THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN AGGREGATE ACCESS SUPPLY MODEL

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

NEIL BARRY HILSEN

Norman, Oklahoma

1973

AN AGGREGATE ACCESS

SUPPLY MODEL

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the individuals who directly or indirectly contributed to the various phases of this research. Professor Antti Talvitie, my thesis advisor, made an invaluable personal contribution and is deserving of special recognition and thanks for his expert guidance and constant encouragement.

The initial opportunity to enter a doctoral program and the continued support and advice received from Professor George Reid also warrants a special note of thanks. Professors Ed Crim and Bill Southard, the other principle members of my doctoral committee, provided very useful ideas and suggestions.

I appreciate the help of Professors Chong Liew and Thomas Wilbanks for serving on my committee on such short notice during Professors Crim and Southard's absence from campus for academic year 1973-1974.

I would also like to thank my colleagues at the University of Oklahoma. Valuable discussions with Peter Liou in particular have helped to clarify many aspects of the research.

Finally, I would like to extend my warmest thanks to my wife, Ruth, our sons, Scott and Mitchell, and our parents whose moral support was always there.

ABSTRACT

AN AGGREGATE ACCESS SUPPLY MODEL

In this study supply models are estimated for the access portions of rail and bus trips. The models are designed to predict aggregate zonal travel times as a function of transportation system, zone size, and volume related characteristics of a zone. Trip segments not requiring geographic aggregation such as waiting, transferring, and linehaul time are not evaluated.

Three models are estimated that deal with a rail trip. These are access walking time, access driving time, and access riding time in bus. The walking time to a bus stop is modelled for the bus trip. Corresponding models are developed for the within zone variance or standard deviation of the access times. A specific model for the drive access trip to a bus stop, could not be formulated due to a lack of observations in the trip data.

The basic objective of these models is to provide an input to the existing travel demand analysis to improve their accuracy and help reduce the bias currently present in travel estimates.

The data for the empirical estimation of the

models comes from a 1969 survey conducted by the Southward Transit Area Coordination Committee in Chicago. The method of ordinary least squares regression is employed for estimating the coefficients in the models. The independent variables describing the zonal characteristics constitute such things as zone size, number of rail stations per zone, average travel distance, and population density.

The predictive accuracy of the final models is evaluated in terms of standard indices of forecasting accuracy. The results show that the estimation of these types of access models can be produced with reasonable accuracy. The coefficients of determination (R^2) for the walk, drive, and bus mean models are very high and the standard errors are relatively low. However, the walking time to bus stop model has a low R^2 value due to a lack of variation in the data. Also, the values of other error indicators in this model as well as the standard deviation models indicate that the forecasts might be subject to some uncertainty.

It remains to be seen whether these supply models can improve the forecasting ability of existing travel demand model analysis. It is therefore recommended that this model structure be applied to travel demand model analysis as well as related areas.

TABLE OF CONTENTS

Chapter page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The analysis of transportation systems involves dealing with some basic principles. One of these principles, developed by Manheim[1], says that "a transportation system is a particular form of market in which supply and demand reach equilibrium within the constraining channels of the transportation network". In the economic market the theory behind supply and demand is given as functions of price only. Whereas, in transportation we are dealing with a vector of interaction variables. These variables are characteristics of the transportation service provided, and are called level of service variables. Level of service, designated L, can be expressed in terms of travel time, trip costs, safety, comfort, convenience, and other characteristics. However, travel time and cost have traditionally played the most important roles in assessing the level of transportation service. One reason is that factors such as convenience, reliability, and schedule delays are directly related to

---

[1]Manheim, M. L., Principles of Transport Systems Analysis, Professional paper P67-1 (Cambridge: Massachusetts Institute of Technology, 1967), Vol. XVI.

travel time while safety, comfort, and aesthetics, although conceptually important, have not been often observed and are very difficult to express in quantitative terms.

In order to give some insight to the importance of predicting level of service variables, the core of the transportation analysis problem will be briefly explained. This consists of the prediction of network flows.

As discussed in another paper by Manheim[1], and presented here, the problem of predicting network flows can be expressed in terms of the following variables:

T = Specification of transportation system options in terms of technology, networks, link characteristics, vehicles, and operating policies.

A = Specification of the system's activity

F = The pattern of system flows, which is defined as the volumes and service levels (L) that actually will occur for a given T and A.

L = Level of service characteristics.

---

[1] Manheim, M. L., Ruiter, E. R., and Bhatt, K. V., "Search and Choice in Transport Systems Analysis", <u>Highway Research Record 293</u> (Washington, D.C.: Highway Research Board, 1968), pp. 54-78.

V = Volume of flows.

Supply functions indicate the following:

$L = S(T, V)$

The level of service (L), of a transportation system is a function (supply function S), of the transportation options (T), and the volume of flow (V).

Demand functions are shown as:

$V = D(A, L)$

The volume of flow (V) demanding transportation is a function (D), of the activity system options (A), and the level of service (L).

The flow pattern (F) consists of the origin and destination volumes, (and their traveled paths) and the level of service experienced by the travelers.

For a particular setting of the options T and A, the flow pattern which will actually occur is the equilibrium solution to the supply and demand functions.

It can be seen that the key concept behind this formulation is the level of service vector of variables. As Manheim says, "These service variables both characterize the transportation system and serve as the basis for the demands for transportation".

So far very little attention has been paid to developing supply models. Attention has been almost exclusively directed to the relationship between the attenuation of vehicular volume and travel time over a distance. This has been largely due to the requirements of the modeling system employed in travel demand analysis.

The objective of this study is to focus attention on the supply models and on the various components of which the level of service vector is composed. Specifically models will be developed for aggregate access supply models for both rail and bus transit trips. These will be relatively simple models that estimate the mean and variance of a traffic zone's level of service variables. This would provide better inputs to existing direct demand models[1], for the purpose of forecasting equilibrium urban passenger travel.

---

[1]Direct or explicit demand models predict the demand for trips from an origin zone to a destination zone for a specific mode, and for a specific trip purpose. The indirect or sequential approach that has been taken in analyzing travel demand, separates the problem into trip generation, attraction, distribution, assignment, and modal split models. Explicit demand models can be used in a "direct" approach to computing equilibrium, while sequential models require the use of an "indirect" approach.

CHAPTER II

AN AGGREGATE ACCESS SUPPLY MODEL

Background: Current Methods of Supply Model Estimation

Capacity-Restraint Functions

The approach usually taken in the analysis of forecasting urban travel involves four separate and sequential steps: trip generation, trip distribution, mode split, and trip assignment. This is known as the urban transportation model system (UTMS). The trip assignment phase of the process consists of assigning generated traffic to links in a network. Normally this involves the selection of minimum travel time paths between zones.

The supply models that have been most recently utilized to characterize the links for trip assignment highway planning are known as capacity-restraint functions or curves. These functions simulate a reduction in speed when traffic volume exceeds capacity. Therefore, for a link having an assigned volume greater than its practical capacity, the travel time is increased making it less desirable in route selection. Conversely, for a link having an assigned volume less than its capacity, the travel time is decreased to make it more desirable.

Four capacity-restraint methods are discussed here: Bureau of Public Roads (BPR); Schneider; Smock; and the Irwin, Dodd and Von Cube. The procedure for all of these techniques involves first selecting minimum time paths between network nodes and assigning trips to these routes. As these trips are "loaded", some may exceed the capacity of the facility and new minimum paths must be computed using adjusted travel times. These adjustments are made according to predetermined relationships.

The BPR Method[1]. The relationship between travel time and volume for each network link is as follows:

$$T_A = T_O \ (1.0 + .15(V/C)^4)$$

where: T_A = adjusted travel time

T_O = travel time at practical capacity multiplied by .87

---

[1]U.S. Department of Commerce, Bureau of Public Roads, Traffic Assignment Manual (Washington, D.C. Government Printing Office, June, 1964).

V = assigned volume

C = practical capacity[1]

Travel times are adjusted after each cycle of iteration.

The procedure is iterated until a balanced network is obtained.

The Smock Method[2] is quite similar to the BPR approach. The function is:

$$T_A = T_o e^{(V/C-1)}$$

where: T_o = original travel time or travel time on a link when V = C.

The Schneider Method[3] performs only one iteration of the network but adjusts the travel times for each zone separately as it is randomly added to the network. The times are adjusted as follows:

---

[1]Practical capacity is defined as the maximum number of vehicles that can pass over a given section of a lane or roadway during a given time period under prevailing traffic and roadway conditions. It is the maximum rate of flow that has a reasonable expectation of occurring.

[2]Smock, R. B., "A Comparative Description of a Capacity-Restrained Traffic Assignment", Highway Research Record 6 (Washington, D.C.: Highway Research Board, 1963), pp. 12-41.

[3]Huber, M. J., Boutwell, H. B., and Witheford, D. K., "Comparative Analysis of Traffic Assignment Techniques with Actual Highway Use", Highway Research Record (Washington, D.C.: Highway Research Board, 1968), pp. 18-23.

$$T_A = T_O(2)^{(V/C-1)}$$

where: T_O = travel time at free flow conditions.

The network times are constructed only once. Travel times are adjusted after loading each zone and when finished no further travel time adjustments are made.

In the Irwin, Dodd and Von Cube[1] approach to capacity restraint there is a family of curves representing various speeds and types of vehicles. This technique contains a feedback mechanism which allows the adjusted travel times to affect alternate route generation, trip distribution, and trip assignment phases.

These supply functions were developed basically for planning the linehaul facilities of urban automobile networks. They are not appropriate for transit trips. Neither are they useful for explicit demand models. For explicit demand models and especially for transit models, it becomes extremely important to disaggregate the service components and estimate all elements of door-to-door travel time and cost.

---

[1]Irwin, N., Dodd, N., and Von Cube, H. G., "Capacity Restraint in Assignment Programs", Highway Research Board Bulletin 297 (Washington, D.C.: Highway Research Board, 1961), pp. 109-127.

However, at present, techniques do not exist that properly predict the times and costs to and from transit stations, for a zone. In order to do this we have to not only model the travel time on access links but also be able to obtain an average value of the access time (or cost) for a zone. This average value is needed for forecasting purposes or for the estimation of aggregate travel demand models.

Existing Access Supply Methodology

Rassam and Ellis[1] have developed a framework for describing the access phenomenon and for estimating access characteristics for intercity person movements. The system analyzes common carrier terminals, such as rail and air, for large area districts in the northeastern metropolitan corridor of the United States. The access characteristics or travel impedances considered are time, cost, and distance. Only the problem of estimating the impedances from an origin (to a destina-

---

[1]Rassam, P., and Ellis, R., Access Characteristics Estimation System, Report to Department of Transportation Office of High Speed Ground Transportation, Northeast Corridor Transportation Project (Washington, D.C.: December, 1969), Vol. I.

tion) to (from) a terminal is considered.

In order to arrive at average access characteristics for each district, the computerized process includes the following:

1. Distribution of travel demand among subdistricts
2. Estimation of access impedances
3. Access mode mix weighting
4. Allocation of demand among competing terminals
5. Aggregation of subdistrict impedances into district impedances

These functions will be briefly described.

Distribution of Travel Demand Among Subdistricts. The districts in the study were divided into subdistricts. This geographic disaggregation was necessary to obtain an accurate estimate of district-to-terminal impedances. A weighted average is obtained for each district. The weights that are used should be the percentage of district travel that originates in each subdistrict. Therefore, it was assumed that the relative travel demand for a given subdistrict, is equal to the percentage of the district's residential population located in that subdistrict. The weights are designated P_j (j=the subdistricts).

Estimation of Access Impedances. There are two impedance models, one for urban districts and the other for rural districts. These models are highway oriented and predict access characteristics for private automobiles. It is assumed that the impedances of other highway submodes (taxi, bus, or limousine) are linear functions of the predicted automobile values.

A centroid is associated with each subdistrict. In the rural impedance model, the distance between the terminal and a subdistrict centroid is estimated by the following linear function:

$$d_{jk} = A_i \cdot d^*_{jk} + B_i$$

where:

i = a given rural district

j = a subdistrict of i

k = a terminal

d_{jk} = road distance between k and the centroid of j

d^*_{jk} = air distance between k and the centroid of j

A_i, B_i = parameters calibrated for district i

The travel time between a terminal k and a subdistrict j is:

$$t_{jk} = \frac{d_{jk}}{V_i}$$

V_i is the average travel speed for district i.

The access cost is estimated by the relation:

$$C_{jk} = C(V_i) . d_{jk}$$

where:

$C(V_i)$ = the unit cost of travel at speed V_i.

The development of impedances for the urban model is similar to these relationships, but more complex due to the fact that speed is assumed to be variable within urban districts and constant within rural districts. Also, distances are computed from a minimum time path composed of radial and/or circumferential routes.

Access Mode Mix Weighting. In order to obtain a single estimate of impedance time, cost, and distance, from a subdistrict to a terminal, the weighted sum of the characteristics was determined for each access mode, (i.e., each access mode has a value for d_{jk}, t_{jk}, and C_{jk}). The weights are the proportion of intercity passengers from the subdistrict, who use a given access mode.

Competing Terminals Model. If there is more than one terminal serving a district, it becomes necessary to weigh the access impedances for each subdistrict terminal pair. These weights, designated W_{jk}, were developed on the hypothesis that the proportion of intercity travelers using each competing terminal is a function of the travel time to the terminals from each subdistrict and the intercity transportation service provided by the terminals. The resulting impedances are given by:

$$d_j = \sum_k W_{jk} \cdot d_{jk}$$
$$t_j = \sum_k W_{jk} \cdot t_{jk}$$
$$c_j = \sum_k W_{jk} \cdot c_{jk}$$

Aggregation Procedures. Now, to arrive at a single district impedance vector, it is necessary to weigh these subdistrict impedances by the population weights developed in step 1. This is indicated by the following:

$$d_i = \sum_j P_j \cdot d_j$$
$$t_i = \sum_j P_j \cdot t_j$$
$$c_i = \sum_j P_j \cdot c_j$$

The above procedures are executed for all intercity main line modes that are identified between two districts.

Discussion. It is believed that this system has made considerable progress toward quantifying the access problem. However, the techn'que just described doesn't explicitly confront the aggregation problem. Therefore, much research needs to be done. It will be shown that the variance of the access characteristics must be analyzed to produce better predictions. Also, access impedances of all access modes should be studied.

The proposed research described later in this chapter has similar characteristics to the Rassam and Ellis technique, but is significantly different in that more variables are included which characterize the transportation system of a zone or district.

