74-6969

LIOU, Peter Shek, 1945-
DISAGGREGATE ACCESS MODE AND STATION SELECTION MODELS FOR RAIL TRIPS.

The University of Oklahoma, Ph.D., 1973
Engineering, civil

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DISAGGREGATE ACCESS MODE AND STATION SELECTION MODELS FOR RAIL TRIPS

A DISSERTATION

SUBMITTED TO THE GRAUDATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

PETER SHEK LIOU

Norman, Oklahoma

1973

DISAGGREGATE ACCESS MODE AND STATION SELECTION MODELS FOR RAIL TRIPS

APPROVED BY

[illegible]

[illegible]

Hillel Kumin

Ronald R. [illegible]

DISSERTATION COMMITTEE

ACKNOWLEDGEMENT

I wish to express my sincere gratitude and appreciation to Dr. Antti Talvitie, my major professor and Chairman of the doctoral committee, for his continual guidance, encouragement, and many constructive criticisms during the preparation and development of this study.

Appreciation is also extended to Dr. Moshe Ben-Akiva of M.I.T. for supplying the computer program necessary for this work.

I am particularly appreciative of the encouragment and understanding of my wife, Mimi, which is of great value to me.

I also wish to thank Professor George Reid for supporting me through the early part of my research.

This research is partially supported by the Urban Mass Transit Administration of the U.S. Department of Transportation.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd.)

LIST OF TABLES

LIST OF ILLUSTRATIONS

DISAGGREGATE ACCESS MODE AND STATION SELECTION MODELS FOR RAIL TRIPS

By

PETER SHEK LIOU

In this study disaggregated probability choice models are developed for the access mode and for the access station selection. In each of these models, there are at least two alternatives from which an individual traveller is allowed to make a choice.

The mathematical model used in this study is the multinomial logit model based on the axiom of the "independence of irrelevant alternatives". It basically assumes that the odds of choosing one alternative over the other in a system containing both are independent of the presence of a third alternative in the system.

Two methods of approach concerning the travellers' decision-making processes are used in constructing the mode and station choice models. The first is the simultaneous approach which assumes that the access mode and the access station choices are made simultaneously. The second is the sequential approach which assumes the individual traveller may make the access mode and station choice decisions in one of two sequences: the station-mode sequence - station choice preceeding mode choice, or the mode-station sequence - mode choice preceeding station choice. In the case when the sequential assumption is adopted, the choices of the access mode and the access station are modeled separately.

The results of this study suggest that the travellers' decision-making process for the mode and station choices of an access

trip is behaviorally of a separate nature. In particular, it is in a sequence of the station choice followed by the access mode choice. The study also shows that the travellers do not value or assign the same weights to the set of transportation system attributes when making decisions of the access mode and station choices. Furthermore, it is learned from this research that the walk and the bus modes are the more preferred access modes as compared to the auto mode. For the station choice, the accessibility of the train station has the greatest influence over the traveller's station selection decision.

A small test was also conducted to investigate the sensitivity of the model to the way in which the "relevant alternatives" are defined. The results of this test, reported in Appendix III, suggest that the coefficient of the model are significantly affected by the definition of the "relevant alternatives."

CHAPTER I

INTRODUCTION

The Study

In this study, disaggregated probability choice models are developed for the access mode and for the access station selection. In each of the models, there are two or more alternatives from which an individual traveller is allowed to make a choice; therefore the regular binary choice models such as the logit and probit models are insufficient to describe the systems under investigation. The model employed in this study is the multinomial logit model.

A person travelling the major part of his journey by rail generally has to make a number of choices. These choices include the selection of the stations he intends to board and alight from the train; the choice of travel modes to go to and from the stations; and so forth. These decisions of stations and modes have a bearing on transportation planning. In particular, it is of interest what types of transportation and socio-economic attributes affect travellers' choices and how much.

A basic assumption in formulating the access mode and station selection models is that all the individuals who have chosen travel by rail rationalize their choices of access mode and station by choosing that alternative whose utility is the highest. The utility of an alternative travel choice is represented in terms of the attributes of the alternative, e.g. travel time, travel cost, and the socio-economic attributes of the individual, e.g. income.

There are two aspects of objectives in this study. The first is the modelling strategy aspect. The second is the transportation planning aspect.

In the modelling strategy aspect, there are three objectives. The first objective is to find out whether the choices of the access mode and the access station are made simultaneously, or are they made sequentially, e.g. the mode first and then the station, or vise versa. The second objective is to find out whether the individual traveller values the various attributes which influence choice differently when making his decision. The third objective is to study the techinque and the application of the multinomial logit model.

In the transportation planning aspect, there are two objectives. The first objective is to develop disaggregate access mode and access station choice models, providing the basis for the projection of future demand by various modes to various stations. These disaggregate choice demand models should be applicable to any of the metropolitan and urban areas even though these models are formulated on Chicago data.

The second objective is to supply information to the transportation planners which transportation characteristics are most important for the access/egress part of the rail transit trips.

CHAPTER II

EXISTING MODE CHOICE MODELS AND THE MODELS FOR THE PRESENT STUDY

Background: Current Methods and Models

During the last decade, some work has been done in applying the various types of probability choice models on travel mode decisions. Most of the early work is based on aggregated data. In these models the mode choice "probabilities" are expressed in terms of shares; the share being equal to the pereentage of people choosing a certain mode. A good example of these models is the TRC[1] mode split model in which curves were developed for the share of interzonal person trips choosing the transit mode from among two alternative modes, auto and transit. In this model, the modal shares are looked upon as being influenced by three system attributes: the ratio of the total travel time by the two modes; the ratio of the travel cost by the two modes; the ratio of the excess or out-of-vehicle time by the two modes; and also by the travellers' socioeconomic variable, the median income. The trip purpose (work, non-work) was also believed to have a certain effect on the modal choice.

In the TRC model as well as in many other models, all the observations are aggregated to the traffic zones. The models appear

---

[1]Deen, T. B., Irwin, N. A. and Mertz, L. "Application of A Modal Split Model to Travel Estimated for the Washington Area", *Highway Research Record 38*, pp. 97-123 (1963).

to be sensitive to zoning and can hardly be transferred geographically to apply to any other area.

In recent years, researchers have formulated a number of disaggregated choice probability models which relate an individual's choice behavior to the alternative travel attributes and the user characteristics. These models can be grouped into two categories, the binary choice models and the multinomial choice models. The former deal with situations in which an individual has only two alternatives from which to make a choice, the latter when two or more than two alternatives exist.

In the disaggregate binary choice models the earliest work was done by Warner[1], who used the methods of discriminant and modified regression analyses to arrive at an S-shaped curve to assign probabilities of choice as a function of the variables affecting the choice. Other disaggregate binary choice models include the probit[2] and logit[3] analyses. These methods, as did Warner's, also employ an S-type curve, which is cumulative normal or approximately so in shape, to model the probabilities of choice.

---

[1]Warner, S. L., "Stochastic Choice of Mode in Urban Travel", *Northwestern University Press* (1962).

[2]Lisco, T. E., "The Value of Commuter's Travel Time: A Study in Urban Transportation", *PhD Dissertation,* Department of Economics. University of Chicago (1967). Lave, C. A., "Modal Choice in Urban Transportation: A Behavioral Approach", *PhD Dissertation,* Department of Economics, Standord University (1968).

[3]Stopher, P. R., "A Probability Model of Travel Mode Choice for the Work Journey", *Highway Research Record 283,* pp. 57-65 (1969)

Common to all these disaggregate mode choice studies is their analysis approach. This approach is based on the behavior of the individual traveller. The behavior is assumed to be influenced by the attributes of the alternative choices and the attributes of the individual himself. Thus, the individual looks at the attributes of travel choices, such as the travel time and travel cost, as a "price" he has to pay in order to use either transportation service. He then assigns weights to each attribute based on his own preferences, which in turn are supposed to be influenced by his income, and subsequently chooses that mode which requires the least amount of "combined prices", or in other words, which provides the most utility to the traveller.[1]

Mathematically the analysis problem is to relate the choice probability, P(X), to a utility function, U(X,b) where X is a $K \times 1$ vector of explanatory variables describing all mode and user attributes and b is a $1 \times K$ vector of their coefficients:

$$U(X,b) = b_0 + b_1X_1 + b_2X_2 + \dots \qquad (2\text{-}1)$$

The mode characteristic variables in the utility function are usually expressed as the difference of the attributes of the two modes, but they need not be so. When the choice model is estimated, the marginal utility, or the coefficients, of the attributes are inferred from the observed data on the actual choice made. The value of P(X) is taken to be either 1 or 0 depending on whether the

[1]It is realized that the mode choice models of the type discussed here are actually conditional probabilities for choosing a mode given that a certain destination was chosen and that the trip was in fact made. Whether this is an appropriate probability to model from the viewpoint of travel behavior is not addressed.

mode is chosen (1) or not chosen (0). In general, the values of the variable coefficients are estimated by the maximum likelihood method.

