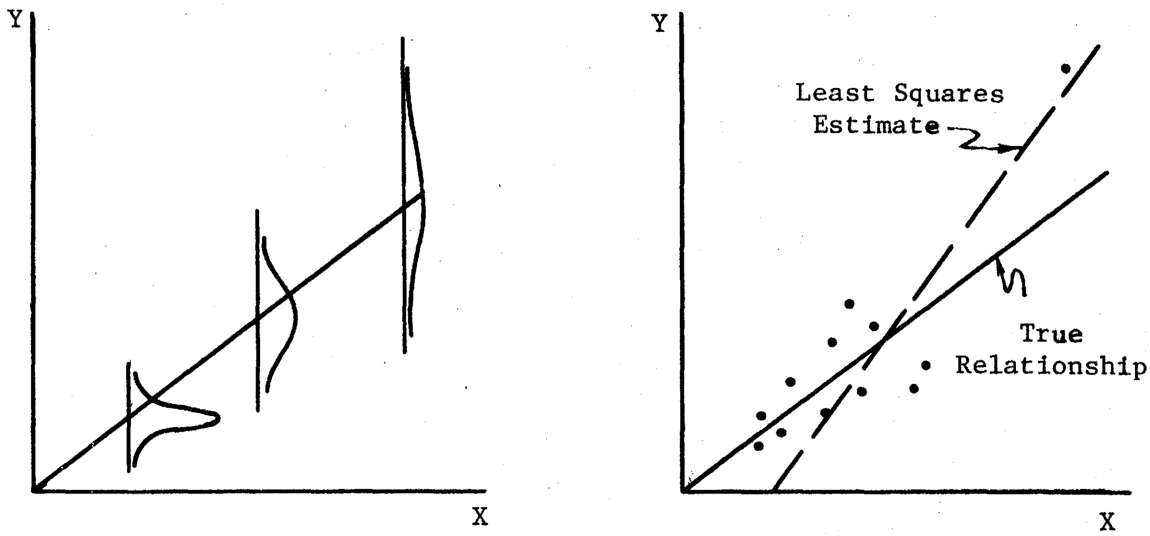
The relatively poor quality of the regression models in estimating trip attractions, as opposed to productions, might be due in part to erroneous or incomplete measurement of employment from secondary sources, errors in measuring trip attractions from the home interview information, or a combination of these two. From this set of data, it was not possible to ascertain which of these sources or to what degree they were responsible for the poor quality of the regression estimates.

Omission of Variables

Failure to properly account for the effects of all independent variables affecting the value of the dependent variable of a regression model (by omitting the variable altogether or omitting some appropriate non-linear function of the variable) may result in a biased estimate. Such specification bias is generally evidenced by systematic deviation of residuals (differences between observed and estimate values) from zero. For certain of the regression models developed here, such bias in the form of overestimation for the small zones and underestimation for the larger zones was noted.

Violation of Assumption

The assumption of greatest concern here is that of homeoscedasticity (constant variance of the dependent variable throughout the observed range of values). Schuldiner has pointed out that violation of this assumption for residential trips may be quite marked,[3] and indeed a definite increase in variance for larger zones was observed in this study, especially for Non-Home Based and Non-Work trips. Under certain conditions, this factor can significantly affect the usefulness of a model as indicated by the following hypothetical example.



Known Distribution of Hypothetical Population

Samples from Hypothetical Population

Figure 1. Hypothetical example illustrating estimation error based on a sample for a heteroscedastic population.

For the population shown above, the Y is directly proportional to X with a considerable increase in variance as X (also Y) increases. For the sample of this population shown in the opposite sketch, the points are generally clustered around small values of X and Y with very few points at the upper end of the range.

From this example, it can be seen that the possible effect on the least squares regression line of a single point sampled with large error at a high value of X is very pronounced where other points are concentrated at lower values. The least squares estimate will almost invariably be forced through or near such a point with little regard to the relationship of the other points. Since the variation of such sample points about the line describing the actual relationship is due to chance alone, with the least squares line is unbiased, but grossly inaccurate. Where data tend to be concentrated at the low end of the observational range, the conventional least squares estimate is unduly sensitive to normal variations in Y for larger values of X.*

This trend was generally noted for the trip generation data used from the Waco study and was particularly pronounced for those dependent variables for which the regression models provided poor estimates.

Zones With No Residential Activity

Although not included here, regression models were also developed using the data with the zones of no residential activity included. These equations included in the same independent variables given in the

---

*This discussion assumes that all sample points are in the population considered. For outliers or sample points which are not part of the population considered, the effects are likely to be more serious. Such might be the case where certain zones included in the sample exhibit some unusual characteristics not typical of other zones.

models presented earlier. No appreciable differences were noted with respect to R^2, SE, the regression coefficients, or the residuals when the two sets of results were compared. This indicated, at least in this case where the proportion of zones having no residential activity was small, that it made little difference in the ultimate regression equation whether or not they were included in the analysis.

COMPARISON OF RATES AND REGRESSION

Using the conventional or unweighted least squares criterion (SE and R^2) for judging the adequacy of the estimating equations, it appears from Table 3 that the regression models developed are superior to the rates in all cases considered. In fact, it can be shown mathematically that when the independent variable used for the rates calculation is included in the standard linear regression model, the regression model will give the minimum standard error.

TABLE 3

COMPARISON OF STANDARD ERRORS FOR ESTIMATION BY REGRESSION AND RATES METHODS

TRIP TYPE	STANDARD ERRORS*	
	Regression	Rates
HBW-P	52	92
HBNW-P	170	268
H-W(Out)	28	51
H-NW(Out)	87	134
W-H(In)	28	44
NW-H(In)	84	134

The question in making the comparison, however, is not which method gives the best least squares estimate, but whether least squares as applied here is the most appropriate technique for estimating trip

(*) For Regression, Standard Errors were taken from the models shown. For Rates, Standard Errors were obtained using the "Ratio of Averages" method of calculation and Population 5 Years and Older as the independent variable.

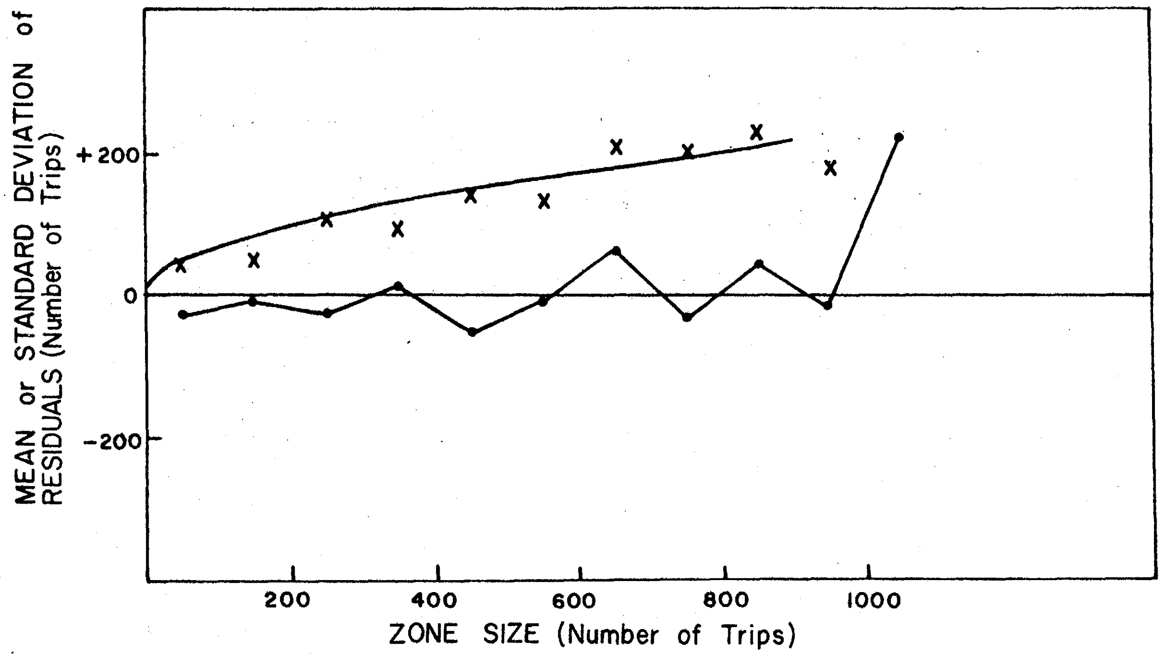
generation. It has been pointed out earlier that where clustering of data occurs, with a small percentage of the values falling outside the range of the cluster, these isolated points may seriously reduce the reliability of the regression estimate.

Heteroscedasticity

To indicate the relationship between the variance of the independent variable (number of trips) and zone size (number of trips per zone), residuals for the regression models previously developed were examined. For each of these models, the range of the dependent variable (number of trips) was divided into increments of 100 observed trips per zone and the residuals were grouped accordingly. For each such group of residuals, the mean and standard deviation were computed. Figures 2-4 show graphically the relationships between these two paramenters and the observed number of trips per zone.

From these figures, it appears that a parabola through the origin reasonably approximates the relationship between the standard deviation and the dependent variable, or equivalently, it appears that the variance is proportional to the dependent variable. This trend is especially true for Non-Home Based trips and Home Based Non-Work trips. For the Home Based Work trips, it appears that a constant might also provide a reasonable approximation of the variance over most of the range of data, excepting only the zones for which a small number of trips was reported. This constant approximation appeared somewhat better for the Productions than the Attractions.

