AN ANALYSIS OF THREE METHODS OF TRIP GENERATION

Thesis for the Degree of M. U. P.
MICHIGAN STATE UNIVERSITY
JOHN RICHARD STONE
1976

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 01008 8155
LIBRARY
Michigan State
University
BINDING BY
HOAG & SONS'
BOOK BINDERY INC.
LIBRARY BINDERS

NOV 0 6 2000

G19197

ABSTRACT

AN ANALYSIS OF THREE METHODS OF
TRIP GENERATION

By

John Richard Stone

Comprehensive transportation planning is an established component of the total planning effort in most areas today. Indeed, Federal law requires that all of the major urbanized regions in the United States have a transportation plan formulated on the basis of an extensive study of the transportation components and the system flows indigenous to that region. The prescribed plan must be integrated with all the other planning being undertaken by the region.

A major transportation study includes: collecting a wealth of information about the system and verifying the validity of that data; using the data to simulate the real world system by employing a set of explanatory variables; calculating future year magnitudes of these variables and introducing them to the model; analyzing the resultant output of future needs in terms of present supply of facilities; and then, on the basis of the foregoing, formulate a transportation plan for implementation. Typically, portions of the process are repeated every three to five years. In the continuing phase the social, economic and physical systems upon which the predictions were predicted are constantly monitored and necessary updates to the plan are effectuated.

John Richard Stone

One of the most crucial steps in the transportation planning process is the generation of traffic for the future or target year. Historically, the trip generation procedure was nothing more than extrapolation of past traffic growth curves. Time has engendered more sophisticated procedures. Today, trip-making is widely recognized as a function of:

1. the transportation facilities which are available in terms of the type of facility, the accessibility, and the efficiency;
2. the type of land use around these facilities including the intensity; and
3. the demographic characteristics of the population.

Several methodologies have been developed and applied in an effort to have the above parameters simulate trip-making behaviors and produce estimates of traffic generation. As with most procedures in embryonic stages, there is lacking unanimity in methodology and even basic procedural design. Methodological disagreements exist as to whether data should be used directly from households or whether the data should be aggregated to traffic zone totals prior to being used in trip generation. Other questions exist as to whether multiple regression analysis provides any greater degree of accuracy in formulating a model than does a simple equation based on rates of average trips per some variable such as people or dwelling units. This thesis addresses itself to precisely the aforementioned questions.

John Richard Stone

The data used in this thesis is taken from a major transportation planning study being conducted by the Michigan Department of State Highways and Transportation. Trip generation equations by three different methods for each of three trip types are developed and the equations are compared in an effort to determine which method does the better job of estimating trip productions. The three trip types are Home-Based Work, Home-Based Shopping, and Home-Based Other. The first type of equation is formulated by a multiple linear regression method on zonal totals of socioeconomic variables. The second type uses an average-rate method such as the number of trips per zone per auto. The third type also uses the multiple regression analysis method but does so on disaggregated data. That is, data derived directly from households with no areal qualifications or aggregation having been performed.

The three sets of equations are compared statistically, graphically, and subjectively for the magnitude of variables, bias in the prediction range, and the validity of the equations per se.

In general, the conclusion is that zonal level multiple regression analysis is the better of the three methods. Not only is regression analysis more difficult on the dwelling unit level, but it is slightly inferior to each of the other methods. An additional conclusion is that there may not be significant differences between the three methods and that all three have potential application.

AN ANALYSIS OF THREE METHODS OF
TRIP GENERATION

By

John Richard Stone

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER IN URBAN PLANNING

School of Urban Planning and Landscape Architecture

1976

ACKNOWLEDGMENTS

Appreciation is extended to many of my work associates in the Bureau of Transportation Planning, Michigan Department of State Highways and Transportation for their motivation and experienced guidance, without which this study would not have been conjured or brought to fruition.

Special thanks go to my supervisor, Ken Underwood, for his tolerance and assistance throughout the many months this study was being compiled. Also, to my wife for her diligence and patience in typing rough drafts. A deep feeling of gratitude goes to Wayne Meyerowitz, our staff statistician, who provided valuable assistance in the design of this research and analysis of the results. And finally, I must thank my Thesis Committee for lending their time to this venture: Professors Donn L. Anderson and Donald W. Bradley; and the following personnel from Michigan Department of State Highways and Transportation: Robert S. Boatman, Division Administrator; Frank DeRose, Manager; Maynard A. Christensen, Manager; and Louis H. Lambert, Transportation Planning Analyst.

TABLE OF CONTENTS

		Page
LIST OF TABLES		v
LIST OF FIGURES		vi
CHAPTER		
I.	INTRODUCTION	1
	Transportation Planning: An Evolutionary View	2
	Transportation Problems	3
	Comprehensive Planning of Urban Transportation Systems	6
	Trip Generation Philosophy and Methodology	15
	Trips and Trip-Making in Transportation Planning	16
	Evolution of Trip-End Modeling	19
	Purpose and Scope of the Research	30
II.	STUDY PROCEDURES	32
	Description of Estimating Methods of Investigation	32
	Zonal Multiple Regression Method	33
	Average-Rate Method	37
	Dwelling-Unit Multiple Regression Method	38
	Description of the Study Area	39
	Preparation of the Data	41
	Analysis Procedures	45
	Analysis Criteria	49
	Statistical Analysis	49
	Graphical Analysis	53
	Subjective Analysis	55
III.	STUDY RESULTS	59
	Three Zonal Estimating Equations	60
	Three Average-Rate Estimating Equations	66
	Three Dwelling-Unit Estimating Equations	70
	Comparison of Equations by Analysis Method	74
IV.	STUDY CONCLUSIONS	93

APPENDIX Page

A. GLOSSARY OF TERMS 97

B. STUDY AREA MAP 99

C. ZONE BOUNDARY MAP 100

BIBLIOGRAPHY 101

LIST OF TABLES

TABLE		Page
1.	Variables in Zonal Data File	42
2.	Variables in Household Data File	43
3.	Vehicle Trips by Trip Purpose	44
4.	Home-Based Work Production Equation at the Zonal Level	62
5.	Home-Based Shopping Production Equation at the Zonal Level	64
6.	Home-Based Other Production Equation at the Zonal Level	65
7.	Home-Based Work Production Equation Using an Average-Rate	67
8.	Home-Based Shopping Production Equation Using an Average-Rate	68
9.	Home-Based Other Production Equation Using an Average-Rate	69
10.	Home-Based Work Production Equation at the Dwelling-Unit Level	71
11.	Home-Based Shopping Production Equation at the Dwelling-Unit Level	72
12.	Home-Based Other Production Equation at the Dwelling-Unit Level	73
13.	Comparison of the Estimating Methods for Home-Based Work Productions	76
14.	Comparison of the Estimating Equations for Home-Based Shopping Productions	77
15.	Comparison of the Estimating Equations for Home-Based Other Productions	78

LIST OF FIGURES

FIGURE		Page
1.	The Transportation Planning Process	9
2.	Trip-Making Variables	20
3.	Least-Squares Regression Line	34
4.	Flow-Chart of Analysis Procedures	46
5.	Linearity of the Data for Home-Based Work and Residential Labor	54
6.	Residuals Versus An Independent Variable (X)	56
7.	Residuals Versus the Observed Dependent Variable (Y) . . .	57
8.	Zonal Residuals Versus Home-Based Work Productions	83
9.	Rate Residuals Versus Home-Based Work Productions	84
10.	Dwelling-Unit Residuals Versus Home-Based Work Productions	85
11.	Zonal Residuals Versus Home-Based Shopping Productions . .	86
12.	Rate Residuals Versus Home-Based Shopping Productions . .	87
13.	Dwelling-Unit Residuals Versus Home-Based Shopping Productions	88
14.	Zonal Residuals Versus Home-Based Other Productions . . .	89
15.	Rate Residuals Versus Home-Based Other Productions	90
16.	Dwelling-Unit Residuals Versus Home-Based Other Productions	91

CHAPTER I

INTRODUCTION

Transportation is the act of moving physical objects from one place to another. Those objects may be people or goods. The relocation process may involve vehicles such as a railroad boxcar full of washing machines or a bus carrying passengers. The vehicles may be motorized or nonmotorized such as bicycles or horses. Transportation may also take place without vehicles such as pedestrian travel. The inherent tangibleness of the transported objects distinguishes transportation from communication. The mode of transportation ranges from interplanetary to pedestrian movements within an office building or dwelling-unit. As an institution, there is perhaps none more pervasive to our society. Some people even proclaim that mobility is one of our basic freedoms. This thesis addresses itself primarily to the urban highway mode of transportation though the statements and findings may be applied to other modes and other levels as well.

Transportation planning is the study of, and possible affectation of, a transportation system. This specialized branch of planning has its own body of knowledge, tools, and techniques, but is not a totally independent branch as will be illustrated by this chapter. The term transportation planning has become synonymous with a set of procedures used to do highway-oriented urban planning. Although this popular usage is far too narrow, it is adequate for this thesis and this chapter will expand upon the transportation planning process.

Trip generation is one component of the transportation planning process. Travel on a transportation system is quantified by trips and the description of the number of trips starting or ending in a particular area is a key element to an understanding of the system. Other components of the transportation planning process are intimately related to trip generation such as distributing the trips after they are generated, assigning a route, and deciding upon a mode of travel. This chapter discusses trip generation in more detail and concludes with a study design appropriate for furthering the understanding of the concept.

Chapter II details the study procedures including the data and the methods employed. Chapter III presents the results of this trip generation study, and Chapter IV offers conclusions for and implications of the findings.

Transportation Planning: An Evolutionary View

Transportation planning is a discipline marked by rapid development in theory, procedure, and application which is nothing more than a reflection of the rapidly evolving domain encompassed by the discipline. Technologically, transportation has progressed from the invention of the wheel to space ship travel. Beginning with footpaths, transportation systems now include planetary orbits and trajectories. Planning is as simple as laying out the path of least resistance between two villages to the complex mathematical modeling process employed today in urban area simulations. Just as technology has provided us with improved systems, it has also provided improved means of studying and designing those systems. There is no reason to assume the evolutionary nature of transportation planning will cease. This thesis per se represents an attempt to further the evolution of the process by proposing more exacting standards to trip generation techniques. Almost all attempts to improve a given situation start with the recognition and statement of a problem.

Transportation Problems

A problem is the difference between the desired and possible state of a given situation and the actual state at a point in time for an individual or a group of individuals. Solutions to problems are perhaps unattainable, unattainable within constraints of time or finances, or if attainable create another problem. Problems have hierarchical orders on the basis of the severity, people affected, breadth, and direct or indirect relationship to transportation service.

In 1956 Congress passed the National Interstate and Defense Highway Systems Act. This bill has provided most of the limited access highways in use today. In less than 20 years these roads have virtually changed the social, economic, and physical structure of neighborhoods, communities, and even whole cities. Many of the manifestations of these roads were unplanned and unimagined. Large, regional shopping centers have been built at major interchanges. Towns have grown to cities and new towns have emerged at the confluence of interstate routes. Cities have altered their urbanscape for new orientations and the social compositions and physical attributes of places have been radically redesigned. Changing travel patterns have had marked effects on personal and family lives. Consumer markets are more widespread precipitating large-scale business activities for even remote areas. New kinds of markets have occurred; for example, the weekend (second) home. People commuting 30 to 50 miles to work is fairly common. These examples could be increased many-fold. Each of these manifestations give rise to new problems and queries. For example, when a man works in one city, lives in a second city, and perhaps shops in a third city, where do his allegiances lie; to what political entity? More recently, a serious introspection has confronted the American society as to whether our commitment to the private use of the automobile is a conscious and directed occurrence and whether, in any event, we ought to reconfirm these convictions or allow for public funding of new modes.

Because of the foregoing, our transportation problems are presently manifold. They first received national attention on April 15, 1962, when President John F. Kennedy delivered to Congress his Special Message on Transportation. In addition to being the first time in the history of the United States that a President had delivered his own message dealing exclusively with problems of transportation to Congress, this action emphasized the urgency of squarely meeting the serious problems of today's transportation system.

One of the most important of these problems, and certainly the most widely appreciated, is urban congestion and the high cost of meeting urban transportation demands. Estimates by various studies have derived the extent of land in downtown areas of our major cities devoted to rubber tired vehicle transport at 66 to 80 percent. The ecologists have taken a nearly militant stance in an effort to warn the public of the impending threat to our urban ecosystems by these high concentrations of vehicles. Exhaust emissions pollute the air and the fact that the urban area is completely "paved over" contributes to the pollution of our water bodies by rain water run-off replacing absorption. In some cities as much as 5/6 of additional public capital investments is for streets, roads, and highways. The other 1/6 goes for water supply, flood control, sewerage systems, schools, hospitals, parks, and recreational facilities. The socially conscious people argue that the automobile has denigrated the responsibility of man. Nowhere is this more obvious than in the competition between human and machine for space. People wait for vehicles before entering a common space. Children playing football in an empty parking lot would yield to the motorist who wished to park. Despite the tremendous outlays of space and money for transportation facilities, travel time between work and home, in many instances, amounts to two hours per day. In Los Angeles the transportation problem has even received some of the blame for the riots which occurred in the overcrowded, underemployed Watts district in 1965. New transportation facilities are usually built to have at least a 20-year life before they reach full capacity.

And yet, in many cases, by the time the facility is planned, programmed, and constructed, it is outdated.

These and other problems suggest that urbanized areas are experiencing and will continue to experience severe transportation problems. A transportation planning process has evolved which seeks solutions to some of these problems. As will be documented in the next section, this process has had a changing focus and at no time has it attempted to address all transportation problems.

Comprehensive Planning of Urban Transportation Systems

Transportation planning prior to the early 1950s concentrated on the costs and benefits to the system user. Straight-line projections were made of traffic counts and these forecasted volumes were compared with existing capacities. Surveys of traffic and parking were performed but modeling endeavors were nonexistent. The evaluation of alternatives was almost entirely in economic terms.

The 1950s was a period where urban transportation studies were done on a large scale. Millions of dollars were spent for extensive home interviews to determine travel patterns and demographic characteristics. Land use data and transportation system parameters were also collected. That such massive amounts of data could be processed owes to the development of the digital computer. The emphasis during this period was on the technical process and the results provided fixed plans for investment over the succeeding 20 to 25 years.

The winds of change blew strongly during the 1960s. First, the Highway Act of 1962 required that every metropolitan area having more than 50,000 persons have a comprehensive, coordinated, and continuous transportation plan implemented by 1965 or no Federal funds, especially those provided by Congress in 1956 for the Interstate system, would be forthcoming. Second, the deplorable state of mass transit was recognized. Several pieces of legislation appropriated over $10 billion for loans, grants, subsidies, and studies. There was a renewal of interest in transit technologies and an increased recognition that significant numbers of people were not being served by the auto-highway mode of transportation. Third, the National Environmental Policy Act of 1969 required environmental impact statements for all federally sponsored projects. This Act expanded the factors of concern from economic and engineering to impacts from social, aesthetic, and air, water, and noise pollution. Fourth, metropolitan areas or regions, as opposed to local communities, were given review and consent power for federal-aid projects.

The 1970s continue to identify new problems and concerns for the transportation planner. The bankruptcy of the Penn Central and subsequent government support necessitates an incorporation of urban goods movement into the planning process. The continuing clamor by the American public for participatory democracy brings decision-making in all areas out from behind bureaucratic doors and into the public arena. The requirements for public information and the determination of exactly who are the decision-makers have the transportation planners scurrying for further techniques.

The complexities of the transportation problems and the high costs of transportation facilities make detailed planning a necessity. The transportation planning process must provide quantitative information on the travel demands created by various combinations of land uses and transportation facilities. The transportation planning process was not designed to cure all the ills of urban America. It probably will not even provide solutions to all of our urban problems which are transportation related. The transportation planning process will, however, yield a systematic analysis of the transportation system and all its components.

Figure 1 illustrates a typical transportation planning process. There are four major divisions to a complete study. Not all studies proceed in exactly the same manner. Certain steps may be deleted, appended, or placed in a different position. The remainder of this section discusses briefly a typical major transportation study.

The first of the four major divisions is involved with collecting the basic data to be employed throughout the remainder of the study. This task is both arduous and expensive. Many of the inventories are collected by sampling, so the question of validity becomes extremely pertinent. The whole study and the resultant plan is based upon these base data. Thus, the output of the study is only as good as the input.

The economic activity of the study area is measured by the numbers of people employed at certain locations. Net sales in dollars are possibly collected for commercial areas. Socioeconomic measurements

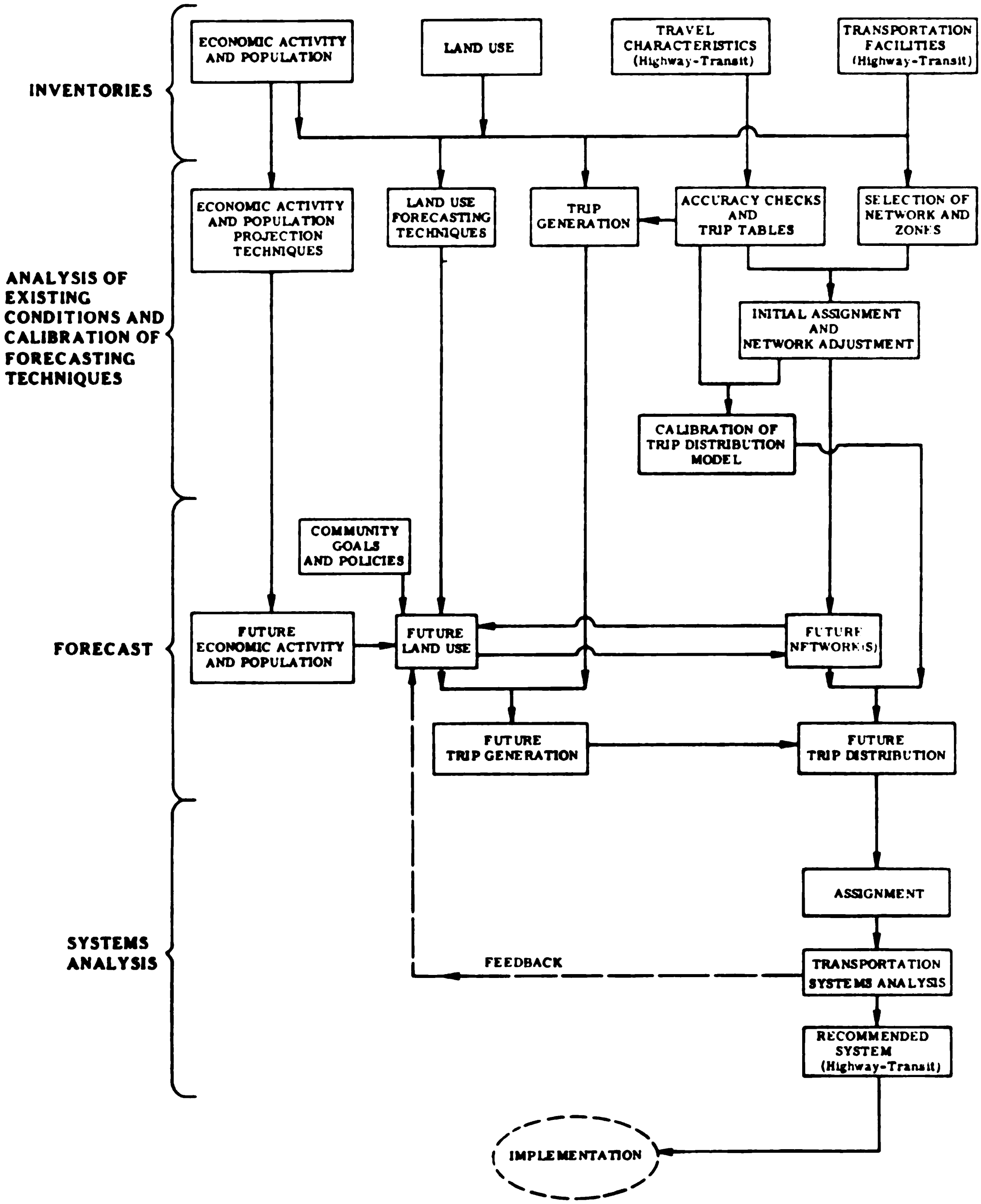


Figure 1. The Transportation Planning Process.

Source: Trip Generation Analysis, U.S. Department of Transportation, 1967.

are made at the dwelling unit level. Data recorded includes the number of residents, their ages, places of employment, household income, value of their home, length of residency, number of cars, and the number of people employed. This is by no means an exhaustive list. Again, the data collected vary with each study. The specific land use of all parcels in the entire study area is recorded.

The travel characteristics inventory records the trips made by the study area residents for a given 24-hour period. Additionally, specific characteristics of those trips are garnered. For example, the mode of travel relates whether the trip was made as an automobile driver, an automobile passenger, a truck driver, a truck passenger, a transit rider, a school bus passenger, or by walking to work. Determination is made as to the number of people in the vehicle, the purpose of the trip, the origin of the trip, the destination of the trip, the times of starting and ending the trip and where the vehicle is garaged. The concept of a trip is of paramount concern for a full understanding of a transportation study and this thesis. A discussion in the next section provides a common reference to this concept.