The Basis for Access Supply Modeling

Supply Estimation for Direct Demand Models

New and better approaches are being developed for aggregate travel demand modeling. These models are based on conventional economic theory; the theory of consumer, behavior. These models, so called direct demand models predict trip generation, trip distribution, and mode split in the same equation, and include three sets of variables: system attributes (travel times and costs for all competing modes), socioeconomic attributes of the travelers, and the activity system (attraction) variables.

Research by Domencich, Kraft, and Valette[1] has shown that travelers react differently to different components of travel time and cost. The time components were separated into access and linehaul portions. The access time was the time spent outside the principal mode.

The important fact here, is that, although the level of service was expressed by many variables (e.g., access and linehaul components) the average (access) times for zone pairs were computed using only the sampled travelers and this does not properly represent the access supply characteristics of the zones. Trips that could be made by potential travelers, wherever they reside in the zone, must be considered.

A recent study by Talvitie[2] concluded that transit access times for bus and rail trips are especially important in forecasting travel demand. However, the average zonal access values that were used for estimating these demand models, were also computed from

---

[1] Domencich, T., Kraft, G., Valette, P., "Estimation of Urban Passenger Travel Behavior: An Economic Demand Model", Highway Research Record 238, (Washington, D.C.: Highway Research Board, 1968).

[2] Talvitie, A., An Econometric Model for Downtown Work Trips, Ph.D. dissertation, Northwestern University, Evanston, Illinois, December, 1971.

just actual travelers.

It is clear that access supply models describing all the people in the zone need to be formulated to improve the overall transportation systems analysis. A model to predict the supply characteristics would also provide the analytical methodology of properly representing the level of service for each zone used in a study, without the relatively expensive acquisition and processing of large amounts of data. For forecasting purposes such a model is a necessity, of course.

Need for Model Development

As discussed, little has been done modelwise, to evaluate or quantify the access and egress portions of a transit trip. Furthermore, it has been discussed and shown in the literature that these trip segments constitute a major determinant in the choice of a travel mode and whether the traveler will make a trip at all to a particular destination.[1] A transit trip from an origin to a destination normally involves several alternative decisions. For example, in making

---

[1]Kraft, G., and Wohl, M.,"New Directions for Passenger Demand Analysis and Forecasting", <u>Transportation Research</u> (London: Pergamon Press, 1967), Vol. I, No. 3, pp. 213-214; Talvitie, A., <u>An Econometric Model for Downtown Work Trips</u>.

a trip the following must be determined; access mode; access station (rail) or stop (bus); linehaul mode; line or path for linehaul segment; egress station or stop; egress mode. These decisions are primarily based on travel times and costs. In other words the level of service of the trip is evaluated by the traveler; therefore, in the analysis of forecasting travel demands and choices of modes these transportation system components must be quantified.

In order to accurately predict aggregate zone to zone travel demand by mode and path, we must have good estimates of aggregate values of the level of service for the trip segments. However, existing supply models do not estimate these components with functions that produce aggregate values from zonal characteristics.

The (aggregate) level of service characteristics that are usually associated with direct demand models are total access time, linehaul time, total egress time, and total costs. These are total average values for a zone. It is not meaningful to develop supply models to determine <u>total</u> access and egress values for a zone. It is better to estimate average values by access mode. However, total zonal values can be determined by utilizing multi-mode split models.[1] These mode split models produce modal shares[2] that serve as weights for computing the average total access and egress times and costs for a zone.

Another fundamental need for the development of mean and variance supply models is to provide a necessary input for using and/or estimating disaggregate and aggregate travel demand models in an unbiased manner. It has been shown[3] that besides the zonal means, the within zone variances of explanatory variables are needed in both estimating and forecasting aggregate travel demand models. Means and within zone variances are also needed if the disaggregate models are to be used in forecasting travel demand.

The need for mean and variance models in aggregate modeling arises as follows: The error term in an

---

[1] Liou, P., Disaggregate Access Model and Station Selection Models for Rail Trips, Ph.D. dissertation in progress, Department of Civil Engineering, University of Oklahoma.

[2] The modal shares themselves, are a function of the access and egress times and costs to each rail station or bus stop in a zone.

[3] Talvitie, A., "Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models", Transportation Research Forum (October, 1973).
The discussion above basically follows the presentation in this article.

aggregate travel demand model has at least three components. The first is the sampling error (v) in the dependent variable V (volume). The second component of the error is the random error (e) of the model; and the third component of the error is due to zonal aggregation ($\bar{u}$) of the explanatory variables ($\hat{X}$). It is important to note that this aggregation error is denoted $\bar{u}$ because it is the error in explanatory variables which are zonal averages.

If these three errors are included, the following demand model will result:

$$V = a + b\,\tilde{X} + v + e$$

$$\hat{X} = \tilde{X} + \bar{u}$$

The coefficients a and b can be estimated using the maximum likelihood technique if v, e, and $\bar{u}$ are independent and normally distributed and if var (v) and var ($\bar{u}$) are known.[1] The prediction made with the above model is of particular interest. Johnston shows that

$$E(V/\hat{X}) = a + b\left[\frac{\mathrm{Var}(\bar{u})\cdot\bar{X} + \mathrm{Var}(\tilde{X})\cdot\hat{X}}{\mathrm{Var}(\bar{u}) + \mathrm{Var}(\tilde{X})}\right]$$

where: $\bar{X}$ is the mean of $\tilde{X}$

---

[1]Johnston, J., Econometric Methods, (New York: McGraw-Hill, 1972).

It may be observed that the prediction $V = a + b\hat{X}$ is unbiased only if $\hat{X} = \bar{X}$. Thus, the within zone variance of the explanatory variables is needed for both unbiased estimation and application of the aggregate travel demand model.

It should be noted here that the access supply models developed in this research are not, of course, models for var $(\bar{u})$ (within zone variance of the mean) but for var (u) (within zone variance of the variable). Talvitie suggests the following relation between var (u) and var $(\bar{u})$:

$$\mathrm{Var}\ (\bar{u}) = \mathrm{Var}\ (u) \cdot \frac{M-V}{M-1}$$

where: M is the size of the market (e.g., population) in the origin zone.

Disaggregate travel demand models do not need the aggregate mean and variance supply models in the estimation phase because they are based on individual travel information and the error in the utility function G(x) can be assumed to be strictly random. However, the aggregate mean and variance supply models are needed if these demand models are intended for use in forecasting aggregate travel demand. This comes about as follows: Disaggregate travel demand models are often estimated using logit analysis. In logit analysis the following model is fitted to data.

$$P_k = \frac{e^{G(x_k)}}{\sum_{i}^{n} e^{G(x_i)}}$$

In this expression P_k is the probability of an individual chosing alternative k among the i = 1...n relevant alternatives, and $G(x_i)$ is a function of the level of service, socioeconomic, and activity system variables characterizing the utility of alternative i.

The expected value of P_k, or the share of people chosing alternative k, is as follows:[1]

$$E(P_k) = \bar{P}_k + \mathrm{Var}(G(x))\ \bar{P}_k(\bar{P}_k-1)(\bar{P}_k-\tfrac{1}{2})$$

where: $\bar{P}_k$ is the value of P_k evaluated at the mean of G(x).

This result indicates that for a disaggregate model to give unbiased travel forecasts, the within zone variances of the explanatory variables need to be taken into account. Only in trivial cases (i.e. $\bar{P}_k$ = 0,1, or ½) will no bias result from not considering the within zone variances whose weighted sum forms the variance of the utility function G(x).

---

[1]Only the binary choice situation is considered here. Extension to multinomial cases is similar, and described in Talvitie's article.

The conclusion to be drawn from the above discussion is that there is an urgent need for supply models which provide not only the zonal averages of the level of service variables, but also their variances. These models are needed regardless of whether travel forecasting is done using aggregate or disaggregate travel demand models.

Characteristics of Rail and Bus Trips

The access portion of a rail or bus trip consists of the time spent on walking, driving, or busing from the trip origin to the linehaul station. The linehaul portion is the trip time from the origin zone station to the rail or bus destination zone station. The egress portion comprises the time spent walking, busing, or taking a taxi to the final trip destination.

Trip costs are usually estimated by analyzing both out-of-pocket and operating costs for all trip segments.

Tables 1 and 2 show the characteristics that are most often considered for each segmented portion of a rail or bus trip.

The trip characteristics for time can be aggregated in numerous ways. Table 3 presents the rail and bus aggregations that were considered for this research.

TABLE 1

RAIL CHARACTERISTICS

Access		
Walk:	Time:	Walk, wait for train
	Cost:	----
Drive-Park:	Time:	Walk to car, ride to parking lot, walk to station platform from lot, wait for train.
	Cost:	Parking, operating
Drive-Drop:	Time:	Walk to car, ride to station, walk to platform, wait for train
	Cost:	Operating
Bus:	Time:	Walk to bus stop, wait for bus, ride bus, walk to station, wait for train
	Cost:	Fare
Linehaul		
Rail:	Time:	Riding time, transfer time
	Cost:	Fare
Egress		
Walk:	Time:	Walk to destination
	Cost:	----
Bus:	Time:	Walk to stop, wait for bus, ride bus, walk to destination
	Cost:	Fare
Taxi:	Time:	Walk to curb, wait for taxi, ride taxi, walk to destination
	Cost:	Fare

TABLE 2

BUS CHARACTERISTICS

Access			
Walk:	Time:	Walk to stop, wait for bus
	Cost:	----
Drive-Park:	Time:	Walk to car, ride to parking space, walk to stop, wait for bus
	Cost:	Parking, operating
Drive-Drop:	Time:	Walk to car, ride to stop, wait for bus
	Cost:	Operating
Linehaul		
Bus:	Time:	Riding time, transfer time
	Cost:	Fare
Egress		
Walk:	Time:	Walk to destination
	Cost:	----
Taxi:	Time:	Wait for taxi, ride taxi, walk to destination
	Cost:	Fare

TABLE 3

RAIL AND BUS TIME AGGREGATIONS

RAIL	
Access	
Walk:	$\boxed{\text{Walk}}$[1], wait
Drive-Park:	Walk, $\boxed{\text{ride}}$[1], wait
Drive-Drop:	Walk, $\boxed{\text{ride}}$[1], wait
Bus:	$\boxed{\text{Walk}}$, $\boxed{\text{ride}}$[1], wait
Linehaul	$\boxed{\text{Ride}}$[2], wait
Egress	
Walk:	$\boxed{\text{Walk}}$
Bus:	Walk, $\boxed{\text{ride}}$, wait
Taxi:	Walk, $\boxed{\text{ride}}$, wait
BUS	
Access	
Walk:	$\boxed{\text{Walk}}$[1], wait
Drive-Park:	Walk, $\boxed{\text{ride}}$, wait
Drive-Drop:	Walk, $\boxed{\text{ride}}$, wait
Linehaul	$\boxed{\text{Ride}}$[2], wait (transfer)

[1]The estimation of these trip characteristics constituted the major portion of this research.

[2]Even these components can be considered to be deterministic (i.e., not subject to geographic aggregation) if the access and egress terminals are modeled explicitly.

The segments that are blocked are variable and subject to geographic aggregation. Therefore, they could be modeled, while all others are assumed to be strictly deterministic or simply constants and do not warrant any aggregation process. It also appears that trip costs (per interchange) can satisfactorily fit this category.

Proposed Supply Models

In this study, access supply models are developed which produce aggregate values for selected transportation system attributes. The attributes under study are the travel times for access portions of rail and bus trips. Three models are estimated that deal with a rail trip. These are the following: access walking time to station; access driving time; and access riding time in bus.

One access model is developed for a bus trip. This is the walking time to a bus stop; designated bus/walk. However, this bus/walk time can also be considered a segment of a rail access trip if a traveler walks to a bus stop in order to go to a rail station.

In addition to the above models, which estimate the zonal mean access time, corresponding models for the within zone variance or standard deviation of the access times were developed. It should be mentioned here that these models are for the variance of the access time in a zone and not, of course, for the zonal mean access time.

Two reasons existed for estimating the within zone standard deviation or variance. The first, was to simply be able to determine a measure of accuracy for the access times provided in a zone for planning purposes. The second and most important reason was explained earlier.[1] This was to provide an input to the existing travel demand model analysis and/or forecasting that would hopefully help reduce the bias in present travel estimates.

The selection of these models for research was based on many considerations. An assumption was made that the egress travel times can be estimated by the access models. Taxi invehicle time can be equated with automobile driving time. The drive-drop and drive-park segments of a rail access trip are the same except for the time required to find a parking space and walk to the rail platform. This time is a function of each terminal's characteristics, and will not be evaluated here.

The drive by car access portion of the trip to a bus linehaul station appears to be quite important. Many

---

[1]See subsection on Need for Model Development in this Chapter.

cities are making large parking lots available for centrally located express bus service to downtown and business district areas. However, a specific model could not be formulated due to a lack of observations in the data. It can be argued, however, that drive models designed to access a central bus station are very similar to drive models designed to access a rail station. Therefore, the drive models developed for this research could be used for this bus access purpose.

The linehaul riding time for rail and bus trips can be estimated quite accurately using the following variables; speed, distance, volume and number of stops. It is not a trip segment that requires geographic aggregation to represent the zones involved. Since many research studies have concentrated on this aspect of a transit trip, linehaul times will not be estimated in this research.

As stated earlier, very little research has been done on the access portions of mass transit trips. It is quite difficult to accurately estimate aggregate[1]

---

[1]The assumption is made here that the average value of the access time is equal to the aggregate value.

zonal access time by submode. However, this research attempts to properly quantify the indicated models. Also as described, these models should be extremely useful to either aggregate or disaggregate travel demand analysis for providing proper forecasting estimates.

The supply models developed in this study have the following functional form:

$$\text{Access Time}_j^m (\hat{L}) = S(\text{Zone Size variables}_j, \text{Transportation System variables}_j, \text{Volume variables}_j) + e_j$$

where:

j = origin zone

m = access mode of travel

$\hat{L}$ = estimate of access time

S = supply model

e = error term

As described in the basic theory for predicting networks flows[1], the level of service is a function of the transportation system and the volume of flow. In order to apply this theory for predicting the level of service of a zone, we must include variables that characterize the zone.

---

[1] Manheim, et al, "Search and Choice in Transport Systems Analysis".

Therefore, three types of variables are considered. They represent the size of zone, the transportation system in a zone, and the trip volume in a zone. The zone size variables are composed of such factors as the area of the zone and the average travel distance. The transportation system variables are represented, for example, by the number of rail stations, the amount of bus route miles and the percentage of zonal area from which people can realistically walk to a bus or train. The trip volume is described by the population density in a zone.