The logit model eventually emerged as the most popular disaggregate binary choice model. Part of the reason is due to the fact that the binary logit model can be extended to the multilogit model which is capable of handling systems with multiple travel choice alternatives.

The logit analysis tries to fit a logistic curve from a set of observations. The logit model has the form,

$$P(X) = \frac{e^{U(X,b)}}{1 + e^{U(X,b)}} \tag{2-2}$$

To cite an extremely simplified example, suppose the individual's choice of travel mode is solely influenced by one attribute, the travel time for instance; then the utility function is a function of only one independent variable, i.e.:

$$U(t,b) = b(t_A - t_C) \tag{2-3}$$

where b is the marginal utility of the attribute; t_C is the travel time by the chosen mode and t_A is the travel time by the alternative mode.

By substituting Eq. (2-3) into Eq. (2-2), it can be noted that Eq. (2-2) has the following three characteristics:

Firstly, as the travel time of the chosen mode, t_C, becomes increasingly greater than that of the alternative mode, t_A; the chosen mode becomes more and more undesirable and therefore the probability of choosing it, $P_C(t)$, approaches zero.

Secondly, conversely when t_A is much greater than t_C, then mode C becomes more attractive than mode A and therefore $P_C(t)$ approaches 1.

Thirdly, when the travel time by the two modes are exactly equal to each other, there can be no preference as to the choice of travel mode. Therefore, $P_C(t)$ is expected to be .50.

Graphically, the logistic curve has an S-shape and are rotationally symmetric along the axis where the difference of the attributes of the two modes is zero (Fig. 1).

The binary choice models are quite restricted in their application. In an urban area there often exist more than two modes of travel. In order to study the modal choices in such a circumstance, transit modes are customarily lumped together to preserve the binary choice situation. However, most often the several transit modes represent real alternatives and may not be well lumped together. Thus, a multinomial choice model is needed to conduct an adequate modal choice analysis. Of course, if also the other travel choices, trip frequency, choice of destimation etc., are modeled then a multinomial choice model is an absolute necessity.

A number of researchers[1] have used various multinomial logit models to describe situations with multiple alternatives. The existing multinomial logit models are very similar to one another.

---

[1]Theil, H. A., "A Multinomial Extension of the Binary Logit Model", *Rept. 6631 (1966)*. Stopher, P. R., *A Multinomial Extension of the Binary Logit Model for Choice of Mode of Travel*, University of Chicago. Unpublished paper, Northwestern University (1969). Rassam, P. R. *et al.*, "The N-Dimensional Logit Model: Development and Application", *Highway Research Record 369*, pp. 135-147 (1971).

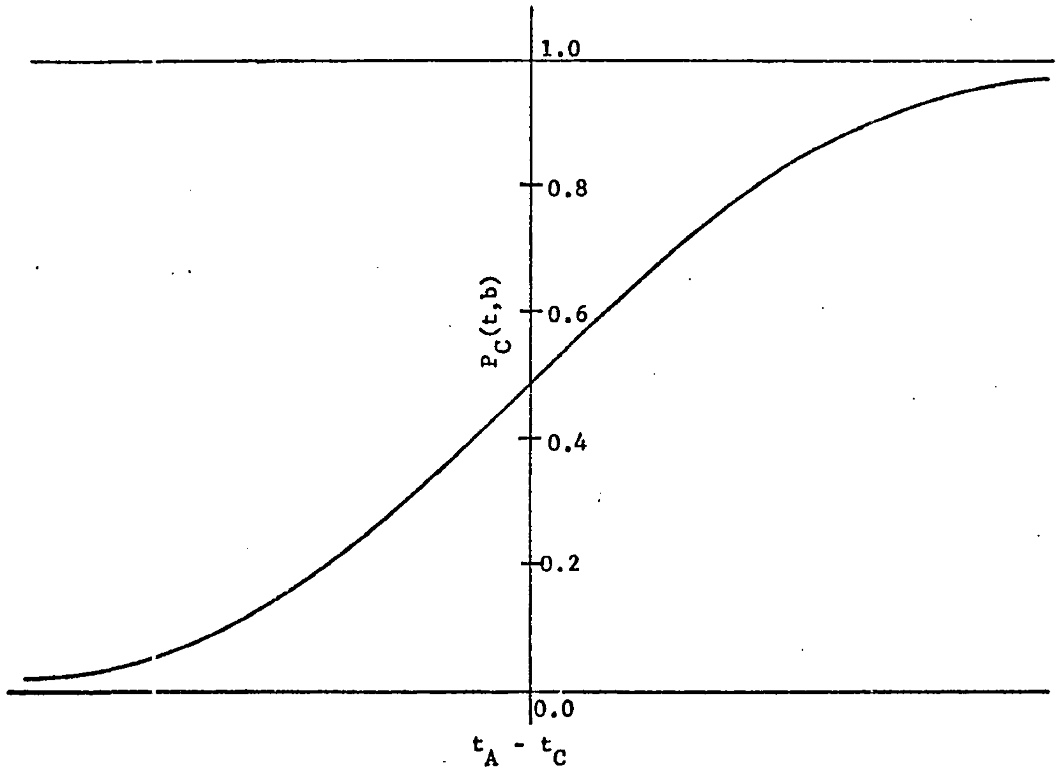

FIGURE 1. Logit Curve

They all have the following mathematical expression

$$P_i = \frac{e^{U_i}}{\sum_j e^{U_j}} \qquad (2\text{-}4)$$

where U_i is utility function of alternative i, and $i \varepsilon \{j\}$.

The various multinomial logit models are different in the approach and the initial assumptions in obtaining the general form, Eq. (2-4). Consequently, they are usually associated with their own peculiar restrictions.

The multinomial logit model which is of interest in this study is based on the "Independence of Irrelevant Alternatives" assumption.[1] This model was chosen because it allows individuals to have a different number of (unranked) alternatives to be considered in each choice.

In addition to the regular assumptions that the choice probability of a certain alternative has to be between 0 and 1, and that the sum of the probabilities for all the alternatives is unity, this model assumes the odds that one alternative will be chosen over the other in a system containing both are independent of the presence of a 'third alternative' in the same system. It is this condition that allows the functional models to be applied to any number of modes of transportation alternatives for the various individual travellers in the system.

---

[1]McFadden, D., *The Revealed Preferences of a Government Bureaucracy*, University of California, Berkeley (1968).

The Axiom of Independence of Irrelevant Alternatives

Let i and j be two elements in a set A which contains elements other than i and j; and let P(i:A) denote the probability that i is chosen from the set A. Furthermore, let A' be a subset of set A such that A' contains elements i and j only, i.e.; A'εA and A' = {i,j}; then the independence of irrelevant alternatives assumption can be expressed mathematically as follows:

$$\frac{P(i:A)}{P(j:A)} = \frac{P(i:A')}{P(j:A')} , \qquad (2\text{-}5)$$

where P(i:A') and P(j:A') are the probabilities of choosing i and j, respectively, in the binary set A'εA.

For set A', P(i:A') + P(j:a') = 1 and P(i:A')/P(j:A') = γ, where γ is a constant.

Suppose a new alternative, k, is introduced into A'; then $0 \leq \underline{P}(k:A') \leq 1$ and $\underline{P}(i:A') + \underline{P}(j:A') + \underline{P}(k:A') = 1$. The $\underline{P}$'s indicate the probabilities for the new A' set. Note that, if alternative k is competitive to the alternatives i and j, then $\underline{P}(i:A')$ and $\underline{P}(j:A')$ are expected to be smaller than P(i:A') and P(j:A'), respectively. However, according to the above theorem, the ratio of $\underline{P}(i:A')$ and $\underline{P}(j:A')$ in the new A' set remains the same as that in the old A' set, i.e.

$$\frac{\underline{P}(i:A')}{\underline{P}(j:A')} = \frac{P(i:A')}{P(j:A')} = \gamma \qquad (2\text{-}6)$$

Based on this assumption, McFadden derived his multinomial logit model.

$$P_i = \frac{e^{U_i}}{\Sigma_j e^{U_j}} \qquad (2\text{-}7)$$

where the utility function has a linear form:

$$U_i = B_1X_{i1} + B_2X_{i2} + B_3X_{i3} + \cdots \qquad (2\text{-}8)$$

In the above expression, the X's are the independent variables, or transportation attributes for person i, and the b's are their corresponding coefficients, or weights. A detailed derivation of McFadden's multinomial logit model is presented in Appendix I.