The fourth inventory is composed of information requisite in describing the transportation network. This network is broken into a series of links with the ends of each link identified by a number called a node number. Data recorded for each link include distance, average effective vehicle speeds (impedances), capacity, traffic count, the type of facility, a general geographic area in which the link is found and whether it is a one-way or two-way link. The foregoing is primarily for highway networks. If the study area has transit facilities, additional data is recorded; for example, headway times, transfer points, and fare structure.

The second major division in the transportation planning process involves analyzing the collected base data and the formulation of models simulating the real world system for the base year. First, techniques for ascertaining the level of economic activity for the target year and for estimating absolute numbers of people and their demographic variabilities must be specified. These techniques may consist of the utilization of projected figures from some organization which specializes in this function. Alternatively, tabulation techniques may be developed in-house. Or, a computer model may be obtained which will provide the figures. The same alternatives exist for land use forecasting techniques.

The transportation network does not utilize all the roadway facilities in the study area. Local streets are not directly represented on this network. Therefore, some selection process in network definition must be made. Traffic zones are delineated in conjunction with network definition. A traffic zone is an areal unit of homogeneous land use. It is the basic unit with which the study is conducted. Origins, destinations, productions, attractions and magnitudes of other variables are all in terms of zones. Generally, network links will not traverse a zone but rather bound the zone. Zone size will vary proportionately with land development density.

The inventories which yield the travel characteristics must somehow be verified. That is, the information is collected on a percentage sample of the total population (universe). If the sample is 20 percent, then all the collected data items are multiplied by five to produce the universe. Selected attributes of the interview data can then be compared to the corresponding figures from an unrelated source; for example, the United States Bureau of Census. The comparisons will verify the validity of the figures per se and identify any possible geographic biases in the sample selection process. An example of this procedure would be a zonal comparison of dwelling units. When sufficient accuracy checks have been completed trip tables are generated. A trip table is a square matrix of origins versus destinations (or, sometimes, productions versus attractions). This matrix provides a tabulation of the travel interchanges between zones. No routes are specified. The interchange values may represent person trips or vehicle trips. The tables may stratify the trips by mode of travel or purpose for travel.

The zonal interchange values from the vehicle trip tables are then assigned to the network in an effort to replicate the known quantities of traffic on the links. Assignment to a network is accomplished by a computer model which calculates the minimum path between two zones (an origin zone and a destination zone) on the basis of time. The most common assignment model in use today assigns all trips from zone A to zone B along the same least-cost path. Hence, it is known as an "all-or-nothing" assignment model. Having only one possible route from A to B is unrealistic. This is a serious limitation of this model. Other types of assignment models are being perfected which make assignments on the basis of various percentages of the total interchange value to alternate paths. Still others will make the assignment to alternative paths by a probability function.

The assigned traffic volumes on each network link are compared to the observed volumes. If serious discrepancies occur, the network can be redefined. This redefinition will, of course, create new minimum time paths (trees) which in turn will alter a subsequent assignment. When this repetition has produced close coordination between the assigned volumes and the observed volumes the network is considered to be calibrated. That is, it will satisfactorily simulate the real system.

Given a working network simulation, a technique must be devised for simulating the trip table. This is accomplished by another computer model known as a trip distribution model. This model is predicated on a loose interpretation of Newton's Law of Gravity. The assumption is that the trips are generated and attracted in direct proportion to the size of the attraction and in inverse proportion to the spatial separation of the two trip ends. Certain aberrations of this model attempt to include other parameters, such as friction factors between neighborhoods. When the distribution model allocates base year productions and attractions so that they correspond to base year origins and destinations (the trip table), this model is also calibrated.

The remaining forecasting technique which must be developed in this second division of a transportation study is a method for producing the trip origins and trip destinations that were distributed and assigned above. In other words, the residents of the study area were asked to enumerate the trips they made, producing origin and destination values for the base year of the study. Another simulation technique must be developed which will predict future productions and attractions for each zone. This phase of the process is called trip generation. The fundamental position in the total transportation planning process which trip generation occupies should be evident from the foregoing discussion and the illustration in Figure 1. If future estimates of zonal trips are in error, the distribution and assignment of trips will obviously be invalid. The concept of trip generation is more fully developed in the next section.

The third major division of a transportation study utilizes all the techniques developed in the second division to develop a simulated transportation network for a future year. The assumption is made that the calibrated models are as valid in some future year as they are in the base year for which they were developed. The basic input to this whole part of the study is local area goals and policies, that is, their decisions as to the future development of their community. The forecasting process creates an intricate fabric of future study area data. Woven together are projections of economic activity and population, land use, and transportation networks. The development policies become the catalyst. Each aspect of the total forecast is an interrelated contributor. Changing the land uses alters the network and planning for fewer people requires different land use allocations. With a given land use and network, future trips can be generated, distributed, and assigned to the network. Actually, several assigned future networks based on various land use alternatives would be developed.

These alternate networks must then be compared and contrasted. The fourth division of the study is concerned with this task. Analysis of one network provides feedback indicating other networks and land use possibilities. Evaluation of the networks are continued until one recommended network emerges. A plan for the future transportation system of the study area can then be devised and documented. That plan meets the Federal requirements as specified in the 1962 legislation. Steps are then taken to implement the plan and monitor the variables used as input to the models for possible discrepancies. Should discrepancies arise the plan is revised accordingly.

Trip Generation Philosophy and Methodology

The mathematical aspects of transportation models are commonly described under the headings of trip generation, trip distribution, modal split and trip assignment. Trip generation concerns the estimation of the number of trips into and out of various areal units; trip distribution deals with the estimation of the number of trip interchanges between pairs of traffic zones; modal split determines the proportion of trips by the various means of travel; and trip assignment allocates the trips to the road network links.

The trip distribution, modal split, and assignment stages of the traffic forecast have attracted considerable research work and are generally regarded as more complex problems than that of trip generation. It is important to establish, however, that current trip distribution techniques are primarily designed to produce the relative rates of attraction between zones and that the volumes of inter-zonal movements are obtained only when these rates are applied to the trip generation estimates. The trip generation stage thus determines the scale of the traffic forecast. Before proceeding with a discussion of trip generation, the concept of a trip needs to be explored and a common frame of reference established.

Trips and Trip-Making in Transportation Planning

A trip as defined in transportation planning is rather unique. Trips are made for specific, identifiable purposes and they are unidirectional. A trip to the grocery store to purchase a loaf of bread is considered two trips by the transportation planner, one to the store and one returning home. An extended shopping trip could possibly be composed of three or more trips. When several trips are linked together the purposes and the origins and destinations may change with each trip. For instance, a woman could leave home in the morning with a child as an occupant of her vehicle. The child could be taken to school. The trip would have an origin at the home and a destination of the school location and a purpose of serving a passenger. The woman may then decide to visit her mother. A second trip is made with an origin at school and a destination at the mother's house with a social-recreational purpose. The third trip may have a destination of some shopping center and a purpose of shopping. If her next stop is a bank she would have made a trip with a new destination and a purpose of conducting personal business. Assuming the time to be approaching the noon hour, she may then drive from the bank (origin) to a restaurant (destination) for a purpose of eating a meal. This is the fifth trip. If she has a job afternoons, her sixth trip would be for the purpose of work and would have the appropriate destination. At five o'clock she drives to her husband's place of employment to transport him home. The eighth trip is made from her husband's location as an origin to their home as a destination for a purpose of returning home. There are two people in the vehicle. This, of course, would not be the only trips for that household for that day. In fact, they may not be the woman's entire trip record for the day. This example illustrates the precise nature of trips as used by the transportation planner. Two further distinctions must be made.

Trips may be classed as vehicle trips or person trips. A vehicle traveling from point A to point B makes one vehicle trip. However, there may be four occupants in the vehicle for a total of four person trips and one vehicle trip. In the above example of eight vehicle trips there were ten person trips made as the first and eighth trip had two occupants in the vehicle.

Trips are also differentiated on the basis of whether they are home-based or nonhome based. That is, did the trip originate at the residence or at some other origin? Furthermore, the transportation planner becomes even more discriminate by identifying trip-ends. Any trip has two ends. The ends are given names as to whether the trip-end is a production or attraction. Productions and attractions are not the same as origins and destinations. The following is offered to aid in this clarification:

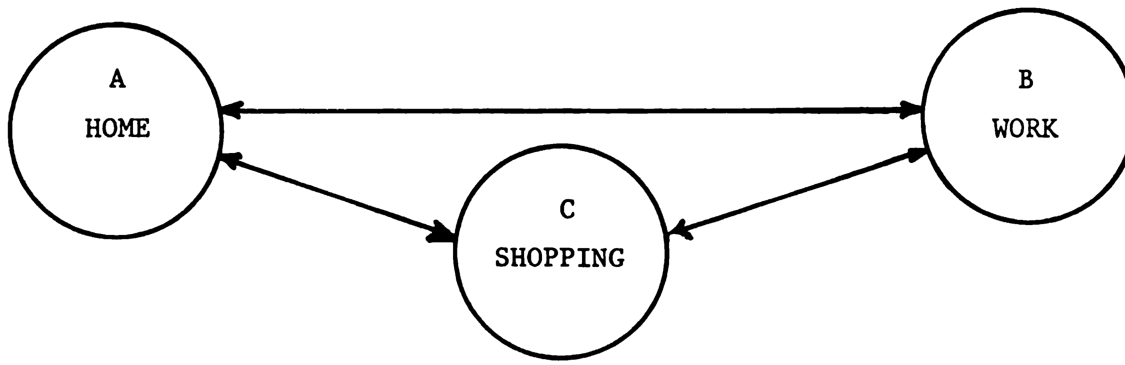
A. Home-Based Trips

1. Origin or destination at place of residence
2. Production zone is always the zone of residence; the other zone is always the zone of attraction.

B. Nonhome-Based Trips

1. Trips having neither their origin or destination at the place of residence.
2. Production zone is always the zone of origin and the attraction zone is always the zone of destination.

The following example may be helpful.



Trips

A to B B to A	A is the production zone; B is the attraction zone.
A to C C to A	A is the production zone; C is the attraction zone.
B to C	B is the production zone; C is the attraction zone.
C to B	C is the production zone; B is the attraction zone.

Evolution of Trip-End Modeling

Trip generation or trip-end modeling is a function of three basic factors; land use variables, system variables, and socioeconomic variables. These variables are all somewhat related and could easily be illustrated as a three-dimensional entity as portrayed in Figure 2.

Until recently, very little was done to quantify the above variables or to attempt any descriptions of their relationships. In the immediate post-World War II era, the concern of origin-destination studies was to tabulate traffic flows into trip tables of origins and destinations. This data was also presented as "band width maps" where the traffic on particular streets was scaled and represented as bands

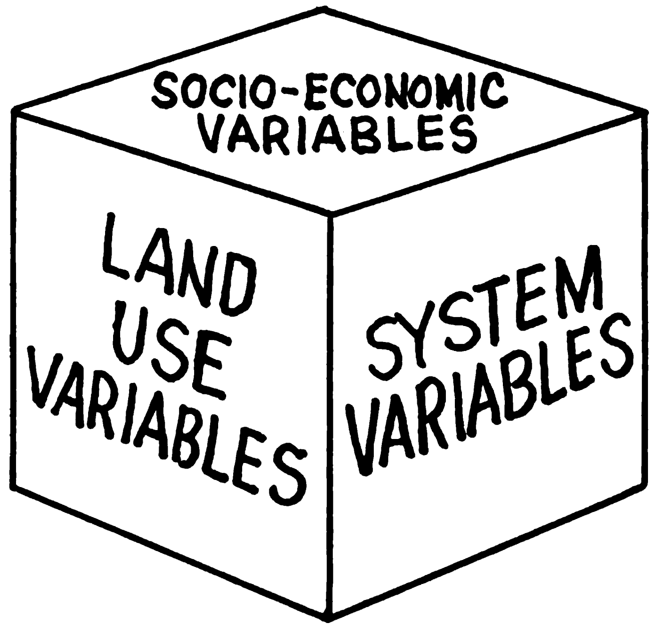


Figure 2. Trip Making Variables.

of varying widths. The data was also portrayed as "desire lines." These are also scaled bands, but only linking origins and destinations and not following streets. Any projections of future trips were made by an extrapolation method on historical traffic records.

> Beginning in the early 1950's, analytical techniques were used in an attempt to quantify urban trip volumes in terms of measurable characteristics of the people making the trips or the land use associated with the trip ends. Here, existing land uses were categorized by type of activity, location, and intensity of use. Trip generation rates were developed for each of the categories and applied to a land use plan for the forecast year. These pioneering efforts, although unpolished, were important in that, for the first time, data had been collected for the purpose of developing future land use-travel relationships. The most noteworthy of the early studies with respect to the

> use of sound land use-travel relationships were the San Juan, Puerto Rico (1948), and Detroit, Michigan (1953), transportation studies.[1]

These early studies were followed by a number of individuals in the planning and engineering fields who conducted studies which improved the sophistication of the land use-travel relationship. The widespread application of the digital computer to studies of this nature, in the late 1950s, was the most significant contribution to increased analytical expertise. These machines allowed for a vast amount of data to be intricately analyzed by rather complicated calculations.

Today, trip generation or trip-end modeling represents a sophisticated procedure. Most of the Federally required transportation studies have completed this stage of their study. Thus, these studies represent a great deal of experience with selecting and organizing variables to predict trip-ends. A great many variables have been developed, tried, tested, rejected, and accepted. Whereas the first efforts at forecasting trips sought simple linkages between land use and socioeconomic characteristics, the present emphasis is on causal relationships. Some of the methodology in current practice will be discussed below.

Land area trip rate analysis attempts to relate the data collected in the land use survey or the socioeconomic survey to the

---

[1]U.S. Department of Transportation, Guidelines for Trip Generation Analysis, Federal Highway Administration, Bureau of Public Roads, Washington, D.C., June 1967, p. 1.

data collected in the origin-destination survey. Rates can be developed for individual zones (or some other areal configurations), for groups of zones or for the total study area. Trip-ends per residential acre, or trip-ends per registered auto, or trip-ends per $1,000 of family income are all examples of simple rates. Trip-ends can be either productions or attractions. Rates can be developed for various trip purposes. As unsophisticated as this process appears, it nevertheless produces rather satisfactory results.

A second technique of trip generation employs the concept of regression. Actually, all trip-end models are regression models of some type. The regression model simply assumes that some variable "Y" responds to changes in other "X" variables. The Y-variable (hereafter referred to as just Y) is the quantity under study and is known as the "response" or "dependent" variable. The X-variables are those which exhibit a causal effect on the value of Y and are known as the "explanatory" or "independent" variables. The trip generation model employing a rate, as described in the preceding paragraph, is technically a regression model. However, in the literature, in the transportation planning field, and in this research the term "regression" applies to only a specific case, which is least-squares regression. Therefore, a discrimination will be made between regression and rate modeling even though technically this distinction may be invalid.

Least-squares regression assumes that Y, or the response surface, is linear. The dependent variable is proportional to the independent variable(s). This response (Y) is represented by a mathematical equation. The attempt is made to have the response surface pass through all the observed data as closely as possible. Each of the differences between the surface and the actual observed value is squared to remove negative values. When the sum of the squares is minimized, a surface has been defined which is the most accurate. Providing such a "best-fit" equation manually would be cumbersome and time-consuming, but sophisticated computer programs produce this equation and many inferential statistics describing the equation with relative ease on the part of the planner.

The independent variables in formulation of such an equation are the model parameters. If a model is desired which will predict zonal trip-ends, then these measurements are zonal totals. The total number of registered cars in a zone, the number of dwelling units, the number of people, the number of governmental employees are a few examples of independent variables. In this application each transportation zone is treated as one observation. This is the most commonly used trip generation model in the United States today.

A variation of the zonal least-squares regression procedure is to use dwelling unit data for the independent variables and create a model which predicts trip-ends at the dwelling unit (DU) level. Using this application every dwelling unit becomes an observation. The arguments for this type of regression procedure are that a more efficient use of the survey data is made and that interpolation between the data can be realized. In other words, instead of using the mean of aggregations of a great amount of data as single observations each survey or interview is one observation and the uniqueness of the total data set is preserved. Thus, it would follow that smaller interview samples could be utilized without any loss in reliability. Additionally, the problems of dealing with qualitative variables and the assumption of a linear response surface can be overcome by using a dummy variable technique. This technique will be explained later in the text. The problem of a nonlinear response surface in the zonal regression procedure can be overcome by other techniques but, not as easily.

A final procedure which has become more popular in the United Kingdom than in the United States is known as cross-classification (or category analysis). It is a newer application of the rates procedure. Just as dwelling unit regression attempts to use data in its rawest form, cross-classification develops individual and highly specific rates.

> The cross-classification technique is a correct application of the regression concept. It can deal easily with qualitative variables and does not make any assumptions about the shape of the response surface.
>
> The technique is based on determining the average value or average response of the dependent variable for defined categories of the independent variables. The categories are defined by a multidimensional matrix where each dimension represents an independent variable, and where the independent variables are themselves stratified into categories. In applications to trip end modelling each observed analysis unit (e.g., household or zone) is assigned to a particular category or cell of the matrix depending on its values of the independent variables (typically such measures as number of cars, household income, family size, etc.) and average trip rates are subsequently determined for each cell.[2]

---

[2]A. A. Douglas and R. J. Lewis, "Trip Generation Techniques: Introduction," Traffic Engineering and Control 12 (November 1970): 364.

In the Puget Sound Study,[3] the zonal approach was used allocating average socioeconomic values of the zones to a cell in the matrix. Ironically, the cross-classification concept, debuted in this United States' study, has had negligible use. Some rather significant disadvantages may account for this, one being that sufficient base year survey data are required so that every category has enough observations to render the rates valid and reliable. Since the more frequented categories may shift from base year to target year, the complete matrix must be well established. Another disadvantage of the cross-classification technique is that there is a lack of statistical tests with which to analyze the models, particularly the measurement of the extent to which the model accounts for the variance in the data.

Although cross-classification is still considered as a potentially viable trip generation technique, it was eliminated for consideration as a method for analysis in this thesis because of general disuse in the field, problems with statistical testing, and the author's belief that problems with real-world applications will preclude any widespread acceptance. The remainder of this section reports recent trip generation research having to do with rates and regression analysis.

---

[3]John R. Walker, "Rank Classification: A Procedure for Determining Future Trip Ends," Highway Research Record No. 240, Highway Research Board, Washington, D.C., 1968, pp. 88-89.

In an extremely extensive study, Parsonson and Horn[4] used the following four methods of estimating zonal values of auto-driver trip production and attraction for the Raleigh, North Carolina urban area: (1) Average Rate Method; (2) Modified Iowa Method; (3) Modified North Carolina Method; and (4) Multiple Linear Regression Method. In addition, for each of these methods they used varying sample rates and tested the validity of sample size as well.

The first method was used only with trip productions. It utilized rather simplistic parameters of area-wide average rates of trip production per dwelling unit. Each zone using this rate was multiplied by the number of dwelling units.

The Modified Iowa Method is actually a rate method whereby the area-wide trips, either productions or attractions of various purposes, are divided by correlated variables. For example, work trip productions were subdivided into zonal estimates on the basis of the labor force residing in each zone.

The Modified North Carolina Method is a quasi cross-classification technique. Zones were grouped on the basis of socioeconomic level and an average trip rate per dwelling unit was calculated. These rates multiplied by the number of dwelling units in each zone provided the estimates.

---

[4]P. S. Parsonson and J. W. Horn, "Comparisons of Techniques for Estimating Zonal Trip Productions and Attractions," University of North Carolina, School of Engineering, Raleigh, June 1966.

The fourth method was a departure from "rates" and employed the multiple linear regression concept of fitting a planar surface to observed data to provide a mathematical equation which estimated rates as has been discussed.

The study provided a rather good analysis of various kinds of rate techniques. The inclusion of linear regression into this analysis was cursory. The only criterion used to analyze the results was root-mean-square deviations. In this instance, the standard errors of estimate might have been a better analytical tool.

In a Texas study,[5] a comparison of the use of rates with conventional multiple regression models was also investigated. The rates were based upon a single variable describing land use or population characteristics. The regression phase developed several alternate models. Some of the conclusions of this study were:

1. The multiple regression models provide relatively accurate and unbiased estimates for the HB Work productions. Regression estimates for the HB Non-Work productions also were unbiased, but less precise.

2. For HB Work productions regression models were judged superior when the zone is developed and future projections are made with confidence. Otherwise, (undeveloped zones) rates are likely to provide better estimates.

3. For HB Non-Work productions regression is probably better but they conclude that rates would probably do as good a job.

---

[5]John C. Goodknight, "A Partial Analysis of Trip Generation Analysis," Texas Transportation Institute, Texas A & M University, College Station, August 1968, pp. 1-5.

4. For HB attractions (both Work and Non-Work) rates were judged superior.

5. For NHB productions and attractions rates are also superior.

6. Rates calculated as "Ratio of Average" (total number of trips in all zones divided by total number of units of the independent variable for all zones) were superior to rates calculated as the "Average of Ratios" (average of the individual rates for all zones).