The specification of these variables provides a predictive relationship that can be estimated using statistical techniques.

The sample data, model variables, and estimation technique that were utilized in developing the four mean models and four standard deviation models are detailed in the following chapter.

CHAPTER III

DATA AND METHOD

Before the estimation of the model parameters can be performed, three things must be done. First, the sample data on which to base the estimation of model coefficients must be chosen. Second, the derivation of values for each dependent and independent explanatory variable in the sample data has to be explained. And third, the estimation technique along with the criteria for evaluation of model accuracy must be specified.

Data Source

The trip data used in this study came from an origin-destination (O-D) survey, conducted in March and April of 1969 by W. C. Gilman and Company, Inc. for the Southward Transit Area Coordination (STAC) Study in Chicago.

The study area is shown in Figure 1 (inside the heavy line). It consists of 205 traffic zones, ranging in size from one square mile to 26 square miles.

The STAC study area is served by 10 public transit carriers, including 5 railroads and 5 bus carriers.

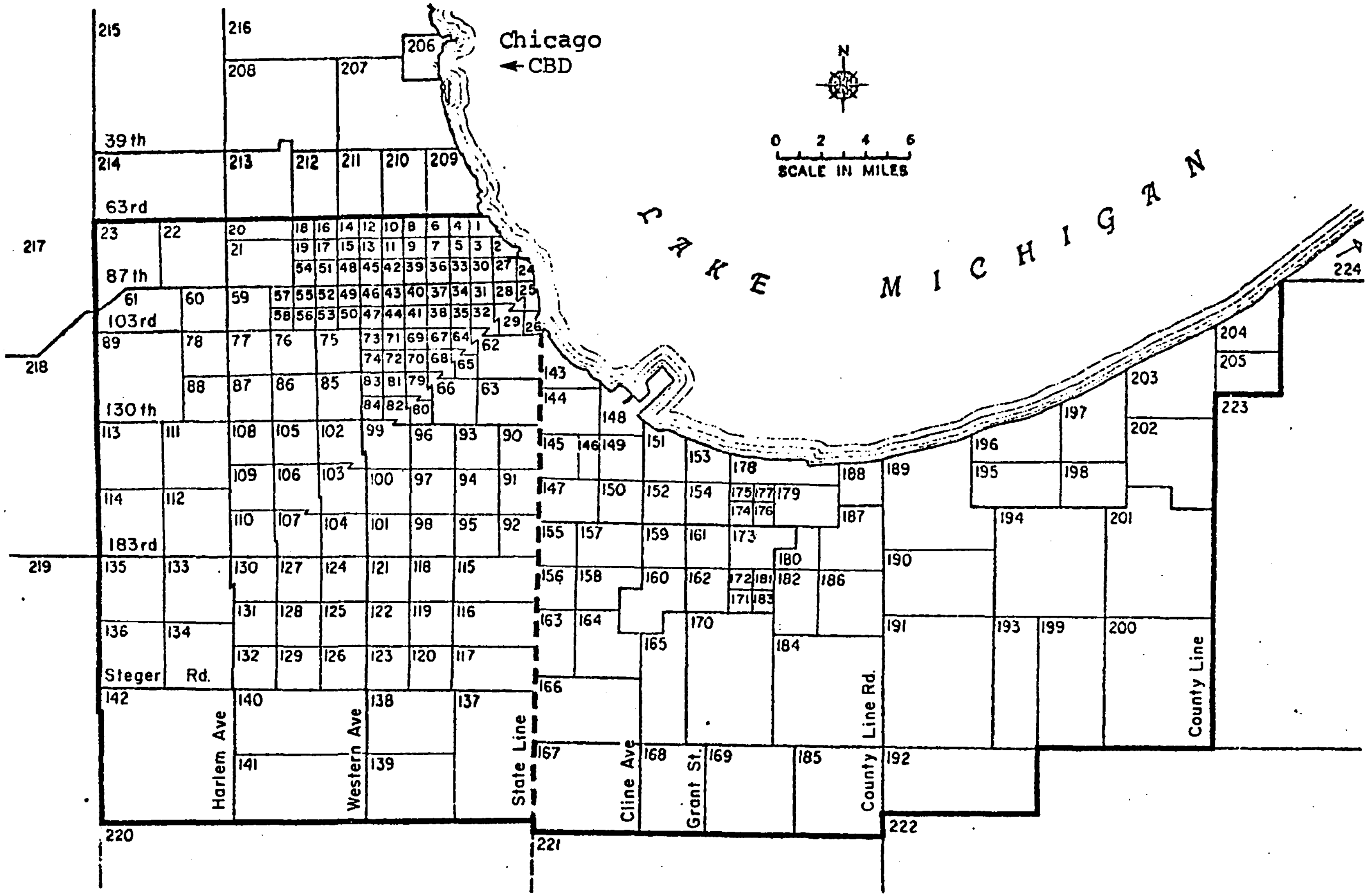

Fig. 1.-- Study Zones

Ninety-five bus routes and eight rail lines are operated each weekday. A total of 2,981 bus trips and 91 train movements are made during the morning peak hours (6:00 a.m. to 9:00 a.m.).

Trip questionnaires were distributed to travelers on-board the transit vehicles and returned on a voluntary basis. The data collected consisted of items such as trip origin, trip destination, purpose of trip, trip time, access and egress mode, and access and egress station. There are 34,088 individual transit trips that have been coded and stored on magnetic tape.

Choice of Sample

The data source, as described, does not provide an ideal sampling frame from which to choose a sample; the reason being the fact that if we only consider travelers who actually made transit trips (as in the on-board survey), we cannot properly estimate the supply relationship for the entire zone population. The ideal frame would be a record of all household trips not merely transit trips[1].

---

[1]The Chicago Area Transportation Study has performed a home interview survey, which would better represent all potential travelers in a given zone. Unfortunately however, the data was not available in a usable form for this research.

The basic concern is that the trip data does not contain individuals who used their automobiles for a linehaul trip or the locations of non-tripmakers. However, it was assumed for the sampling frame used, that these people are evenly dispersed throughout a zone, and cause no bias in the chosen sample. The basis behind this assumption is that each person had the same chance of being selected.

The actual sample frame used consisted of work trips made during the morning peak hour period (6:00 a.m. to 9:00 a.m.) on the Illinois Central (IC) and Chicago South Shore and South Bend (CSSB) railroads. Zones with very few observations and the trips which were not for the work purpose were also eliminated. This resulted in selecting 108 zones where trips originated.

There are two reasons for considering only IC and CSSB travelers for the sample. The first is that in order to properly estimate an access (or egress) supply model from interview data, the effect of the egress (access) part of the trip on choice of rail line, which naturally affects the values of the access (egress) times, must be eliminated. The IC and CSSB utilize the same tracks and stations to the Chicago CBD while the Rock Island, Norfolk and Western, and Penn Central rail lines go to different stations each.

The second reason is that it was assumed that the walking time to a bus stop for a bus linehaul trip could be estimated by using the rail trip table. This is valid because normally the same bus line or another bus line on the same street would be used regardless of whether an individual rides to a rail station or to his destination location.

Thus the models of this study estimate average zonal values for selected access time components given that a certain linehaul mode and line (path) is chosen.

The data set that was used for model calibration, consisted of means and standard deviations of times for selected zones along with average explanatory characteristics of these zones. Therefore, two things had to be determined. Of the 108 available traffic zones, how many should be chosen to produce statistically valid models. And, from these chosen zones, how many individual observations should be used to compute the means and standard deviations of the access times.

These determinations were made by applying the basic procedure used for choosing a sample size. The procedure will be random sampling from a relatively large population. The formulation of this approach is:

$$N = \frac{Z^2_{\alpha/2} \cdot \sigma^2}{E^2}$$

where N = sample size (number of zones)

$Z^2_{\alpha/2}$ = the Z value corresponding to the upper tail of the Z distribution

σ = the population standard deviation

E = estimate of the error

It was assumed that the highest expected population standard deviation (σ) would appear in the walk mode case. Further, the value was estimated to be 10 minutes[1]. The sample design was based on this access walk mode estimation, since it was expected to yield the highest error. This error (E) was defined to be $\pm$ 2.5 minutes. Also, the probability that the mean access walk time lies between these error limits was set at 90 percent.

It was resolved that a 90 percent confidence interval would be sufficient, based on the intended use of the predicted estimates and the resources available to quantify the larger data set that would result from considering a 95 or 99 percent limit.

---

[1]This was believed to be a rather high value.

Also the estimate for the error variance was considered to be high rather than low.

The solution in the above equation with $Z_{\alpha/2}$ equal to 1.645, is approximately 44 zones. It was decided, therefore, that 50 randomly selected zones would be selected to estimate the models. However, the zones were of varying geographic size. And, it had to be determined how many zones of each size should be selected, since an a priori assumption was made to analyze the significance of zone size.

The study zones vary in size from 1 to 8 square miles, with the major portion of the zones consisting of 1 and 4 square mile areas. After considerable analysis of proper study area representation, the following number of zones were randomly selected from each classification: 25 zones of one square mile, 15 zones of four square miles, and 10 zones of greater than four square miles.

The next task was to select the number of individual observations to be used in computing values for the dependent variables for each of the 50 zones. It was assumed that this would be a function of the size of the zone; the larger the zone the greater the number of individual observations.

The equation above was again applied to the walk case. The estimate of the error was set at $\pm$ 2.5 minutes.

Also, the population standard deviation was set at 5 minutes, and 90 percent confidence limits again imposed. This resulted in the need to select approximately 11 observations.

The decision was made to randomly select 15 observations from the 1 square mile zone. To determine the number of observations selected for the larger zones, the area in square miles was multiplied by 12. The limit was set at 48 observations per zone. This resulted in a data set of 1469 individual trip observations.

It should be clarified that these random samples were chosen from the rail trip table in each zone, which consists of drive, walk, and bus access trips. Therefore, a different proportion of drive, walk, and bus trips existed in each of the zone samples. This fact was not particularly important because in order to properly estimate the supply relationship for the entire zone population, the selected 1469 origin-destination observations had to be synthesized. This included the assignment of access trip distances for each of the 1469 observations for the alternative access modes, in addition to the actual chosen mode.

This essentially meant that each original observation actually represented a walk, drive, bus, and bus/walk access trip to a rail linehaul station. Therefore, a total of approximately 5876 observations (1469 for each model) were utilized. Due to the data problem stated earlier, this technique was devised and found to have considerable merit.

The method used to convert assigned distances to trip times is explained in the next section.

Dependent Variables

Chapter II (Proposed Models) discussed the isolation of eight dependent variables to be estimated. This included four mean access time variables plus the standard deviations of each. The prediction of trip access time will be for a one-way trip.

The mean values were computed in the following manner. From the random selections of individuals, a zonal average was determined. The origins of these individuals[1] were first located on detailed street maps of the study area. For the drive and walk access mode, a straight line or air distance (in miles) was measured to the actual rail station chosen. These distances

---

[1]A traveler's trip origin is reported to be within a specific 1/4 square mile area.

were then multiplied by an assumed factor of 1.25 to account for the total access distance traveled. For the walking time to a bus stop and the bus riding time to a rail station, a factor was not used due to the straight line travel directions of these modes.

As noted, each observed traveler was first assigned a measured distance to the actual station he chose. The traveler was then assigned a determined distance for the three alternative access modes. It was assumed that each of these alternative access trips was made to the nearest accessible station[1]; not necessarily to the chosen station.

These distance values were then transformed to time (in minutes) by assuming a speed for each mode. A value of 3 miles per hour was used for the walk modes. The drive and bus mode speeds were derived by developing a linear relationship with the density of each zone. The speed range for the drive mode was assumed to be 12 to 22 miles per hour. The bus range was 8 to 14 miles per hour. The lower speed corresponds to the highest zone density, while the

---

[1]The determination of accessible stations will be explained in the following section.

higher speed corresponds to the lowest density.

The average values for all selected and assigned trip times in each zone, for the four access modes, are the dependent variables for the mean models. Consequently, the standard deviations of the selected and assigned trip times became the dependent variables for the standard deviation models.

Explanatory Variables

The explanatory variables to be considered for this study are those that specify a zone's aggregate characteristics. Many variables were studied to determine a justifiable association-causation relationship with the dependent variables, and to select those that could be quantified.

The independent variables that were tested for each rail model and the methods of quantification are explained below:

Access Walking Time

Zone Size (ZSIZE) is the number of square miles in each zone.

Accessible Stations/Zone (TSTA-W) consists of the total number of stations within or in close proximity of a zone. A specific station is considered not accessible if every traveler in a zone has a closer station available; otherwise it is identified as being accessible (see Average Distance below).

Average Distance (DISTDW) is the average zonal straight line measured distance in miles. This value was estimated by a method of weighting zone sections. Consider the following zone diagram:

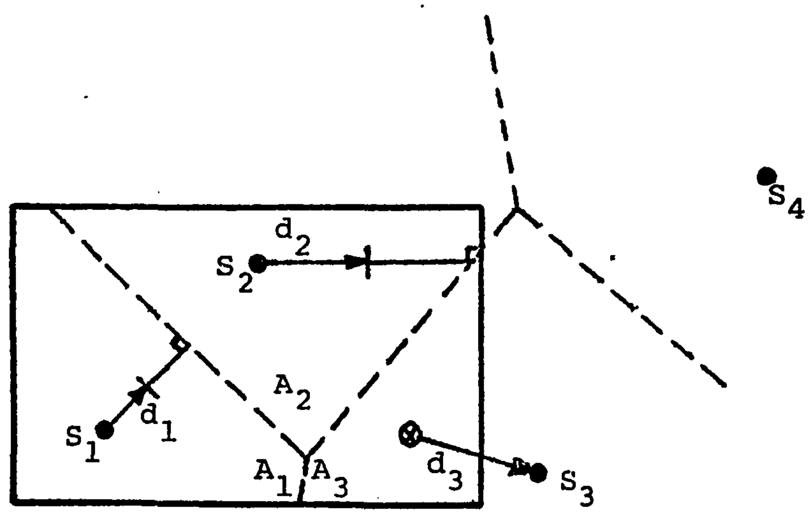

Fig. 2.--Zone Diagram

The zone is divided into sections by drawing bisecting lines between stations that appear to be accessible. The accessible stations are then determined as indicated above. Station S_4 is not accessible to this zone. An average distance (d_i) is determined for each section. This was assumed to be 50% of the distance from the stations (S_i) to the farthest perpendicular section boundary[1]. The average distance for the zone is then:

$$\text{Avg. Dist.} = \sum_{i=1}^{\#\ \text{Stations}} (A_i d_i)$$

where: A_i is the percentage of the zonal area for each section.