Application of the Independence of Irrelevant Alternatives Assumption to Travel Demand Modelling

The "independence of irrelevant alternative multilogit model" has been applied recently in travel demand studies.

The Charles River Associates has made a number of choice models for a shopping trip in which a traveller has to make a number of choice decisions in trip frequency, travel mode, time of day and destination, in that order.

Morshe Ben-Akiva, as a case study in his dissertation, has applied the multilogit model to formulate shopping-trip travel demand models. Two choices, mode and destination, are considered. Ben-Akiva suggested that the traveller's decision-making proeess of a number of choices pertaining to the trip can be modelled on two different behavioral assumptions. They are the sequential assumption and the simultaneous assumption. The sequential assumption states that the decisions of the various choices are made separately

---

[1]Charles Rivers Associates, Inc. (CRA), A Disaggregate Behavioral Model of Under Travel Demand, Federal Highway Administration, U.S. Department of Transportation, Washington, D. C. (1972).

[2]Ben-Akiva, M. E., Structure of Passenger Travel Demand Models, Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology (1973).

according to a certain sequence, as in the CRA study. The simultaneous assumption simply states that the choices are made simultaneously.

Ben-Akiva further pointed out that two more assumptions are necessary when constructing the sequential (recursive) models. The first assumption concerns the specific order of the choices. For example, for the choice of mode and destination, it is possible to assume that the choice of the destination is made before the choice of mode is made. In this case the joint probability of mode and destination is the product of a conditional probability of mode given destination and a marginal probability of destination, e.g.

$$P(m,d) = p(m|d) \cdot P(d) \quad , \qquad (2\text{-}9)$$

where m and d indicate mode and destination, respectively. On the other hand, if the mode choice decision is made first and then the destination choice, the joint probability of mode and destination is

$$P(m,d) = P(d|m) \cdot P(m) \quad . \qquad (2\text{-}10)$$

The second assumption necessary for the sequential modelling process is the separability of choice assumption. It states that an individual's conditional probability of, for example mode for given destination, is dependent only on those alternative model available to the person to go to the given destination, and independent of the alternative modes for other destinations. Ben-Akiva concluded in his study that theoretically the simultaneous models are more superior to the sequential models, and that empirically the coefficients are different depending on whether a simultaneous or sequential model is assumed. Furthermore, the choice of sequence also affects the coefficient estimates.

Formulation of the Access Mode and Station Selection Models

In this research, the access part of the work-journey by rail transit will be studied.

The formulation of the conditional probability models for access mode choice and station selection are based on the assumption that the choice of main mode (and destination and trip frequency) are of higher hierarchial order. Thus the traveler is first assumed to make a choice among modes (and destination and frequencies) and, then conditional that a certain rail line is chosen he will make a choice among alternative access modes and stations.

The formulation of the access mode and station selection models in this study will be approached from two different behavioral assumptions. The first is the sequential choice assumption which holds that station selection and access mode choices can be further decomposed into subsets in such a way that the traveller decides first on one choice, e.g. station, before making the choice on the other, e.g. access mode. The second assumption is the simultaneous choice assumption which holds that no such decomposition of access mode and station selection problem occurs but both are decided simultaneously.

The multinomial logit model which relates the choice probabilities of access mode, of the station, or both, of an individual traveller to the transportation system attributes has the general mathematical form expressed in Eq. (2-7). The utility functions, of course, differ from one another according to which choice the particular model pertains to.

The Sequential Assumption. As indicated this assumption holds that the decision regarding the two choices are made sequentially which in turn implies a two-level decision structure. The relevant question now is which choice comes first, that is, whether the mode choice or the station choice is made first. Apparently, both possible sequences can be justifiably assumed, and therefore both will be studied.

The joint probability of the access mode, m, and the access station, s, is the product of a conditional probability of the mode for predetermined stations and a marginal probability of the station choice[1], i.e.,

$$P(m,s) = P(m|s) \cdot P(s) \quad . \tag{2-11}$$

For the sequence identified above, i.e. station selection precedes access mode choice (station-mode sequence), the access mode models - the models for the conditional probability $P(m|s)$ - will be estimated first followed by the estimation of the station selection models - the models for the marginal probability $P(s)$. These models will now be discussed in greater detail.

Three alternative modes are considered in the access mode model: auto, walk, and bus modes. The probability of a traveller choosing one particular access mode for a given station is expressed

---

[1]Of course, also, this joint probability, P(m,s), is in itself a conditional probability to that the rail mode was actually selected. However, for simplicity the above notation is used instead of the conditional probability notation.

as a function of the level-of-service variables of all the available access modes and the socio-economic attributes of the traveller.

Let L denote the level-of-service variables and S the individual's socio-economic attributes, and define subscripts for L and S as follows:

j represents the individual traveller, j = 1,2,3, ...

s represents the access station, s = 1,2,3, ...

m represents the access mode, m = 1,2,3, ...

then,

$$P_{jm'} = P_j(m'|s) = f(L_{jm}, S_j) \tag{2-12}$$

where $P_{jm'}$ is the conditional probability that the person, j, chooses mode $m'\varepsilon\{m\}$ as his access mode, given that he has chosen access station s. S_j is the socio-economic variable of person j.

Since this study involves only individuals, the subscript j will be eliminated from now on, and multinomial logit model of the access mode is,

$$P_{m'} = \frac{\exp\{U(L_{m'},S)\}}{\sum_m \exp\{U(L_m,S)\}} \tag{2-13}$$

where $U(L_m,S)$ is the utility function of mode m. This utility function is assumed to have a linear form,

$$U(L_m,S) = b_1 L_m + b_s S \ . \tag{2-14}$$

In this equation the b_1 and b_s are the coefficients to be estimated.

It can be seen from Eq. (2-13) that, if the individual income or the household income is used as the socio-economic variable, the S term will be cancelled from the multinomial logit model. In order for the socio-economic information to remain in the model, it is necessary to modify it to be mode dependent, i.e.

$$P_{m'} = \frac{\exp U(L_{m'},S_{m'})}{\sum_m \exp U(L_m,S_m)} \tag{2-15}$$

where S_m denotes the mode-specific socio-economic attribute for mode m. As a matter of fact, because of its mode dependency S_m can be treated as one of the level-of-service variables.

Several access mode models will be constructed using slightly varying utility functions. These will be discussed in detail in later chapters.

The choice of access station is affected by three types of level-of-service variables. These are the general accessibility of the stations; the level-of-service associated with the facilities of the train stations (e.g. parking); and the difference in on-the-train time resulting from choosing one particular station instead of any other stations in the vicinity of the trip origin. This last variable is illustrated in Fig. 2 as $L_{ss'}$.

The choice probability of the access station may then be expressed as a function of the general accessibility L_s; the on-the-train level-of-service difference variable, $L_{ss'}$; and the intrinsic station variable L_{is}. Note that the effect of the socio-economic variable is included as part of the level-of-service variables. Expressed in functional form, the marginal probability for the station choice is,

$$P_{s'} = f(L_s,L_{ss'},L_{is}) \tag{2-16}$$

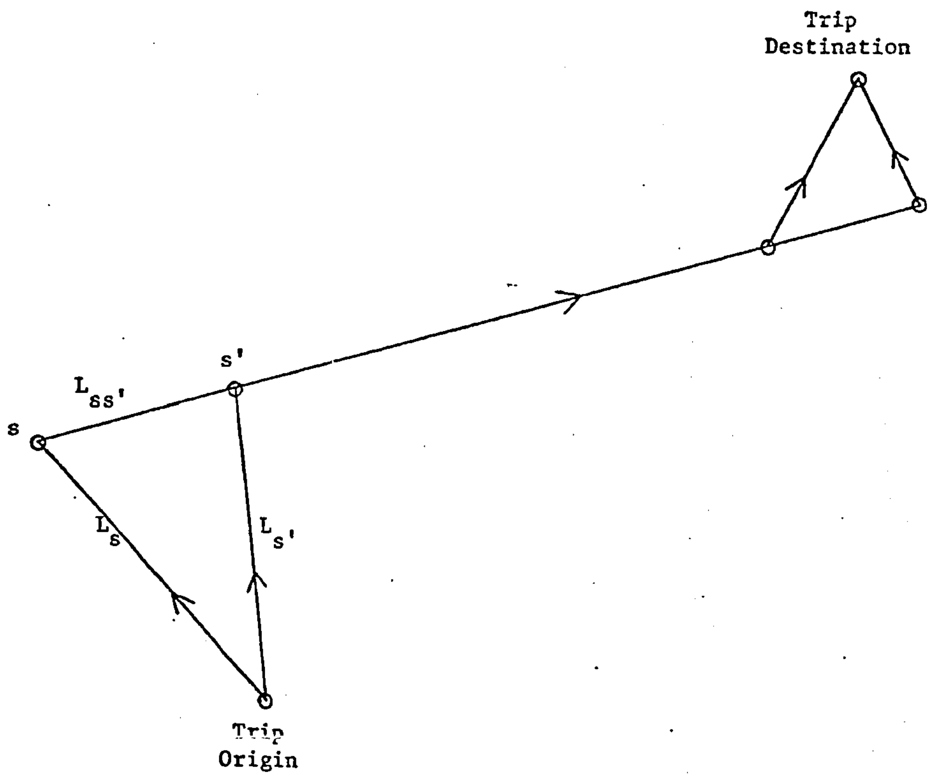

FIGURE 2. On-Train Level of Service Between Stations

where $P_{s'}$ is the probability that a person chooses station s' as his access station given that he has chosen a rail mode. And the multinomial logit model at the access station level is,

$$P_{s'} = \frac{\exp U(L_{s'}, L_{s's'}, L_{is'})}{\sum_{s} \exp U(L_s, L_{ss'}, L_{is})} \qquad (2\text{-}17)$$

and the utility function for this model is of the form

$$U(L_s, L_{ss'}, L_{is}) = b_s L_s + b_{ss'} L_{ss'} + b_{is} L_{is} \ . \qquad (2\text{-}18)$$