One English study[6] concludes that trip-end models, based on zonal data, should be rejected as such models have shown to be unstable from one study area to another. The English authors' deduce that instability over time will follow this geographical instability. They fail to provide an explanation for this connection of time and space. Trip-end models based on disaggregated data, using the dwelling unit as a basis, is the recommendation of these authors for future work.

Both multiple regression and cross-classification techniques are recommended. However, with the latter, there are appreciable deficiencies which have already been discussed.

Kassoff and Deutschman[7] made a two-part study on alternate approaches to trip generation. First they examined the use of aggregated data (zonal) and the performance of relationships based on aggregate totals (such as trips per zone) as opposed to the use of

---

[6]A. A. Douglas and R. J. Lewis, "Trip Generation Techniques: Category Analysis and Summary of Trip Generation Techniques," Traffic Engineering and Control 12 (February 1971): 535.

[7]Harold Kassoff and Harold D. Deutschman, "Trip Generation: A Critical Appraisal," Highway Research Record No. 297, Highway Research Board, Washington, D.C., 1969, pp. 29-30.

aggregate rates (such as trips per household per zone). Second, they examined the implications of using disaggregated data (data not combined and averaged according to predefined areal units) versus aggregated data. All of these methods were accomplished using multiple linear regression. They found that the aggregate-total method had a slight statistical advantage over the aggregate-rate equation. However, because of the flexibility of using rates (the rate equation is "not tied to the data scheme to which it was developed"), the authors recommend that aggregate rates be employed. In comparisons of disaggregate and aggregate procedures, the disaggregate equations proved slightly superior after statistical comparisons. They recommend the use of these equations.

McCarthy[8] found that the zonal sample means are not representative of all the households in the zone. Rather, the distributions are skewed. The author's findings led him to refute the validity of the assumption of zonal homogeneity. Aggregation of data to the zonal average caused a major percentage of the total variation in individual household automobile ownership, family size, and total home-based trip generation rates to be lost in aggregation. McCarthy did not, however, use zone totals of the independent variables. There may be some differences in using the totals as opposed to the means.

---

[8]Gerald M. McCarthy, "Multiple Regression Analysis of Household Trip Generation--A Critique," Highway Research Record No. 297, Highway Research Board, Washington, D.C., 1969, pp. 41-42.

Furthermore, he finds that the coefficient of determination for the total home-based trip generation equation developed from zonally aggregated data is deceptively high. The data utilized in developing the basic trip generation equation contain only 12.3 percent of the total variation existing in the individual household total home-based trip generation rate data.

Fleet and Robertson,[9] after reviewing research findings concerning data variation and aggregation, conclude that aggregation should follow analysis (the trip generation phase). The "fine-tuning" of a multiple regression trip generation analysis with aggregated data may be of marginal value. Also, they find that too much validity is given to statistically derived procedures and possibly the wrong ones are chosen.

Purpose and Scope of the Research

Since there is no concensus among transportation planners as to the appropriate methodology regarding trip generation, this study was designed to investigate and evaluate three alternative methods of performing the trip generation phase of the transportation planning process. First, the zonal estimates of Home-Based Work, Home-Based Shopping and Home-Based Other productions for the residents of the study area were made by a standard zonal level, stepwise multiple regression analysis procedure. Second, estimates were obtained for

[9]Christopher R. Fleet and Sydney R. Robertson, "Trip Generation in the Transportation Planning Process," Highway Research Record No. 240, Highway Research Board, Washington, D.C., 1968, pp. 11, 25-26.

the same type of trips by simple zonal rates. Third, a sufficiently detailed procedure was devised to use disaggregated dwelling unit data with the multiple regression procedure to eventually produce zonal estimates. The question attempted to be answered was: "Which of the three procedures investigated gives the most accurate estimates of home-based trip productions?"

The estimates obtained by each method were judged for accuracy by comparing them with the actual home-based productions as obtained from a comprehensive origin-destination survey conducted in the study area in the spring of 1967. The comparisons were made by employing several common statistical tests, by graphing the data to make a visual interpretation of the results, and by subjectively judging the results for logic.

This study was limited to three types of Home-Based productions as Home-Based attractions, Non-Home-Based trips, and truck/taxi trips are not conducive to estimation using dwelling-unit data. Also, more data are available to relate to these kinds of trips. Since Home-Based productions comprise the majority of trip purposes in any given area, the results could be considered to exhibit a high level of confidence.

CHAPTER II

STUDY PROCEDURES

The design of this research project was formulated to make use of existing data. The scope of the project did not warrant collection of any new data. The estimating methods analyzed are commonly applied techniques within the transportation planning field. No attempt was made to devise new methods or improve upon the old. The analytic tools are mostly procedures and routines in general use, although some original programming was necessary to organize the data for use in this study. The analysis criteria are also commonly applied tests. This chapter begins with a discussion of the three estimating methods indicating the theory behind each and nuances between them. (The study from which the data was taken is herein described because the uniqueness of the area, in part, led to the selection of the study area.) A flowchart of the study steps is presented which illustrates the continuity and relationship of the investigational sequence. This chapter concludes with an enumeration of the criteria used in comparing the estimates.

Description of Estimating Methods of Investigation

The methods of estimating Home-Based trip productions which were used in this study are the following: (1) Zonal Multiple

Regression Method; (2) Average-Rate Method; and (3) Dwelling-Unit Multiple Regression Method. Each of these methods was used to obtain estimates of each of the three trip types for each of 125 zones or 1,932 dwelling-units. Therefore, a total of nine equations were developed and analyzed in this study.

<u>Zonal Multiple Regression Method</u>

The concept of regression, in a general sense, assumes that some variable "Y" responds to changes in another variable(s) "X." The Y variable is the quantity under study and is known as the dependent variable. The X variables are those which exhibit a causal effect on the value of Y and are known as the "explanatory" or independent variables. The response surface is fitted to the data by means of the least-squares technique. For computational purposes the response surface is assumed to be linear. The "best-fit" line or plane through the data is the one which has the smallest sum of squares of the residuals or residual sum of squares (RSS). This fact is illustrated in Figure 3 for a simple regression of one independent variable. The estimating equation can be written as:

$$\hat{Y} = a + b\,X_1$$

for simple regression, and:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 \quad \ldots \quad b_kX_k$$

for multiple regression.

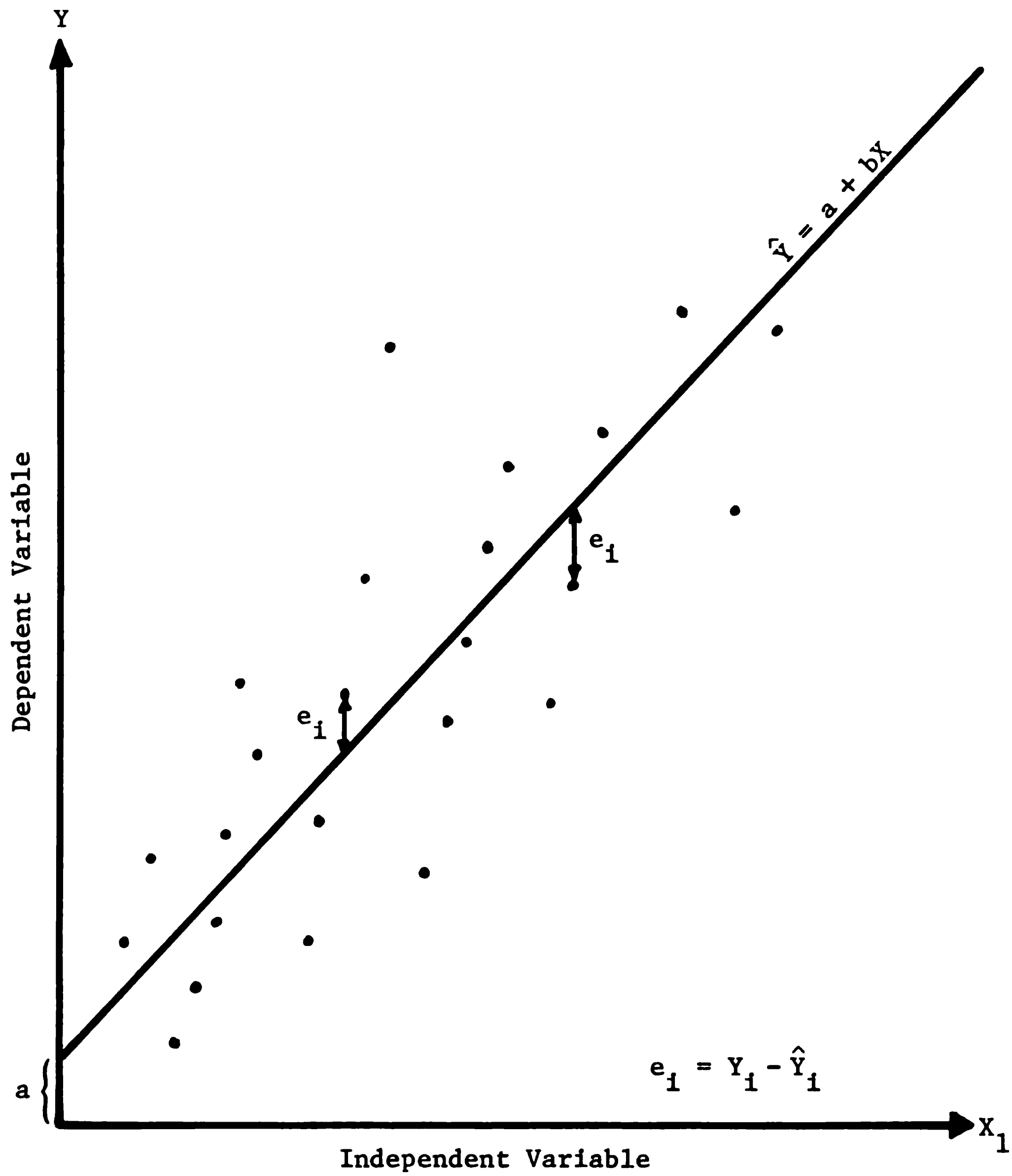


Figure 3. Least-Squares Regression Line.

The residuals (i.e., differences between observed (Y) and estimated ($\hat{Y}$) or predicted values) are denoted by:

$$e_i = Y_i - \hat{Y}_i$$

The equation parameters (a and b_1, b_2, b_3, etc.) are chosen such that e_i^2 (the sum of all the residuals squared) is minimal. Squaring the residuals removes negative signs. The extensive use of multiple linear regression has been made possible via sophisticated computer programs. Such programs can rapidly provide a series of "best-fit" equations which automatically select various combinations of the independent variables.

The program used in this study is part of a standard package of statistical analysis programs developed by the Burroughs Corporation for use on the B-5500 computer as used by the Michigan Department of State Highways and Transportation. The program builds up the regression equation by successively adding explanatory variables. The variable added at any step is the one which produces the greatest reduction in the residual sum of squares. At each step, however, the program checks to determine whether the independent variable has a "significant" effect on the prediction of the dependent variable. Variables already in the regression set may become nonsignificant and removed. The test of significance is achieved by calculating t-values for each regression coefficient. In this study t-values had to be above ± 1.96 for acceptance, which is significant at the 95 percent level of confidence.

Requisite to the use of the least-squares method are a number of important assumptions. The various statistical tools associated with this model may become invalid if these basic assumptions are violated. The main assumptions of the least-squares model as used in transportation planning trip generation are as follows:

1. Constant error variance. The mean and covariance of the residuals is zero, their variance is constant and their distributions are normal.
2. Multicollinearity. The correlation between X's must be kept low as the effect of each on the equation cannot be judged accurately otherwise. In this study if two independent variables had a correlation of more than 0.80, only one was used.
3. Errors in variables. The least-squares model estimates the mean value of Y for given values of X. Measurement errors in X are not, therefore, allowed.
4. The shape of the response surface. The model assumes Y is a linear function of the X's. The X's may not be in their original form and transformations such as logarithms, reciprocals, square roots, and exponential powers are sometimes used to provide this linearity. No transformations were made in this study.

The zonal multiple regression method has been used widely in past transportation studies. In its application each traffic zone is treated as one observation (Appendix C contains a study area zone map). The Y is a measure of a specific type of trip as taken from the O-D survey. The X's are zone totals of such measures as number of people, number of cars, number of households, etc. The model is then "fitted" as described above so that the mathematical equation takes known socioeconomic data to estimate known trips.

In using the model as a forecasting tool, it is assumed that the derived relationship will remain stable over time, and that future values of the independent variables can be estimated accurately outside the model. The typical forecasting periods are 20 and 25 years.

Average-Rate Method

Rates, as the name implies, are a number of trips "per" some standard. The standard could be dwelling units, cars, people, families, etc. There are several distinct procedures for developing rates. The "Average of Ratios" is found by calculating the trip generation rate for each zone individually and then averaging these individual rates. This might be called a zonal average rate. The "Ratio of Averages" is calculated by dividing two study area totals. For example, dividing the Home-Based Work trips for all 125 zones in this study by the number of dwelling units in the study area would yield a rate for that trip purpose. This might be called a (study) area rate. A third procedure is to rank all zones by the increasing value of the individual zone rate and then divide them into quartiles and calculate the "Ratio of Averages" for each quartile. A Texas study[10] documented these procedures and concluded the "Ratio of Averages" was the most satisfactory.

---

[10] John C. Goodknight, "A Partial Analysis of Trip Generation, Texas Transportation Institute, Texas A & M University, College Station, August 1968.

Therefore, this thesis used a rate for each trip type based upon the "Ratio of Averages" for the same independent variables as the Zonal Regression equation produced.

Dwelling-Unit Multiple Regression Method

Zonal multiple regression only attempts to explain variation in trip-making behavior between zones. A greater amount of variation may exist within zones. Reducing zone size decreases within-zone variation particularly when the smaller zones are homogeneous with respect to socioeconomic characteristics. However, smaller zones introduce the problem of high sampling errors which cannot be allowed for in the least-squares model. The logical extension of reducing zone size is to develop disaggregate models which make no reference to zone boundaries. Analysis at the person level may appear to be the best choice. However, the dominating influence of the head of household implies that the trip-making activity of the household members can only be accurately predicted through a knowledge of total household characteristics. It would appear that the household can reasonably be considered a behavioral trip-making unit and therefore, treated as the basis for the trip-end estimating procedures. It should be noted, however, that the ultimate aim is to produce an estimate of total zonal trip-ends for input into the trip distribution model. Consequently, any disaggregate model must be capable of expansion to the zonal level.

The practical side of the argument in favor of dwelling-unit analysis considers the large sums of money spent for household interviews. Collecting household data and then reducing it to zonal totals seems a waste of resources. Should this method be valid and reliable, smaller survey samples may be acceptable which would even produce a monetary savings.

The problems of dealing with qualitative variables such as type of dwelling unit or stage in family life cycle can be overcome by using dummy variables. Some (or all) of the independent variables are stratified into categories. The categories are then represented in the regression model by a system of dummy variables which take either a 1 or 0 depending on whether the observation (household) falls into the particular category. Therefore, a great many more variables, as well as observations, are able to be input to the model for a given set of data.

With the exceptions noted thus far, the development and operation of this method is as was previously described for zonal regression. In this part of the study another regression program was actually used. Its operation was the same, the only difference was the first program had a limitation on observations which made it incapable of handling the large dwelling-unit data file.

Description of the Study Area

The basic data used in this thesis was supplied by the Adrian-Tecumseh Area Transportation Study, which was initiated in 1967 by the Michigan Department of State Highways and Transportation. The study was a major Origin-Destination study, but is somewhat unique in that it has two urban concentrations which are not at all contiguous. These are the cities of Adrian and Tecumseh which are separated by about 10 miles. The study area is located 40 to 50 miles southwest of Detroit. The study area comprises all of two townships and part of four other townships. About 216 square miles in the center of Lenawee County is within the cordon line. The study area is divided into 125 internal analysis zones and 23 external zones or stations. The cordon line cuts highway trunklines in seven places and 50 additional roads cross it. During the spring of 1967 the home and vehicle interviews were taken as well as all the vehicle counting and classifying at the various stations. This base year data yielded a population of 46,050 residing in 14,085 dwelling units for a person per dwelling unit average of 3.27. There were 17,567 passenger cars tabulated producing 1.25 passenger cars per dwelling unit and 2.62 persons per car. There was an average of 8.80 vehicle trips per dwelling unit and 14.15 trips per dwelling unit. The number of passenger cars in a zone varies from 0 to 565. The number of employed people in a zone varies from 0 to 3,233.

There are no limited access facilities in the Adrian-Tecumseh study area. Two state trunklines traverse the total area: M-52 bisects the study area into east and west halves, and M-50 passes east and west across the top of the study area. Another state trunkline, US-223, crosses east and west in the southern half of the study area. M-50 passes through Tecumseh and US-223 passes through Adrian. A fourth state trunkline, M-34, emanates from Adrian and moves in a westerly direction to the western cordon line.

Preparation of the Data

After all the field work was completed, the interviews were keypunched to standard IBM cards and a magnetic tape was created. This tape was then edited for errors and expanded by appropriate factors. The expansion was done to increase the interviews from the sample rate to the universe. This expansion was done under carefully controlled conditions and accuracy checks were made on the expanded data. This final magnetic tape for the Adrian-Tecumseh Area Transportation Study contained 59,232 records. Each record was in effect an interview or a trip. These records included 2,930 household interviews, 25,156 internal trips, 26,772 trip records from cordon station interviews, and 4,374 trip records from truck and taxi interviews. The household interviews were taken on a 20 percent sample rate. The external interviews, which are interviews of people passing into and out of the study area, were about a 60 percent sample. The truck sample was 50 percent while 100 percent of the taxi trips were recorded.

In order to do the research for this thesis, two data files had to be constructed from this base data file. The first was used in zonal multiple regression analysis. In each case, the trips with dependent variables "Y's" would be the same. For analysis at the zonal level, the socioeconomic variable or the independent variable would be zonal totals. These totals are contained in one file. The second data file contains actual household information. The composition of these two data files is summarized in Tables 1 and 2, respectively.

Table 1. Variables in Zonal Data File

Independent Variables

Population
Dwelling Units
Resident Labor Force
Autos per Dwelling Unit
Autos
Employment:
- Manufacturing
- Wholesale/Retail
- Service
- Government
- Other

Dependent Variables

Home-Based Work Productions
Home-Based Shopping Productions
Home-Based Other Productions

Home-Based Other productions for each of the two files is a composite of all the Home-Based production trips excluding work and shopping. These trips could have been made for transacting personal business, school, social recreation, changing mode of travel, eating a meal, medical or dental or serving a passenger. It is felt that none of these categories by themselves constitute enough trips to warrant creating separate trip generation equations. Table 3 summarizes the study area trips as used by both data files and gives the percentages of each type of trip compared to total trips. Trips, as shown here, are really trip-ends as has been previously explained.

On the zonal data file, there were 125 observations for each of the variables. These observations are the zones in the study area. On the household data file, there were 1,932 observations. This is

Table 2. Variables in Household Data File

Independent Variables

Structure Type:

- Single-Double
- Group Quarters
- Residential Hotels
- Mobile Homes
- Multiple (Apartments)
- Other

Cars Available:

- 0
- 1
- 2
- 3
- 4 or more

Persons at the Address:

- 1
- 2
- 3 and 4
- 5 or more

Years at the Address:

- 1
- 2-5
- 6-10
- 11 or more

Resident Labor Force (RLF):

- 0
- 1
- 2
- 3 or more

Income:

- $0-$3,999
- $4,000-$5,999
- $6,000-$7,999
- $8,000-$9,999
- $10,000-$14,999
- $15,000 or more

Family Life Cycle (FLC):

- 1 person over 45 years
- 2 persons over 45 years
- 1 person under 45 years
- 2 persons under 45 years
- 3+ people, youngest less than 5
- 3+ people, youngest 5-18
- 3+ people, youngest over 18

Dependent Variables:

Home-Based Work Productions

Home-Based Shopping Productions

Home-Based Other Productions

Table 3. Vehicle Trips by Trip Purpose (1967)

Component of Analysis	Vehicle Trips	
	Number	Percentage
Residents		
HB Work	17,573	15.70
HB Shopping	15,998	14.30
HB Other	41,705	37.27
Non-HB (All)	36,619	32.73
Total	111,895	70.79
Non-Residents		
Cordon Trips	28,372	17.95
Truck-Taxi	17,792	11.26
Total	158,059	100.00

comparably fewer than the 2,930 household interviews that were contained on the base file. The criterion for selecting these records of the base file was that each record had to be complete for the variables in question. Therefore, 1,932 of those data were complete records.

It can be seen by comparing Table 1 and Table 2 that the type of independent variable and the organization of these independent variables was somewhat more detailed on the household file. Also, each record is an interview, therefore, that interview can fit only one stratification of each of the independent variables. For instance, record number 1 must be one structure type but it can only be one structure type. Therefore, this data file simply contains a 1 or 0 in the appropriate column depending on whether the observation was represented by that variable or not. In the zonal data file, however, for each of the 125 observations, there exists a number. For example, in zone 1 there were 284 dwelling units and 369 autos; for zone 80 there were 266 people residing in that zone, 20 of which were employed. For each independent variable in the zonal data file, there exists a value, not just a 1 or a 0.