From these figures, it is again apparent that the assumption of

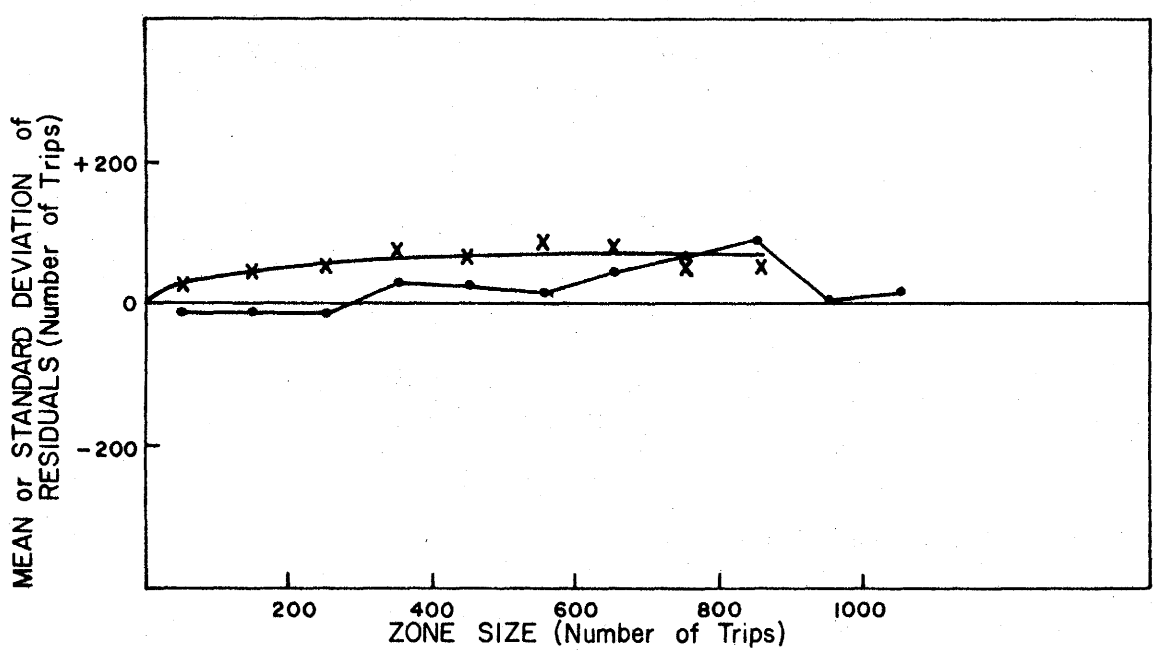


HBNW - P

LEGEND:

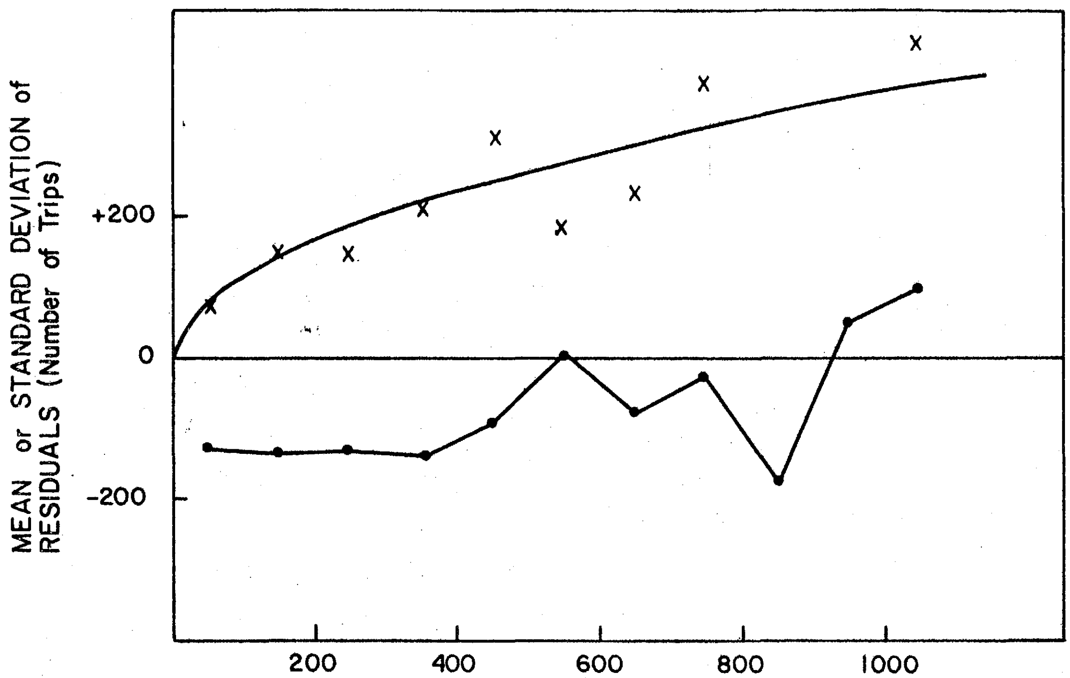
• MEAN DEVIATION OF RESIDUALS

X STANDARD DEVIATION OF RESIDUALS



HBW - P

FIGURE 2



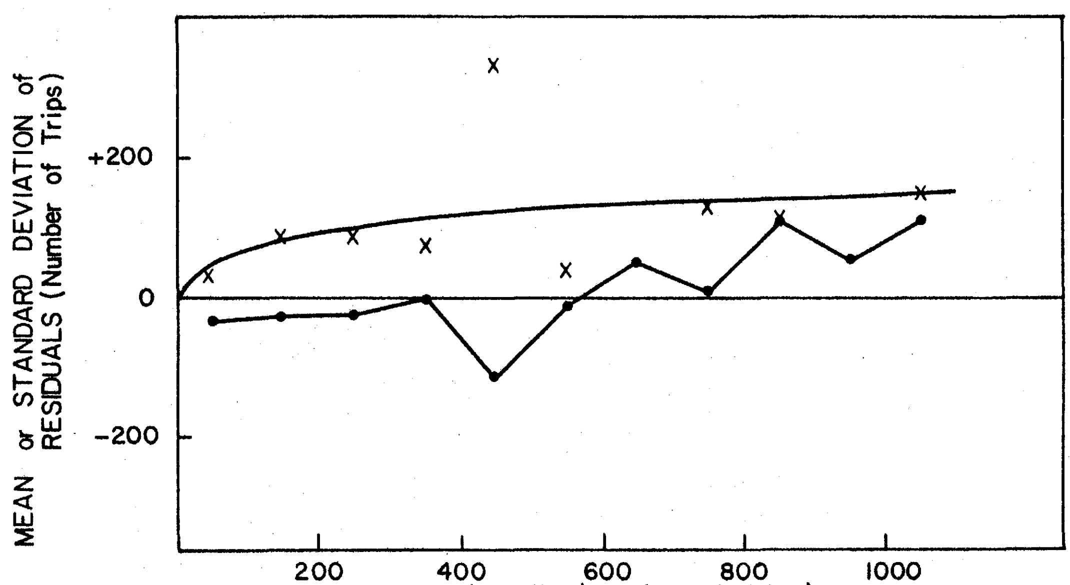
ZONE SIZE (Number of Trips)

HBNW - A

LEGEND:

• MEAN DEVIATION OF RESIDUALS

X STANDARD DEVIATION OF RESIDUALS



ZONE SIZE (Number of Trips)

HBW- A

FIGURE 3

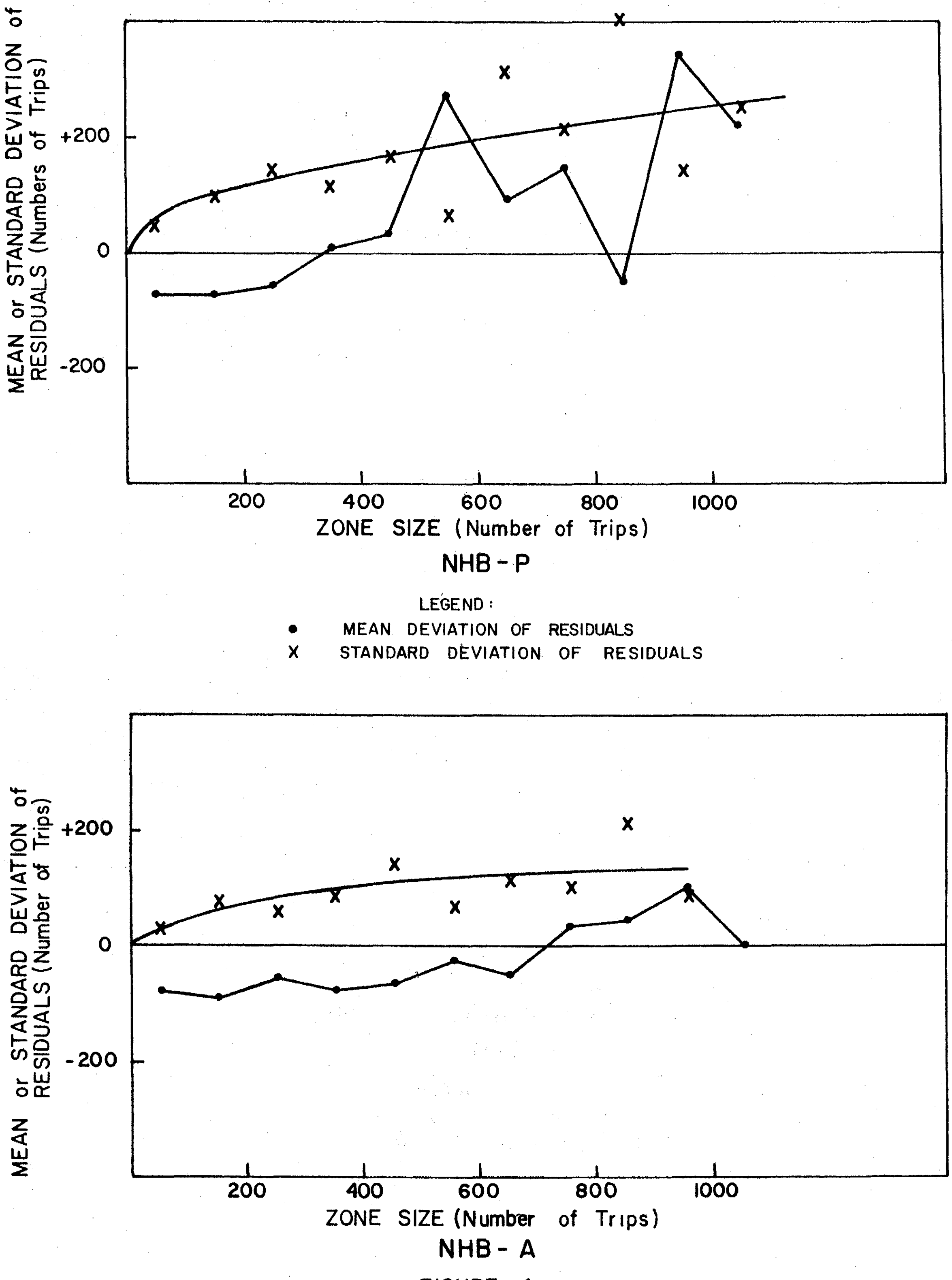

MEAN or STANDARD DEVIATION of RESIDUALS (Numbers of Trips)
+200
0
-200
200
400
600
800
1000
ZONE SIZE (Number of Trips)
NHB - P
LEGEND:
• MEAN DEVIATION OF RESIDUALS
X STANDARD DEVIATION OF RESIDUALS
MEAN or STANDARD DEVIATION of RESIDUALS (Number of Trips)
+200
0
-200
200
400
600
800
1000
ZONE SIZE (Number of Trips)
NHB - A

FIGURE 4

uniform variance made in applying the multiple regression models has been violated. To provide the optimum trip generation estimates using regression, therfore, some compensation for non-uniform variances should be made. This compensation can be provided by the weighted regression procedure.

Weighted Regression

The application of weighted regression has been presented in detail elsewhere, and is considered only very briefly here (4, 5). Although weighting is applicable to the general regression problem, specific consideration is given here to the simple regression model for a single independent variable with a zero Y-intercept. This model is given by:

$$Y_i = \beta X_i + \varepsilon_i$$

where Y_i = value of the dependent variable for the ith observation;

X_i = value of the independent variable for the ith observation;

ε_i = error in measuring Y_i.

In general, the regression coefficient in the above equation is given by:

$$\beta = \frac{\sum_{i=1}^{n} W_i X_i Y_i}{\sum_{i=1}^{n} W_i X_i^2}$$

where: W_i = weight assigned to the ith observation;

n = total number of observations.