Two of the three types of explanatory variables reasoned to be a part of the station selection model can be easily derived. The on-the-train level-of-service difference variable and the intrinsic station variable may be directly observed. However, the variable describing the general accessibility of a station is related to the effort which is required of the individuals to reach the station by the available modes. Two ways will be used in this study to compute the general accessibility of a station. The first is the weighted price method, and the second is the weighted inclusive price method.

In the weighted price method, the individual's general accessibility to a station is represented by the weighted level-of-service variables (the "weighted price" variables). These variables are computed by weighting the level-of-service variables of each access mode for the individual to go to the station by the respective access mode choice probabilities of the individual. Of course, the access mode choice probabilities are those which are obtained from the access mode model. For example, if auto and bus modes are the available modes for a traveller to reach a certain station, and if the level-of-service variable under consideration is out-of-vehicle time (OVT), then, the weighted out-of-vehicle time is

$$OVT_s = \sum_m P_m^s OVT_m^s = P_A^s OVT_A^s + P_B^s OVT_B^s \quad . \qquad (2\text{-}19)$$

It may be noted that the level-of-service attributes describing the access to a station by different modes also describe the general accessibility to the access station. However, the estimated models at the two choice levels, mode and station, may have different values for the coefficients of the attributes. From the behavioral standpoint, this means that the traveller simply values the same level-of-service attribute differently when making his mode choice and his station choice decisions.

In the weighted inclusive price method all the level-of-service variables describing the access to a station are combined into a single index. This index is computed by first calculating the value of the utility function (the "inclusive price") in order to reach a station for each mode e.g. $U_m = \sum_l b_l L_l$, and then weighting the "inclusive price" of each mode by the probability of choosing that mode, e.g. $L_s = \sum_m P_m U_m$. The coefficients b_l and probabilities P_m are, of course, obtained from the access mode model.

The significant difference between the weighted price method and the weighted inclusive price method of deriving values of the modal level-of-service variables[1] is that, in the latter method the values of the coefficients for the modal level-of-service variables remain unchanged. This is because the entire utility function in the access mode model is weighted in the weighted inclusive price method. From the behavior viewpoint, the travellers are assumed to view the relative importance of these modal level-of-service atrributes, which influence their choices, equally at the two choice levels. An interesting consequence of this method of approach, as reflected in the estimated weighted inclusive price station model, is that the coefficient of the inclusive price variable should be 1. The reason is that the various modal level-of-service variables in the inclusive price have already been "weighted" before the parameters of the station model is even estimated. Consequently, only when the coefficient of the inclusive price variable in the station selection model is equal to one would a unit change in any of the modal level-of-service attributes yield the same amount of influence to the traveller's station selection decision as to his mode choice decision.

The other access mode, access station choice sequence is the sequence of access mode choice prior to the station choice (mode-station sequence). In this case, the joint probability of the access mode and access station is a product of the conditional probability of the station for predetermined mode and a marginal probability of the mode choice, i.e.,

---

The term "modal level-of-service variables or attributes" always refer to those level of service attributes that describe the individual's access trip between his trip origin (home) and the station.

$$P(m,d) = P(s|m) \cdot P(m) \quad . \tag{2-20}$$

In this sequence, models for the conditional probability $P(s|m)$ will be estimated first followed by the estimation of the models for the marginal probability $P(m)$. The input variables of the utility functions for the probability models of the two choices are the same as in the station-mode sequence. However, the values of the input variables for the $P(s|m)$ models, specifically the values of those variables describing the accessibility of station, are obtained directly from the original data. On the other hand, values of the level-of-service variables describing each mode for the $P(m)$ models are obtained by using the weighted price method, for example also consider the out-of-vehicle time (OVT),

$$OVT^m = \sum_s P_s^m \cdot OVT_s^m \tag{2-21}$$

where superscript m denotes the mode and subscript s denotes the station.

The Simultaneous Assumption. It was indicated in the early part of this chapter that it is possible to estimate the joint probability for the access mode choice and station selection directly. This model results from the behavioral assumption that a traveller makes these two decisions simultaneously rather than in any particular sequence.

In this model structure, the probability that a traveller chooses access mode m and access station s is a function of the level-of-service variables of each available mode; the on-the-train level-of-service difference variables; and the intrinsic station variables. Again, the socio-economic attributes of the traveller is included in the level-of-service variables.

The multinomial logit model for the joint probability of the access mode, station choice is

$$P_{m's'} = \frac{\exp\{U(L_{m'},L_{is'},L_{s's'})\}}{\sum_{m,s}\exp\{U(L_{m'},L_{is},L_{ss'})\}} \qquad (2\text{-}22)$$

and the utility function is

$$U(m,s) = b_1 L_1 + b_{ss'} L_{ss'} + b_{is} L_{is} \qquad (2\text{-}23)$$

All these model structures will be empirically examined as part of this research.

The Level-of-Service Variables and The Utility Functions

The time and cost characteristics of the transportation systems will be represented by five variables. The following notations are used to represent these basic variables;

OVT for out-of-vehicle time

AT for auto in-vehicle-time

BT for bus riding time

OC for operating cost

PC for out-of-pocket cost

OVT is the time the person spends outside a vehicle while on his way to the station. The value of the OVT depends on which one of the four possible access modes is the choice mode of the individual traveller. If the traveller drives to the train station by himself, the OVT is the time he spends walking from his trip origin, home for instance, to his car plus the time he spends to walk from the train station parking lot to the station. If the traveller is driven to the station, then the OVT is simply the time to walk to his car. For the walk mode, OVT is the walk time. For the bus mode, OVT is the sum of walk time and wait time.

AT is the time a traveller spends in the automobile while on his way to the train station, is zero if bus or walk mode were used.

BT is the time a person spends onboard a bus. It is zero for auto mode and walk mode.

OC is the operating cost for the automobile such as gas, oil and so on. Again it is zero if bus or walk mode were used.

PC is the fare if the mode is bus, or the parking cost if the traveller drives to the station. It is zero for walk mode.

Both AT and BT represent the time spent inside of a vehicle. They are made mode specific because there may be a preference, on the part of the individual traveller, between spending the same amount of travel time in the automobile and in the bus. It is therefore of interest to see as to what degree this preference affects the individual behavioral responses which are evidenced by the individual choice of travel mode or travel pattern.

In the sequential models, the general expression of that part of the utility function pertaining to the level-of-service of an access mode, m, is

$$L_m = f(OVT_m, AT_m, BT_m, OC_m, PC_m) \quad (2\text{-}24)$$

where m includes walk (W) and bus (B), auto driver (AD), auto passenger (AP), i.e.

$$\begin{aligned} L_{AD} &= f(OVT_{AD}, AT_{AD}, 0, OC_{AD}, PC_{AD}) \\ L_{AP} &= f(OVT_{AP}, AT_{AP}, 0, OC_{AP}, 0) \\ L_W &= f(OVT_W, 0, 0, 0, 0) \\ L_B &= f(OVT_B, 0, BT_B, 0, PC_B) \end{aligned} \quad (2\text{-}25)$$

The individual socio-economic variable is taken as the ratio of the total cost of the mode to the household income of the person. For example,

$$S_m = TC_m/HHI \quad \text{for the access mode model}$$

where TC = OC + PC and HHI is the household income of the traveller.

For the station models, in addition to the basic level of service variables, the on-train time difference variable called line-haul time (LHT) along with the parking facility variable are used to completely describe the station. The parking variable is a dummy variable and is called a parking dummy (PD).