The variables in the household file which indicate Family Life Cycle had to be calculated at the time the file was created from the base file. The Michigan Department of State Highways and Transportation had never had occasion to use such a variable in any analytical work and the question per se was not asked during the interview process. The author wrote a program which would read relevant fields from the base data file and make the appropriate calculation for each observation.

Analysis Procedures

The research procedures discussed thus far are preparatory towards resolution of this thesis which states that there are significant differences in the accuracy of the trip generation model when the basis of that model is a regression on zonal data, or a regression on dwelling-unit data, or an application of a study area average rate. This section presents and discusses Figure 4 which illustrates in flow chart style the analytical procedures to investigate model differences.

Figure 4 reads from left to right in a series of chronological steps. The chart is constructed in three tiers which are the three

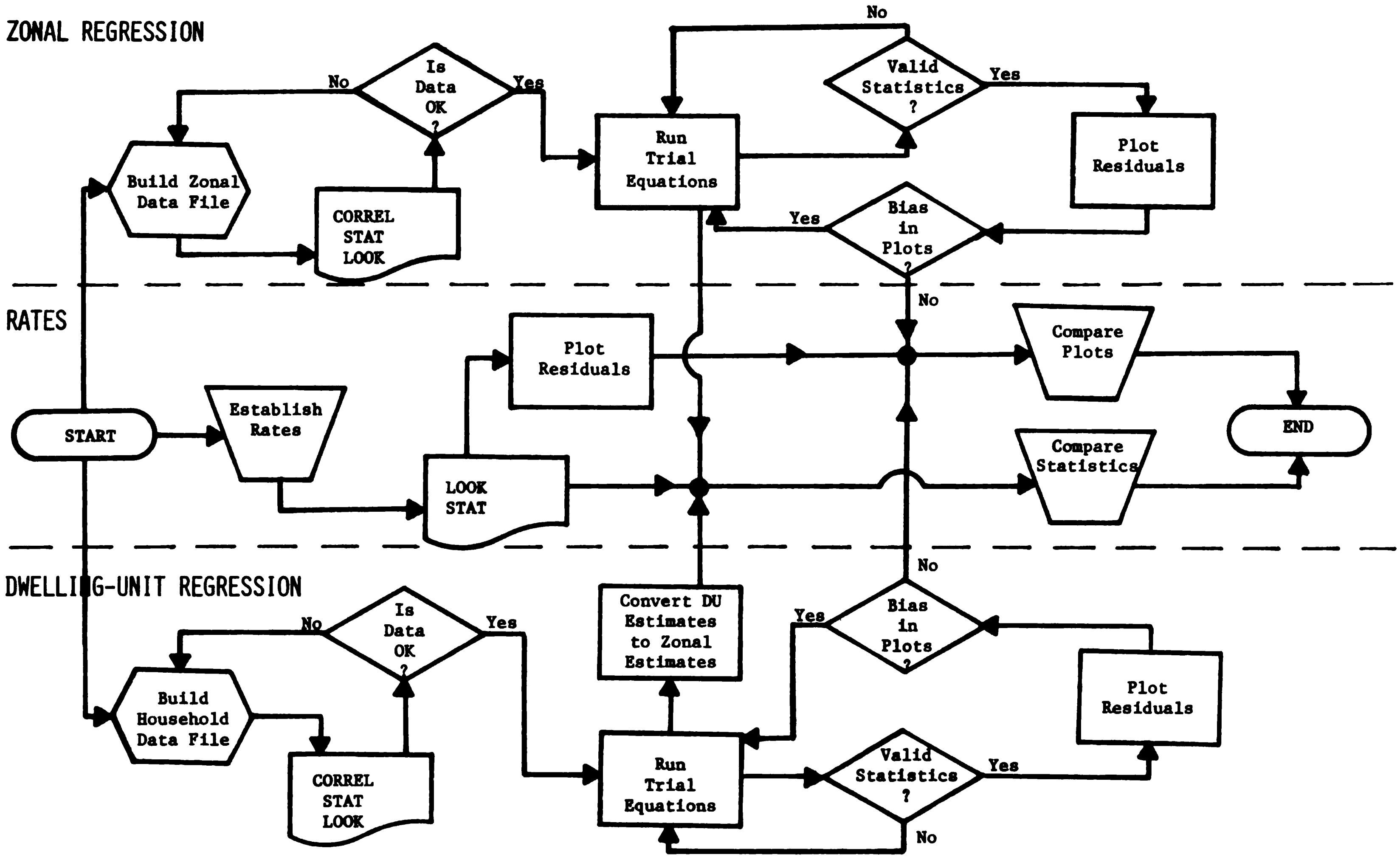


Figure 4. Flow Chart of Analysis Procedures.

analysis methods following their own paths and eventually coming together for graphic and statistical comparisons. The results of these procedures are presented in Chapter III and discussed further in Chapter IV.

The building of the two data files has previously been discussed. A computer program (CORREL), which calculates correlations between all variables, was then run on both files. Another program (STAT) was used which calculated basic statistics (mean, sums, sum of squares, standard deviations, etc.) on all the variables. Finally, a third program (LOOK) displayed certain parts of the data. All the data were analyzed to be certain they were logical and rational in the perspective of analysts acquainted with the study area.

Next, the zonal equations were formulated by repetitive runs of the regression program. All variables and any plausible combinations of variables were tested. A computer program (PLOT) which plots Cartesian points was used to plot the residuals from each of the runs to aid in selecting the best equation. One equation for each trip purpose (Work, Shopping, Other) was selected as being superior after applying statistical, graphical, and subjective judgments. The independent variables in these equations became the variables upon which the rates were constructed.

The middle tier of Figure 4 was performed after the upper tier. After the final equations using the zonal regression method were indicated, the author wrote a program to tabulate study area totals of these variables from the base data tape. Each rate is some number of trips for one unit of some socioeconomic characteristic. Using the display program (LOOK) with a computational option, the rates were applied to each zone and the zonal estimates printed. The residuals were calculated for plotting and the statistics obtained. Each was available for comparison with the regression methods.

The Dwelling-Unit (DU) file was input to another stepwise regression program (one that could handle more observations) with all the variables as possibilities. An analysis was made of the output to determine at which step the t-value of the newly added independent variable was <u>not</u> ±1.96. This became the established "cut-off" point and the program was run again with only those variables. This was done for each of the trip types to produce the three superior equations.

The dwelling-unit estimates were converted to zonal estimates by multiplying by the number of dwelling-units in each zone. At this point there were statistics and residual plots for the dwelling-unit method to compare with the statistics and plots from the other two analysis methods. This bottom tier on Figure 4 was accomplished concurrently with the former two methods (tiers).

All of the estimates and residuals for each of the nine equations, the trip data for the three types of trips, and the socioeconomic data from the zonal file were put together on one file. Final plots and the basic statistics routine were then run again on this complete data file for the analysis and comparisons of results.

Analysis Criteria

A model is only as good as the validity of the tests performed. Evaluations of the equations in this study consisted of three levels applied at two different phases. First, a myriad of zonal multiple regression equations were produced and determination had to be made as to which was the best. Several different DU equations were also produced from which the superior had to be selected. Second, after the final nine equations were ascertained the same sort of analysis had to be made to determine which of the three methods was superior. These evaluations, in each of the two phases, were made by the following three analysis procedures: (1) Statistical Analysis; (2) Graphical Analysis; and (3) Subjective Analysis.

Statistical Analysis

The analysis procedures illustrated by Figure 4 create statistics on dependent and independent variables and on total equations. Certain statistical parameters become tests when theorems are applied to the values. The statistics and the statistical tests used in this thesis are described below.

a. Mean (u, $\overline{y}$, $\overline{x}$)--the most common measure of central tendency of a set of numbers. It is also referred to as the arithmetic average for it is simply the arithmetic sum of all the values divided by the total number of values.

b. Standard Deviation(s)--is a measure of dispersion about the mean. The formula used is:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X - \bar{X})^2}{n - 1}}$$

The standard deviation is given in the same units as the original data. The variance is the square of the standard deviation.

c. Simple correlation coefficient (r)--is computed for each pair of variables (both independents and dependents) and measures the degrees of linear association between them. The value of r lies between ±1 with +1 being perfect, direct correlation and -1 being perfect, indirect correlation. A condition of multicollinearity exists when two independent variables are not only highly correlated with Y but highly correlated with each other. When that situation was evident, only one of the X's was allowed to enter the equation (see comments on subjective analysis).

d. Multiple correlation coefficient (R)--is an empirical measure of the "goodness of fit" between the regression estimates and the observed data. It may be regarded as the simple correlation coefficient between the observed (Y_i) and the estimated ($\hat{Y}_i$) values.

e. Coefficient of multiple determination (R^2)--a measure of the amount of variation in the dependent variable that is beyond that accounted for by the variation in the independent variable(s). A common formula is:

$$R^2 = \frac{SSR}{SST}$$

where SSR = sum of the squares for regression, and

SST = sum of the squared deviations of the dependent variables about its mean.

R^2 multiplied by 100 gives the percentage of variation explained by the regression. It follows that a high value of R^2 is desired. The R^2 values are useful for comparing successive regression equations, but they must be used with caution when comparing substantially different equations. High R^2 values by themselves must not be taken as evidence of a good predictive equation.

f. Standard error of estimate (SEE)--The standard error of estimate (SEE) is a measure of the standard deviation of the estimated values of Y for a given X. It can provide confidence limits for the estimates. One SEE provides a 67 percent confidence limit the same as standard deviation. The standard error of estimate is also used in the detection of outliers. An outlier is an observation with an associated residual that lies three or four standard errors of estimate from the mean of the residuals. In general, the value of the SEE is computed from the following formula:

$$SEE = \frac{(\text{actual} - \text{estimated})^2}{n - k}$$

where n = the number of observations, and

k = the number of parameters in the equation (constant and coefficient(s)).

g. Coefficient of variation (V)--is a measure of the relative dispersion of the distribution. In transportation planning the term "% SEE" is used synonymously. Because it is a normalized statistic, it can be used to compare the dispersion of different frequency distributions or the relative error in different equations. This statistic is computed from the following formula:

$$V(\%SEE) = \frac{SEE}{\text{the mean of } Y}$$

h. Constant as a percent of the mean of Y--in general, the constant should be a small number related to the values of zonal trips. When the constant approaches or exceeds 25 percent of the mean of Y, its presence in the equation is too dominant and the dependent variable loses its dependence on the independent variable. The constant is further important, especially in this thesis, in that it is the only parameter in the regression equation distinguishing that equation from a rate equation.

i. Signs--of constants and coefficients should be positive (+). An equation with a negative constant, for example, will produce a negative amount of trips for an analysis unit which has zero of the independent variables. This concept of negative trips is invalid. Positive signs are not mandatory but advisable.

j. t-tests on the regression coefficients. The statistical significance of the individual explanatory variables is measured by computing standard errors and t-values for each regression coefficient.

$$t = \frac{\text{the regression coefficient for the variable}}{\text{the standard error of the regression coefficient}}$$

As already noted, a variable was rejected if its t-value was less than 1.96.

Graphical Analysis

A data point defined by Cartesian coordinates (an x-value and a y-value) can be plotted with a routine program on a line printer. This procedure is useful in several ways. First, as a check on the assumption in least-squares regression of a linear function between the dependent and independent variables. Plotting these two variables against each other can illustrate whether linearity exists. An example of this procedure is shown by Figure 5. This type of plot was made for all the variables. Second, the residual errors are the disturbance terms in a regression equation. They must always be examined to verify that the basic assumptions (as previously outlined) have not been seriously violated. The principle means of plotting the residuals (e_i) in trip generation are:

a. Overall plots. When the residuals are plotted, we obtain knowledge of their normality. In this study the residuals were input to a basic histogram program for these plots.

b. Residuals against $\hat{Y}$. When the residuals are plotted against the estimated value ($\hat{Y}$), a configuration of the residuals other than a horizontal band would indicate nonconstant variance in the residuals, error in analysis, or that the model is inadequate and possibly needs another term. These plots for the final equations will be displayed in the next chapter.

c. Residuals against the X's. This is the same as the previous plot; the appearance should be the same. Other configurations would show nonconstant variance, an error in calculations of the X's, or a

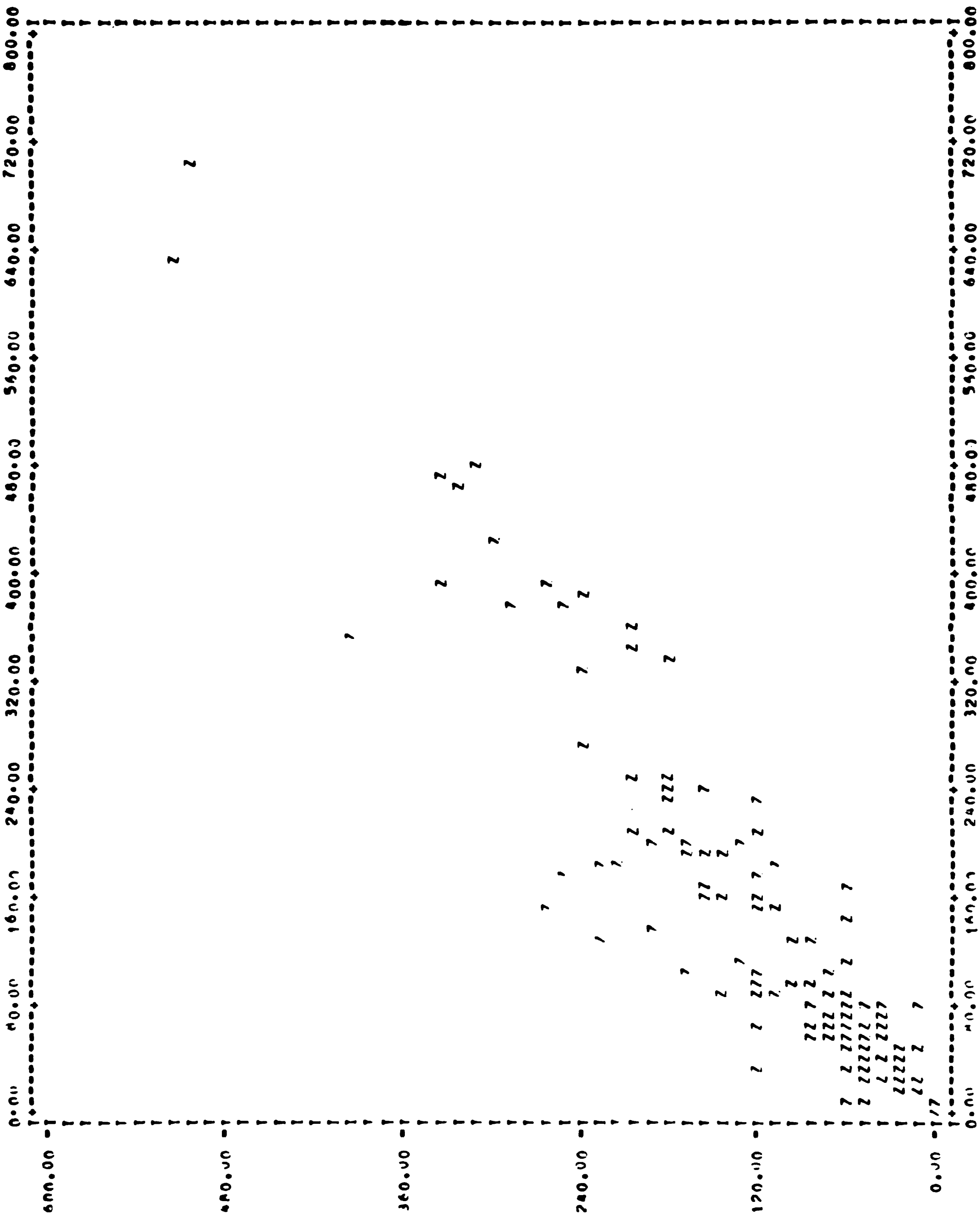


Figure 5. Linearity of the Data for Home-Based Work and Residential Labor.

need for extra terms in the equations. Figure 6 illustrates this procedure.

d. Plots against the Y. When the residuals are plotted against the actual (observed) trips, the configurations should again be a horizontal band. Anomalies which may be observed may be bias in the model from unaccounted geographic or zone size factors. Figure 7 is an example of this procedure.

In the plots a-d outliers may be detected. An outlier among residuals is one that is far greater than the rest in absolute value and perhaps lies three or four standard deviations or further from the mean of the residuals (assumedly zero). Outliers are peculiarities and need careful attention. If there is some sound reason for uniqueness, they might be removed and the equation rerun with the absence of that observation.

Subjective Analysis

The objective of trip-end modeling is to provide a reliable forecasting tool. The process of data fitting is a means towards that end, but statistically well-fitted models may not provide reliable and valid forecasts. Judgment, the subjective element, should be applied to check the forecasting implications of the model. Specific attention should be given to the following areas:

a. Reasonableness of the independent variables. Explanatory variables should be logically related to the dependent variables. Most desirable is the case where the X(s) measure(s) factors directly

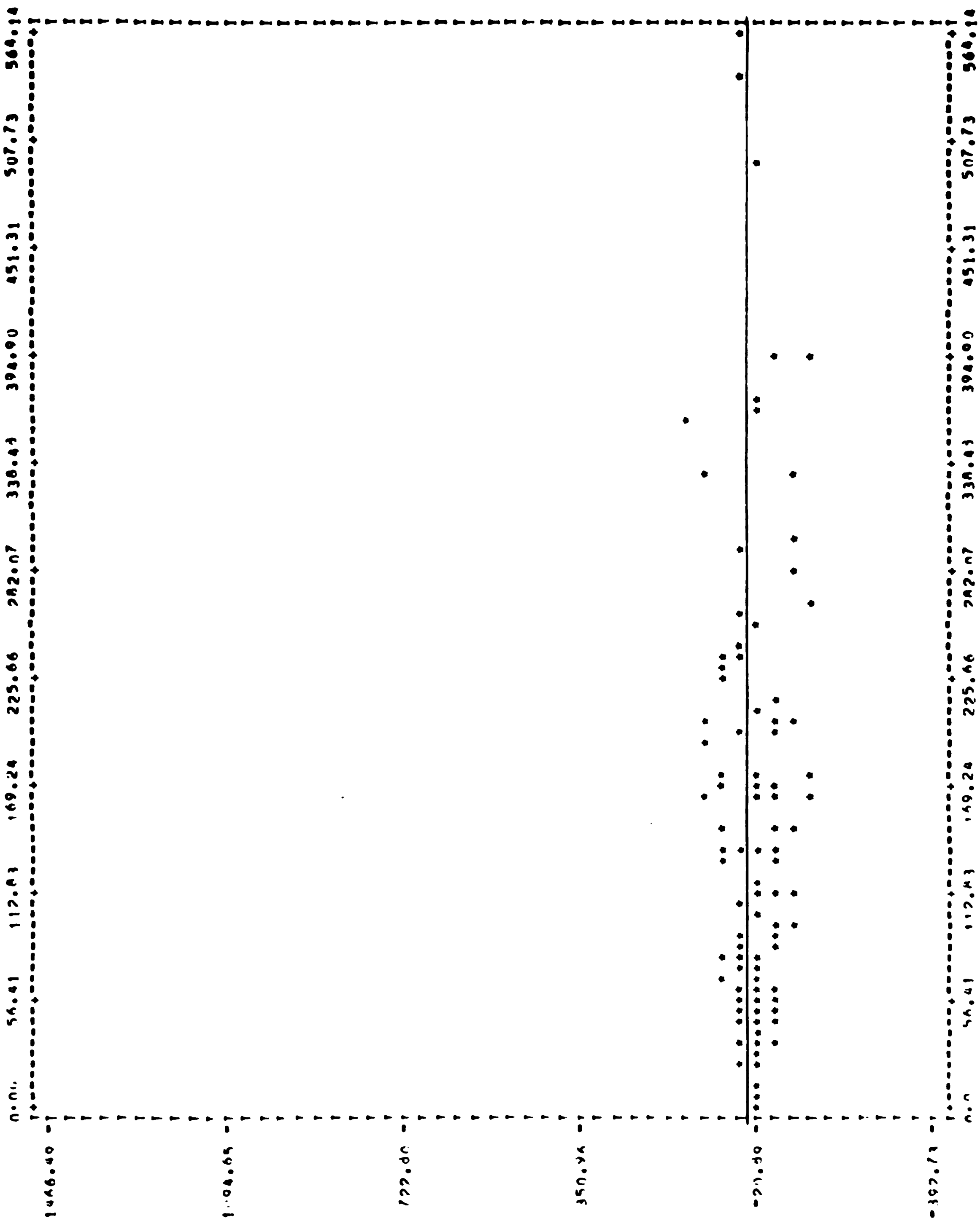


Figure 6. Residuals Versus An Independent Variable (X)