In the case where the variance is constant, $W_i = 1/k$ and the best estimate of β is then reduced to:

$$b = \frac{\Sigma X_i \; Y_i}{\Sigma X_i^{\;2}}$$

which is the form used to estimate the coefficient in the conventional least squares method. When the variance of Y_i is proportional to X_i, then $W_i = 1/kx_i$ and the best estimate of β is:

$$b = \frac{\Sigma Y_i}{\Sigma X_i}$$

Finally, when the variance of Y_i is proportional to $X_i^{\;2}$ (standard deviation is proportional to X_i), $W_i = 1/kw_i^{\;2}$ and β is best estimated by:

$$b = \frac{\Sigma \; (Y_i/X_i)}{n}$$

As was pointed out earlier, the variance of Y for the models considered here appears to be approximately proportional to the value of Y. Assuming that Y is proportional to X implies that the variance of Y is also approximately proportional to X. This incidates that for trip generation estimates where the chacteristics of the data are similar to the Waco data, the best estimate of β is approached by $b = \Sigma Y_i/\Sigma X_i$, which is exactly the estimator selected on the basis of earlier consideration or rates.

It can further be seen that the "Ratio of Averages" and the "Average of Ratios" methods of rate calculations discussed earlier are simply applications of the weighted regression with zero intercepts, a single independent variable, and weights of $1/kx_i$ and kx^2, respectively.

In general $b = \Sigma Y_i/\Sigma X_i$ ("Ratio of Averages") is the simplest of the three methods discussed and appears to be the most efficient means of calculating trip generation rates at the traffic zone level.

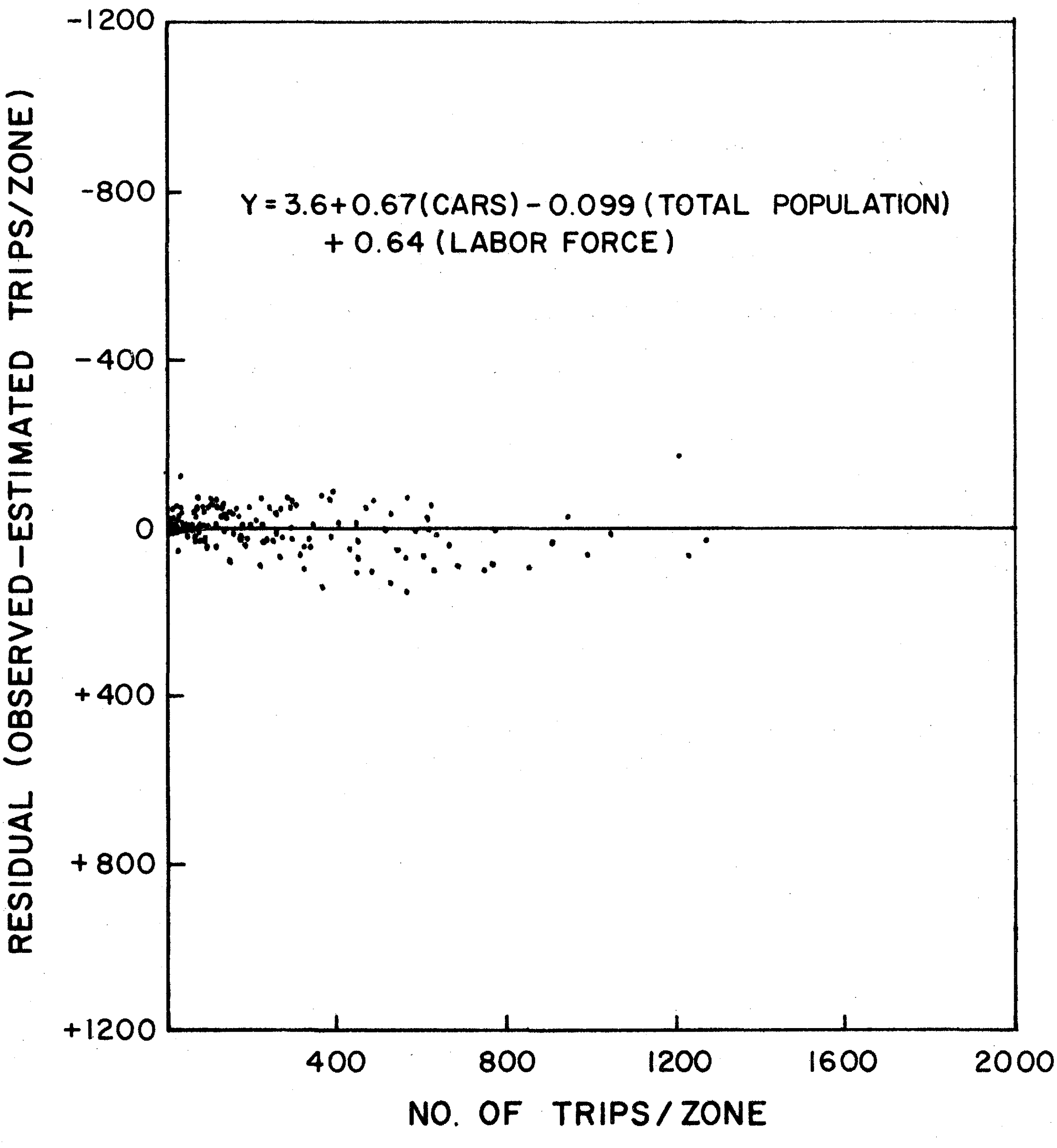
Comparison of Residuals

Referring again to Figure 2, it appears that the regression estimates for Home Based Work Productions are reasonable. From the residuals shown in Figures 5 and 6, it appears that neither the rate nor the regression model gives biased estimates. With respect to the deviations of estimates from observed values, however, the regression estimates are clearly superior.

From Figures 2 and 7 no appreciable bias was noted for the regression estimates of the Home Based Non-Work Productions. Although rate estimates also appeared unbiased, the residuals were somewhat greater, indicating again that estimates of Home Based Non-Work Productions might be improved by using multiple regression rather than rates.

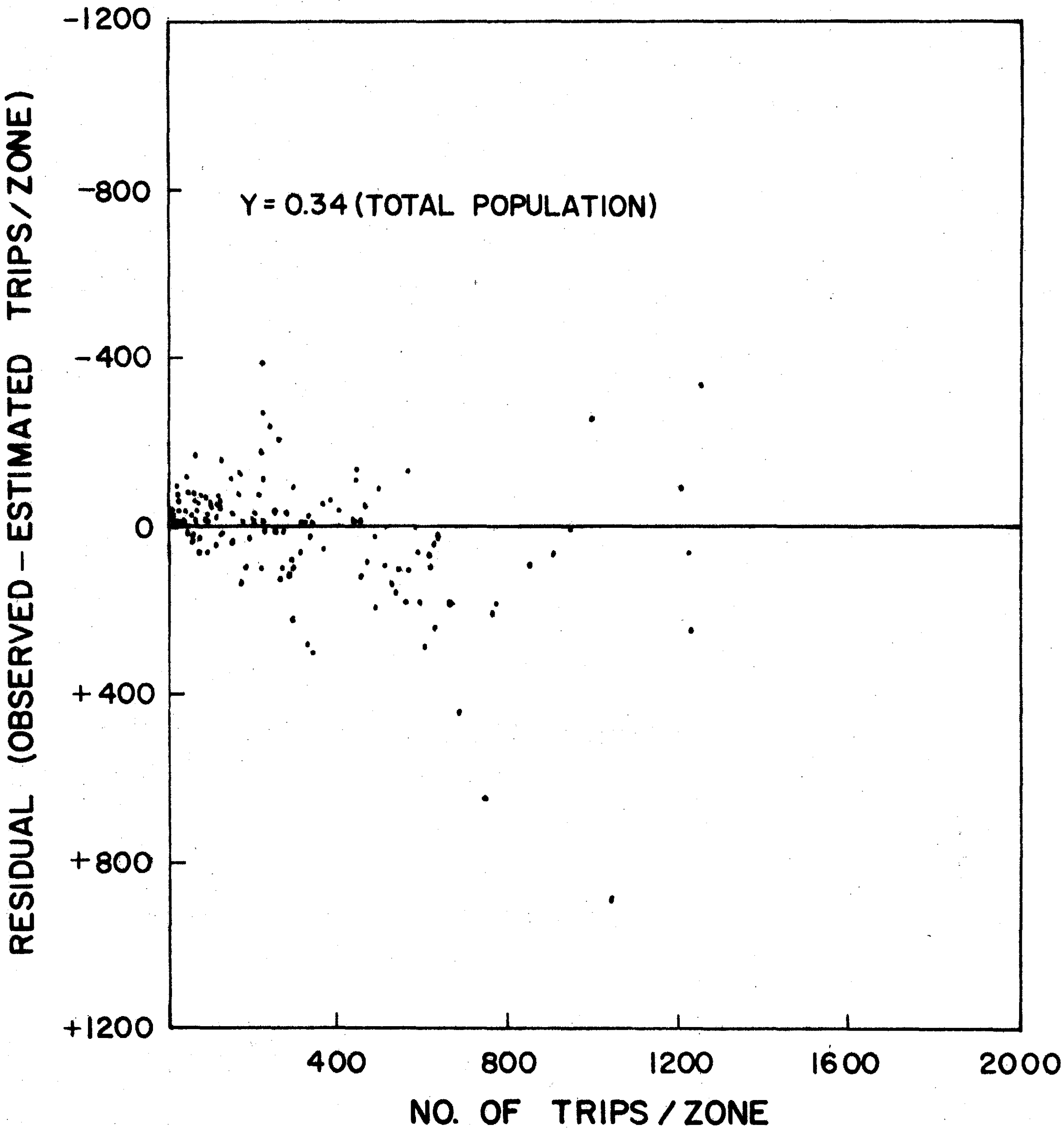
For the Home Based Attractions, the regression estimates tended to underestimate trips for large zones and overestimate for small zones (Figure 3). This trend was more pronounced for Non-Work than Work trips. In comparing residuals for rate and regression estimates for Home Based Attractions (Figures 9-12), it appears that for both Work and Non-Work trips, the rate estimates are less biased and have only a slightly greater variability than the regression estimates. This suggests that rates are more appropriate than the regression models used here for the estimation of Home Based Attractions.

For the Non-Home Based trips, residuals of rate and regression estimates were not compared directly. However, it is to be noted from Figure 4 that the same trend (i.e., overestimation for small zones and underestimation for large zones) occurred for the Non-Home Based trips as for the Home Based Attractions. This suggests that rates might also provide superior estimates over regression for Non-Home Based trips.