Thus,

$$U_s = f(OVT_s, AT_s, BT_s, OC_s, PC_s, LHT_s, PD_s, S_s) \quad (2\text{-}26)$$

Finally, the same variables described previously can be employed in the simultaneous model. The utility function is therefore as follows:

$$U_{m,s} = f(OVT_m, AT_m, BT_m, OC_m, PC_m, S_m, LHT_s, PD_s, S_s) \quad (2\text{-}27)$$

It must be pointed out that although the utility functions discussed up to this point include all the possible explanatory variables, it is not necessarily true that all these variables will appear in the utility functions of the actual operational multi-logit models.

The derivation of the values of each of these variables will be explained in the following chapter. The following chapter also describes how the various estimated models will be evaluated.

CHAPTER III

DATA AND METHOD

Introduction

The access mode choice model, the station selection model and their utility functions in functional forms was presented in the previous chapter. Three more matters have to be settled before the model estimation may begin.

The first is to pick sample of travellers, observe their choices of mode and station as well as alternative modes and stations. The second is to compute values of the system attributes of the choice mode and station as well as alternative modes and stations. The third is to explain the method of estimation of the attribute coefficients and evaluation of the resulting models.

Data Source and Choice of Sample

The trip data used in this study are taken from the origin-destination survey conducted in 1969 by W. C. Gilman Company and Inc., for the Southward Transit Area Coordination Committee (STAC) in Chicago.[1]

---

[1]The trip data for this study are obtained from the Chicago Area Transportation Study and the STAC Report.

Only transit (rail and bus) information was assembled in the STAC survey. The survey questionnaires were distributed on-board the transit vehicles and returned on a voluntary basis.

The trip data was collected on a disaggregated basis. A trip is classified as a rail trip if the individual was surveyed on-board a train, or if the egress mode of a survey taken on-board a bus is rapid transit. The assumption is that rapid transit is the predominant mode of a bus-rapid transit combination trip. This was based on the opinion of the local engineers.

The specific survey information which is of interest in this study is the observed rail line, the trip origin, the access mode, the access station and the access distance to the access station. The trip origin may be obtained from the survey by a ¼-square mile. The access modes considered in the survey are auto driver, auto passenger, walk and bus.

A total of 150 individuals making a work trip are randomly selected from the Illinois Central (IC) Railroad surveys. Those samples with incomplete information are thrown out and replaced with valid samples, also randomly picked. Another set of 25 samples each are selected in a similar manner from the Rock Island Railroad and the South Shore and South Bend Railroad samples. The number of observed travellers of each access mode for the three rail lines are given in Table 1.

The various access mode and station selection models will be constructed using the data associated with those travellers who take Illinois Central to work. The set of observations taken from

Rock Island and South Shore will be used only for the purpose of testing the various operational models.

The Dependent Variable

The dependent variable of a multinomial logit model is the choice probability, P_i, where the subscript i indicates one of the alternatives in the choice set. The probabilities are not observed but only the actual choice. Thus, when estimating the model parameters the value of P_i is equal to one for the chosen alternative and zero otherwise.

Table 1. Rail Samples

Mode Rail	IC	RI	SS
Auto Driver	27	4	4
Auto Passenger	20	4	4
Walk	50	9	9
Bus	53	8	8
TOTAL	150	25	25

The Alternatives

For the access mode choice models the alternatives considered are auto, bus, and walk[1]. However, each person in the sample was not always considered as having all alternatives available to him.

[1]Even though it was observed in the survey whether an auto access to station was "drove" or "driven" type, the detail of the data did not permit further that detail in modeling the access mode choices.

Auto ("driven" or "drove") was always considered to be a relevant alternative. Walk access mode was considered to be unavailable to a person if his walking distance to a station was more than 20 minutes. Bus access mode was judged to be available if a bus route did not require more than a total of one-half mile walking.

For the station selection model the alternative stations were chosen on the basis of data and were usually adjacent to the chosen station.

The Transportation System Attributes

The transportation system attributes considered in this study are: Out-of vehicle time; Auto in-vehicle time; Bus riding time; Operating cost; and Out-of-pocket cost;[1] On-the-train travel time between alternative stations; and the Parking dummy. The first five system attributes affect the choice of access mode only while all the variables may affect the choice of the access station. The derivation of the values for each system component is explained below.

Out-of-vehicle time (OVT). The out-of-vehicle time for an auto driver is the total walking time the auto driver spends at the two ends of the access trip. Normally it is approximately 3.5 minutes. The out-of-vehicle time for auto passenger is assigned depending upon the location of the trip origin. In the inner urban area where most people live in apartments, it is assumed to be 2 minutes. In the outer urban area where single houses are the

---

[1]The cost attributes may also be considered together as Total Cost.

predominant dwelling units, it is 1 minute. The out-of-vehicle time by auto mode, for a traveller whose chosen mode is walk or bus mode, is the average of the out-of-vehicle time for auto driver and auto passenger.

The out-of-vehicle time for walk mode is derived by using an average walking speed of 3 mph[1], or 20 minutes per mile. It is therefore the actual travel distance to the station multiplied by 20.

The out-of-vehicle time for bus mode is the sum of the walking times to and from the bus stop and the waiting time at the bus stop. The walking time to and from the bus stop is obtained from the information of the trip origin, the access "blocks walked" information on the survey, and the location of the bus routes. The expected waiting time is one-half the headway, but not more than 8 minutes.

Auto Time (AT). The auto time for auto mode is estimated on the basis of the auto speed on local streets and the distance to the station. The average speed for the auto mode is 20 mph. An additional minute of auto time is given to the auto driver in order to take into account the time spent in the car while looking for a parking space and parking the car. Again, the auto time for a traveller whose chosen mode is walk or bus is the average of the auto time for auto driver and auto passenger.

---

[1]The values of the average speeds for walk, auto and bus modes; along with the auto operating cost estimates are based on the values used in the STAC study.

Bus Time (BT). The bus time is computed on the basis of the travel distance and the average speed. The travel distance is measured along the bus route between the street intersection closest to the origin point and the street intersection closest to the train station. The average bus speed is 11 mph.

Operating Cost (OC). The operating cost for both the auto driver mode and the auto passenger mode is 5 cents per mile. The total operating cost for the auto passenger mode is twice as much as that for the auto driver mode, since the auto passenger car has to be driven back home. The STAC study also includes an additional base cost of 30 cents to each auto driver trip and 10 cents to each auto passenger trip.[1] Both of these cost structures will be examined in this research.

The operating cost for a person whose chosen mode is walk or bus mode is the average of the operating cost for auto driver and passenger.

Out of Pocket Cost (PC). The pocket cost for auto driver is the parking cost. On the average, it costs 25 cents to park the car for one day. Hence, the parking cost for one way trip is one-half as much. The pocket cost for auto passenger is zero. The pocket cost for a traveller whose chosen mode is walk or bus mode is once again the average of the out-of-pocket cost for auto driver and auto passenger.

---

[1]The rationale behind the base costs is that the auto drivers are expected to own more cars than the auto-passengers, who in turn own more cars than persons walking or taking the bus.

The pocket cost for bus mode is the bus fare which is 40 cents for most access trips.

Linehaul Time Difference (LHT). This is the time between an alternative station and a chosen station, and equals the distance between the station pair divided by the average train speed. The average train speed is 35 mph. The chosen station is used as a reference point. If the alternative station is situated behind the chosen station, the distance and the resulting LHT is of a positive value. The positive sign indicates that the chosen station has an advantage over the alternative station (as far as the LHT is concerned). Conversely, if the alternative station is situated ahead of the chosen station, the LHT has a negative value.

Parking Dummy (PD). The parking dummy of a station is decided based on the information of the average parking space per auto driver at the station. A PD value of 1 is given to those stations with 0.3 or more parking space per auto driver. All other stations are given a PD value of 0.

Socio-economic Attributes (S). The socio-economic variable is the ratio of the total cost by a travel mode to the median income of the zone where the traveller originates his trip. This median income of origin zone is not of adequate individual socio-economic information, but is the only information available.

The socio-economic attribute for auto driver is the total cost, sum of operating cost and out-of-pocket cost, for auto driver divided by the median income. The socio-economic attribute for auto passenger is the total cost, operating cost for auto passenger, divided by median income.

For a traveller whose chosen mode is walk or bus mode, the socio-economic attribute by auto mode is the average of the socio-economic attribute for auto driver and auto passenger. The socio-economic characteristic for walk is zero, since there is no cost associated with the walk mode. The socio-economic characteristic for bus is simply the ratio of out-of-pocket cost for bus mode to the median income.