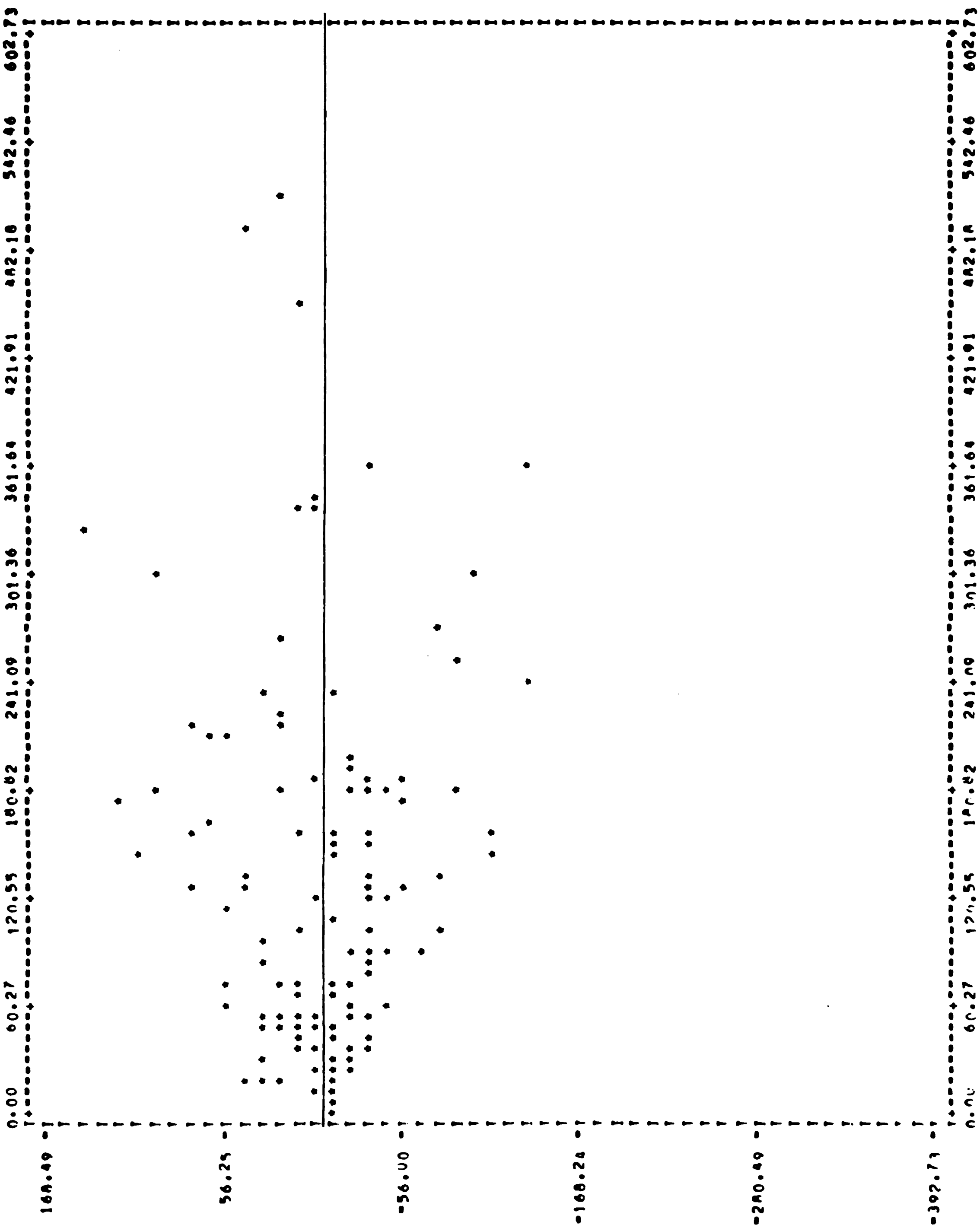


Figure 7. Residuals Versus the Observed Dependent Variable (Y)

"causing" generation of trips. Furthermore, X's should be independent of themselves as was previously mentioned. They should not be redundant measures of the same characteristic.

b. Variable selection. Variables selected must, obviously, be available for the base year of the study and must be relatively simple to keep current. Out of necessity, exacting projections of these variables should be available for future years.

c. Number of variables. The model containing the fewest number of variables is generally preferred, other things being equal.

d. Trip behavioral changes. Known or anticipated changes in trip-making behavior should be reflected in the model. For example, models for vehicle trips must reflect the rising level of vehicle ownership.

e. The range of the data. Usually it will be necessary to estimate beyond the range of data used to develop the model. A linear relationship found in present-day data may not continue in future years. For example, the linear upward trend in car ownership is likely to level off as saturation is reached.

Subjective comparisons were made at all steps of the analytic procedures which were shown in Figure 4.

CHAPTER III

STUDY RESULTS

The results of this study are presented in two parts. The first three sections of this chapter discuss the nine final equations as developed for the three trip types by three different methods. The presentation of these three sections is by analysis method: zonal multiple regression, average-rate, and dwelling-unit multiple regression. Each section, therefore, contains an equation constructed by one of the methods for each trip type: Home-Based Work productions, Home-Based Shopping productions, and Home-Based Other productions. The second part, or fourth section of this chapter, alters the perspective to observe the three equations constructed for a single trip purpose to ascertain which analysis method provided the most satisfactory results in estimating that type of trip production. Chapter IV will discuss the overall findings and conclusions of this research.

The nine final equations were developed according to the statistical, graphical, and subjective evaluation criteria as has been discussed. This chapter contains several tables showing the statistical and graphical results of these final equations as well as other tables showing the contrasted operation of these trip generation models when applied to common base data. These tables illustrate the same statistics and statistical tests as were discussed in Chapter II. The text accompanying these tables will describe the statistical and subjective tests used to analyze from three to a dozen candidate equations in the selection of the final equation in each case. In the last section, "Comparison of Equations by Analysis Method," the plots of the residuals for each of the nine equations are presented in addition to statistical evaluations.

Three Zonal Estimating Equations

The three equations developed by a regression on zonal data are quite similar in most respects. All are rather excellent equations per se as compared to the author's experience with similar modeling on other data bases. The equations,

$$\text{Home-Based Work Productions} = -11.3380 + 1.2526\ (\text{Resident Labor Force});$$

$$\text{Home-Based Shopping Productions} = 0.2182 + 0.9106\ (\text{Autos});\ \text{and}$$

$$\text{Home-Based Other Productions} = -36.4931 + 2.6642\ (\text{Autos}),$$

represent logical and statistically good estimates of trip productions. All of the constants are small; however, in two equations the constants are negative. Each equation has only one independent variable which is, as might be expected, highly correlated with the dependent variable. The coefficients of the independent variables are logical when compared considering each is for a different kind of a trip. The statistics in this section are more straightforward than in the two sections to follow.

The zonal regression equation for Home-Base Work productions ($\hat{Y}_1$) is shown in Table 4 along with the specifics of that equation. The equation says that every person living in a zone who is employed will produce 1.2526 work trips-ends and, when all these trip-ends are summed, 11.338 trips-ends must be subtracted from each zone total. The idea of subtracting trips from the zone total, particularly for a zone which has zero Resident Labor Force, has some difficulty. When such a model is actually applied to future data, the zones with a negative amount of trips will have to be set to zero trips as the trip distribution model would not recognize a negative-trips-per-zone condition. However, the constant is small; only about 8 percent of the mean of work trips, so any adverse impact should be negligible. Considering this equation produced work trip-ends the coefficient of 1.2526 work trips per employed person might seem low, but, the equation is for an "average-day" application which includes work days and nonwork days for the employed people and even on work days there is absenteeism. The independent variable seems the more logically related to work trips than other variables tested from the candidate variables which are shown in Table 1. Moreover, the statistics are better than for other trial equations. Resident Labor Force is more correlated with Work trip productions than the other trip purposes and their independent variable (which is the same in both cases--Autos). Over 87 percent of the variation in trip-making between zones is accounted for by this equation; the

Table 4. Home-Based Work Production Equation at the Zonal Level

$$\hat{Y}_1 = -11.3380 + 1.2526 \text{ (Resident Labor Force)}$$

(Y)	Dependent Variable	HB Work Productions
($\overline{Y}$)	Mean of Dependent Variable	140.4919
(s)	Standard Deviation of Dependent Variable	137.0471
(R)	Multiple Correlation Coefficient	0.9347
(R^2)	Coefficient of Multiple Determination	0.8737
(SEE)	Standard Error of Estimate	48.9112
(V)	Coefficient of Variation	0.3481
(a)	Constant Term	-11.3380
	% of Mean	8.0700
(N)	Number of Observations	124.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Resident Labor Force	1.2526	29.0460	121.2097	0.9347

Special Generator Zone	Rate
Adrian College (Zone #12)	0.573584 Trips/Employee
	0.125205 Trips/Person

most of the three zonal regression models. Conversely, the errors of estimation for this equation were lowest of the three. $\hat{Y}_1$ was constructed on the basis of 124 zones of observations as one zone was removed from the data set because plots of the residuals showed this zone to be an outlier. As can be seen from the trips per employee rate shown at the bottom of Table 4, this college zone produces work trips at less than half the rate of the other zones. The part-time nature of college employment most likely accounts for this reduced rate.

$\hat{Y}_2$, the regression equation for Home-Based Shopping trip-end productions, is given in Table 5. $\hat{Y}_2$ says that for every auto in a zone there will be 0.91 shopping trip-ends plus another 0.21 trip-ends as a zone constant. This equation has, by far, the smallest constant of the three zonal equations. The constant as a percentage of the mean is 0.17 which is negligible. The between-zone variance in trips is the lowest of the three trip types, but the variation accounted for by the equation is also the lowest of the three. The errors of estimation, however, were better than in one of the other zonal models. In other words, the equation predicted rather well and further produced no outliers.

The third zonal regression equation estimates Home-Based Other trip-end productions and is shown in detail on Table 6. This equation says there will be 2.66 vehicle trip-ends for each auto in a zone less 36.49 from the zone total. $\hat{Y}_3$ has both a higher constant value and a higher coefficient value than either of the two previous equations. This is reflected in the fact that "Other" is an aggregation of several

Table 5. Home-Based Shopping Production Equation at the Zonal Level

$$\hat{Y}_2 = 0.2182 + 0.9106 \text{ (Autos per Zone)}$$

(Y)	Dependent Variable	HB Shopping Productions
($\bar{Y}$)	Mean of Dependent Variable	127.9840
(s)	Standard Deviation of Dependent Variable	117.4640
(R)	Multiple Correlation Coefficient	0.9096
(R^2)	Coefficient of Multiple Determination	0.8273
(SEE)	Standard Error of Estimate	49.0074
(V)	Coefficient of Variation	0.3829
(a)	Constant Term	0.2182
	% of Mean	0.1700
(N)	Number of Observations	125.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Autos	0.9106	24.2770	140.3050	0.9096

Table 6. Home-Based Other Production Equation at the Zonal Level

$$\hat{Y}_3 = -36.4931 + 2.6642 \text{ (Autos per Zone)}$$

(Y)	Dependent Variable	HB Other Productions
($\overline{Y}$)	Mean of Dependent Variable	333.1935
(s)	Standard Deviation of Dependent Variable	337.1635
(R)	Multiple Correlation Coefficient	0.9207
(R^2)	Coefficient of Multiple Determination	0.8478
(SEE)	Standard Error of Estimate	132.0944
(V)	Coefficient of Variation	0.3964
(a)	Constant Term	-36.4931
	% of Mean	10.9500
(N)	Number of Observations	124.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Autos	2.6642	26.0642	138.7606	0.9207

Special Generator Zone	Rate
Adrian College (Zone #12)	1.1723 Trips/Auto
	0.320428 Trips/Person

trip types which together are a major trip component. The mean of the dependent variable (trips) is several times greater than either of the two previous trip types. $\hat{Y}_3$ also has more variance in the dependent variable than $\hat{Y}_1$ or $\hat{Y}_2$ evidenced by an (s) greater than the mean. However, this equation has the second best R^2 (85 percent of the variance). This equation also has a negative constant, larger as an absolute number and also as a percentage of the mean than $\hat{Y}_1$ or $\hat{Y}_2$, but still well below the criteria of 25 percent. The errors of estimate are the worst of the three. The independent variable is correlated very highly with the dependent variable. The same college zone as with $\hat{Y}_1$ was deleted as an outlier. Again, the special generator rate is less than half the coefficient reflecting not only a different shopping trip behavior but probably a different auto ownership rate as well in the college zone. Tests of other possible independent variables were deficient in statistical or subjective evaluation for this trip type.

Three Average-Rate Estimating Equations

Tables 7, 8, and 9 summarize the rate equations for the three trip purposes. The determination of these rates was a rather straightforward procedure. After $\hat{Y}_1$, $\hat{Y}_2$, and $\hat{Y}_3$ were determined, the same X as used in these equations was selected and the "Ratio of Averages" for these X's was calculated. This was done by dividing the sum of Y by the sum of X. For example, 24,000 trips for the whole study area divided by 12,000 people in the study area would yield an area-wide rate of 2.0 trips per person.

Table 7. Home-Based Work Production Equation Using an Average-Rate

$$\hat{Y}_4 = 1.148937 \text{ (Resident Labor Force)}$$

(Y)	Dependent Variable	HB Work Productions
($\bar{Y}$)	Mean of Dependent Variable	140.4919
(s)	Standard Deviation of Dependent Variable	137.0471
(R)	Multiple Correlation Coefficient	0.9350
(R^2)	Coefficient of Multiple Determination	0.8740
(SEE)	Standard Error of Estimate	49.8526
(V)	Coefficient of Variation	0.3548
(N)	Number of Observations	124.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Resident Labor Force	1.148937	—	121.9097	0.928

Special Generator Zone	Rate
Adrian College (Zone #12)	0.573584 Trips/Employee
	0.125205 Trips/Person

Table 8. Home-Based Shopping Production Equation Using an Average-Rate

$$\hat{Y}_5 = 0.912190 \text{ (Autos in Zone)}$$

(Y)	Dependent Variable	HB Shopping Productions
($\bar{Y}$)	Mean of Dependent Variable	127.9840
(s)	Standard Deviation of Dependent Variable	117.4640
(R)	Multiple Correlation Coefficient	0.9100
(R^2)	Coefficient of Multiple Determination	0.8280
(SEE)	Standard Error of Estimate	48.8096
(V)	Coefficient of Variation	0.3817
(N)	Number of Observations	125.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Auto	0.912190		140.3050	0.910

Table 9. Home-Based Other Production Equation Using an Average-Rate

$$\hat{Y}_6 = 2.377979 \text{ (Autos in Zone)}$$

(Y)	Dependent Variable	HB Other Productions
($\bar{Y}$)	Mean of Dependent Variable	333.1935
(s)	Standard Deviation of Dependent Variable	337.1635
(R)	Multiple Correlation Coefficient	0.9210
(R^2)	Coefficient of Multiple Determination	0.8480
(SEE)	Standard Error of Estimate	135.7181
(V)	Coefficient of Variation	0.4073
(N)	Number of Observations	124.0000
	Unit of Analysis	Zone

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
Autos	2.377979	--	138.7606	0.913

Special Generator Zone

Adrian College (Zone #12)	1.1723	Trips/Auto
	0.320428	Trips/Person

These equations were not developed by a computer program. There is no constant term in the equation, only a regression coefficient which is the rate. These coefficients will be similar to those of the zonal regression equations assuming the "a" terms (constants) are small.

All these equations had at least 82 percent of the variation accounted for. The SEE in each case is relatively insignificant. Verbally, Work productions will be 1.115 trips for every person employed (excluding zone #12); ($\hat{Y}_5$)--Shopping productions will be .91 trips for every auto registered; and ($\hat{Y}_6$)--Other productions will amount to 2.38 trips for every auto owned except in zone #12.

Three Dwelling-Unit Estimating Equations

The summary statistics for the Dwelling-Unit level regression analysis are presented in Tables 10, 11, and 12. The Home-Based Work equation ($\hat{Y}_7$) has six independent variables, the Home-Based Shopping equation ($\hat{Y}_8$) has five independent variables, and the Home-Based Other equation ($\hat{Y}_9$) has seven independent variables. The three equations have positive, small constants. The constants as a percentage of the mean of Y are large, in fact larger than the means in all cases. This is due in part to the dummy variable technique.

The next discernable item on these tables is the drastic reduction in magnitude of the figures when compared with the two previous sets of tables. This is obviously true, as this method deals with individual households and not with aggregated areas of possibly hundreds of households. The R^2 as given in Table 4 is

Table 10. Home-Based Work Production Equation at the Dwelling-Unit Level

$$\hat{Y}_7 = 1.78684 - 0.56062\ (V40) + 0.33802\ (V41) + 1.37441\ (V42) - 1.21129\ (V47) + 0.60640\ (V48) + 0.84798\ (V49)$$

(Y)	Dependent Variable	HB Work Productions
($\bar{Y}$)	Mean of Dependent Variable	1.6677
(s)	Standard Deviation of Dependent Variable	1.6395
(R)	Multiple Correlation Coefficient	0.42095
(R^2)	Coefficient of Multiple Determination	0.17719
(SEE)	Standard Error of Estimate	1.4895
(V)	Coefficient of Variation	0.8931
(a)	Constant Term	1.78684
	% of Mean	107.1400
(N)	Number of Observations	1,932.0000
	Unit of Analysis	Dwelling Unit

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
1 Car (V40)	-0.56062	-7.53236	0.5652	-0.285
3 Cars (V41)	0.33802	1.96546	0.0487	0.144
4+ Cars (V42)	1.37441	4.21972	0.0114	0.126
0 RLF (V47)	-1.21129	-8.95248	0.0719	-0.265
2 RLF (V48)	0.60640	7.84152	0.3199	0.212
3+ RLF (V49)	0.84798	5.63073	0.0694	0.183

Special Generator Zone	Rate
Adrian College (Zone #12)	0.573584 Trips/Employee
	0.125205 Trips/Person

Table 11. Home-Based Shopping Production Equation at the Dwelling-Unit Level

$$\hat{Y}_8 = 1.21368 - 0.33439\ (V43) - 0.24508\ (V44) - 0.20707\ (V46) + 0.23904\ (V50) + 0.27959\ (V51)$$

(Y)	Dependent Variable	HB Shopping Productions
($\bar{Y}$)	Mean of Dependent Variable	1.14650
(s)	Standard Deviation of Dependent Variable	1.36910
(R)	Multiple Correlation Coefficient	0.12712
(R^2)	Coefficient of Multiple Determination	0.01616
(SEE)	Standard Error of Estimate	1.35980
(V)	Coefficient of Variation	1.18600
(a)	Constant Term	1.21368
	% of Mean	1.05860
(N)	Number of Observations	1,932.00000
	Unit of Analysis	Dwelling Unit

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
1 Person @ Addr. (V43)	-0.33439	-2.49656	0.0585	-0.048
2 Persons @ Addr. (V44)	-0.24508	-3.45145	0.2666	-0.066
1 Year @ Addr. (V46)	-0.20707	-2.38677	0.1501	-0.051
Income (V50) $4,000-$4,999	0.23904	2.40589	0.1108	0.042
Income (V51) +$15,000+	0.27959	2.42764	0.0797	0.056

Table 12. Home-Based Other Production Equation at the Dwelling-Unit Level

$$\hat{Y}_9 = 2.23579 - 0.36465\ (V40) + 3.03453\ (V42) + 0.56451\ (V45) + 0.064691\ (V49) - 0.71222\ (V50) + 0.89926\ (V51) + 1.05110\ (V52)$$

(Y)	Dependent Variables	HB Other Productions
($\bar{Y}$)	Mean of Dependent Variables	2.63920
(s)	Standard Deviation of Dependent Variable	2.86550
(R)	Multiple Correlation Coefficient	0.34566
(R^2)	Coefficient of Multiple Determination	0.11948
(SEE)	Standard Error of Estimate	2.69370
(V)	Coefficient of Variation	1.02070
(a)	Constant Term	2.23579
	% of Mean	0.84710
(N)	Number of Observations	1,932.00000
	Unit of Analysis	Dwelling Unit

Independent Variable	Regression Coefficient	t Value	Mean Value	Simple Correlation With Y (r)
1 Car (V40)	0.36465	-2.74182	0.5652	-0.171
4+ Cars (V42)	3.03453	5.14037	0.0114	0.165
5+ Persons @ Addr. (V45)	0.56451	4.06526	0.2955	0.165
3+ RLF (V49)	0.64691	2.53591	0.0694	0.155
Income (V50) $4,000-$4,999	-0.71222	-3.59998	0.1108	0.164
Income (V51) $15,000+	0.89926	3.81485	0.0797	0.164
FLC 3+ 5-18 (V52)	1.05110	7.75756	0.2754	0.246

Special Generator Zone

Adrian College (Zone #12)	1.172300 Trips/Auto
	0.320428 Trips/Person

accounting for between-zone variance. The R^2 figure in the present tables account for between-household variance. Studies by the Federal Highway Administration have shown that an R^2 of 36 percent on a household level equation jumps to over 95 percent on a zonal basis. Therefore, no direct comparisons between statistics of the different methods can be made. The next section will present a procedure and an analysis for making comparisons.

The simple correlations of the X variables with the Y variables are much lower. There are a great many more independent variables in each of these equations and the t-values are also lower. All this is indicative of the fact that each X contributes less to the estimate. Furthermore, at least one variable in each equation is negative and the inadvisability of such coefficients has been discussed.