This value was adjusted by multiplying by a factor of 1.25 to approximate actual travel distance, prior to estimating the model coefficients.

Accessible Stations Within Zone (STAI-W) is the number of stations within or on the boundary line of a zone.

Accessible Stations Outside Zone (STAO-W) is the number of stations outside of a zone.

Percentage Stations Outside Zone (PSTO-W) represents the proportion of accessible stations outside the zone.

Station Density (SI/Z-W) is the ratio of accessible stations within a zone to the size of the zone in square miles.

---

[1]Average distances to stations outside the zone are measured from the section centroid.

Walker Zone Coverage (COV-W) is the percentage of zonal area comprising one-half mile surrounding each accessible station. Studies have shown that, in general, people will not walk more than one-half mile for the access portion of a trip[1].

Distance Range (RANG-W) represents the spread in distance (miles) of the accessible stations of a zone. The value of the variable is computed by measuring the difference in distance from the zone centroid to the farthest and closest stations.

If only one station is accessible to a zone, the distance range is the difference between the farthest and closest quarter mile area of the zone.

The variable was predominantly used to explain the standard deviation of the access walking time, but was also tested for significance in the mean model.

Access Driving Time

Zone Size (ZSIZE) is the number of square miles in each zone.

Population Density (DEN) is the number of people in a zone divided by the zone size in square miles. The

---

[1]Lisco, T., The Value of Commuters Travel Time: A Study in Urban Transportation, Ph.D. dissertation, Department of Economics, University of Chicago, 1967.

value used in the regression was scaled down by dividing by 10,000.

This variable simulates the effect achieved by traffic volume. The more dense the zone or the higher the volume, the longer the driving time will be. Chicago data was not available on access volume counts.

Accessible Stations/Zone (TSTA-D) was derived in the same way as (TSTA-W) for the access walk model.

Average Distance (DISTDW) was derived exactly the same as (DISTDW) in the access walk model.

Accessible Stations Within Zone (STAI-D) was derived in the same way as (STAI-W).

Accessible Stations Outside Zone (STAO-D) is the number of stations outside the zone identified as accessible for the drive mode.

Percentage Stations Outside Zone (PSTO-D) was derived in the same way as (PSTO-W).

Station Density (SI/Z-D) was derived in the same manner as (SI/Z-W).

Distance Range (RANG-D) was derived in the same way as (RANG-W).

Access Bus Time

The following explanatory variables, unless indicated otherwise, were derived and computed in the same manner as the access drive variables.

Zone Size (ZSIZE)

Population Density (DEN)

Accessible Stations/Zone (TSTA-B)

For the bus model, a rail station could be accessible only if a bus line served the station. Therefore, the accessible stations for each zone were determined by specifying those that were conveniently served by a bus line.

Accessible Stations Within Zone (STAI-B)

Accessible Stations Outside Zone (STAO-B)

Percentage Stations Outside Zone (PSTO-B)

Station Density (SI/Z-B)

Distance Range (RANG-B)

Average Distance (DIST-B)

This value was computed by averaging the distances in miles from the centroid[1] of a bus route to an accessible station. The number of distances to be measured and averaged was determined by the number of accessible bus stations.

---

[1]The centroid of a bus route was assumed to be the midpoint of the route in a given zone.

Access Bus/Walk Time

Zone Size (ZSIZE) is the number of square miles in each zone.

Route Miles (RT-MI) consists of the total miles of bus lines that are in or on the boundary of a specific zone. Only bus routes that serve access trips (i.e., stop at main streets) were counted.

Route Mile Density (RTMI/Z) is the total route miles of bus lines per square miles of zonal area.

Bus/Walk Zone Coverage (COV-BW) is the percentage of zonal area comprising one-quarter mile surrounding each bus line. Studies have shown that, in general, if travelers walk to a bus stop for the purpose of riding to a rail station, they normally do not walk a distance greater than one quarter mile to that bus stop[1].

Accuracy of Input Data

A great deal of time and effort was spent in preparing the sample input data. This is a common requirement for empirical studies of this nature. Therefore, because of the care that was exercised, errors in analysis resulting from errors in the input data, were expected to be minimal.

---

[1]Talvitie, A., An Econometric Model for Downtown Work Trips.

Estimation Technique

The criterion that was used to calibrate the model parameters was to minimize the square of the error. The error is defined as the difference between the actual and estimated value of the dependent variable.

The method of ordinary least squares (OLS) is rather simple to use. Many well developed computer packages exist for its use. Because of this computational efficiency, the OLS method can effectively be used to help select the best combination of the explanatory variables. Also, the standard output of a stepwise ordinary least squares procedure facilitates the examination of the contribution of an individual variable to the coefficient of determination, and allows a check on the stability of the signs of the variables in the equation when a new variable is entered.

In this study, statistical inferences are made about a population based upon information contained in the selected sample. Since populations are characterized by descriptive measures called parameters, we are making inferences about these parameters. In ordinary least squares analysis, we are estimating or predicting values for these parameters. We are trying to determine the relationship or correlation between variables.

The relationship is linear if there is a tendency for the dependent variable to change by a constant amount when the independent variable changes by a given absolute amount. This situation occurs many times in applied research. It has become customary to fit a regression line to the data by the OLS method. The line is fitted in such a way that the sum of the squared differences of the observed dependent values from the corresponding line values is a minimum.

The basic limitation of the OLS method is that if some of the independent or explanatory variables are highly correlated, (i.e., collinear) the coefficient estimates may not be plausible nor the consequent model structurally correct. This is the problem of multicollinearity. An unreasonable allocation of the effect of the independent variables on the dependent variables normally arises if the correlation between two or more independent variables is close to $\pm$ 1, that is, if the values of the variables move together. Very large standard errors will result if this condition exists. This, in turn, may cause important independent variables to be incorrectly removed or rejected from the equations.

Other limitations stem from the assumptions that are placed on the general linear model. Some of these are:

1. The error terms are random variables with zero mean and constant and finite variance.
2. The independent variables are usually fixed variables (i.e., the error is in the dependent variable).
3. The dependent variables are uncorrelated random variables.
4. No correlation exists between the independent variables and the error terms.

When these assumptions are violated, the OLS method produces improper estimates of the parameters. If the independence of the error term is violated, we have the problem of auto-correlation. If assumption (1) is violated, this results in heteroscedastic disturbances. Other problems arise when there are errors of measurement in the explanatory variables (assumption 2).

It was first considered that an alternative estimation technique, such as constrained least squares (CLS) would need to be used to handle the problem of highly collinear variables, a condition common to travel demand analysis. This method of estimation consists of estimating parameters by minimizing the sum of squared deviations, as with ordinary least squares, but performing this minimization while satisfying certain prespecified conditions derived from a priori information. By making use of this additional a priori information, the analyst can assess the individual effects of collinear variables.

However, after applying the OLS method to the data set, and studying the correlation between the variables, it was concluded that collinearity was not a major problem in the estimation of the model coefficients.

Another problem that was initially expected to occur was that the variance of the disturbances or error terms were variable. This is the problem of heteroscedasticity, and would violate an assumption placed on the general linear model. This was suspect because of the manner in which the individual observations were selected in each zone. Initially, many more observations were randomly selected from the smaller densely populated zones, than from the larger less populated zones. However, since the dependent variables for the mean access time models are the means of the zonal observations, the values for the smaller zones would be calculated more accurately than the larger zones. This would create error terms that had variances that were variable. The method of weighted least squares was attempted to relieve this problem. This, however, proved ineffective because estimates of these error variances could not be well determined. The problem was handled by applying a different criteria to the selection of zonal observations. This was discussed in a previous section. It was felt further that assumptions 3 and 4 were not violated.

In summary, the estimation technique utilized was the ordinary least squares method. As presented in following chapters, this method produced models with structurally correct parameters that generally satisfied the probabilistic tests of significance.

Evaluation of Model Accuracy

As indicated earlier, in Chapter II, one of the objectives of this study was to develop supply models as accurately as possible. Normally, models are evaluated from two standpoints: structural accuracy and predictive accuracy. The evaluation of structural accuracy is difficult, because there is no simple standard as to what constitutes the most accurate structural model. Therefore, the method of evaluation that is primarily followed here, attempts to assess the predictive or forecasting accuracy of models.

The structure of the supply models is assumed to be sound. This must be based on intuition and judgement, as there is no empirical evidence of access supply estimation developed from zonal characteristics. As for the structure of model parameters; it will be shown that the coefficients all have proper signs. Also, as previously indicated, the explanatory variables that were considered for significance in each model had a justifiable association-causation relationship with its respective dependent variable.

The criteria used to evaluate the forecasting accuracy is primarily based on the determination of commonly employed statistical measures. For this purpose the following measures are tabulated.

1. The coefficient of determination (R^2)
2. Standard error of estimate
3. Average absolute error = $\frac{1}{N}\sum_i \left|V_i - A_i\right|$;

   where: V_i = the model value

   A_i = the actual value

   N = the number of observations
4. Average error as percent of the actual value

   $= \frac{100}{N} \sum_i (V_i - A_i)/A_i$

5. The Theil U-coefficient[1]

---

[1]Theil, H., Economic Forecasts and Policy (North Holland Publishing Co., 1958); and Meyer and Glauber, Investment Decisions, pp. 206-207.

Following Meyer's presentation, the U-coefficient is estimated as follows:

$$U = \frac{\sqrt{\sum_i (V_i - A_i)^2}}{\sqrt{\sum_i (V_i^2 + A_i^2)}}$$

It has an upper bound of 1, and equal to zero for perfect predictions.

The U-coefficient can be decomposed to three components, U^M, U^S, and U^C. The first two components, U^M and U^S, reflect the fractional loss in forecasting accuracy due to unequal means and variances of the actual and estimated values of the dependent variable, respectively. The third component, U^C, reflects the fractional loss in forecasting accuracy due to unequal covariation of the actual and estimated values of the dependent variable with the values of the explanatory variables. These three components are computed as follows:

defining $\bar{M}$, $\bar{A}$, S_M and S_A as the means, and standard deviations of predicted and actual values, and r as the correlation coefficient between them, and D, the denominator of the U-coefficient, then

$$U_M = \frac{\bar{M} - \bar{A}}{D}, \quad U_S = \frac{S_M - S_A}{D}, \quad U_C = \frac{\sqrt{2(1-r)S_M S_A}}{D}$$

and

$$U^2 = U_M^2 + U_S^2 + U_C^2, \text{ normalizing by } U^2$$

$$1 = \frac{U_M^2}{U^2} + \frac{U_S^2}{U^2} + \frac{U_C^2}{U^2}$$

the three components are obtained as:

$$U^M = \frac{U_M^2}{U^2}, \quad U^S = \frac{U_S^2}{U^2}, \quad U^C = \frac{U_C^2}{U^2}$$

These three indices are quite helpful in understanding the different sources of error in the models.

Another criterion that was used in judging the accuracy or validity of the models, was the plots of the computed or estimated dependent variables versus the residuals. These plots can provide valuable information with respect to, for example, the constant variance assumption of the OLS technique, over-estimation or underestimation of model prediction, and model specification.

For the access models, the plots will be the estimated access travel times in minutes versus the model errors (residuals) or differences between the actual access times and the estimated access times.

The width of the bands from the resulting plots is a measure of the variance. As the band width increases, the variance of the error increases. If the plot is inclined, the model is systematically over or underestimating the computed dependent variable. If the plot is a conical shape, the need for weighted least squares regression or a transformation on the actual observations is indicated. Further, if the plot resembles a convex or concave shape, the model is somewhat inadequate and extra terms in the model could be required.

In short, it is clear that one of the objectives is to obtain a plot with a narrow horizontal band that lies along the zero line.

The main point of this analysis is that if the fitted models are correct, the residuals should exhibit tendencies that tend to confirm the assumptions about the errors. These assumptions are that the errors are independent, have zero mean, a constant variance, and follow a normal distribution. When the residuals are examined we should be able to conclude either that the assumptions appear to be violated, or the assumptions do not appear to be violated.

These residual plots along with the described statistical measures will be analyzed in the following chapters for each model.

CHAPTER IV

DEVELOPMENT OF ACCESS WALK MODEL

In this section the mean and standard deviation access walk models are developed. First, the variables that were tested and those selected to comprise the model specification are discussed. Second, the evaluation of the models is presented by deriving a series of measures and plots to show their predictive accuracy.

Regression Analysis

In order to provide access time models which are relatively easy to use and understand, the mathematical form of the model variables was exclusively linear.

It was specified that the explanatory variables tested for these models had an a priori relationship with one or both of the walk dependent variables. This analysis was done prior to quantifying the variables. Therefore, spurious relationships with the dependent variables were not anticipated.

The specification of the model and associated variables[1] that were tested for the mean and standard

---

[1]See Chapter III for explanatory variable descriptions.

deviation follows:

$$\text{Access Walk Time}_j = S(\text{Zone Size variables}_j, \text{Transportation System variables}_j) + e_j$$

where:

j = origin zone

S = supply model

e = error term

Zone Size Variables

<u>Zone Size</u>

<u>Average Distance</u>

Transportation System Variables

<u>Accessible Stations/Zone</u>

<u>Accessible Stations Within Zone</u>

<u>Accessible Stations Outside Zone</u>

<u>Percentage Stations Outside Zone</u>

<u>Station Density</u>

<u>Walker Zone Coverage</u>

<u>Distance Range</u>

Volume related variables were not assumed to affect the walking time.

It was expected that a negative relationship (negative variable coefficient) would result for the explanatory variables: accessible stations/zone, accessible stations within zone, station density, and walker zone coverage. The remaining variables were expected to exhibit a positive relationship.

Ordinary least squares regression analysis was then performed. The selected models are shown in Table 4 with the Student t values for the selected variables. The only variable selected for the mean model was the average distance. The final regression analysis did include, however, walker zone coverage and zone size, but respectively, the variables had the wrong coefficient sign (positive rather than negative) and were statistically insignificant.

For the standard deviation model the final computer analysis included the distance range variable, in addition to the model selections of zone size and percentage of stations outside of zone. However, this variable also had the wrong sign (negative), lacked significance, and was not included in the model. The t value for PSTO-W (percentage stations outside zone) did not give a very strong indication of significance, but contributed to reducing the standard error of estimate.

TABLE 4

ACCESS WALK MODELS[a]

WALK-MEAN

Model Constant	Average Distance (DISTDW)
0.86136	19.35178
	(47.5)[b]

WALK-STANDARD DEVIATION

Model Constant	Zone Size (ZSIZE)	% Stations Outside Zone (PSTO-W)
3.07423	1.45797	1.11616
	(7.12)	(1.25)

[a]Travel times for all access models are expressed in minutes.
[b]Computed t values for included variables.