<u>Summary of Input Variables</u>

- OVT Out-of-vehicle time, the sum of the walking time and waiting time during the individual's access trip to the station.
- AT Auto-time, the amount of time the individual spends inside of an auto during his access trip to the station.
- BT Bus-time, the amount of time the individual spends inside of a bus during his access trip to a station.
- OC Operating cost, the operating cost of an auto during the access trip to the station.
- PC Out-of-pocket cost, the parking cost for auto driver or the bus fare for the bus mode.
- TC Total cost, the sum of the operating cost and the out-of-pocket cost.
- LHT Linehaul time difference, the on-train travel time difference resulting from choosing the station instead of the alternative stations.
- PD Parking dummy of the available parking space per auto driver.

S Socio-economic attribute, the ratio of the total cost to the median income.

Estimation Technique

The model estimation technique is the Maximum Likelihood Estimation for the Multilogit Model.[1] It is the same estimation method used by Moshe Ben-Akiva from M.I.T.

Evaluation Criteria

The evaluation of the access mode choice and the access station selection models is done in three ways. The first is to examine the statistical significance of each variable in the model and the model as a whole. The second is to examine the reasonableness of the magnitude of the coefficients of the variables in the model. This incorporates the study of the elasticities in order to see the effects of the attributes to the choice probability, and the value of time. The third way is to apply the model to situations different from that from which the model is estimated.

Statistical Tests

Since the choice probabilities are not actually observed, statistical test such as the estimated residue measurements (R^2) ordinarily used for the linear regression analysis does not bear much significance. The statistical tests used for the various disaggregate models in this study are mainly the t-test which indicates the statistical significance of each variable in a model, and the χ-test which indicates the statistical significance of the entire model.

---

[1]McFadden, Revealed Preferences.

Coefficients of the Variables

The coefficient of a variable is examined in two respects. The first is the sign of the coefficient and the second is the magnitude of the coefficient. The sign of a coefficient can be reasoned in a logical way. If one of the attributes of a model, for example the out-of-vehicle time of bus mode, increases while all the other attributes in the model remain constant then the probability for choosing the bus mode is expected to decrease. Conversely, if the out-of-vehicle time of the other modes increase, then the probability for choosing the bus mode should increase. Therefore, the sign of the coefficient for the out-of-vehicle time variable must be a negative sign. Similarly, the coefficients for the auto time, bus time, operating cost, out-of-pocket cost, weighted price, and socio-economic characteristic variables must also have negative signs. The situation for the linehaul time difference variable and the parking dummy variable is opposite of that for the previously discussed variables. This is because an increase in these variables means an improvement of the level-of-service of the station and thus an increase in the choice probability of the station. Therefore, the coefficients for these two variables must have positive signs. The weighted inclusive price variable (WIP), is a negative quantity itself, therefore, an increase in the absolute value of the weighted inclusive price variable will bring about a decrease in the choice probability of the station only if the coefficient of this variable has a positive sign.

The magnitude of the coefficients of the variables may be examined by studying the elasticities of the choice probabilities with respect to each of the variables. The elasticity for a multinomial logit model is the per cent of change in the choice probability in response to one per cent of change in the value of a variable. There are two types of elasticities, the direct elasticity and the cross elasticity. The direct elasticity in this study indicates the elasticity of the choice probability of a mode or station with respect to the attribute(s) describing that mode or station. The cross elasticity in this study indicates the elasticity of the choice probability with respect to the attribute(s) describing the alternative mode(s) or station(s). The mathematical expression for the direct and cross elasticities are[1],

$$E_{X_{1i}}(P_i) = b_1 X_{1i}(1-P_i) \qquad (3\text{-}1)$$

$$E_{X_{1j}}(P_i) = -b_1 X_{1j}(P_j) \qquad (3\text{-}2)$$

where:

P_i is the choice probability of alternative i. X_{1i} and X_{1j} are the 1th explanatory variable describing alternatives i and j, respectively. b_1 is the coefficient of X_1. $E_{X_{1i}}(P_i)$ and $E_{X_{1j}}(P_i)$ are the direct and cross elasticities with respect to X_1, respectively.

The implied value of time may be obtained from the coefficients of the time and cost variables in a model. The value of

---

[1]The mathematical derivatives of the direct and cross elasticities for the multinomial logit model are presented in Appendix II.

time obtained from this research will be compared with the value of time obtained from other studies.

Model Application

The disaggregate access mode and station models are further evaluated by applying each model to different situations. It may be remembered that, during the process of data preparation, a group of 150 samples was selected from the Illinois Central (I.C.) railroad observations and another group of 50 samples was selected from the Rock Island and South Shore (R.I./S.S.) railroads observations.

The Rock Island and the South Shore railroads are different from the Illinois Central railroad in that the operators, the number of rail tracks, the distances between adjacent stations, the train operating frequencies and the type of signal and train facilities for these railroads are different. As far as the service areas are concerned, the Rock Island Railroad mainly services the western part of the study area while the South Shore Railroad operates commuter and intercity passenger services between Chicago and South Bend, Indiana, which are 90 miles apart.

The various models estimated in this study will be based on the set of I.C. data, the base data. The set of R.I./S.S. data (the control data), will be used for the sole purpose of testing the estimated models. Each model will be applied to both the base data and the control data. For each individual sample, comparisons will be made between the expected probability of the chosen mode or station with the expected probabilities of the alternative modes or stations in their respective choice set. If the expected probability of the chosen mode or station is greater or equal to those of the other alternatives, then the model is considered to have made a correct prediction. Otherwise, it is a wrong prediction.

Furthermore, for the access mode model, the expected number of users of each mode is compared with the actual number of users of the same mode.

The estimation and evaluation of the access mode and access station models will be presented in the next two chapters.

CHAPTER IV

DEVELOPMENT OF ACCESS TRIP CHOICE MODELS
DECISION SEQUENCE: STATION-MODE

The access mode and station choice models of this chapter are estimated based on the assumption that the traveller chooses the access station first followed by the choice of the access mode to the chosen station. The estimated access mode and station models are also evaluated in this chapter.

THE ACCESS MODE CHOICE MODEL

Introduction

The access mode models, which predict the conditional probability of choosing the mode for predetermined station $P(m|s)$, are estimated and evaluated in this section.

Initially, examinations were made into the most acceptable model specifications. In particular, the way two of the variables, the cost variables and the in-vehicle time variable, should enter the model were investigated.

As indicated in the previous chapter, the auto operating cost would be entered either at a flat rate of 5 cents per mile or adding a base cost of 30 cents per trip for the auto driver and 10 cents per trip for the auto passenger. The results showed that the latter operating cost structure, the one with an additional base cost, produced models with invalid coefficient signs. Thus, the 5 cents per mile operating cost structure was adopted.

The in-vehicle time was also entered in the model in two ways. The first way was to enter it directly into the model. At the same time, dummy variables were used to indicate the access mode (auto or bus) with which the in-vehicle time information was associated. The second way was to separate the in-vehicle time information into two mode specific variables, auto time and bus time. The first way resulted in models with wrong coefficient signs consequently the mode specific in-vehicle time variables were adopted for use. The remaining model specifications which were acceptable as far as the signs of the coefficients are concerned may include the following variables: out-of-vehicle time (OVT), auto time (AT), bus time (BT), operating cost (OC), out-of-pocket cost (PC) and socio-economic attribute (S).

A number of models were estimated using the above mentioned variables. Two of the estimated access mode models appeared to have the correct signs and statistically acceptable coefficients for each variable. Both of these models include the out-of-vehicle time, auto time and bus time. The cost variable is however, different in these two models. In the first model - Model I, it is the operating cost (OC) and in the second model - Model II, it is the total cost divided by income (S). These models are evaluated below.

Evaluation of the Estimated Models

It was indicated in Chapter III that the models will be evaluated in three ways. The first is the statistical tests of the individual coefficients and the model as a whole. The second is the evaluation of the coefficients in terms of their magnitude (elasticities) and value of time implied. The third is evaluation of the model performance in the base data and the control data. This evaluation of model performance will be carried out in two ways, misclassification among the alternative modes and comparison of the expected and actual mode users.

The coefficients of the two access mode models and other relevant information are given in Table 2.

The statistical tests indicate all the variables in Model 1 as well as the model itself, are significant with .99 level of confidence. In Model 2, the socio-economic characteristic variable is statistically significant only with .75 level of confidence. The bus time variable is statistically significant with .95 level of confidence. The out-of-vehicle time and the auto time variables, along with the model as a whole are statistically significant with .99 level of confidence. Therefore, on the whole both models are statistically acceptable.