The independent variables were chosen by allowing the regression program to input all the variables shown on Table 2 until the t-value of the new variable was outside the ± 1.96 range. The result of this procedure in the logic of some of the variables as a contributor to trip-making is questionable.

Comparison of Equations by Analysis Method

At this point it has been concluded that the nine equations are all reasonably valid. It is then necessary to enter the second phase of evaluations. This will be done by statistical comparisons and graphical comparisons. The nine equations have been regrouped by the type of trip they are estimating. This regrouping appears in

Tables 13, 14, and 15. Table 13 contains summary statistics for $\hat{Y}_1$, the equation for estimating Home-Based Work trips from the zonal multiple regression method; and $\hat{Y}_4$, the equation which was derived from the average-rate method for estimating Home-Based Work trips; and $\hat{Y}_7$, the Home-Based Work trip estimator from the dwelling-unit multiple regression method. Tables 14 and 15 each similarly contain the summary statistics from the three analysis methods for Home-Based Shopping trips and Home-Based Other trips, respectively. In addition, these tables illustrate the appropriate statistics for the residuals of each of these equations, given as $Y_w - \hat{Y}_1$, for example. The basic statistics for the observed trips (Y_w) are also given.

The first column in Table 13, "Variables," lists the dependent variables being considered. The column, "SUMX," gives the total for all the observations (zones) of the variables. For the $\hat{Y}$'s, it should be a close approximation of Y_w SUMX. $\hat{Y}_1$ is one number off. After calculating an estimation for each of the 124 zones and then summing them, $\hat{Y}_1$ was one trip short someplace. This should have provided a residual ($Y_w - \hat{Y}_1$) of one. The table shows zero which is due to rounding off by the computer. Due to the way in which the least-squares technique operates the residuals will always be zero for this method. That condition is not true, however, for the rates method. Here the model ($\hat{Y}_4$) underpredicted by 152 trips (17421-17268) and a residual sum of 152 is shown. The dwelling-unit model ($\hat{Y}_7$), on the other hand, over-estimated by 2,776 trips. Being a least-squares regression model, the residuals from this model should also be zero.

Table 13. Comparison of the Estimating Methods for Home-Based Work Productions

Variable	SUMX	$SUMX^2$	Mean	Standard Deviation	SEE	% SEE	R	R^2
HB Work (Y_w)	17421		140.49	137.05				
$\hat{Y}_1$	17420		140.49	128.10			0.935	0.874
$Y_w - \hat{Y}_1$	0	291859	0.01		48.71	34.67		
$\hat{Y}_4$	17268		139.26	117.49			0.935	0.874
$Y_w - \hat{Y}_4$	152	305876	1.23		49.85	35.48		
$\hat{Y}_7$	20197		162.88	134.19			0.925	0.856
$Y_w - \hat{Y}_7$	-2776	401978	-22.39		52.56	37.41		

Table 14. Comparison of the Estimating Equations for Home-Based Shopping Productions

Variable	SUMX	$SUMX^2$	Mean	Standard Deviation	SEE	% SEE	R	R^2
HB Shop (Y_s)	15998		127.98	117.46				
$\hat{Y}_2$	15997		127.98	106.84			0.910	0.828
$Y_s - \hat{Y}_2$	0	295411	0.004		48.81	38.14		
$\hat{Y}_5$	15998		127.98	107.03			0.910	0.828
$Y_s - \hat{Y}_5$	0	295414	0		48.81	38.14		
$\hat{Y}_8$	15218		121.75	105.34			0.847	0.717
$Y_s - \hat{Y}_8$	779	491322	6.24		62.63	48.94		

Table 15. Comparison of the Estimating Equations for Home-Based Other Productions

Variable	SUMX	$SUMX^2$	Mean	Standard Deviation	SEE	% SEE	R	R^2
HB Other (Y_o)	41316		333.19	337.16				
$\hat{Y}_3$	41315		333.19	310.44			0.921	0.848
$Y_o - \hat{Y}_3$	0	2128777	0.001		131.56	39.48		
$\hat{Y}_6$	40916		329.97	277.09			0.921	0.848
$Y_o - \hat{Y}_6$	399	2266874	3.223		135.72	40.73		
$\hat{Y}_9$	34642		279.37	231.32			0.889	0.790
$Y_o - \hat{Y}_9$	6673	3876277	53.82		169.10	50.75		

However, because the model was not run (for predictions) on the same data set that it was built upon, a residual sum other than zero appeared. This appears to be a rather significant error compared to $\hat{Y}_1$ and $\hat{Y}_4$. From this primary comparison, it could be assumed that $\hat{Y}_1$ is doing the better estimating job. This assumption, however, could be completely erroneous. That the sum of the errors (which are positive and negative numbers) is lower does not necessarily mean that the model is doing the better estimating job. Consider Model A of 10 observations and Model B also of 10. If five of the 10 observations for A were +10 in estimation and five were -10, the total error would be zero. If another rate model (b) was off by only +1 in six observations and -1 in four, it would have a total error of +2. With A at 0 and B at +2, the most consistent and best predictor is not Model A. This is a simplified example of the present circumstance. The need for more tests is obvious.

If the residuals are squared, two things occur. One, the signs all become positive and direction is not a consideration. Two, the larger residuals are given disproportionately greater sums. The column headed $SUMX^2$ shows this operation. The smallest sum of squares for residuals would suggest the better estimator.

The over-predicting nature of $\hat{Y}_7$ is again evident in the "Mean" column of Table 13 as the mean of $\hat{Y}_7$ is considerably greater than Y_w. The slight underpredicting of $\hat{Y}_4$ is apparent also. The means of the residuals being significantly different than zero indicate a normal distribution of the error terms. Given this observation, there is a possibility that $\hat{Y}_7$ does not conform to the basic assumptions of regression analysis as presented earlier.

The standard deviations measure the dispersions of the distributions of trip values. A standard deviation of 128.10 for $\hat{Y}_1$ as opposed to 137.05 for Y_w would indicate that the estimates are bunched around the mean (the mean being the same) tighter in the estimated distribution than in the observed distribution. The distribution for the rate method ($\hat{Y}_4$) is even more restricted.

It is the errors in these predictions that are most interesting. The dispersion of the errors is measured by the standard error of estimate (SEE). The objective is to have as little dispersion as possible. In other words, here it is desirable to have the distribution bunched up around the mean. The mean is, theoretically, a condition of zero error and larger SEE's would indicate larger errors. The % SEE is a normalized SEE that allows various equations to be directly compared. Again, the smallest number is desirable. In both these statistics, $\hat{Y}_1$ is superior while $\hat{Y}_7$ is the inferior.

The final statistic for comparison is R^2. This too provides an indicator of how two equations compare as it measures the amount of variation in the data that are being accounted for by the model. The zonal model and the rate model do equally as well, accounting for 87.4 percent each. The dwelling-unit model is nearly 2 percent lower. It is interesting to compare R^2 in Table 13 with R^2 in Table 10 for $\hat{Y}_7$ operating on the dwelling unit data set. On the earlier table, the model accounts for 17.71 percent of the variation within zones. When the same model is applied to the zonal level, the R^2 increases to 85.67 percent (Table 13).

Table 14 portrays the pertinent statistics for the Home-Based Shopping productions. Area-wide, $\hat{Y}_2$ and $\hat{Y}_5$ predicted the same number of trips as were observed with no residual sums. On the other hand, $\hat{Y}_8$ under-predicted. The means of all the dependent variables are nearly the same with $\hat{Y}_2$ and $\hat{Y}_5$ exactly the same. The dispersions of the three estimated distributions were nearly identical. All, however, were more constricted than the dispersion of the observed data.

Oddly enough, for these equations, the statistics for $\hat{Y}_2$ and $\hat{Y}_5$ were identical throughout. The difference between them and $\hat{Y}_8$ becomes very obvious when considering the two important statistics % SEE and R^2. The dispersion of the residuals of $\hat{Y}_8$ was 10 points higher than the other two. At the same time, the zonal and rate models accounted for 11 percent more of the variation. Generally, the errors in the Home-Based Shopping models were slightly greater than the Home-Based Work models as less of the variance in the data was accounted for by the former.

The Home-Based Other models are illustrated in Table 15. Observe a value of SUMX of 41,316 for Y_o. The rate model slightly under-predicted while the dwelling-unit model was extremely under-predictive (SUMX for $\hat{Y}_9$ residuals is 6,673). The sum of squares of residuals is lowest for $\hat{Y}_3$, a little greater for $\hat{Y}_6$, and much greater for $\hat{Y}_9$. The means again reflect the predictive qualities. $\hat{Y}_3$ is significantly close to the observed, $\hat{Y}_6$ is slightly under, and $\hat{Y}_9$ is considerably under. The mean of the residuals is negligible for zonal regression, shows a moderate amount of skewness for rates, and is significantly skewed for dwelling-unit regression. A value of 53.82 indicates a normal distribution of the errors. The distribution of the estimated trips ranges from fairly constricted ($\hat{Y}_3$) to considerably constricted ($\hat{Y}_9$).

After normalizing the errors of estimate, it can be seen that $\hat{Y}_3$ and $\hat{Y}_6$ are nearly equal at around 40, while $\hat{Y}_9$ is again 10 points higher. There is clearly more variation in the dwelling-unit estimates. Finally, $\hat{Y}_3$ and $\hat{Y}_6$ account for nearly 85 percent of the variation while $\hat{Y}_9$ accounts for 6 percent less.

The analysis of these models from a graphic perspective is now indicated. Certain peculiarities are not apparent from examination of the statistics alone. If these equations, or some outputs of them can be viewed as discrete data points rather than numbers, new knowledge can be gleaned. Several residual plots can be made, as previously discussed. The residuals plotted against the dependent variables are illustrated herein and comments are directed towards them. Figures 8 through 16 are these nine plots for the nine equations, $\hat{Y}_1$ to $\hat{Y}_9$.

Each of these plots ideally should be a band of data points equidistant about the zero line laterally across the plot. The vertical distances between the data points and the zero line represent the magnitude of the errors of estimation (residuals).

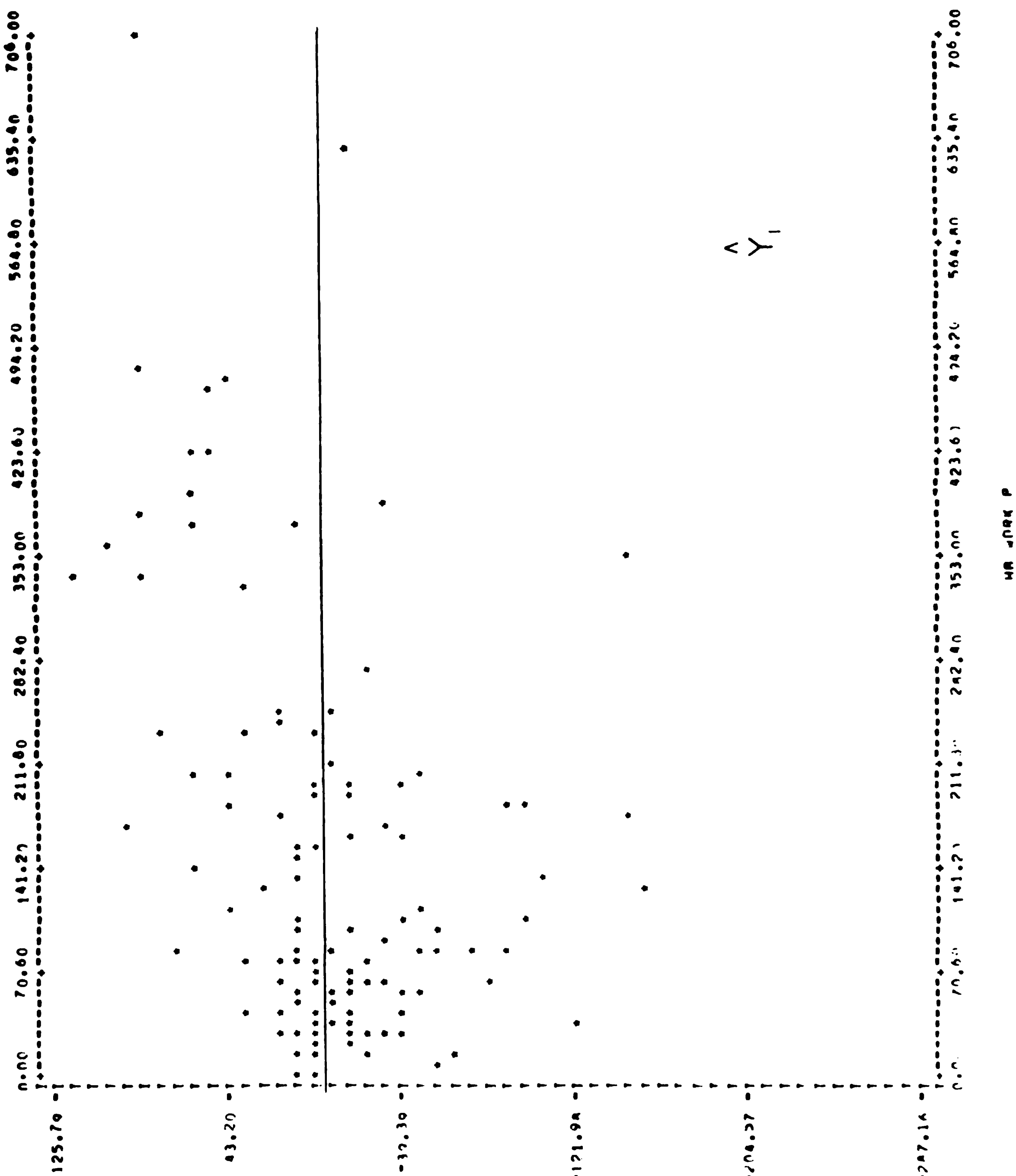


Figure 8. Zonal Residuals Versus Home-Based Work Productions.

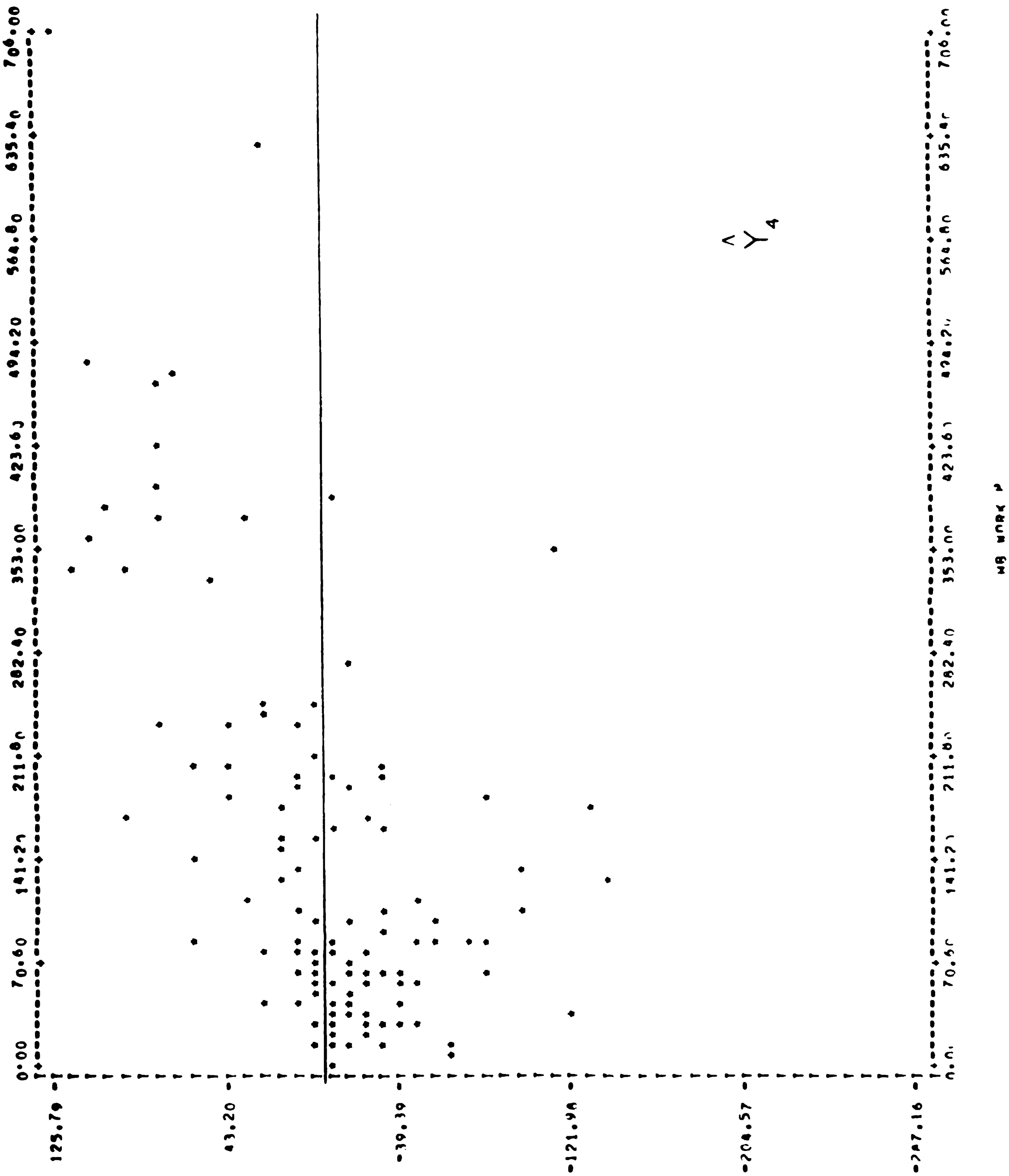


Figure 9. Rate Residuals Versus Home-Based Work Productions.

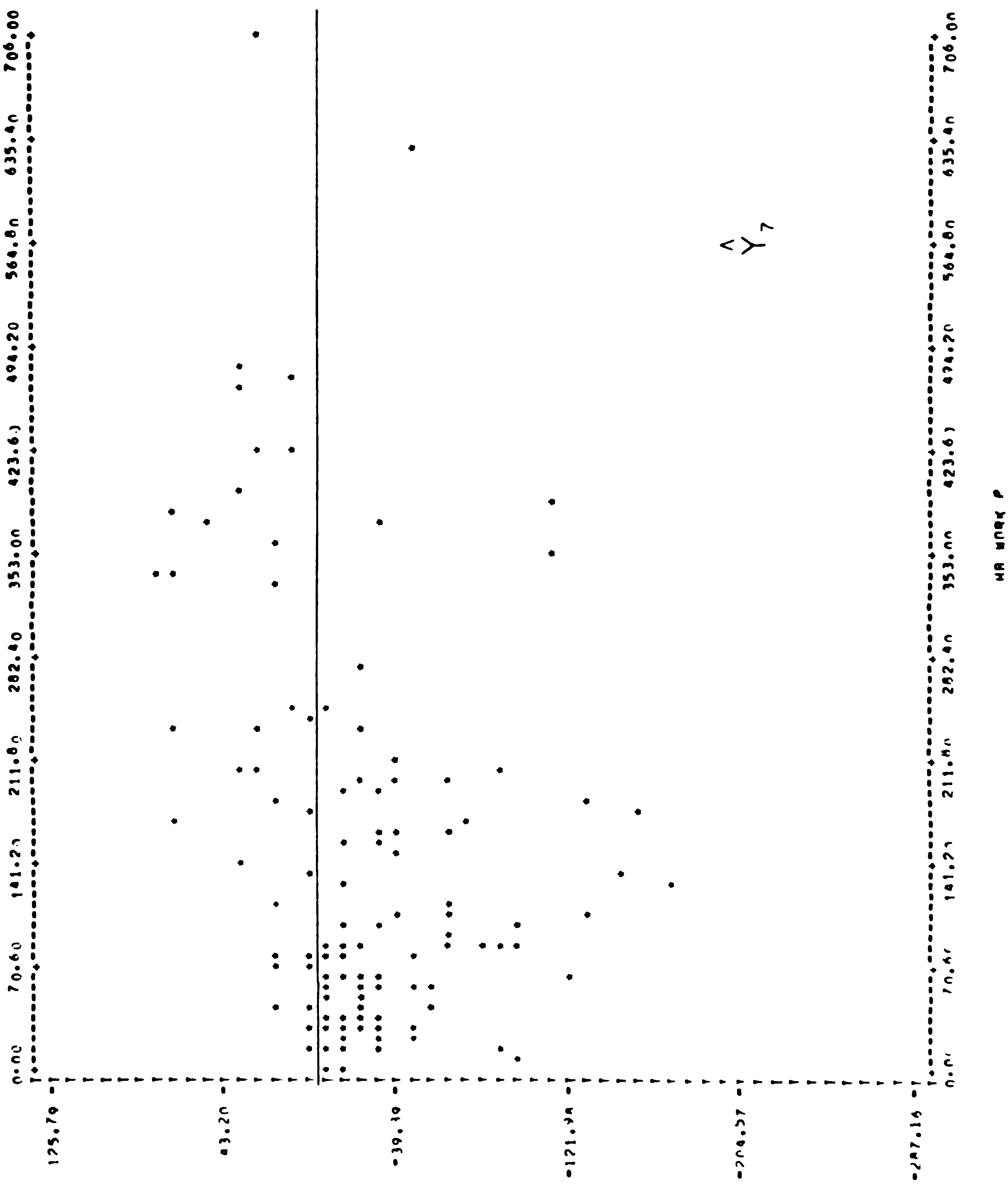


Figure 10. Dwelling-Unit Residuals Versus Home-Based Work Productions.