The regression analysis was run several times for each model, prior to the final analysis, in order to test all variables and select the most representative equation.

Model Evaluation

Measures of the predictive accuracy of the mean and standard deviation models are tabulated in Table 5. An examination of the values of the R^2 indices shows that the mean model is quite high (.981), while the standard deviation model is much lower (.558). This was expected for the standard deviation model, due to the relatively small sample size (for standard deviation estimation) and wide range of trip times.

An assumption was made prior to the selection of the sample size that the largest standard error for all models would occur in the mean walking time model. This turned out to be correct with a value of 5.1 minutes. However, this value is only about 15 percent of the mean zonal value. The standard error for the standard deviation model was 2.61 minutes. This value is larger; about 34 percent of the mean zonal standard deviation value. Also, as implied by the percent average error, the models tend to underestimate the walking time and its standard deviation.

TABLE 5

PREDICTIVE ACCURACY OF ACCESS WALK MODELS

Model	R^2	Standard Error of Estimate (Minutes)	Average Absolute Error (Minutes)	Percent Average Error	Mean Zonal Value (Minutes)	U	U^M	U^S	U^C
Walk-Mean	.981	5.10	3.63	-2.3	33.4	.0107	.0000	.0049	.9951
Walk-Standard Deviation	.558	2.61	1.89	-18.6	7.6	.0316	.0000	.1448	.8552

F value for mean model = 2262.8
F value for standard deviation model = 27.8

From the Theil U test the reason for most of the error appears to be unequal covariation of the actual and estimated values of the dependent variable with the explanatory variables; for both models over 99 percent and 85 percent, respectively of the value of Theil U is contributed by U^C. Only a minor part of the error (U^S) is due to unequal variances of actual and estimated values of the dependent variable.

An examination of the residual plots in Figures 3 and 4 appears to indicate some slight abnormalities. These are plots of the estimated dependent variables versus the model errors. Figure 3 exhibits a concave shape which indicates the possible need for extra terms in the model (e.g., square or cross product terms). Also the variance does not appear to be constant as the computed dependent variable is increased.

The standard deviation model plot in Figure 4 also indicates a variance that is increasing (conical shape), implying the possible use of weighted least squares analysis. The underestimation error is apparent by the slight increase in slope toward the negative residuals.

As a result of this evaluation the mean model appears to be suitably estimated while the standard

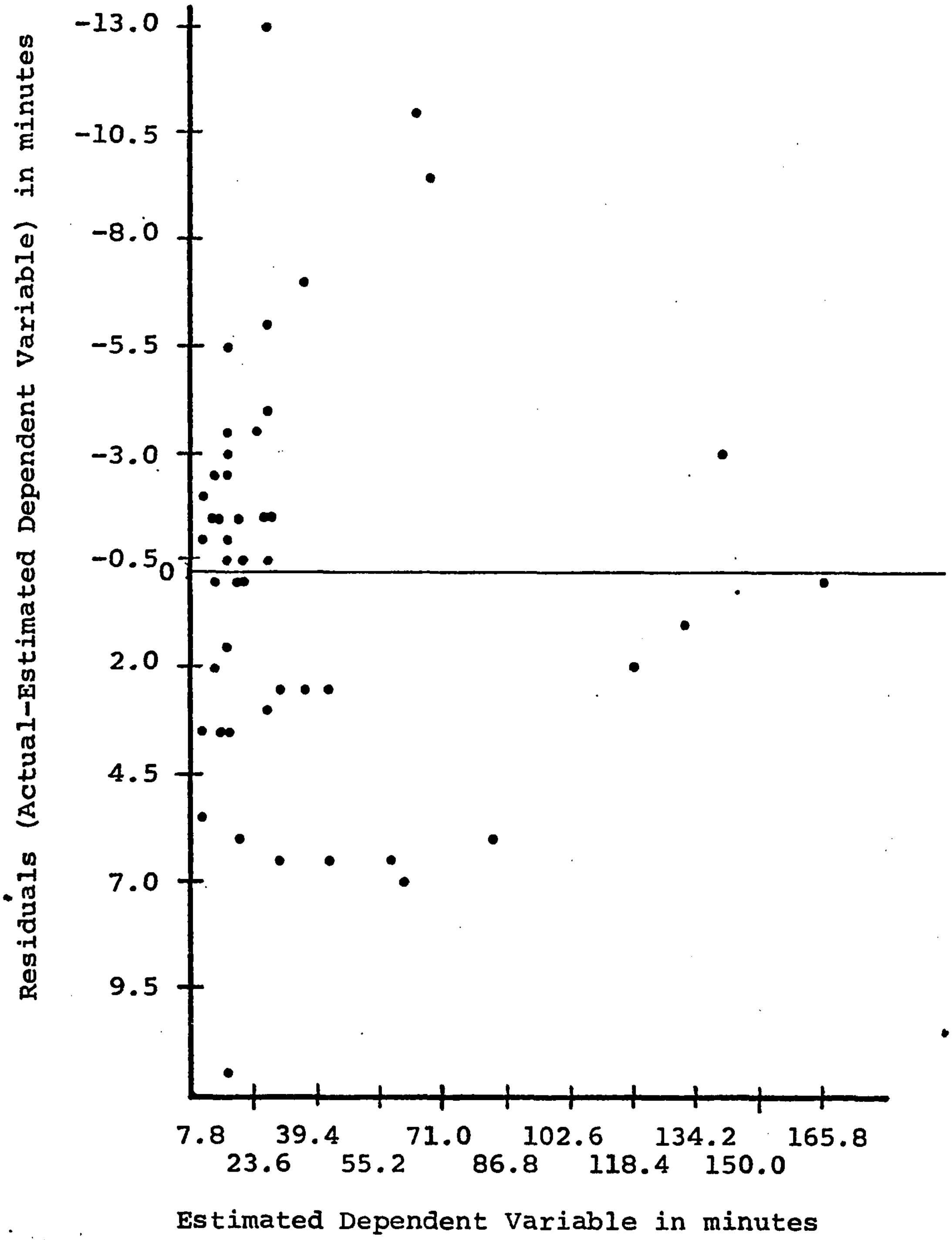

Fig. 3.--Plot of Residuals for Walk Mean Model

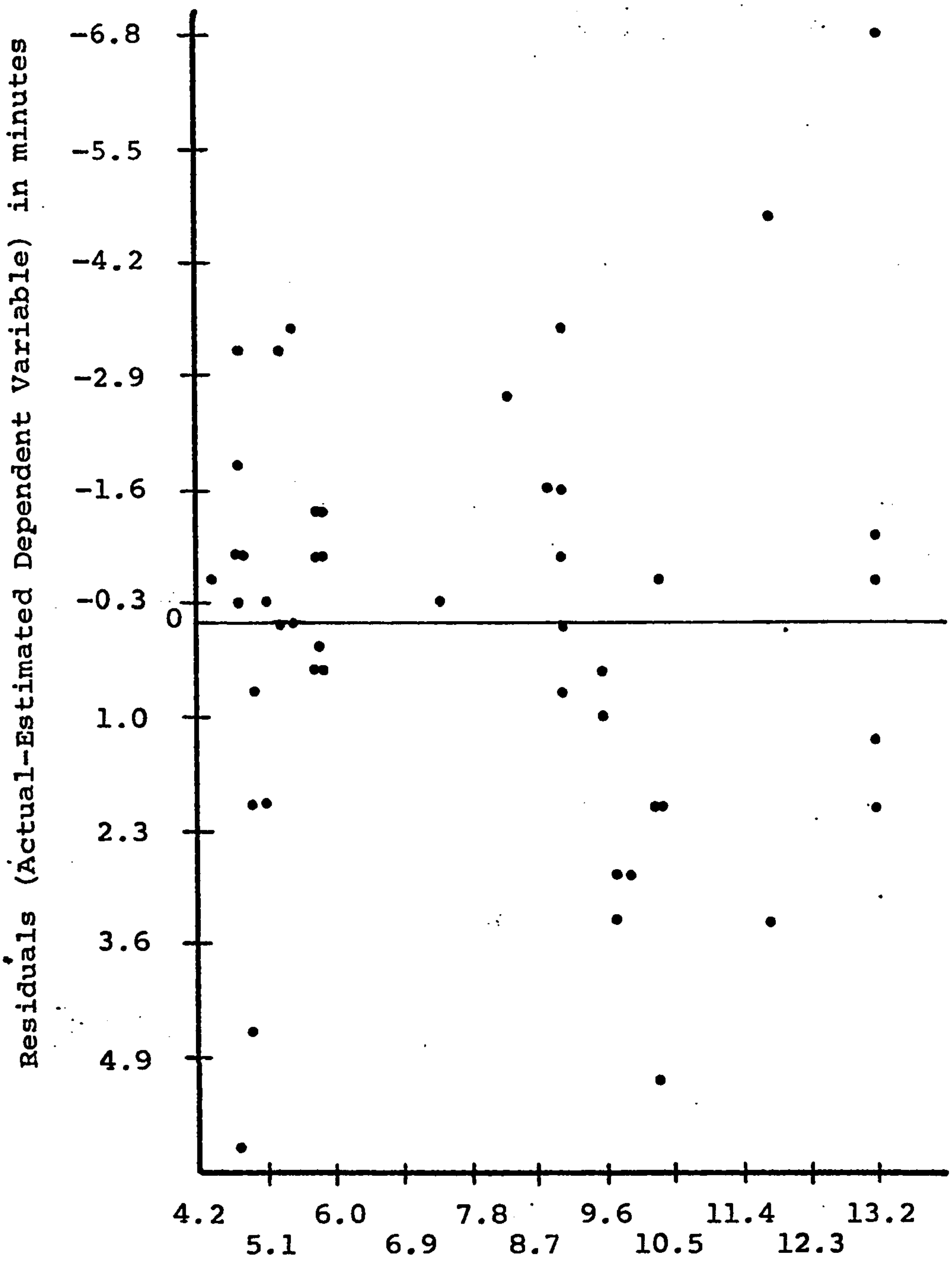

Fig. 4.--Plot of Residuals for Walk Standard Deviation Model

deviation model has a few inherent problems. These problems arose from the relatively small sample size and wide range of trip times. The examination of the residual plots did not show that the least squares assumptions were violated. On balance it cannot be said they were obeyed either; but it could not be specified which way they were violated. Therefore, on the basis of the observed data, we cannot conclude that the assumptions are incorrect.

The direction for future research would be to consider enlarging the sample size and utilize a data source that better represents all potential travelers. This should improve the models and eliminate most of the error. However, for this research on access walk models, considering their intended use, it is felt that the accuracy is quite adequate.

CHAPTER V

DEVELOPMENT OF ACCESS DRIVE MODEL

In this section the mean and standard deviation access drive models are developed. First, the variables that were tested and those selected to comprise the model specification are discussed. Second, the evaluation of the models is presented by deriving a series of measures and plots to show their predictive accuracy.

Regression Analysis

In order to provide access drive models which are relatively easy to use and understand, the mathematical form of the model variables was again linear.

It was previously specified that the explanatory variables tested for these models had an a priori relationship with one or both of the drive dependent variables. This analysis was done prior to quantifying the variables. Therefore, spurious relationships with the dependent variables were not anticipated.

The specification of the model and associated variables[1] that were tested for the mean and standard

---

[1]See Chapter III for explanatory variable descriptions.

deviation follows:

$$\text{Access Drive Time} = S(\text{Zone Size variables}_j, \text{Transportation System variables}_j, \text{Volume variables}_j) + e_j$$

where:

j = origin zone

S = supply model

e = error term

Zone Size variables

<u>Zone Size</u>

<u>Average Distance</u>

Transportation System variables

<u>Accessible Stations/Zone</u>

<u>Accessible Stations Within Zone</u>

<u>Accessible Stations Outside Zone</u>

<u>Percentage Stations Outside Zone</u>

<u>Station Density</u>

<u>Distance Range</u>

Volume variables

<u>Population Density</u>

It was expected that a negative relationship (negative variable coefficient) would result for the explanatory variables: accessible stations/zone, accessible stations within zone, and station density. The remaining variables were expected to exhibit a positive relationship.

Ordinary least squares regression analysis was then performed. The selected models are shown in Table 6 with the Student t values for the selected variables. The final computer analysis for the drive mean model included the zone size variable in addition to those shown. However, this variable lacked significance, had the wrong coefficient sign (negative rather than positive) and was not included in the model.

The final analysis for the standard deviation model also included zone size. The correct sign resulted, but the significance was very low. The t values for PSTO-D (percentage stations outside zone) and SI/Z-D (station density) did not show strong significance, but contributed substantially to reducing the standard error of estimate. Also, these variables were selected because their association--causation relationship with the drive standard deviation was judged to be important.

The regression analysis was run several times for each model, prior to the final analysis, in order to test all variables and select the most representative equation.

<u>Model Evaluation</u>

Measures of the predictive accuracy of the mean and standard deviation models are tabulated in Table 7.

TABLE 6

ACCESS DRIVE MODELS[a]

DRIVE-MEAN

Model Constant	Population Density (DEN)	Average Distance (DISTDW)	Distance Range (RANG-D)
-1.08567	0.69703	3.49496	0.74211
	(3.45)[b]	(27.1)	(3.18)

DRIVE-STANDARD DEVIATION

Model Constant	Distance Range (RANG-D)	% Stations Outside Zone (PSTO-D)	Station Density (SI/Z-D)
1.16196	0.56385	0.75977	-0.20547
	(3.53)	(1.43)	(1.07)

[a]Travel times for all access models are expressed in minutes.

[b]Computed t values for included variables.

TABLE 7

PREDICTIVE ACCURACY OF ACCESS DRIVE MODELS

Model	R^2	Standard Error of Estimate (Minutes)	Average Absolute Error (Minutes)	Percent Average Error	Mean Zonal Value (Minutes)	U	U^M	U^S	U^C
Drive-Mean	.963	1.14	.82	-4.0	5.4	.0143	.0000	.0094	.9906
Drive-Standard Deviation	.439	.90	.67	-29.3	1.9	.0397	.0000	.1787	.8212

F value for mean model = 390.1
F value for standard deviation model = 12.0

An examination of the values of the R^2 indices shows that the mean model is quite high (.963), while the standard deviation model is again much lower (.439). This was expected for the standard deviation model, due to the relatively small sample size (for standard deviation estimation) and wide range of trip times.