From the coefficients in Model 1, the implied values of the out-of-vehicle time, auto time and bus time are \$.48/hr. (S.D. 0.25), \$.74/hr. (S.D. 0.50), and \$.41/hr. (S.D. 0.25), respectively. Comparisons of these values with the sub-mode values of time of other studies are not available. However, comparisons with the values of time of other demand model studies show that the value of the in-vehicle time in this research is approximately the same as the value of in-vehicle time of some recent studies (approximately \$.70/hr)[1],

[1]Talvitie, A. P., "Comparison of Probabilistic Modal Choice Models: Estimation Methods and System Outputs; *Highway Research Record 392*, 1972. Ben-Akiva, M. *Disaggretate Direct Demand Model.*

TABLE 2

Coefficients and Relevant Information of the Access Mode Models p(m|s)

Variable	MODEL 1 Coefficient	MODEL 1 Standard Error	MODEL 2 Coefficient	MODEL 2 Standard Error
Out-of-vehicle time	-.441	.094	-.286	.059
Auto time	-.681	.291	-.112	.260
Bus time	-.382	.120	-.164	.091
Operating cost	-.556	.193	-	-
Socio-economic attribute	-	-	-.084	.111

χ^2 = 99.498 with degrees of freedom = 4 (Model 1)
χ^2 = 87.647 with degrees of freedom = 4 (Model 2)

but the values of the out-of-vehicle time in this research are much lower than the value of the out-of-vehicle time of other studies ($3.00/hr and up). It may be noted that the trips under investigation in this study are access trips whereas the previous studies consider either the major part of the trip or the entire trip.

From the coefficients of the second model, the implied value of the auto time turns out to be $ 80/hr., which is too large to be reasonable. Therefore, Model 2 is considered invalid.

Equations (3-1) and (3-2) of the direct and the cross elasticities of a multilogit model indicate that, aside from the fact that these two types of elasticities have opposite signs, the direct elasticity involves a (1-P) term while the cross elasticity involves only a (P) term instead. In other words, the absolute values of the direct elasticities at a probability value of, for example P', are identical with the absolute values of the cross elasticities at (1-P'). Therefore, to calculate the direct elasticities at probabilities of .30, .50 and .70, is in fact the same as calculating the cross elasticities at probabilities of .70, .50 and .30, respectively. The signs of the direct and the cross elasticities are, of course, still opposite to each other.[1]

The direct elasticities of the access mode model, computed at the means of each explanatory variable, for each fixed probability are listed in Table 3.

TABLE 3

Direct Elasticities of the Access Mode Model

Variable	Mean (minutes or cents)	Direct Elasticity (E)		
		P = .30	P = .50	P=.70
Out-of-vehicle time	6.7	2.08	1.49	.89
Auto time	6.7	3.19	2.28	1.37
Bus time	10.1	2.71	1.94	1.16
Operating cost	15.5	6.03	4.31	2.58

It can be seen from Table 3 that practically all the elasticities of the choice probability with respect to the transportation attributes, out-of-vehicle time (OVT), auto time (AT), bus time (BT) and operating cost (OC), are elastic when computed at the means of these variables. OC, which is exclusively associated with

[1]In the following context of this thesis, the elasticities will be discussed in terms of their absolute values unless otherwise specified.

the auto mode, has by far the greatest elasticity, and AT, OVT, and BT having smaller elasticities in that order. Values of the direct elasticities in this study are not in line with the a priori knowledge of the elasticities from previous studies. Nevertheless, it must be reminded again of the particular type of trips involved in this study. It can also be noticed that, as the probability increases, the elasticities decrease. This suggests logically that the travellers grow more unconcerned with the changes in the transportation attributes if their chosen mode is chosen with a high probability.

Graphs of the direct elasticities between the upper and lower limits of each variable are shown in Figs. 3-5.

Two observations can be interpreted from the elasticity graphs. The first is that the direct elasticities increase proportionally with the increase in the values of their corresponding variables. This is of course obvious from the expression for the elasticity Eq. (3-1). From the travellers' behavioral standpoint, this means that the travellers are assumed to become more sensitive to percentage changes in the transportation system attributes as the values of the system attributes increase. Intuitively, this appears reasonable because the response to a change in the attribute should depend on the absolute level of that attribute. The second is that as the values of the variables do increase, the auto time and the operating cost variables have much greater increase in elasticities than the out-of-vehicle time and the bus time variables. This observation along with the elasticity computations at the means, reveal that the travellers are very sensitive to changes in level-of-service offered by auto mode while changes in bus time and out-of-vehicle time do not appear to affect the travellers' mode choices as much as changes in auto time

Access Mode Model

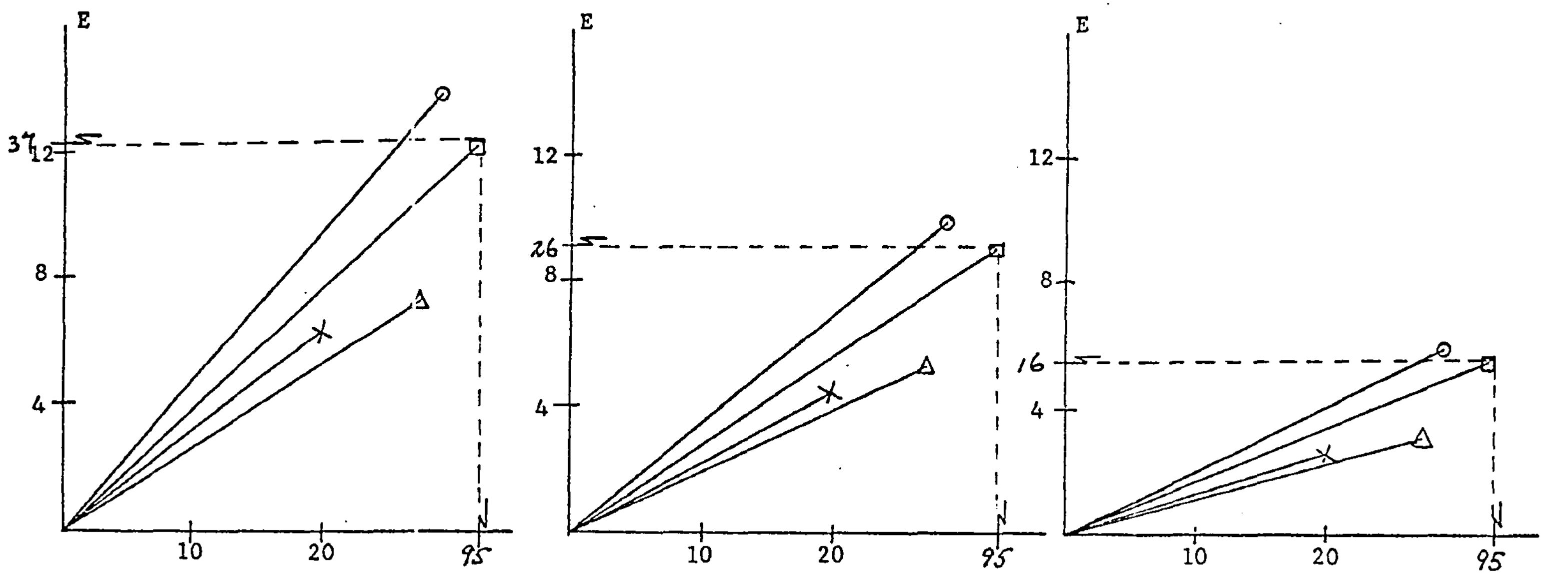

FIGURE 3. Direct E;asticity at P = .30

FIGURE 4. Direct Elasticity at P = .50

FIGURE 5. Direct Elasticity at P = .70

and auto operating cost. One can interpret this to mean that the bus mode and the walk mode are the more popular access modes.

The model is further evaluated by applying it to both the base data (I.C. data) and the control data (R.I./S.S. data) and compare the results.

The numbers of misclassification and the predictive rates are given in Table 4.

TABLE 4

Accuracy of the Access Mode Model

Data	N	M	$\alpha = 1 - \frac{M}{N}$
I.C.	150	12	92%
R.I./S.S.	50	7	86%

N, total number of observations
M, number of misclassifications
α, predictive accuracy

Also, the expected number of travellers by mode can be computed as the sum of the expected probability values of each mode, i.e.,

$$N_m(\text{Expected}) = \sum_i P_m^i \tag{4-1}$$

where P_m^i is the probability of mode "m" being chosen by person "i"; and N_m(expected) is the expected number of travellers to use the mode "m". These expected numbers of mode users are compared with the actual number of mode users, which are obtained directly from the surveys. The comparisons of the expected and the actual number of travellers by mode is given in Table 5.

Results of the comparisons show the expected and the actual values by mode are compatible. For the R.I./S.S. data, the accuracy of the comparison for bus mode is 69 per cent instead of approximately 80 per cent for the others.