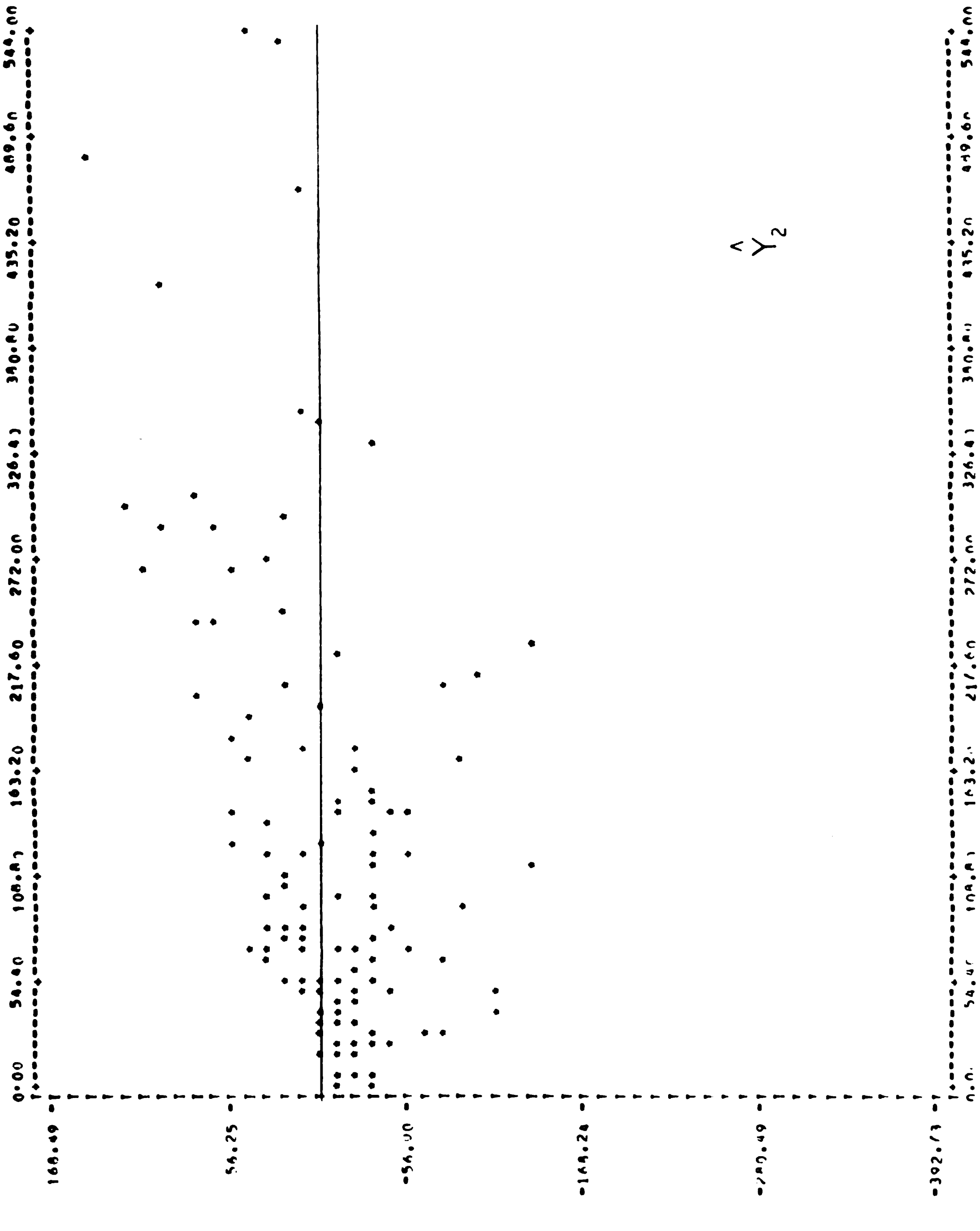


Figure 11. Zonal Residuals Versus Home-Based Shopping Productions.

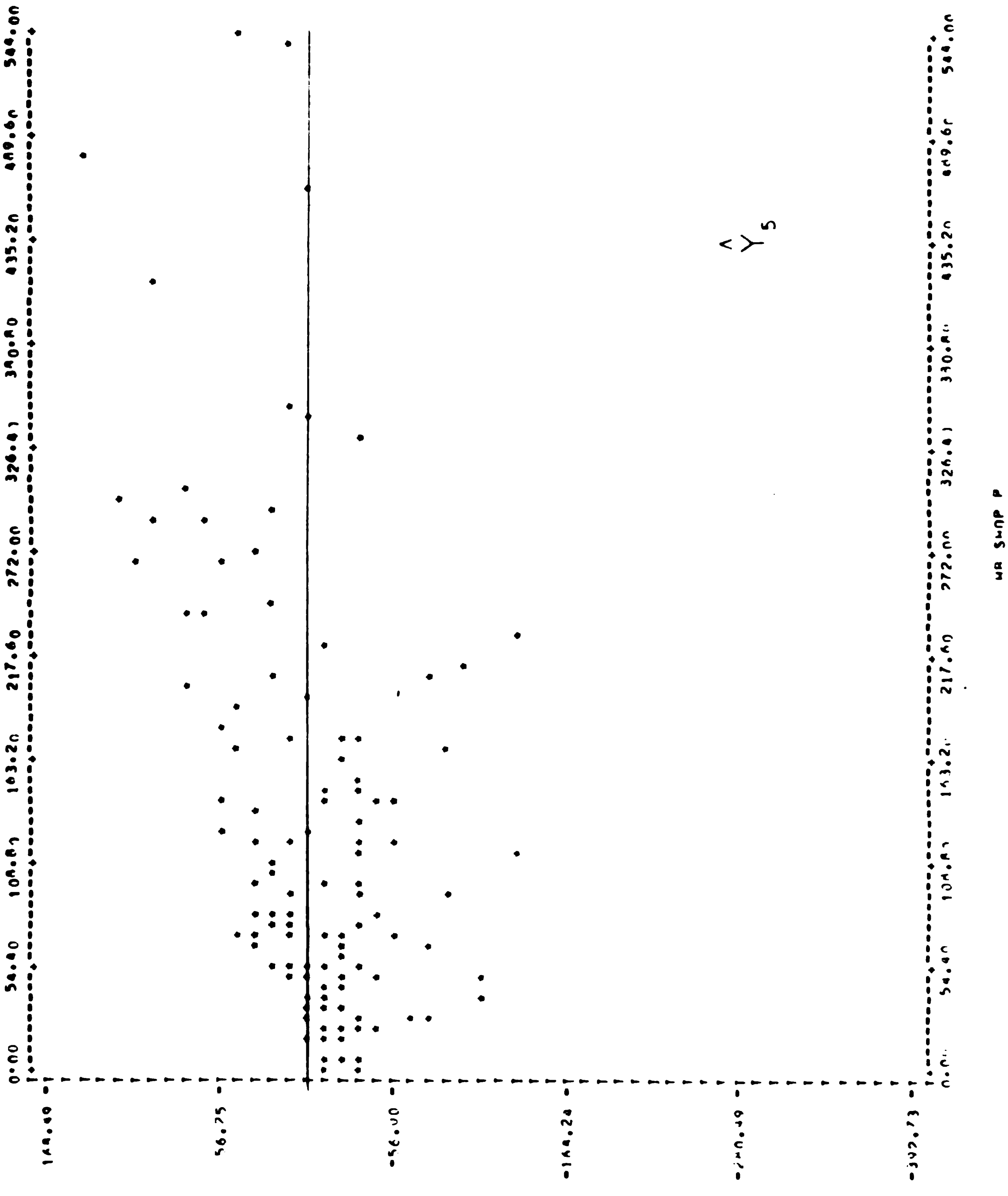


Figure 12. Rate Residuals Versus Home-Based Shopping Productions.

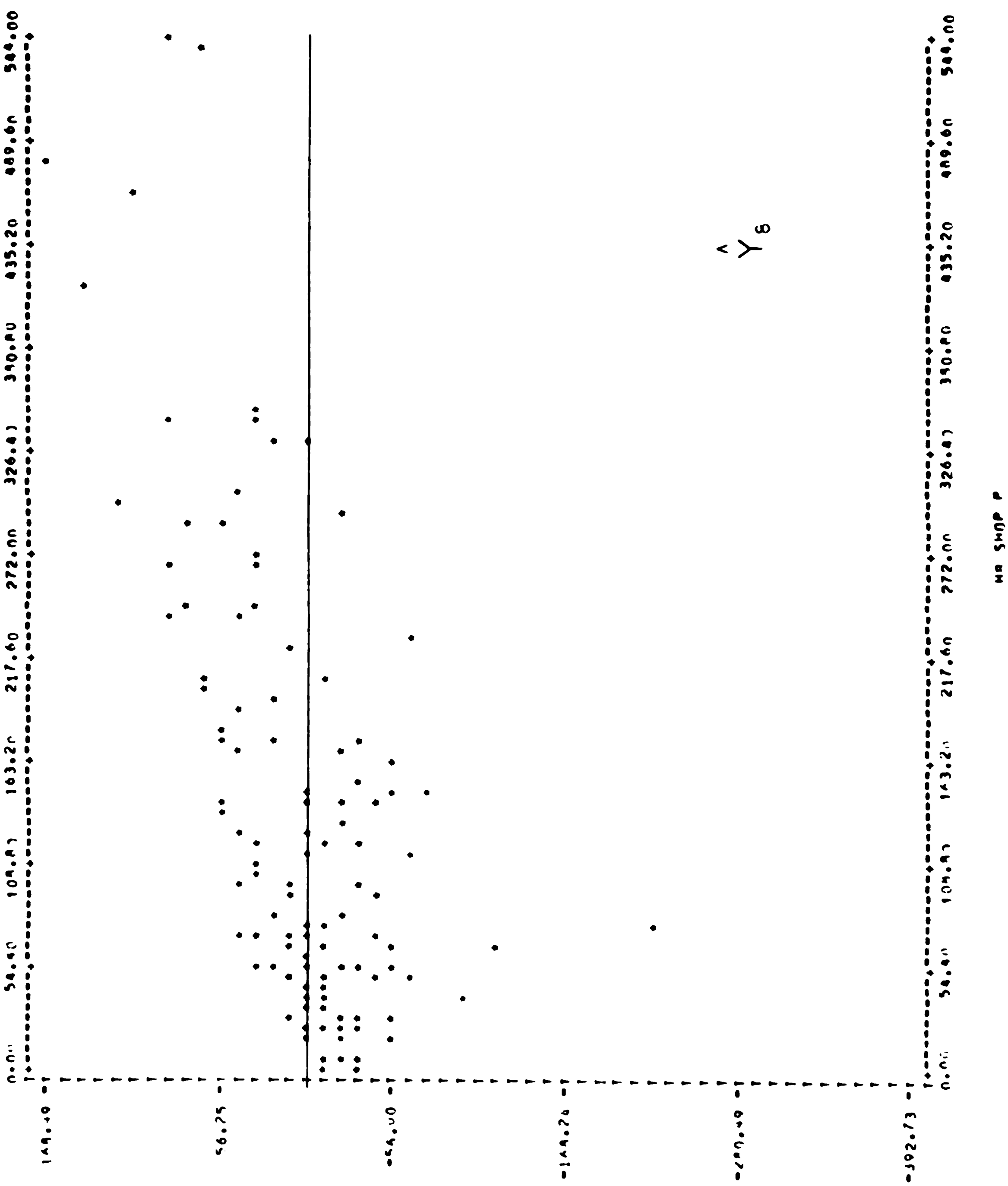


Figure 13. Dwelling-Unit Residuals Versus Home-Based Shopping Productions.

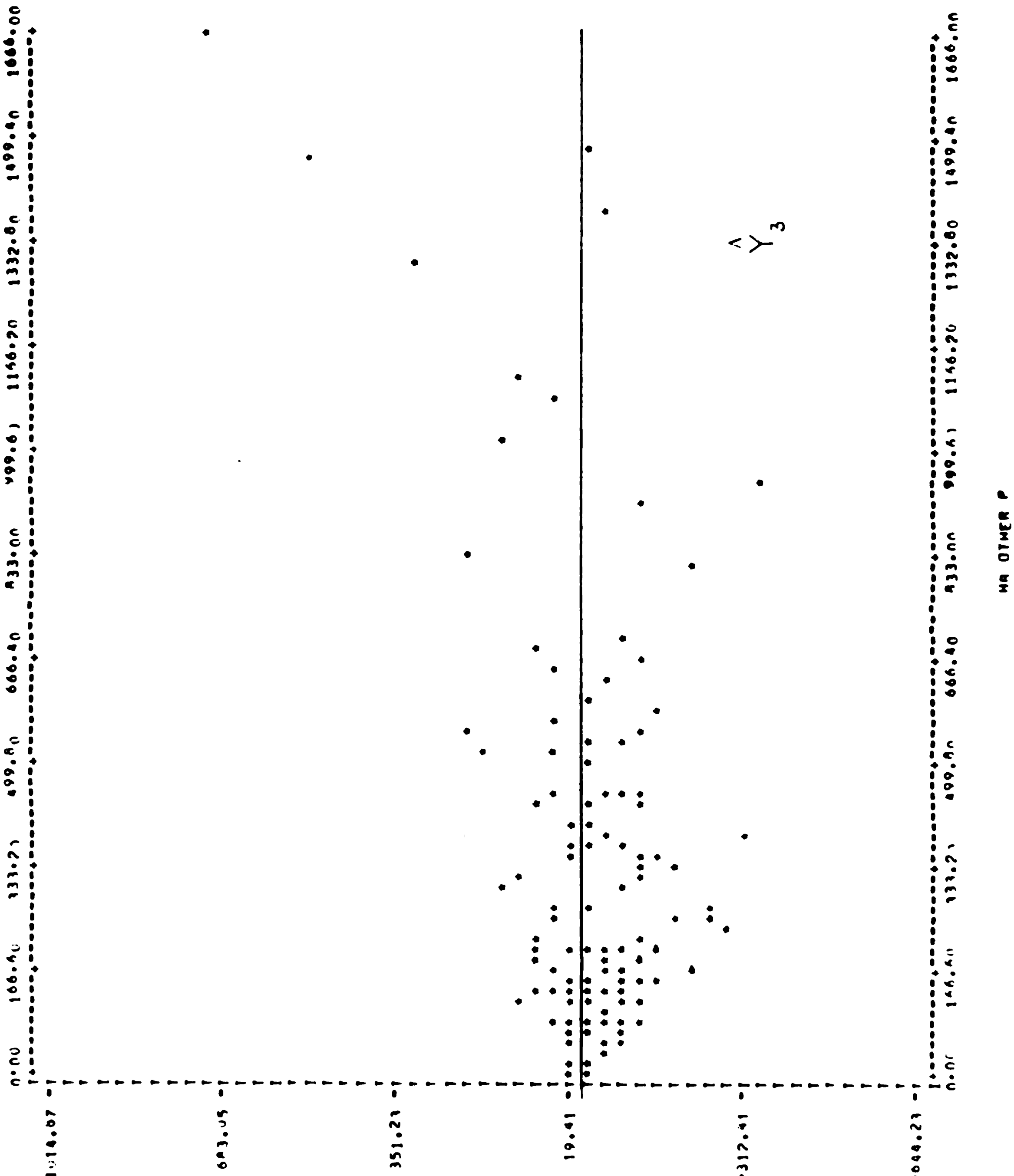


Figure 14. Zonal Residuals Versus Home-Based Other Productions.

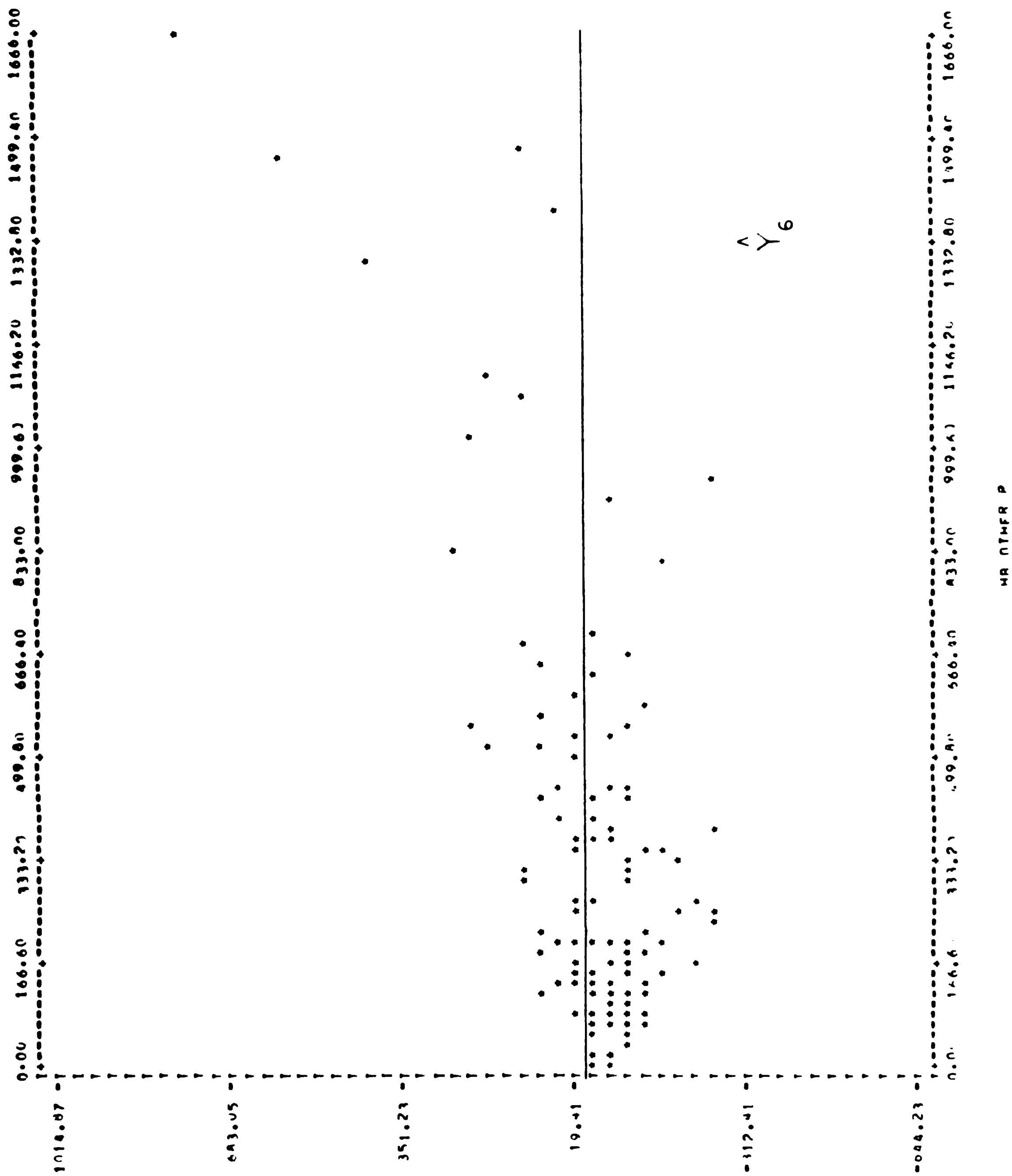


Figure 15. Rate Residuals Versus Home-Based Other Productions.