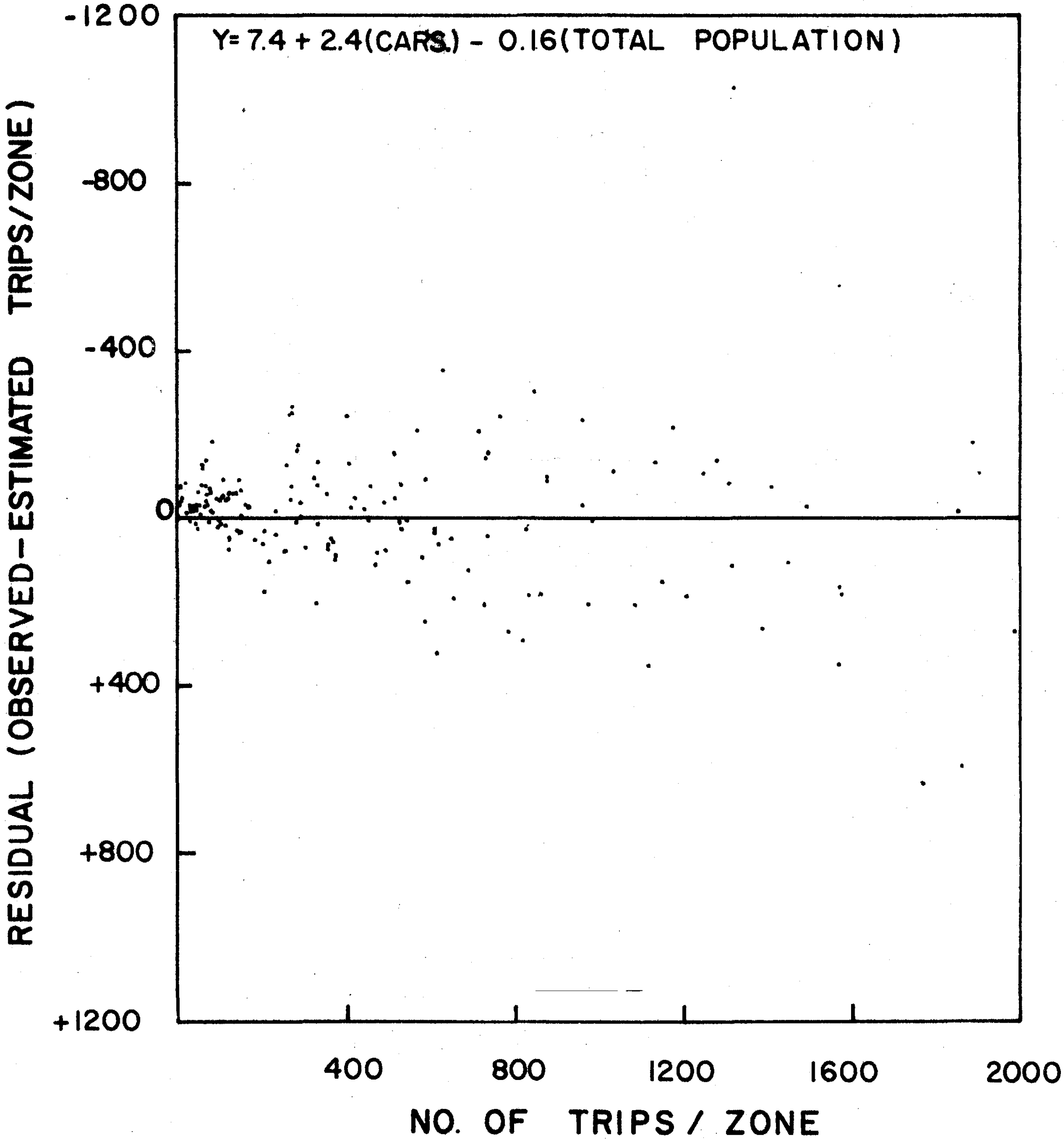


RESIDUALS FOR REGRESSION ESTIMATES
HOME BASED WORK PRODUCTIONS

FIGURE 5

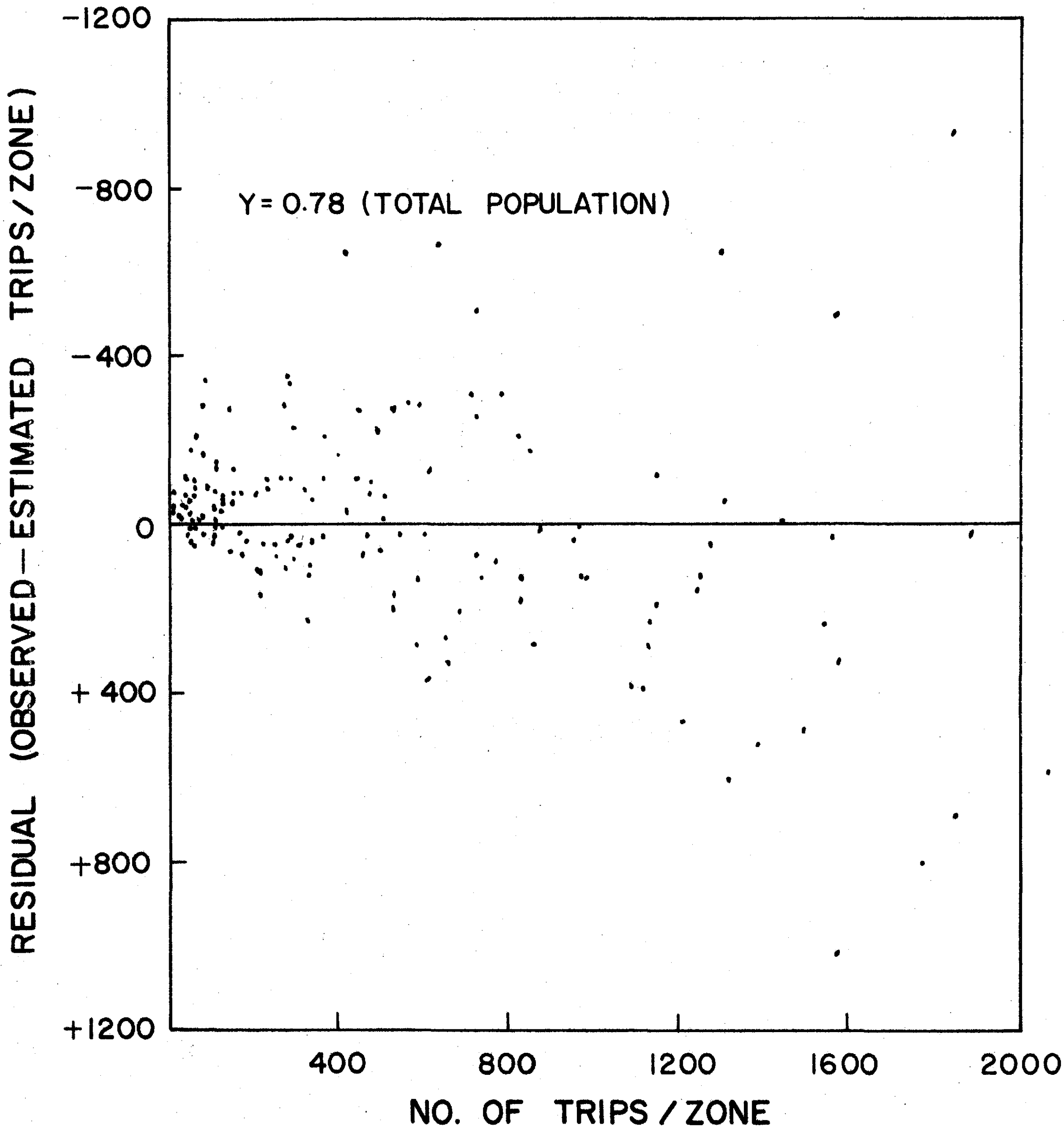


RESIDUALS FOR RATE ESTIMATES
HOME BASED WORK PRODUCTIONS
FIGURE 6



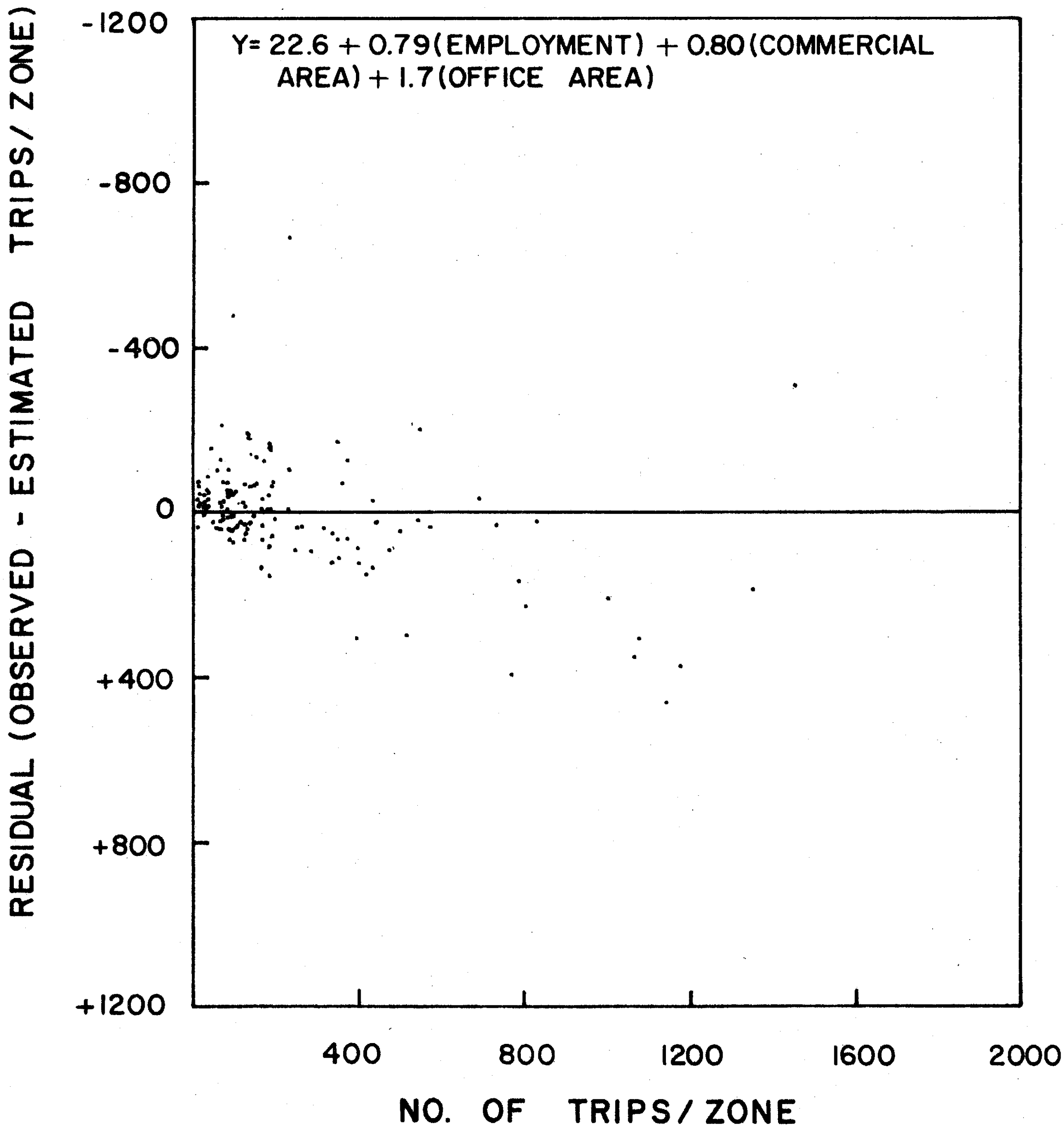
RESIDUALS FOR REGRESSION ESTIMATES
HOME BASED NON WORK PRODUCTIONS

FIGURE 7

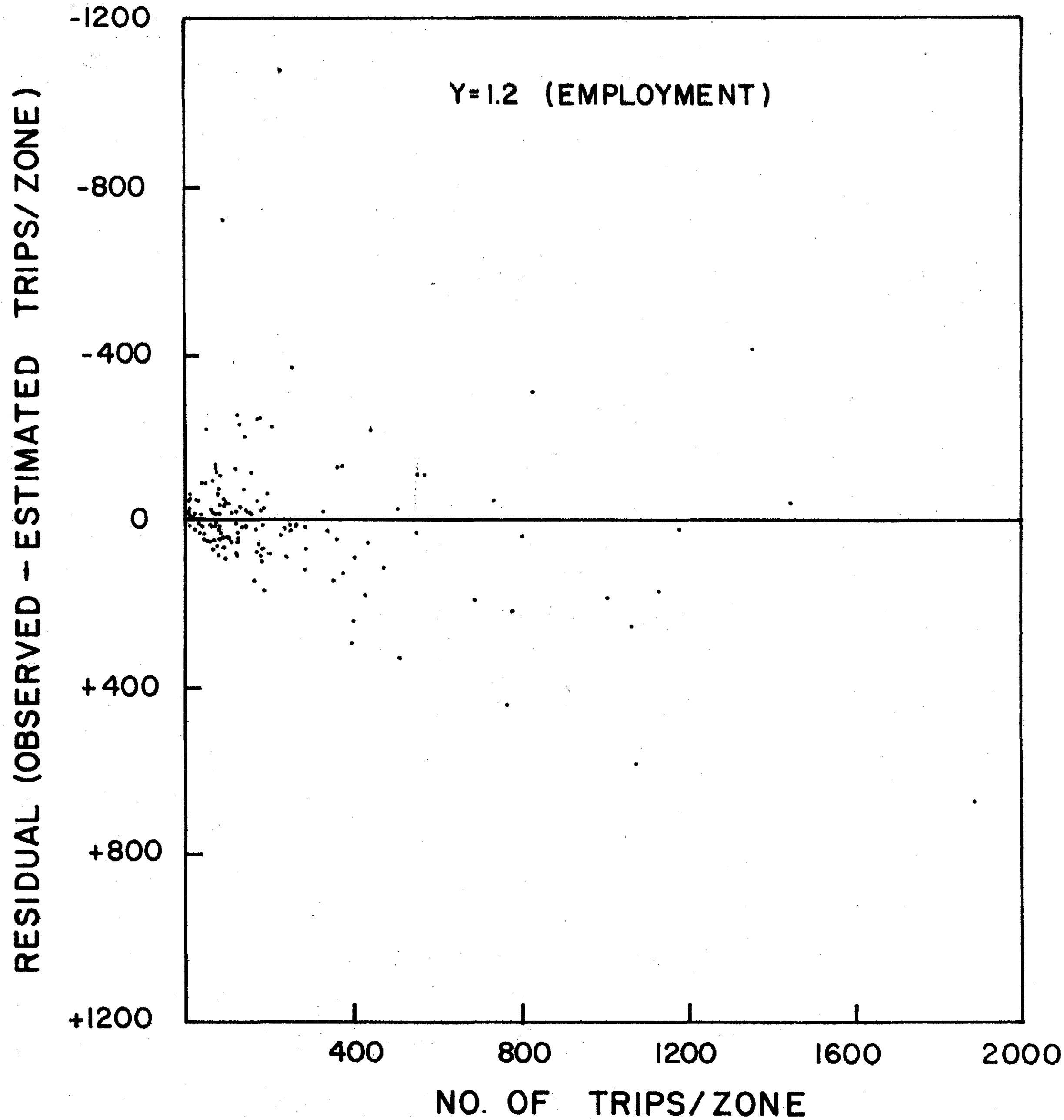


RESIDUALS FOR RATE ESTIMATES
HOME BASED NON-WORK PRODUCTIONS

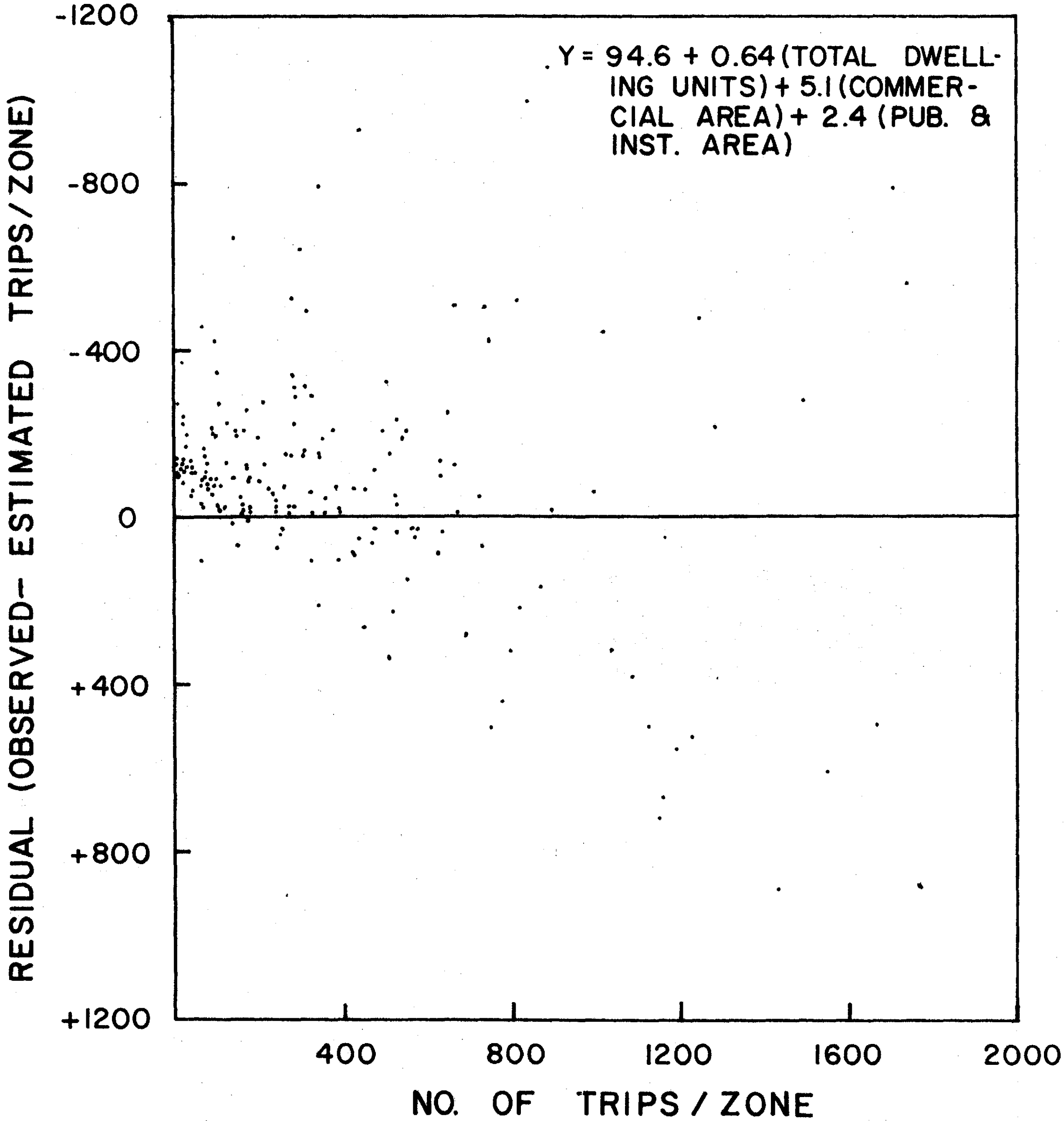
FIGURE 8



RESIDUALS FOR REGRESSION ESTIMATES
HOME BASED WORK ATTRACTIONS
FIGURE 9

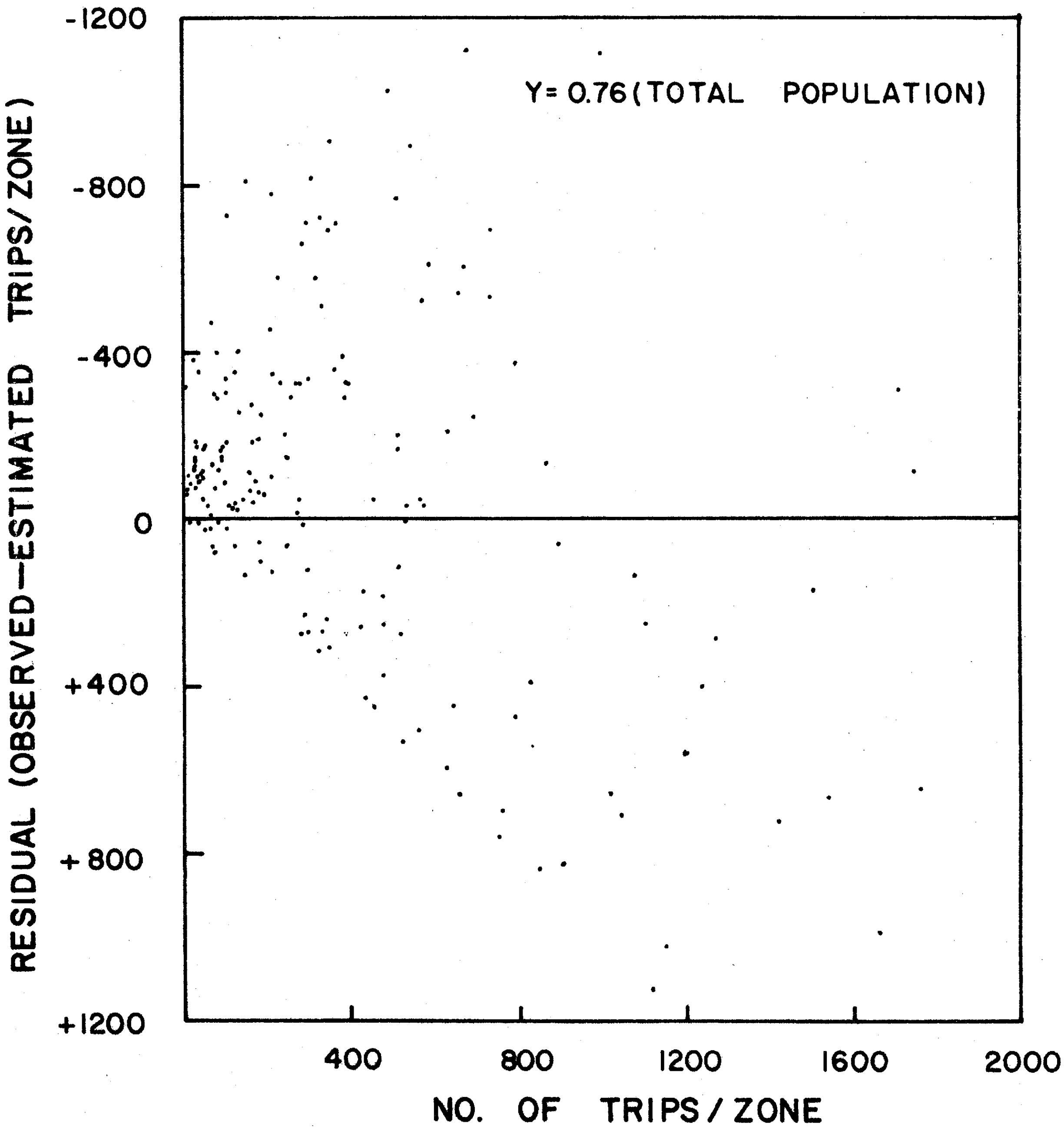


RESIDUALS FOR RATE ESTIMATES
HOME BASED WORK ATTRACTIONS
FIGURE 10



RESIDUALS FOR REGRESSION ESTIMATES HOME BASED NON-WORK ATTRACTIONS

FIGURE 11



RESIDUALS FOR REGRESSION ESTIMATES
HOME BASED NON-WORK ATTRACTIONS

FIGURE 12

TRIP STRATIFICATION

In considering the effects of stratification of trips by purpose and direction, it is recognized that the approach used in the traffic assignment will govern the ultimate stratification of trips. Here, however, consideration was limited to the effects of stratification on the precision of the estimates.

Considering rates, it can be seen that the same estimated number of trips will be obtained regardless of whether the total is estimated or the individuals are estimated separately and combined, as long as the estimates are based on the same independent variable. This is shown below for the "Ratio of Averages" method.

The rates for each of two purposes are given by:

$$b_1 = \frac{\Sigma\ Y_{1i}}{\Sigma\ X_i} \quad ; \qquad b_2 = \frac{\Sigma\ Y_{2i}}{\Sigma X_i}$$

where: b_j = rate for purpose j

Y_{ji} = no. of trips for the jth purpose of the ith zone

X_i = no. of units of the independent variable in the ith zone

Denoting the aggregated rate by B:

$$B = b_1 + b_2 = \frac{\Sigma\ Y_{1i}}{\Sigma\ X_i} + \frac{\Sigma\ Y_{2i}}{\Sigma\ X_i}$$

This reduces to:

$$B = b_1 + b_2 = \frac{\Sigma\ (Y_{1i} + Y_{2i})}{\Sigma\ X_i}$$

For multiple regression estimates the same remarks generally apply, however, it is emphasized that the indifference of the estimate to stratification holds only as long as the same independent variable(s) is used for all trips being estimated.

Stratification by Direction