The standard error appears to be quite reasonable for both models, 1.14 minutes and .9 minutes, respectively. However, these values are actually about 21 and 47 percent of their mean zonal value.

As implied by the percent average error, the models tend to underestimate the walking time and its standard deviation. This was considerable for the standard deviation model.

From the Theil U test, the reason for most of the error appears to be unequal covariation of the actual and estimated values of the dependent variable with the explanatory variables; for both models over 99 percent and 82 percent, respectively, of the value of Theil U is contributed by U^c. Only a minor part of the error (U^s) is due to unequal variances of actual and estimated values of the dependent variable.

The residual plots appear in Figures 5 and 6. These are plots of the estimated dependent variables

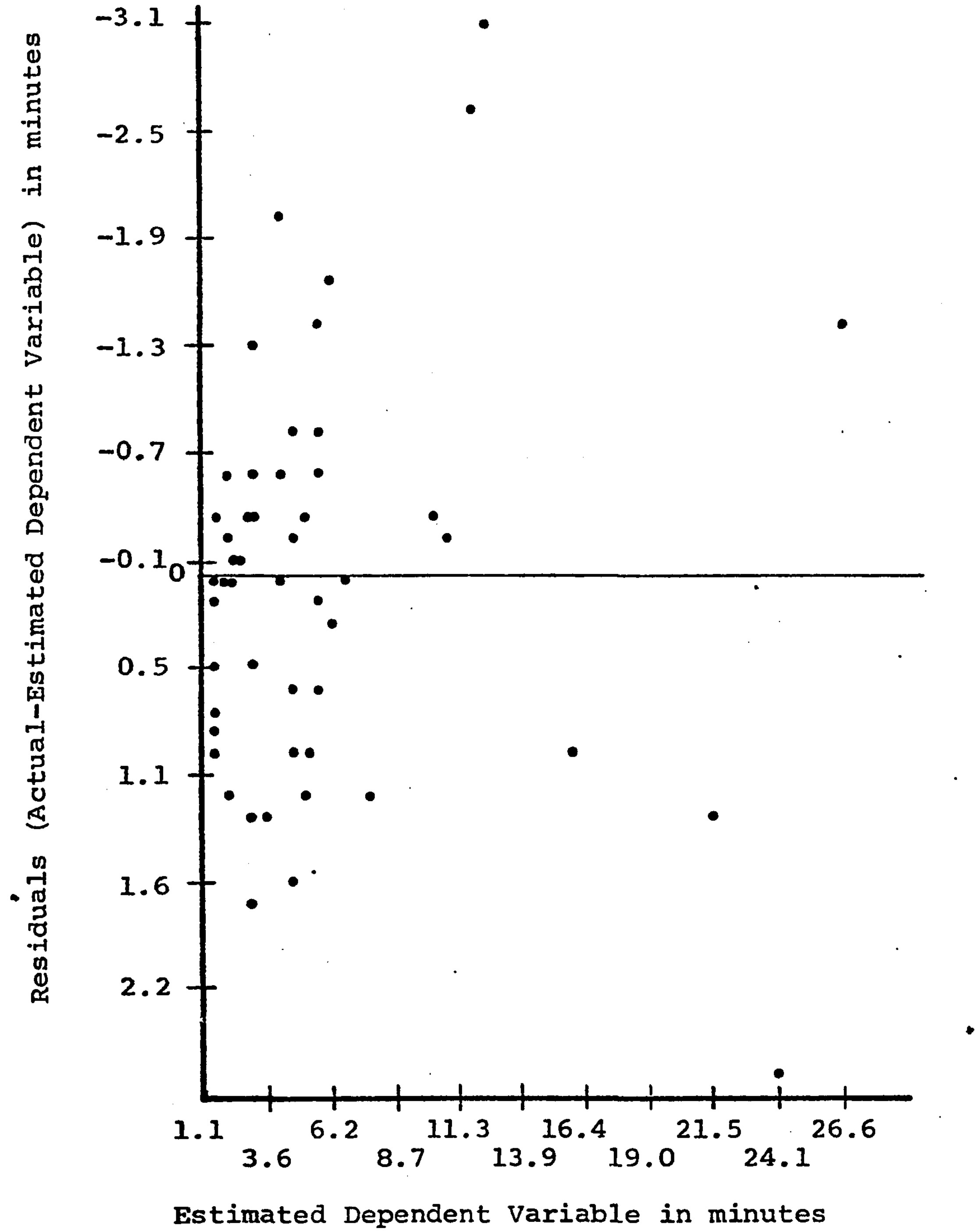

Fig. 5.--Plot of Residuals for Drive Mean Model

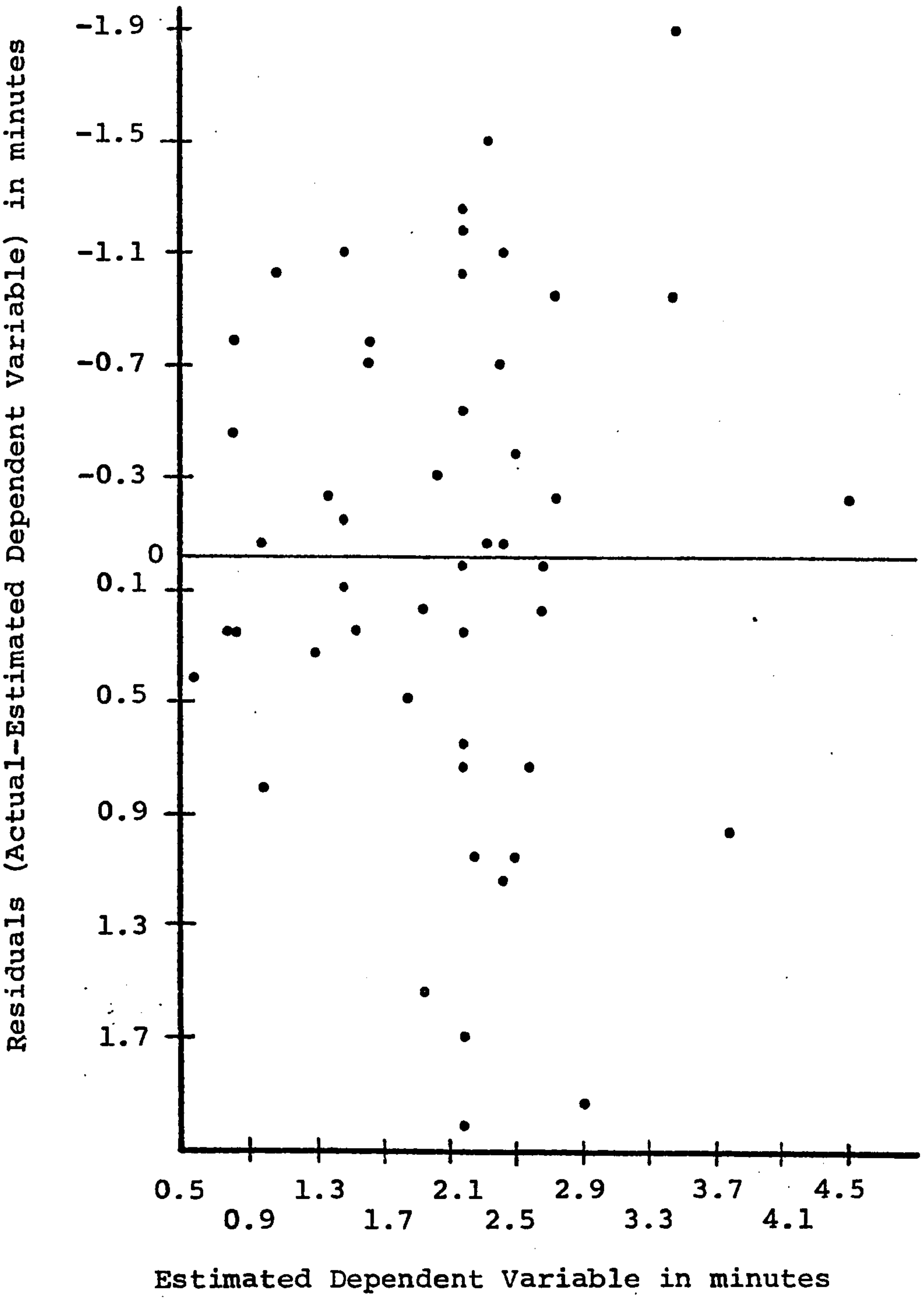

Fig. 6.--Plot of Residuals for Drive Standard Deviation Model

versus the model errors. The mean model residuals show a variance that is constant. The width of the band does not indicate a large variance of the error either. It is a good plot. The standard deviation plot, however, indicates a relatively larger band width with a slightly conical pattern.

As a result of this evaluation the mean model appears to be well estimated while the standard deviation model has a few problems. These problems arose from the relatively small sample size and wide range of trip times. The examination of the plots seems to indicate that the least squares analysis is not invalidated for either model. That is, the least squares assumptions appear not to be violated, even though there is some slight concern for the standard deviation model. In sum, on the basis of the observed data, we cannot conclude that the assumptions are incorrect.

The direction for future research would be to consider enlarging the sample size and utilize a data source that better represents all potential travelers. The data should be made available on the traffic engineering (e.g., street signalization and speed limits) characteristics of a zone and incorporated into the models. Also, actual traffic volume counts would be valuable.

These ideas should improve the models significantly and also make them very useful for travel demand analysis and forecasting.

CHAPTER VI

DEVELOPMENT OF ACCESS BUS MODEL

In this section the mean and standard deviation access bus models are developed. The variables that were tested and those selected to comprise the model specification are discussed. Also, the evaluation of the models is presented by deriving a series of measures and plots to show their predictive accuracy.

Regression Analysis

In the bus access time models which are developed, the mathematical form of the model variables was again exclusively linear. This provides for models which are easily understood and applied.

It was specified that the explanatory variables tested for these models had an a priori relationship with one or both of the bus dependent variables. This analysis was done again prior to quantifying the variables.

The specification of the model and associated variables[1] that were tested for the mean and standard

---

[1]See Chapter III for explanatory variable descriptions.

deviation follows:

$$\text{Access Bus Time}_j = S(\text{Zone Size variables}_j, \text{Transportation System variables}_j, \text{Volume variables}_j) + e_j$$

where:

j = origin zone

S = supply model

e = error term

Zone Size variables

Zone Size

Average Distance

Transportation System variables

Accessible Stations/Zone

Accessible Stations Within Zone

Accessible Stations Outside Zone

Percentage Stations Outside Zone

Station Density

Distance Range

Volume variables

Population Density

It was expected that a negative relationship (negative variable coefficient) would result for the explanatory variables: accessible stations/zone, accessible stations within zone, and station density. The remaining variables were expected to exhibit a positive relationship. Before the results are discussed, it should be indicated that the development of this model included a few assumptions that differed from the walk and drive models. When a traveler was assigned a bus access time, it was assumed that he went to the accessible station that provided him the shortest walking distance to the respective bus line. That is, the criteria used here to select a rail station were based on the distance a person had to walk to a bus stop, not the bus riding distance to a rail station. The walk and drive models were based on the total distance from home to rail station.

A further point to the above criteria was that a station providing direct bus service for a traveler was given a higher priority than a station requiring a transfer. However, this was true only if the walking distance to the bus line for the direct service station was not more than one-half mile greater than the walking distance to the bus line for the transfer service station. This assumption was used in deriving the value of the model dependent variable.

Another assumption states that a traveler is not to be assigned a bus time if another accessible station exists that is closer than the nearest bus line[1].

Turning next to the estimation of the coefficients, ordinary least squares regression analysis was performed. The selected models are shown in Table 8 with the Student t values for the selected variables. The variables shown for the bus mean model were selected each time the analysis was performed, regardless of which additional variables were being tested. The station density (SI/Z-B) variable did not show a very strong indication of significance, but was selected due to its contribution in reducing the models standard error of estimate.

For the standard deviation model the final computer analysis included the distance variable, in addition to the model selections of distance range and station density. However, this variable had the wrong coefficient sign, lacked significance, and was not included in the model. The t value for SI/Z-B (station density) also did not give a very strong indication of significance, but contributed to reducing the standard error of estimate.

The regression analysis was run several times for each model, prior to the final analysis, in order

---

[1] See Chapter VII for further explanation.

TABLE 8

ACCESS BUS MODELS[a]

BUS-MEAN

Model Constant	Population Density (DEN)	Average Distance (DIST-B)	Station Density (SI/Z-B)
-0.0542	0.63533	4.30692	-0.36614
	(2.24)[b]	(36.3)	(1.33)

BUS-STANDARD DEVIATION

Model Constant	Distance Range (RANG-B)	Station Density (SI/Z-B)
1.37795	1.24082	-0.28514
	(5.41)	(1.37)

[a]Travel times for all access models are expressed in minutes

[b]Computed t values for included variables.

to test all variables and select the most representative equation.

Model Evaluation

Measures of the predictive accuracy of the mean and standard deviation models are tabulated in Table 9. An examination of the values of the R^2 indices shows that the mean model is quite high (.974), while the standard deviation model is much lower (.465). This was expected for the standard deviation model, due to the relatively small sample size (for standard deviation estimation) and wide range of trip times.

The standard error of estimate appears to be reasonable for both models, 1.56 minutes and 1.27 minutes, respectively. These values are about 17 and 51 percent of their mean zonal value.

As implied by the percent average error, the models tend to underestimate the bus riding time and its standard deviation. This was again considerable for the standard deviation model.

From the Theil U test the reason for most of the error appears to be unequal covariation of the actual and estimated values of the dependent variable with the explanatory variables; for both models over

TABLE 9

PREDICTIVE ACCURACY OF ACCESS BUS MODELS

Model	R^2	Standard Error of Estimate (Minutes)	Average Absolute Error (Minutes)	Percent Average Error	Mean Zonal Value (Minutes)	U	U^M	U^S	U^C
Bus-Mean	.974	1.56	1.06	-4.7	8.9	.0120	.0000	.0066	.9934
Bus-Standard Deviation	.465	1.27	.88	-30.9	2.5	.0441	.0000	.1890	.8110

F value for mean model = 547.8
F value for standard deviation model = 19.6

99 percent and 81 percent, respectively, of the value of Theil U is contributed by U^c. Only a minor part of the error (U^s) is due to unequal variances of actual and estimated values of the dependent variable.

An examination of the residual plots in Figures 7 and 8 appears to indicate some slight violations of the least squares assumptions. Figure 7 roughly exhibits a concave shape which could possibly indicate the need for extra terms in the model (e.g., square or cross product terms).

The standard deviation model plot in Figure 8 also indicates a variance that is increasing (conical shape), implying the possible use of weighted least squares analysis. The underestimation error is apparent by the slight increase in slope toward the negative residuals.