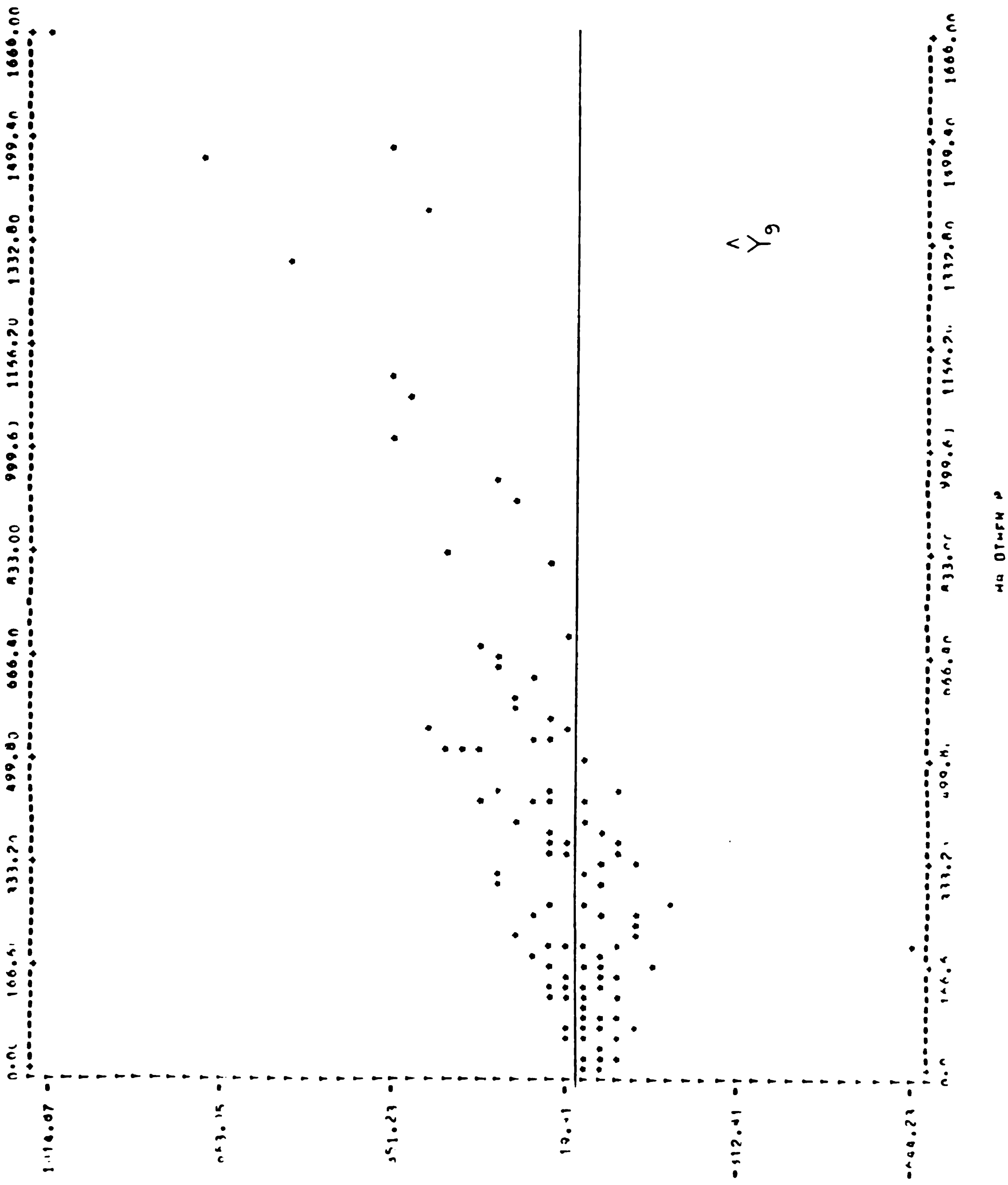


Figure 16. Dwelling-Unit Residuals Versus Home-Based Other Productions.

The narrower this band, the closer all the errors were to zero or no error. The horizontal values are the various values of the observed numbers of trips per zone. Thus, if the band rises or falls, there is some bias to the model. If there were no errors whatsoever, the plot would be a horizontal line at zero.

The plots are arranged as were the previous tables, by trip purpose. The scales within each trip purpose (each three successive plots) are the same but they are not the same between trip purposes.

Comparing Figures 8, 9, and 10 does not reveal any significant differences. All of the models appear rather similar and normal. Comparing Figures 11, 12, and 13 shows the same general normality. All are relatively narrow bands. There may be slightly more bias towards under-predicting in the $\hat{Y}_8$ model for larger zones. Comparing Figures 14, 15, and 16 does show some dissimilarities. The plots of $\hat{Y}_3$ and $\hat{Y}_6$ are quite normal, however, $\hat{Y}_9$ shows a definite bias towards extreme under-prediction of larger zones. This bias is probably pronounced enough to reject this model as invalid and either reconstruct the equation or use another model altogether.

CHAPTER IV

STUDY CONCLUSIONS

This study built trip generation models for three basic types of home-based trip productions--Home-Based Work, Home-Based Shop, and Home-Based Other--by three different methods. The zonal multiple regression analysis method used data which had been aggregated to traffic zones and then used zonal totals of independent variables to construct a predicting equation. The average-rate method used study area rates to predict for each zone. The dwelling-unit multiple regression analysis built equations from unaggregated data of individual households and then applied these equations to the zones, using the appropriate figures, to obtain zonal estimates.

The intent of the study was to determine which of these methods would be the most valid and the most appropriate for predicting each of the three types of trips. In general, the thesis statement that there are significant differences between application of various constructed trip generation models is true. The method producing the most satisfactory results is the regression on zonal data. However, for reasons discussed in this chapter, the author is reticent to proclaim zonal regression as <u>the</u> method of building trip generation models.

The design of this research study could be improved in several ways. There is a possibility that the results may have been different by using data from a larger study area or one with more typical epicenter characteristics. Testing the cross-classification method along with the three methods analyzed would have provided more insight into trip generation methodology. Including cross-classification would have necessitated the development of some statistical testing procedures. Testing other forms of rates, using transformations of independent variables, and regressing on zonal averages versus zonal totals may have produced more differences. Dividing zones by predominant zone type before development of equations could have produced different models. Some studies have used a locational variable such as distance from the CBD. This may have been beneficial. A detailed study of the residuals, zone by zone, to determine the need for new variables would make interesting research. Performing some of these additional tests could provide more valid and reliable models in any future research. On the other hand, any future "fine-tuning" may not produce any better estimates. Another conclusion of this research is that, although statistical test outcomes can be improved, the question most relative is whether a superior trip generation model is being produced.

Modeling an urban area transportation system is a complex and arduous task. Because the transportation planning process is a series of separately applied models, there is lacking a clear understanding of the concept of compounding the errors in these models. For many years research efforts have attempted to create a truly synthetic model that would combine generation, distribution, mode split, and assignment into one set of algorithms. Until that is accomplished, the errors within each model and their contribution to total error may elude the transportation planner. Trip generation is affected by land use, transportation system, and socioeconomic variables as was illustrated in Figure 2. Yet, the concept of this three-dimensional model is rarely used in trip-end modeling. System variables are almost never used as independent variables in trip generation equations. Transportation planners are perplexed that new trip generation equations have to be built in every study area. The differences in zonal accessibilities may account for this.

In summary, this thesis demonstrates the superiority of regression on zonal data when consideration is made that using a disaggregate model means future forecasts of independent variables must be extremely detailed. For instance, the planner is relatively more comfortable projecting autos per zone than the number of households which will have an income level of $15,000-21,000 in 1995. But, this thesis also demonstrates that since models are by definition inexact replicas of some real world phenomena, looking for more accuracy is perhaps superfluous. A statewide (all zones in all studies), three dimensional trip generation model appears to have good indications for future success of trip generation modeling endeavors. This type of model would have characteristics similar to regression, rate, and cross-classification techniques. It would be a model of universal application. The accuracy at the unit of analysis may be inferior to today's models, but this disadvantage would be more than equalled by an increase in reliability and ease of application.

APPENDICES

APPENDIX A

GLOSSARY OF TERMS

APPENDIX A

GLOSSARY OF TERMS

Auto-driver trips: An auto-driver trip is equivalent to an auto trip, because of the obvious fact that a traveling auto has one driver. The term is useful in determining auto trips because person travel can conveniently be classified as auto-driver trips, auto-passenger trips, transit-passenger trips, etc.

Correlation: A mutual or reciprocal relation between variables.

Destination: Terminal end of a trip or the zone in which a trip terminates.

Dwelling unit: A dwelling unit, or DU is defined by the U.S. Bureau of the Census as a house, an apartment, a room, or a group of rooms occupied or intended for occupancy as separate living quarters by a family or other groups of persons living together or by a person living alone.

External cordon or cordon line: This is an imaginary line defining the boundary of the study area. Interviews of drivers may be conducted at stations along this line in order to sample the trips with one or both ends outside the study area.

Home-based auto-driver work trip: This is a trip in either direction between the auto-driver's residence and his place of employment. Herein it is usually referred to simply as a work trip.

Home-based other auto-driver trip. This is a trip in either direction between the auto driver's residence and any place other than his place of employment. This type of trip is sometimes called a Home-Based Non-Work Trip. Herein it is usually referred to simply as an HB Other trip.

Household: A household is defined by the U.S. Bureau of the Census as an occupied dwelling unit. The term "DU" is used in this research to mean either dwelling unit or household.

Interzonal trip: A trip with its origin and destination in different zones.

Intrazonal trip: A trip with both its origin and destination in the same zone.

Origin: The beginning end of a trip or the zone in which a trip begins.

Record: A data processing term for a piece of information on cards, tape, or stored in computer memory.

Standard error of estimate: A statistical measure of the difference as found by the least squares method--within which one would expect to find 68 percent of the cases.

Study areas: The area delimited for the purpose of data collection by a transportation study. This area contains the central city and surroundings which will become urbanized in 20 to 30 years and is the area for which forecasts of travel are made.

Trip: A person or vehicle movement which begins at the origin at the start time, and ends at the destination at the arrival time and is conducted for a specific purpose.

Zone: A subdivision of the study or survey area which is useful in analysis or data collection.

Zone of production and zone of attraction: A home-based auto-driver trip is said to be produced by the zone of the auto driver's residence, regardless of whether the residence is the origin or destination. The other zone involved in the trip is said to have attracted that trip. A nonhome-based auto-driver trip is said to be produced by the zone of its origin and attracted by the zone of destination. Example: A person lives in Zone A and drives to work in Zone B in the morning. Zone A has produced one home-based auto-driver work trip and Zone B has attracted one. That evening he drives back to his home. Again Zone A has produced one home-based auto-driver work trip and Zone B has attracted one.

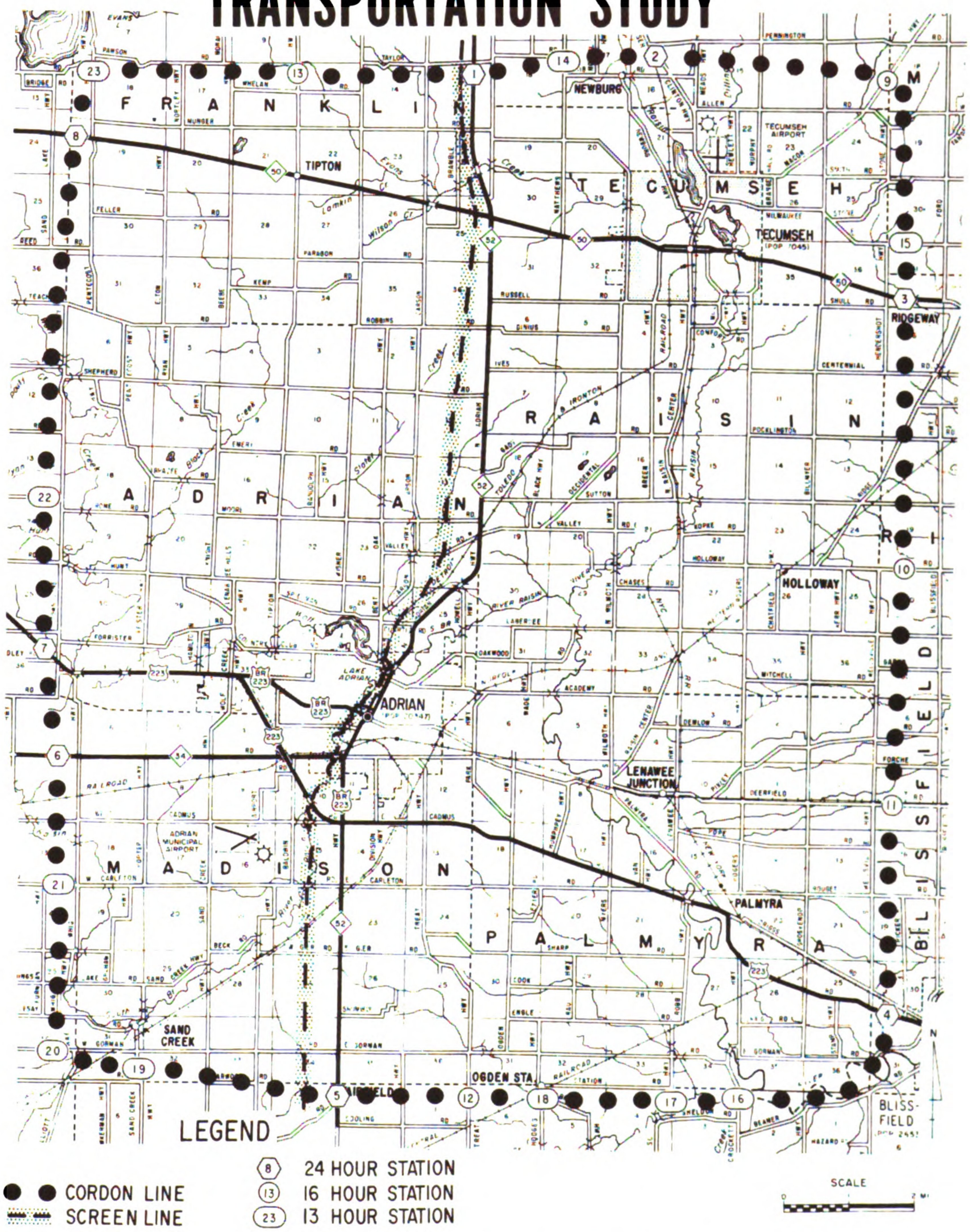
APPENDIX B

STUDY AREA MAP

APPENDIX B

STUDY AREA MAP

ADRIAN-TECUMSEH AREA TRANSPORTATION STUDY



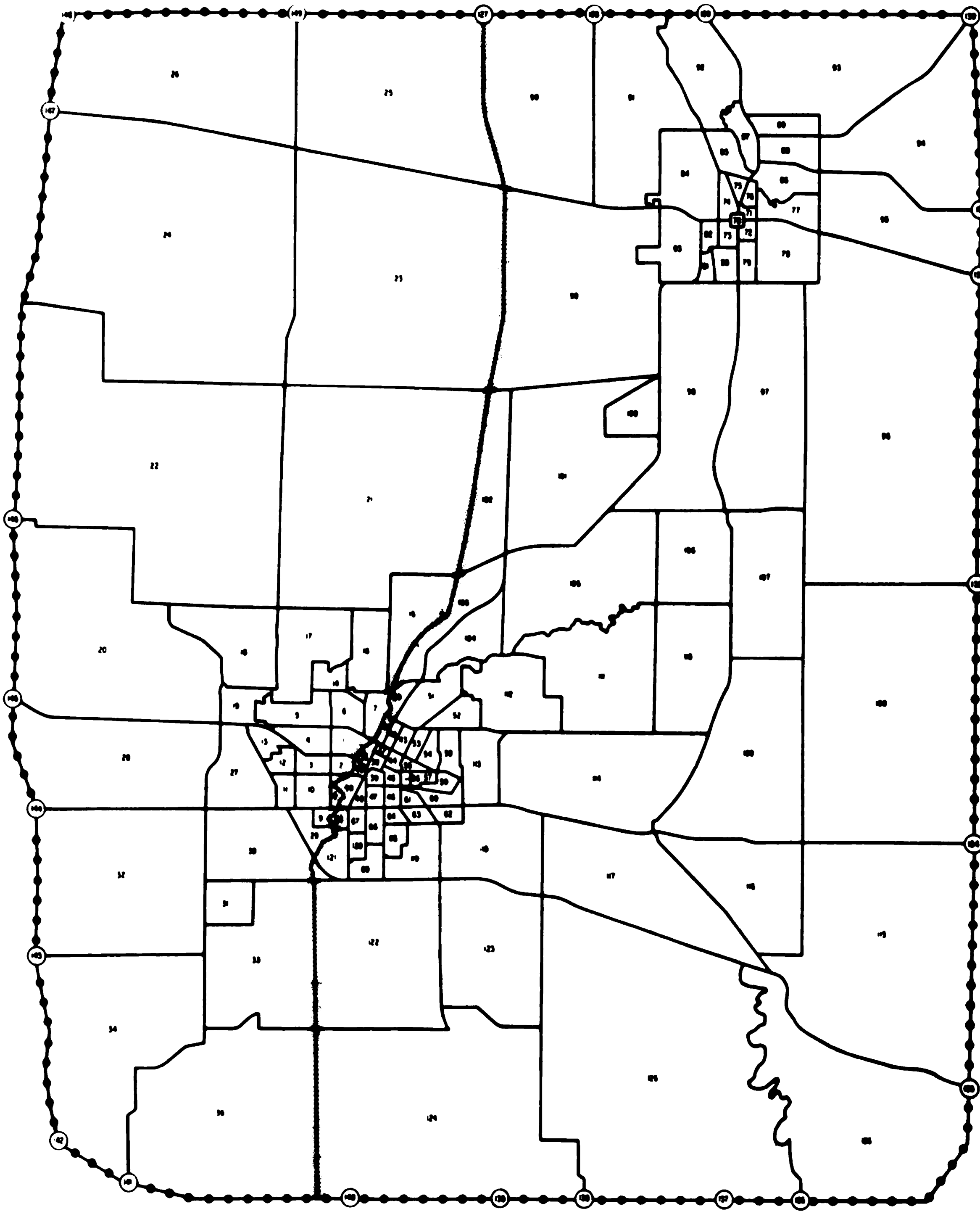
MICHIGAN DEPARTMENT OF STATE HIGHWAYS

APPENDIX C

ZONE BOUNDARY MAP

APPENDIX C

ZONE BOUNDARY MAP



BIBLIOGRAPHY

BIBLIOGRAPHY

Corridino, Joseph C. "The Effect of the Highway System and Land Development on Trip Production." Traffic Engineering, June 1968, pp. 32-39.

Deutschman, Harold D. "Establishing a Statistical Criterion for Selecting Trip Generation Procedures." Highway Research Record No. 191, Highway Research Board, Washington, D.C., 1967, pp. 39-52.

Douglas, A. A., and R. J. Lewis. "Trip Generation Techniques: Introduction." Traffic Engineering and Control 12 (November 1970): 362-365.

________. "Trip Generation Techniques: Zonal Least-Squares Regression Analysis." Traffic Engineering and Control 12 (December 1970): 428-431.

________. "Trip Generation Techniques: Household Least-Squares Regression Analysis." Traffic Engineering and Control 12 (January 1971): 477-479.

________. "Trip Generation Techniques: Category Analysis and Summary of Trip Generation Techniques." Traffic Engineering and Control 12 (February 1971): 532-535.

Fleet, Christopher R., and Sydney R. Robertson. "Trip Generation in the Transportation Planning Process." Highway Research Record No. 240, Highway Research Board, Washington, D.C., 1968, pp. 11-29.

Goodknight, John C. A Partial Analysis of Trip Generation. Research Report 60-12, Texas Transportation Institute, 1968.

Harmelink, M. D., G. C. Harper and H. M. Edwards. "Trip Production and Attraction Characteristics in Small Cities." Highway Research Record No. 205, Highway Research Board, Washington, D.C., 1967, pp. 1-19.

Jeffries, Wilbur R., and Everett C. Carter. "Simplified Techniques for Developing Transportation Plans: Trip Generation in Small Urban Areas." Highway Research Record No. 240, Highway Research Board, 1968, pp. 66-87.

Kassoff, Harold, and Harold D. Deutschman. "Trip Generation: A Critical Appraisal." Highway Research Record No. 297, Highway Research Board, Washington, D.C., 1969, pp. 15-30.

Kolifrath, Michael, and Paul Shuldiner. "Covariance Analysis of Manufacturing Trip Generation." Highway Research Record No. 165, Highway Research Board, Washington, D.C., 1967, pp. 117-128.

Maricopa Association of Governments. Trip Generation by Land Use, Part I: A Summary of Studies Conducted. Urban Area of Maricopa County, Arizona, April 1974.

Martin, Brian V., Frederick W. Memmott, III, and Alexander J. Bone. "Principles and Techniques of Predicting Future Demand for Urban Area Transportation." Massachusetts Institute of Technology Report No. 3. Cambridge: Massachusetts Institute of Technology Press, 1961, pp. 36-60.

McCarthy, Gerald M. "Multiple Regression Analysis of Household Trip Generation--A Critique." Highway Research Record No. 297, Highway Research Board, Washington, D.C., 1969, pp. 31-43.

Meyerowitz, Wayne. Trip Generation Procedural Manual. Lansing, Mich.: Michigan Department of State Highways, 1971.

Parsonson, Peter S., and Paul D. Cribbins. "Estimating Urban Trip Production and Attraction." Journal of Highway Division, November 1967, pp. 115-135.

Parsonson, Peter S., and J. W. Horn. "Comparison of Techniques for Estimating Zonal Trip Productions and Attractions." Highway Research Program, School of Engineering, North Carolina State University, June 1966.

Shuldiner, Paul W. "Land Use, Activity and Non-Residential Trip Generation." Highway Research Record No. 141, Highway Research Board, Washington, D.C., 1966, pp. 73-87.

Shuldiner, Paul W., and Walter Y. Oi. An Analysis of Urban Travel Demands. Evanston, Ill.: Northwestern University Press, 1962.

U.S. Department of Transportation. Guidelines for Trip Generation Analysis. Federal Highway Administration, Washington, D.C., 1967.

Walker, John R. "Rank Classification: A Procedure for Determining Future Trip Ends." Highway Research Record No. 240, Highway Research Board, Washington, D.C., 1968, pp. 88-99.

MICHIGAN STATE UNIV. LIBRARIES
31293010088155