Table 4 includes a summary of the average number of trips stratified by purpose and direction for each zone. From these values it is apparent that the sum of the rates for the Ins and Outs of each purpose will be equal to the rate of the unstratified trips for that purpose, assuming the same independent variable to be appropriate for all estimates; from Table 1 it is clear that this is the case. For the regression models estimating Home Based Productions and Attractions, it was observed that the same independent variables were generally applicable for the Ins, Outs, and totals and that the numerical values of the coefficients in the Ins and Outs models, when added, very nearly approximated the corresponding coefficients in the models for the totals.

Therefore, as long as Ins, Outs, and total trips of a given purpose are estimated by the same variable or combination of variables, it appears that given either the Ins, Outs, or Totals, one could estimate the other two from:

$$\text{Ins} \cong \text{Outs} \cong 1/2 \text{ (Total)}$$

TABLE 4

ZONAL TRIP AVERAGES

Trip	Average at Production End	Average at Attraction End
Home Based Work	218	183
Home to Work	113	95
Work to Home	104	87
Home Based Non-Work	483	470
Home to Non-Work	241	235
Non-Work to Home	241	235
Non Home Based	289	290

For all Home Based trips considered here, this condition was reasonably satisfied. For Non Home Based trips, the different variables of the multiple regression models indicate that regression estimates for such trips might best be stratified by direction; however, if Non Home Based trips are estimated by rates it is likely that the same independent variable will be appropriate for both directions, eliminating the need for directional stratification.

Stratification by Purpose

For the single-variable rates considered earlier, there was no gain by stratification of Home Based Productions by purpose since the same independent variable provided the best estimates for both Work and Non-Work trips. However, it is quite possible that by stratifying Non-Work Productions and Attractions, the use of different independent variables would provide more efficient rate estimates.

For the multiple regression models, estimates for different trip purposes were based on different combinations of independent variables. For this reason, one would generally expect increased efficiency if trips estimated by multiple regression equations were stratified by purpose.