As a result of this evaluation the mean model appears to be suitably estimated while the standard deviation model has a few problems. The examination of the plots seems to indicate that the least squares analysis is somewhat invalidated for both models. That is, the least squares assumptions appear to be partially violated. However, it could not be clearly specified in exactly which way they were violated.

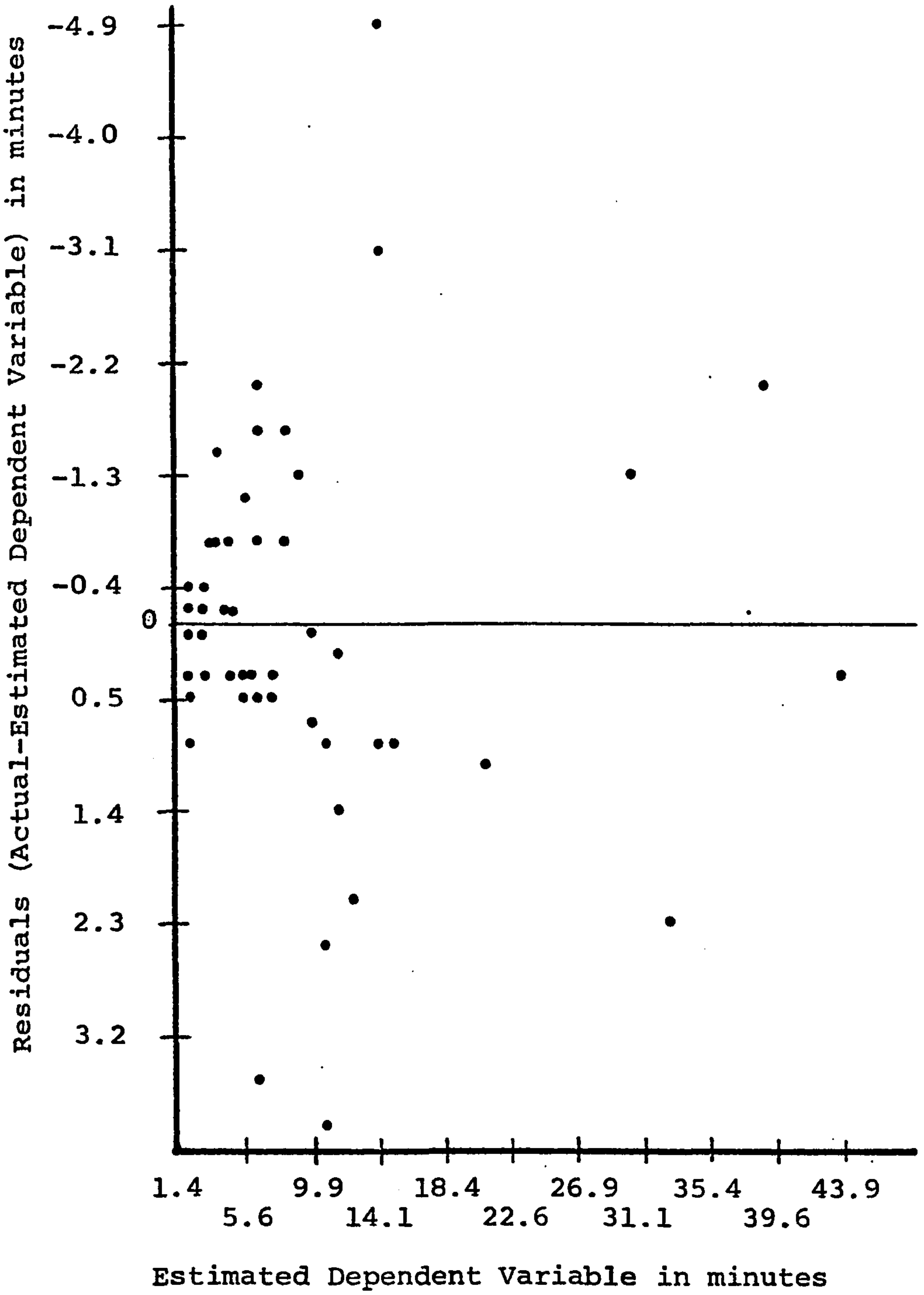

Fig. 7.--Plot of Residuals for Bus Mean Model

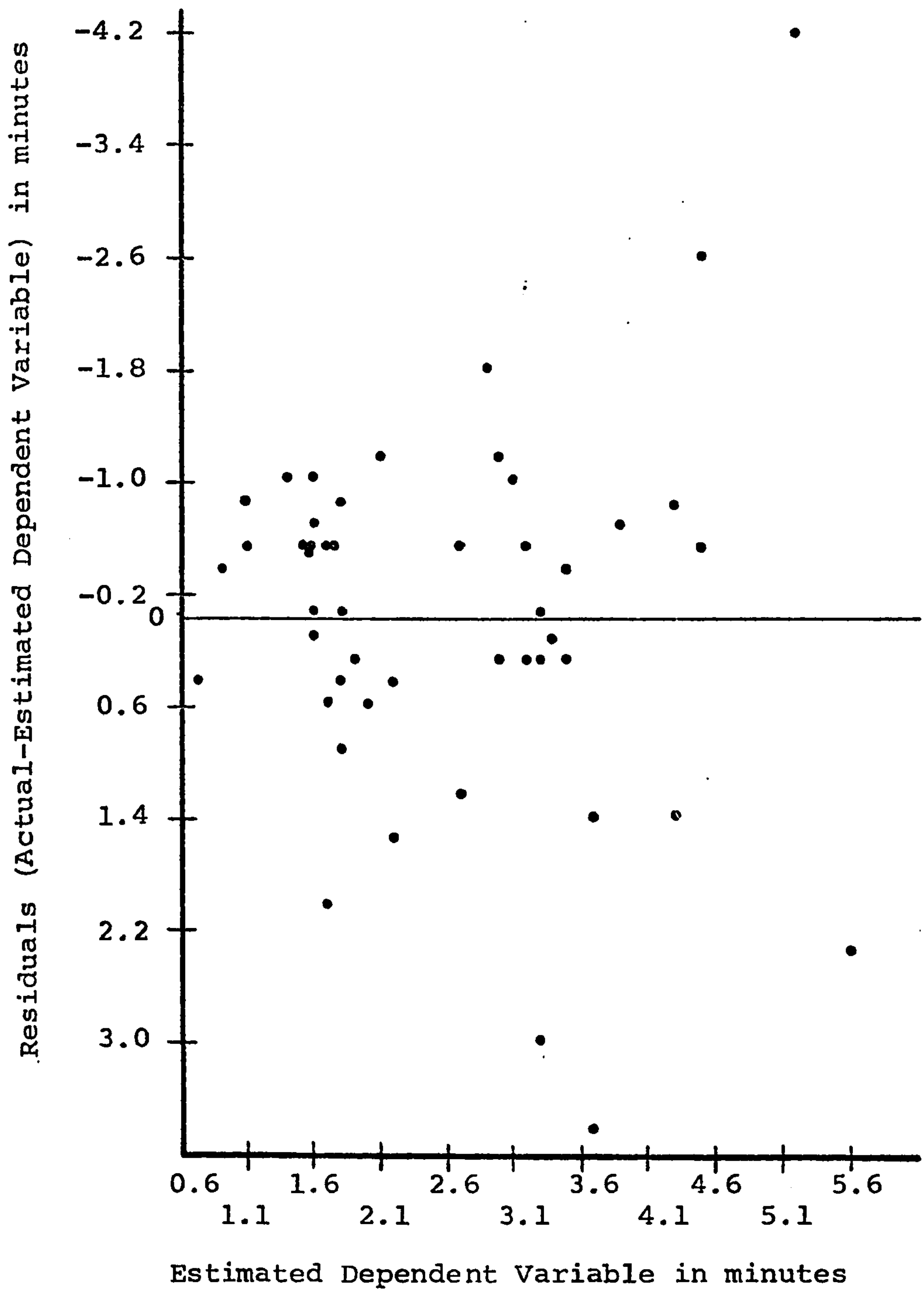

Fig. 8.--Plot of Residuals for Bus Standard Deviation Model

The direction for future research would be to consider enlarging the sample size and utilize a data source that better represents all potential travelers. Also, if data were available on the acceleration and deceleration times of buses, along with the average number of stops in a zone, an improved average bus speed could be determined. The number of passengers per bus stop should also be included in the model.

These concepts should improve the models and eliminate most of the error. However, even as the models stand now their accuracy is quite adequate.

CHAPTER VII

DEVELOPMENT OF ACCESS BUS/WALK MODEL

In this section the mean and standard deviation access bus/walk models are developed. The mean model estimates the walking time to a bus stop. The variables that were tested and those selected to comprise the model specification are discussed. Also, the evaluation of the models is presented by deriving a series of measures and plots to show their predictive accuracy.

Regression Analysis

In the bus/walk access time models which are developed, the mathematical form of the model variables was entirely linear. This provides for models which are easily understood and applied.

It was specified that the explanatory variables tested for these models had an a priori relationship with one or both of the bus/walk dependent variables. This analysis was done prior to quantifying the variables. Invalid relationships with the dependent variables are, therefore, not anticipated.

The specification of the model and associated variables[1] that were tested for the mean and standard deviation follows:

$$\text{Access Bus/Walk Time}_j = S(\text{Zone Size variables}_j, \text{Transpotation System variables}_j) + e_j$$

where:

j = origin zone

S = supply model

e = error term

Zone Size variables

<u>Zone Size</u>

Transportation System variables

<u>Route Miles</u>

<u>Route Mile Density</u>

<u>Bus/Walk Zone Coverage</u>

Volume related variables were not assumed to affect the walking time to a bus stop.

It was expected that a negative relationship (negative variable coefficient) would result for the three transportation system explanatory variables. The zone size variable was expected to exhibit a positive

---

[1] See Chapter III for explanatory variable descriptions.

relationship.

The development of this model included an assumption that was briefly stated in the previous chapter. This assumption indicated that a traveler walks to a bus stop only if this walking distance is less than the distance to another accessible station, not served by a bus line. In other words, it was assumed that a traveler will not walk to a bus stop, and take a bus to a rail station, if another station is closer than the bus stop. These travelers, therefore, were not assigned a walking time to a bus stop, and were deleted from the models trip table. However, this situation occurred only in a small number of cases.

Turning now to the estimation of model coefficients, ordinary least squares regression analysis was performed. The selected models are shown in Table 10 with the Student t values for the selected variables. The final computer analysis for the mean model included the zone size variable in addition to those shown. However, this variable had the wrong sign (negative rather than positive), lacked significance and was not included in the model. The t value for RT-MI (route miles) did not show very strong significance, but was included because it had the correct sign and helped reduce the

TABLE 10

ACCESS BUS/WALK MODELS[a]

BUS/WALK-MEAN

Model Constant	Route Miles (RT-MI)	Zone Coverage (COV-BW)
7.7247	-0.11429	-3.46991
	(0.926)[b]	(4.02)

BUS/WALK-STANDARD DEVIATION

Model Constant	Zone Size (ZSIZE)	Route Miles (RT-MI)	Zone Coverage (COV-BW)
3.26964	0.64415	-0.36119	-1.50782
	(3.29)	(2.36)	(1.22)

[a]Travel times for all access models are expressed in minutes.
[b]Computed t values for included variables.

standard error of estimate.

The variables shown for the standard deviation model were selected each time the analysis was performed, regardless of which additional variables were being tested. The zone coverage (COV-BW) variable was included, despite the lack of a strong t value, because it also possessed the correct sign for the coefficient and reduced the standard error.

The regression analysis was run several times for each model, prior to the final analysis, in order to test all variables and select the most representative equation.

Model Evaluation

Measures of the predictive accuracy of the mean and standard deviation models are tabulated in Table 11. An examination of the values of the R^2 indices shows that the mean model is quite low (.346), while the standard deviation model is much higher (.747). It appears that the reason the coefficient of determination (R^2) was low for the mean model stems from the fact that the variance of the actual dependent variables was also very low (2.89). That is, as the explanatory variables were varied from zone to zone, the dependent variable remained relatively constant. Therefore, a lack of fit resulted.

TABLE 11

PREDICTIVE ACCURACY OF ACCESS BUS/WALK MODELS

Model	R^2	Standard Error of Estimate (Minutes)	Average Absolute Error (Minutes)	Percent Average Error	Mean Zonal Value (Minutes)	U	U^M	U^S	U^C
Bus/Walk-Mean	.346	1.38	.95	-6.0	4.8	.0288	.0000	.2595	.7405
Bus/Walk-Standard Deviation	.747	.85	.61	-23.5	2.1	.0325	.0000	.0457	.9543

F value for mean model = 10.6
F value for standard deviation model = 36.7

The reason that this standard deviation model had a higher R^2 than the walk, drive or bus standard deviation models, was due to the much smaller range of trip times.

The standard error of estimate was considered reasonable for both models, 1.38 minutes and .85 minutes, respectively. Actually these values are about 29 and 40 percent of their mean zonal value. Also, as implied by the percent average error, the models tend to underestimate the bus/walk time and its standard deviation.

For the Theil U test, it can be seen that most of the error in the standard deviation model appears to be from unequal covariation of the actual and estimated values of the dependent variable with the explanatory variables; over 95 percent of the value of Theil U is contributed by U^c. Only a minor part of the error (U^s) is due to unequal variances of actual and estimated values of the dependent variable. The Theil U test for the mean model shows a much larger portion of the error due to U^s.