Using least squares, overstratification can result in unsatisfactory estimates due to instability of statistical relationships which are based on an insufficient number of observations (6). However, because the rate estimate is constrained to coincide with the origin of axes as well as the mean, it is inherently more stable than the conventional regression estimate. This can be illustrated by the example in Figure 13.

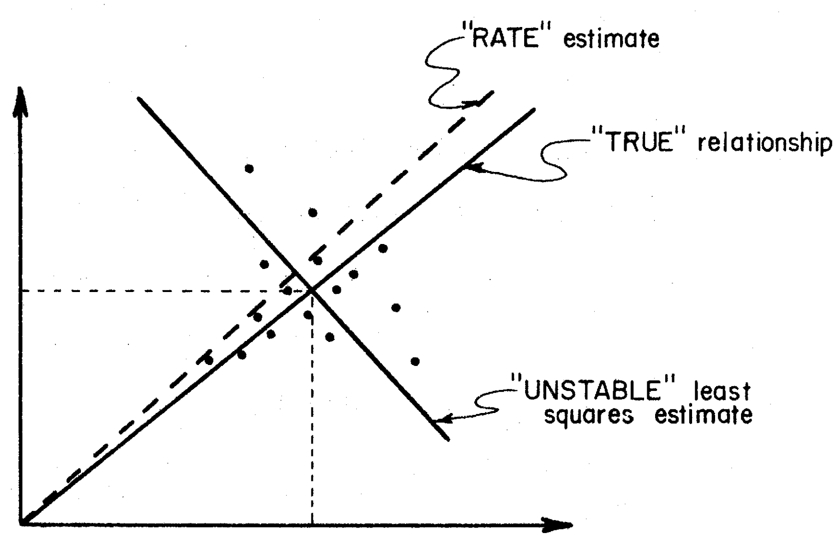


Figure 13

Although this example is admittedly extreme to illustrate this point, it does indicate the greater stability of the rates estimate in cases where the data are too limited to define some relationship (using conventional least squares) which can be assumed to exist. Thus, it appears that in using rates, the precision of the estimates can be improved by stratifying by purpose as long as trips of each purpose are best estimated by different independent variables.

LIST OF REFERENCES

1. Thomas, Edwin N., Frank E. Horton, and John W. Dickey, "Further Comments on the Analysis of Non-Residential Trip Generation," Research Report, Transportation Center, Northwestern University, 1966.

2. Schuldiner, Paul W., Joseph DeSalvo, John Dickey, and Frank Horton, "Non-Residential Trip Generation Analysis," Research Report, Transportation Center, Northwestern University, 1966.

3. Schuldiner, Paul W., and Walter Y. Oi, An Analysis of Urban Travel Demands, Northwestern University Press, 1962.

4. Snedecor, George W., Statistical Methods, The Iowa State College Press, 1956.

5. Draper, Norman and Harry Smith, Applied Regression Analysis, John Wiley and Sons, Inc., 1966.

6. Guidelines for Trip Generation Analysis, U.S. Department of Transportation, Federal Highway Administration, Bureau of Public Roads, June 1965.

APPENDIX A

DEFINITION OF VARIABLES

APPENDIX A

DEFINITION OF VARIABLES

Independent Variables

Number	Designation	Definition[1]
1	1F DU	Number of single family dwelling units
2	2F DU	Number of two family dwelling units
3	3MF DU	Number of three or more family dwelling units
4	CARS	Number of cars
5	TOT POP	Total number of persons
6	POP 5+	Number of persons 5 years old and older
7	CIV LF	Number of persons in civilian labor force (home end)
8	P15-79	Number of persons of driving age (15-79 yrs.)
9	AD CBD	Airline distance to CBD (miles)
10	MP CBD	Minimum path from CBD centroid to zone centroid (minutes)
11	EMPLMT	Number employed (work end)
12	TOT DU	Total number of dwelling units
13	COMMER	Area in commercial use (100 sq. ft.)
14	REC SV	Area in commercial, recreation, and service use (100 sq. ft.)

[1]Values for these variables refer to the number of indicated units for each zone.

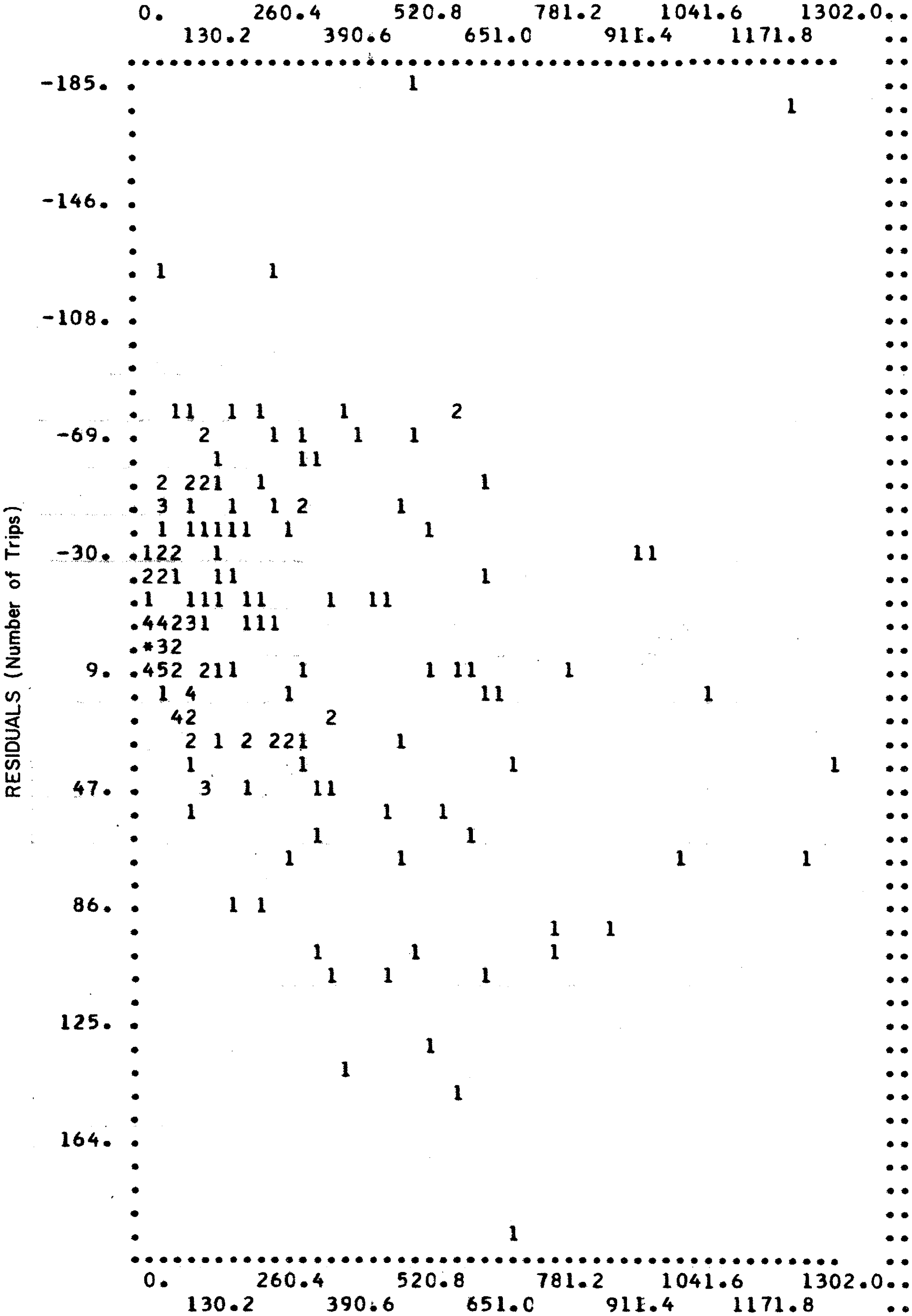
APPENDIX A

DEFINITION OF VARIABLES
(cont'd)

Number	Designation	Definition
15	OFFICE	Area in office use (100 sq. ft.)
16	INDUST	Area in industrial use (100 sq. ft.)
17	TCUTIL	Area in transportation, communication, and institutional use (100 sq. ft.)
18	WHSALE	Area in wholesale and warehouse use (100 sq. ft.)
19	PUB-IN	Area in public and institutional use (100 sq. ft.)
20	AGOPEN	Area in agriculture and open land (100 sq. ft.)
21	NTRESD	Net residential area (tenths of an acre)
23	LABOR	Number of persons in total labor force (home end)