An examination of the residual plots in Figures 9 and 10 appear to indicate some abnormalities only for the mean model. The mean model plot in Figure 9 indicates an error variance that is increasing (conical shape),

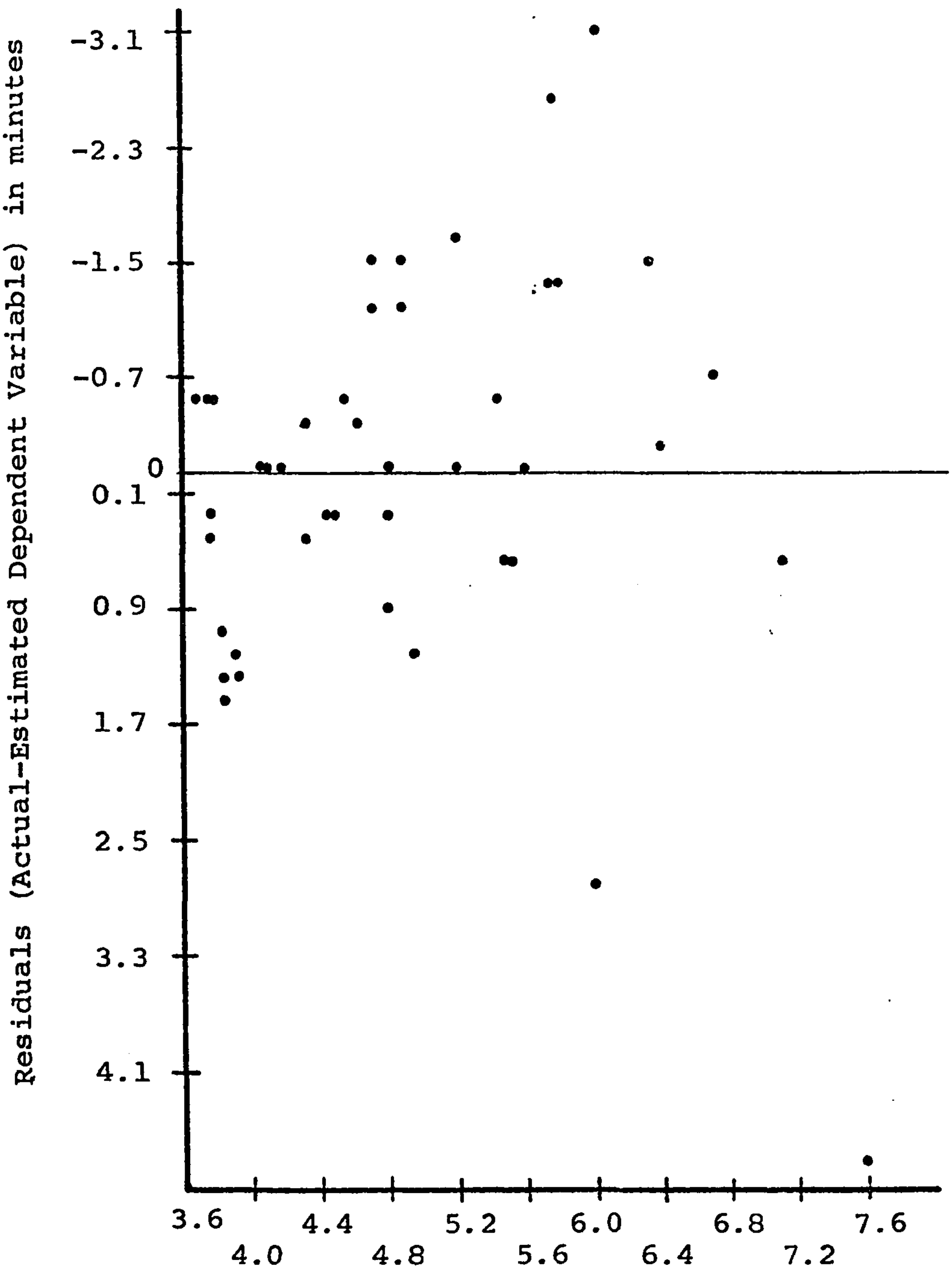

Fig. 9.--Plot of Residuals for Bus/Walk Mean Model

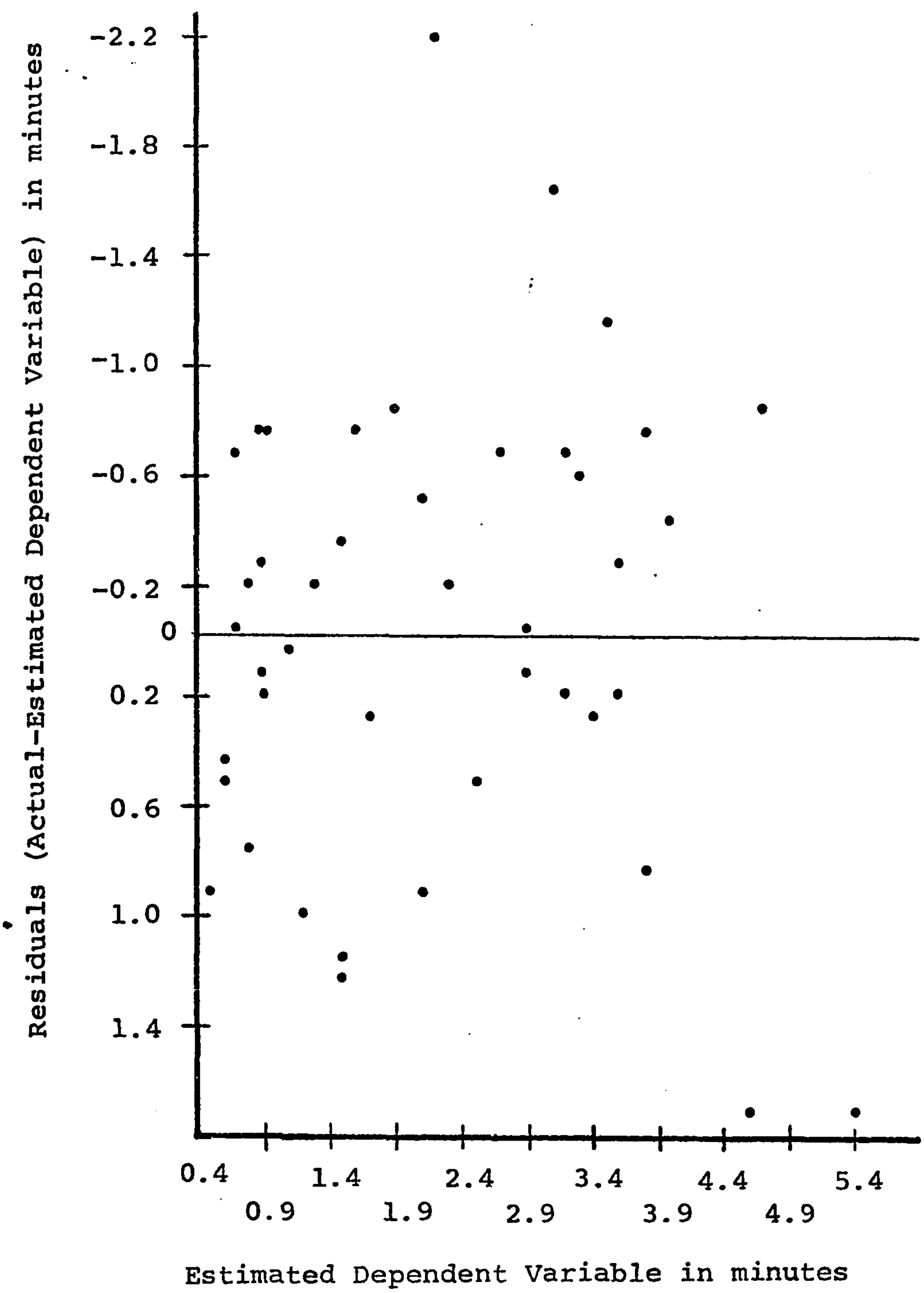

Fig. 10.--Plot of Residuals for Bus/Walk Standard Deviation Model

implying the possible use of weighted least squares analysis.

The standard deviation plot in Figure 10 shows a band width which does not indicate a large variance. However, the plot is inclined upward, but only a little. This upward inclination of the model is confirmed by the systematic underestimation of the computed dependent variable.

The examination of the plots seems to indicate that the least squares assumptions are somewhat invalidated for the mean model. However, it could not be clearly determined in exactly which way and why they were violated. On the whole, if we exclude the effect of a few outlying points in the plots, we cannot conclude that the assumptions are incorrect.

The direction for future research would be to again consider enlarging the sample size and utilize a data source that better represents all potential travelers. Also, it could be significant to include in the model the specific locations of actual bus stops. This model assumed a bus stop existed at all major intersections. Therefore, actual trip distances were not always represented.

These ideas should improve the models and their accuracy. However, the developed models that estimate the walking time to a bus stop and its standard deviation are sufficiently accurate to warrant their further evaluation by application in practice.

CHAPTER VIII

SUMMARY OF FINDINGS AND IMPLICATIONS OF THE STUDY

This study estimated supply models for the access portions of rail and bus trips. The models were designed to predict aggregate zonal travel times as a function of transportation system, zone size, and volume related characteristics of a zone.

The motivation for this study came from existing aggregate travel demand model analysis. Background research in this, and the lack of current methods of supply model estimation, clearly showed a need to develop a technique that would produce level of service values for the access trip segments. The background research revealed that transit access times are extremely important in forecasting travel demand. Further, a technique to estimate these values would accomplish three things: (1) improve the accuracy of demand estimates, (2) reduce the difficulty and expense in compiling level of service variables, and (3) help to reduce the bias currently present in travel estimates.

It was necessary to investigate which level of service variables should or could be analyzed, and which access modes to estimate. This led to selecting access time for walk, drive, and bus submodes. It was shown that these are the critical items to be studied.

The general findings related to this research are as follows: First, the estimation of these relatively simple types of access models can be produced with reasonable accuracy. This was an objective of this study. Also, the results show that, while error is present, a relationship clearly exists between the zonal characteristics (e.g., zone size, accessible rail stations, average distance, and density) and the level of transportation service in a zone.

Second, the criterion that was used to estimate the supply models was the basic ordinary least squares procedure. It was thought initially that an alternative estimation technique, such as constrained least squares or weighted least squares would need to be used. These procedures would handle the problems of highly collinear variables and heteroscedasticity. However, after applying the ordinary least squares method to the data set, it was realized that collinearity was not a major problem. As described earlier, the weighted least squares method was tried, with little success. The conclusion was that these alternative procedures could not improve the models, nor were they actually necessary for this research.

The third general conclusion relating to the development of access supply models is suggested from the analysis of the sample data. The source of useable data did not consist of a home interview survey, which would have better represented all potential travelers in a given zone. Therefore, to compensate for this, the tripmaking of observed transit riders for alternative modes had to be synthesized. If better data were available access supply models could be developed that were more accurate. Until then, the proper alternative modes or stations should be hypothesized.

The specific findings that are related to the access models have to do with their validity. The coefficients of determination (R^2) for the walk, drive, and bus travel time mean models were very high. This was significant because the explanatory variables that were tested were reasoned a priori to have a relationship with the appropriate dependent variable. The bus/walk (walk to bus stop) mean time model, however, had a relatively low R^2 value. This was because a lack of variation existed in the observations. Again, a good home interview study would improve these results.

To improve the drive models performance additional variables could be included. It would be interesting to test the traffic engineering characteristics of a zone. Also, traffic volume counts would provide a better speed input. The bus model could be improved by specifying variables that better describe the operation of a bus system. These include acceleration and deceleration characteristics, number of bus stops per zone, and average volume of passengers. The bus/walk model should consider the specific locations of actual bus stops when determining trip distances.

The values of the error indicators, e.g., percent average error, average absolute error, and others for the mean models indicate that the forecasts might be subject to some uncertainty. The standard errors comprise between 15 and 29 percent of the average zonal access time. On the average zone the percent error ranges between 2 and 6 percent and the absolute error about 1 to 4 minutes.

The standard deviation models exhibit some larger errors due to the relatively small sample size and the wide range of individual access times. The standard errors range between 34 and 51 percent of the mean zonal value. For the average zone the percent error is between 18 and 30 percent. However, the average absolute error is only about 1 to 2 minutes.

The important question is, should these models be used in actual planning and analysis despite the errors and uncertainties that are present. It was discussed earlier that models of this type are presently not available and the approach being used to quantify level of service variables is undesirable. It would certainly be valuable to test the practicality and accuracy of this method on the ability of demand models to improve upon forecasting urban passenger travel. This empirical study could also readily be applied to mode split analysis, preparation of transportation networks, and other special studies.

The implications of utilizing these models or this procedure could lead to great reductions in manpower and money. Most transportation studies involving models are faced with the necessity of computing level of service variables. This is an extremely difficult task that is often done carelessly or done in a biased fashion. It is also a task that is basically done by hand, taking large amounts of time. The possible improvements to travel forecasting models from utilizing these types of supply models, could be significant. Considerable cost savings might also eventually result from better policy recommendations.

Given the primative state of aggregate access supply modeling, the study reported here can be considered successful. Supply models were estimated at reasonable effort and cost.

BIBLIOGRAPHY

Domencich, T., Kraft, G., and Valette, P., "Estimation of Urban Passenger Travel Behavior: An Economic Demand Model", Highway Research Record 238. Washington, D.C.: Highway Research Board, 1968.

Draper, N.R., and Smith, H., Applied Regression Analysis. New York: John Wiley and Sons, Inc., 1966.

Huber, M.J., Boutwell, H.B., and Witheford, D.K., "Comparative Analysis of Traffic Assignment Techniques with Actual Highway Use", Highway Research Record. Washington, D.C.: Highway Research Board, 1968.

Irwin, N.A., Dodd, N., and Von Cube, H.G., "Capacity Restraint in Assignment Programs", Highway Research Board Bulletin 297. Washington, D.C.: Highway Research Board, 1961.

Johnston, J., Econometric Methods. New York: McGraw-Hill, 1972.

Kraft, G., and Wohl, M., "New Directions for Passenger Demand Analysis and Forecasting", Transportation Research, Volume I, Number 3, 1967.

Manheim, M.L., Principles of Transport Systems Analysis, Professional paper P67-1. Cambridge: Massachusetts Institute of Technology, Volume XVI, 1967.

Manheim, M.L., Ruiter, E.R., and Bhatt, K.V., "Search and Choice in Transport Systems Analysis", Highway Research Record 293. Washington, D.C.: Highway Research Board, 1969.

Meyer, J., and Glauber, R., Investment Decisions, Economic Forecasting and Public Policy, Harvard Business School, Boston, Massachusetts, 1963.

Rassam, P., and Ellis, R., Access Characteristics Estimation System, Report to Department of Transportation, Office of High Speed Ground Transportation, Northeast Corridor Transportation Project, Volume I, Washington, D.C., December, 1969.

Smock, B., "A Comparative Description of a Capacity-Restrained Traffic Assignment", Highway Research Record 6. Washington, D.C.: Highway Research Board, 1963.

Snedecor, G.W., and Cochran, W.G., Statistical Methods. Iowa State University Press, Ames, Iowa, 1967.

Southward Transit Area Coordination Study, Final Report, Chicago, Illinois, 1970.

Talvitie, A., "Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models", Transportation Research Forum, October, 1973.

Talvitie, A., An Econometric Model for Downtown Work Trips, Ph.D. dissertation, Northwestern University, 1971.

U.S. Department of Commerce, Bureau of Public Roads, Traffic Assignment Manual. Washington, D.C.: Government Printing Office, June, 1964.