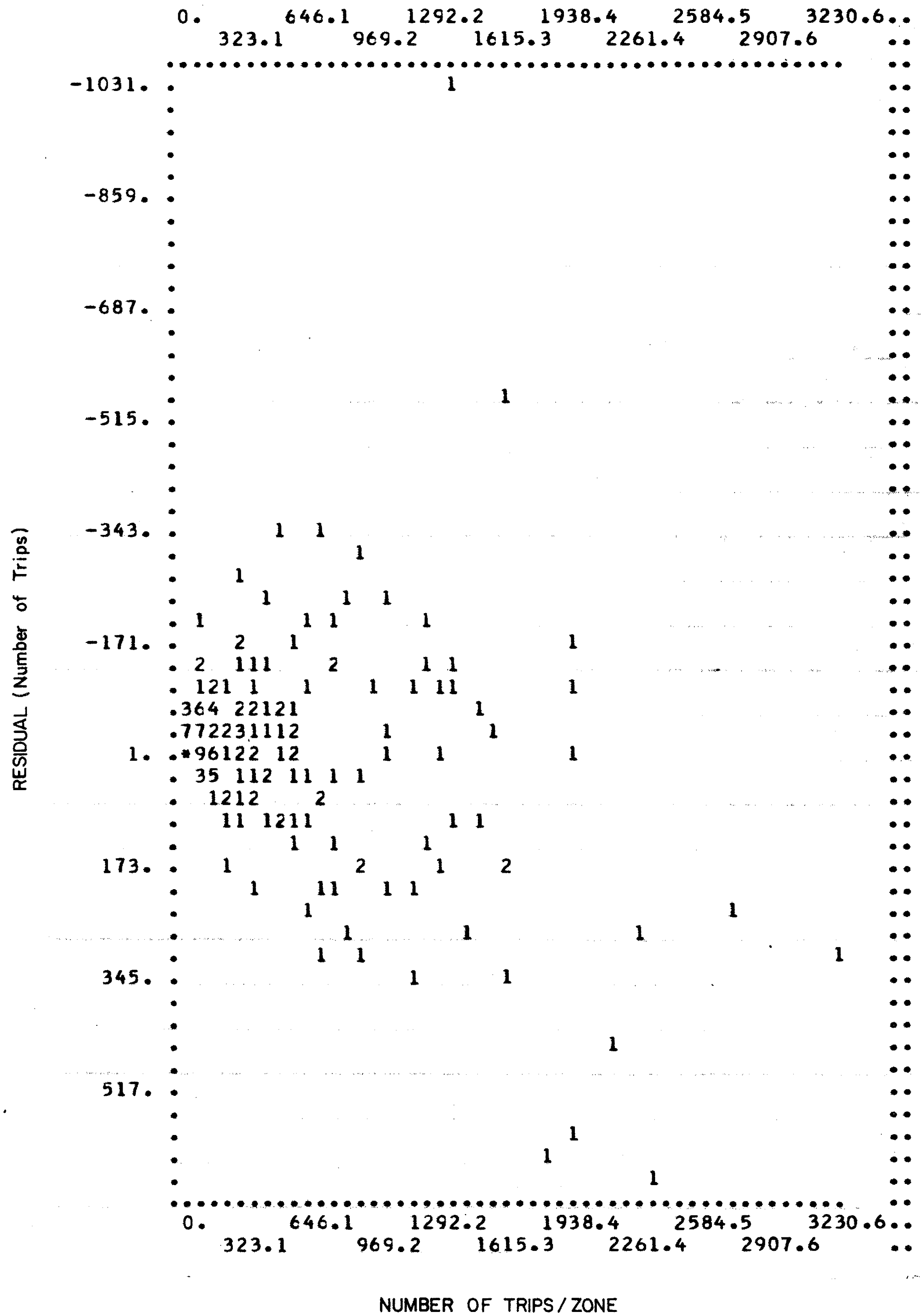
APPENDIX B

PLOTS OF RESIDUALS VS INDEPENDENT VARIABLES FOR REGRESSION MODELS

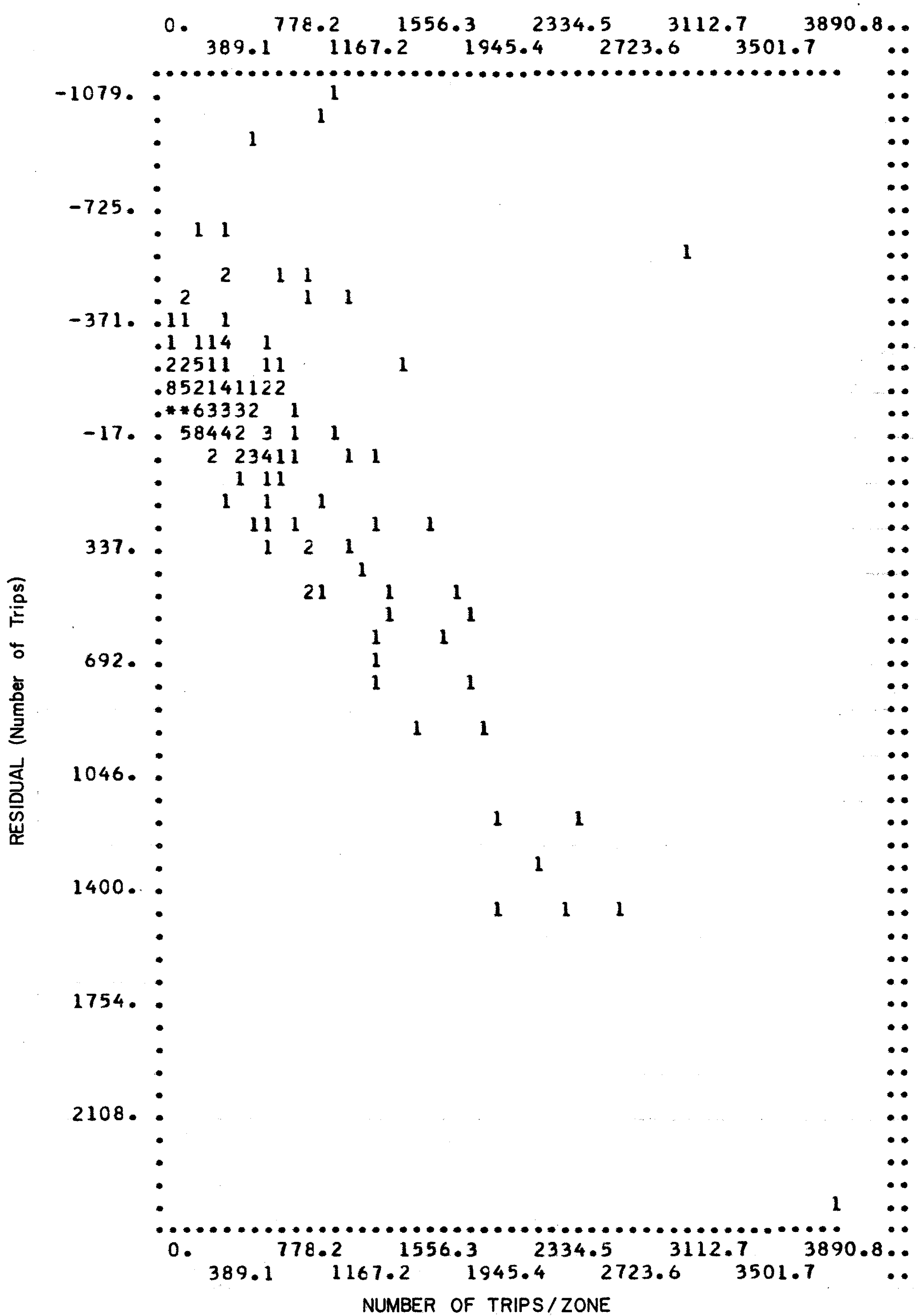


NUMBER OF TRIPS / ZONE

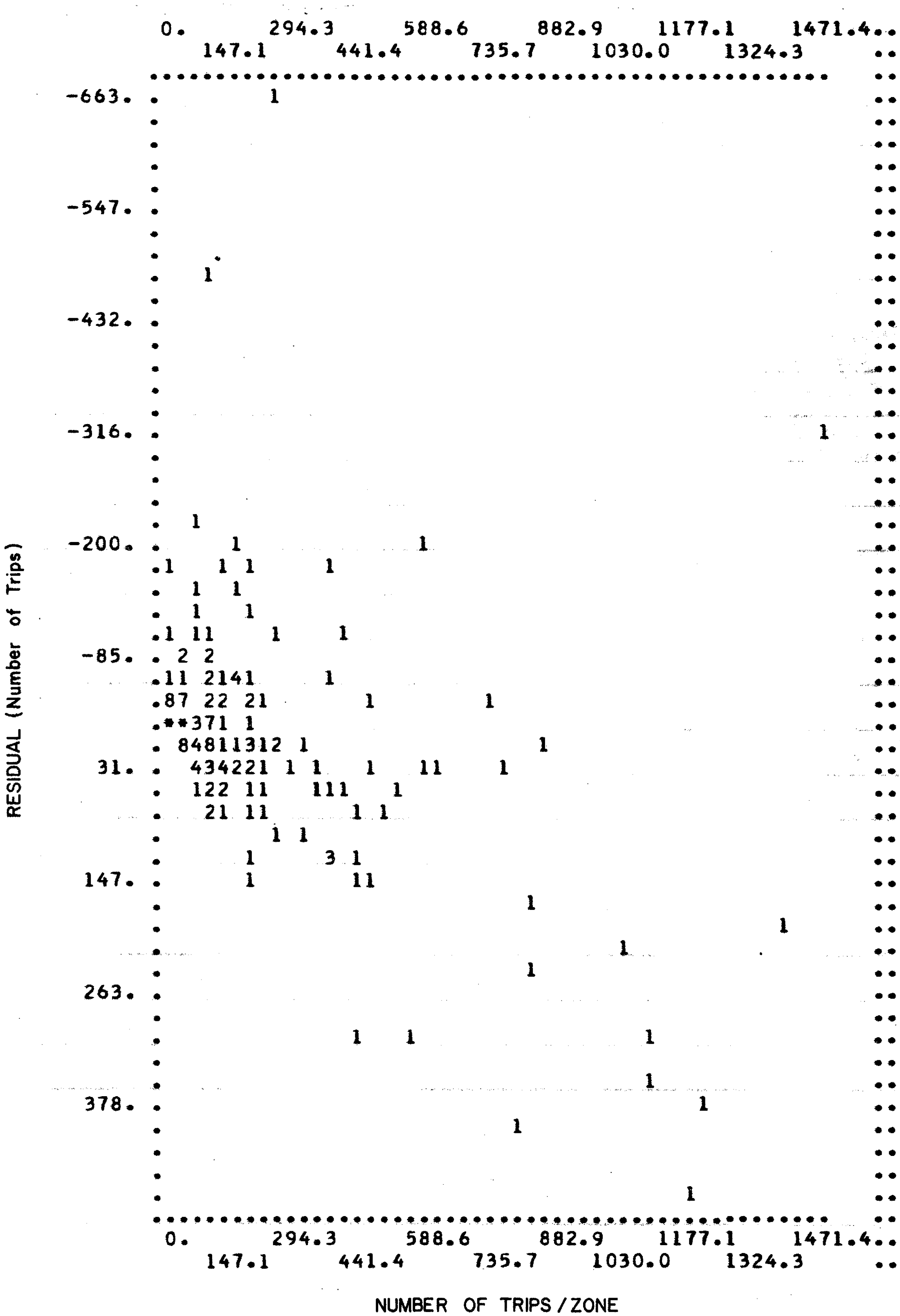
HOME BASED WORK PRODUCTIONS



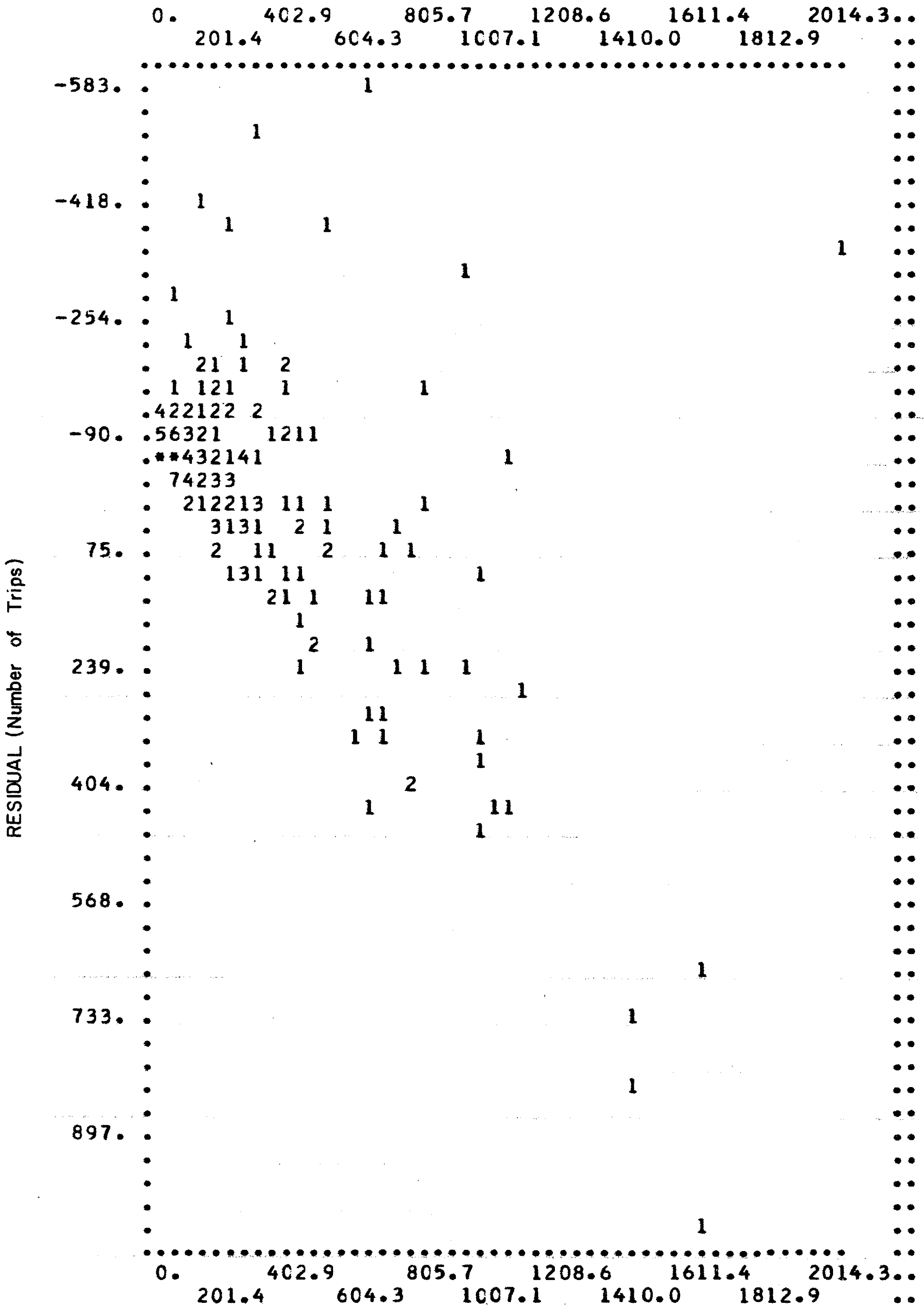
HOME BASED NON WORK PRODUCTIONS



HOME BASED NON WORK ATTRACTIONS

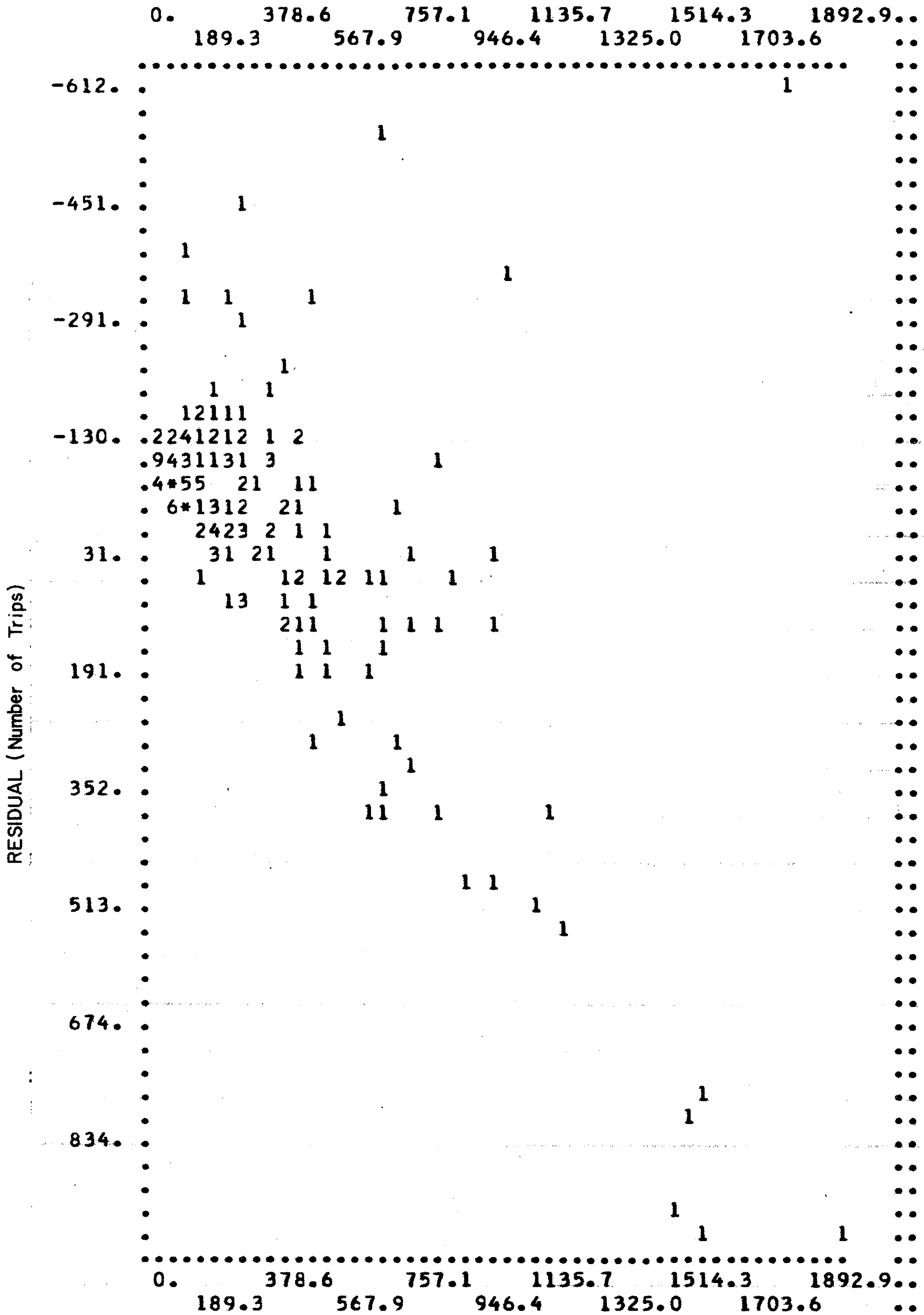


HOME BASED WORK ATTRACTIONS



NUMBER OF TRIPS / ZONE

NON HOME BASED PRODUCTIONS



NUMBER OF TRIPS / ZONE

NON HOME BASED ATTRACTIONS