TABLE 5

Comparison of Number of Mode Users

Data	Mode	N_E	N_A	Rate
I.C.	Auto	56	47	81%
	Walk	40	50	80%
	Bus	54	53	98%
R.I./S.S.	Auto	13	16	81%
	Walk	15	18	83%
	Bus	21	16	69%

N_E, expected number of travellers

N_A, actual number of travellers

This may be contributed to the fact that the bus frequencies in the surrounding areas of the R.I./S.S. railroads are often quite low. Though the waiting time for bus has been set to no more than 8 minutes during the process of data preparation, it is not uncommon to have a 30-minute headway for the buses in these areas. Thus, even though the actual waiting time at the bus stop is not more than 8 minutes, some people have to settle for another tvpe of access mode in order to arrive at the train station, and eventually the trip destination, by a certain time.

The results of the evaluation of the chosen access mode choice model show that the model has reasonable coefficients, is statistically significant and has rather impressive predictive performances. It is therefore judged as a good model.

THE ACCESS STATION SELECTION MODELS

Introduction

The access station models are estimated and evaluated in this section. The modal level-of-service variables which describe the access to a station were aggregated by both the "weighted price" method and the "weighted inclusive price" method. As discussed in Chapter II, the "weighted prices" describing the accessibility to a station are the values of the modal level-of-service attributes weighted by the access mode choice probabilities.

The "weighted inclusive price" was also discussed in Chapter II. It was the value of the utility function weighted by the mode probabilities. The utility of each mode was produced by combining the level-of-service variable in the access mode model with their respective coefficients. Unlike the previous method in which a number of variables are employed to describe the access to a station, in this method the weighted inclusive price (WIP) is the only variable that describe the access to a station.

A number of models were estimated using either the weighted prices or the weighted inclusive price, along with the on-train time difference between stations or the line-haul time (LHT) and parking dummy (PD) variables which also describe the level-of-service pertaining to the station.

Two of the estimated models appeared to have correct coefficient signs and statistically acceptable indications. One of the models employed the weighted prices and the other employed the weighted inclusive price as part of their level-of-service variables.

Evaluation of the Estimated Weighted Price Station Model

The variables included in this station model are weighted out-of-time, weighted auto time, linehaul time difference and parking dummy. The coefficients are given in Table 6.

This model will also be evaluated in three ways. The first is the statistical evaluation of each variable and the model. The second is the evaluation of the coefficients and the elasticities. The third is the evaluation of the model performance in the base data and the control data.

TABLE 6

Coefficients and Relevant Information of the Weighted Price Station Model P(s)

Variable	Coefficient	Standard Error
Weighted out-of-vehicle time	-.385	.102
Weighted auto time	-.957	.172
On-train line-haul time difference	.138	.167
Parking dummy	.827	.469

χ^2 = 90.391 with degrees of freedom = 4

The statistical tests indicate that the weighted out-of-vehicle time and the weighted auto time are significant with .99

level of confidence. The parking dummy is significant with .95 level of confidence. The on-train time difference is statistically significant at .80 level of confidence. The model as a whole is statistically significant with .99 level of confidence.

The direct elasticities are computed at the means for the weighted out-of-vehicle time and the weighted auto time, at 4 minutes for LHT and at 1 for PD. They are listed in Table 7.

TABLE 7

Direct Elasticity of the Weighted Price Station Model

Variable	Variable Value	Direct Elasticity (E)		
		P = .30	P = .50	P = .70
Weighted out-of-vehicle time	6.44	1.73	1.24	.74
Weighted auto time	2.16	1.44	1.03	.62
On-train line-haul time difference	4.00	.39	.28	.17
Parking dummy	1	.57	.41	.25

From the above table, it can be seen that the elasticities of weighted out-of-vehicle time and weighted auto time are elastic at probabilities of .30 and .50. They become inelastic at probability of .70. This indicates again that the travellers are relatively insensitive to the changes in the attributes as the choice probability increases. At the means of the two variables the choice probabilities appear to be somewhat more sensitive to out-of-vehicle time than to auto time. The elasticities of LHT and PD are inelastic.

Travellers' station choice do not appear to be as sensitive to these variables as the two other attributes.

Graphs of the direct elasticities plotted between the upper and lower limits of the WOVT, WAT, and LHT variables are shown in Figs. 6-8. These graphs show that as the values of these variables increase, the WAT has the greatest increase in its elasticity, the WOVT has much smaller increase in its elasticities than the WOVT, and the LHT has the least increase in its elasticities. This indicates that, when selecting the access stations, the travellers are most sensitive to the auto time, This was also observed in choosing the access mode. The results also show that the travellers are relatively unconcerned to the extra amount of time spent (or saved) inside the train in choosing the access station. In spite of the incompleteness of the parking availability variable, it appears clear that it has an effect in choosing the access station. The value of time information is not available, since no cost variable exist in this model.

The model is further evaluated by applying it to both the base data (IC) and the control data (RI/SS) and compare the results. The numbers of misclassifications and the predictive accuracy rates of the model are given in Table 8.

TABLE 8

Model Application Results of the Weighted Price Station Model

Data	N	M	$\alpha = 1 - \frac{M}{N}$
I.C.	150	32	78.7%
R.I./S.S.	50	1	98.0%

N, total number of observations
M, number of misclassifications
α, predictive accuracy

Weighted Price Station Model

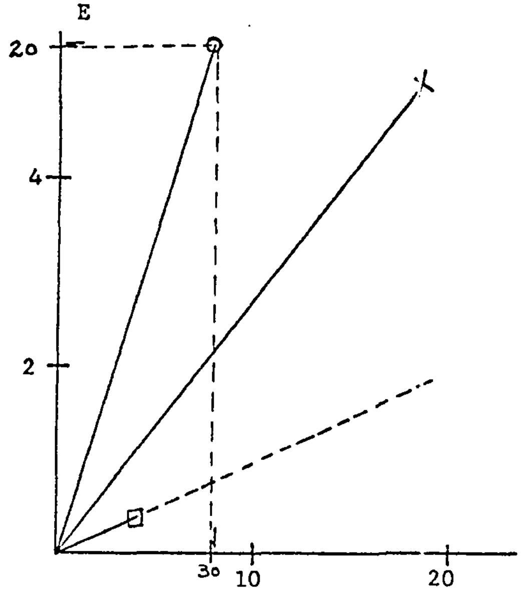

FIGURE 6. Direct Elasticity at P = .30

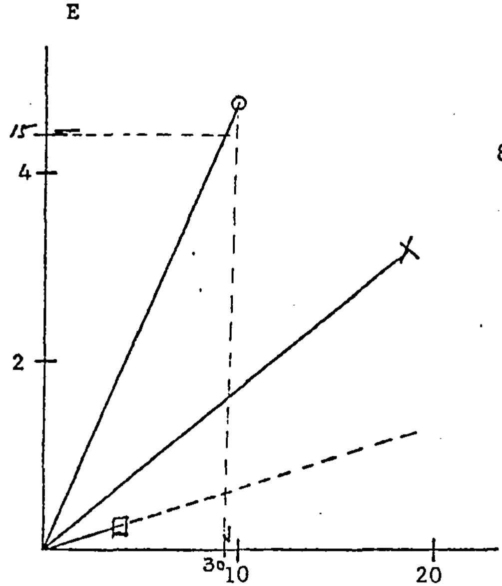

FIGURE 7. Direct Elasticity at P = .50

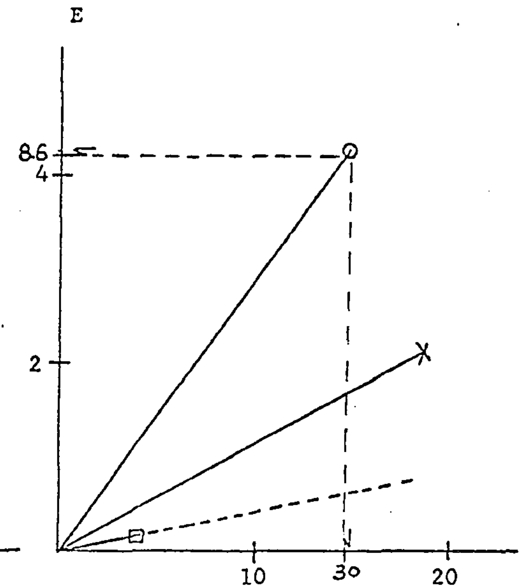

FIGURE 8. Direct Elasticity at P = .70

X -- WOVT ⊙ -- WAT □ -- LHT

The comparisons of the actual number of travellers choosing a certain train station with the expected number of the travellers choosing that station is not made for the access station model. The reason for not making such comparisons is that the small amount of observed travellers are distributed to a relatively larger number of alternative stations. For example, the base data include 37 alternative stations while there are only 150 observed individuals. Similarly for the control data, there are 34 alternative stations on the Rock Island and South Shore railroads, but only 50 observed individuals. The actual number of travellers going to any station is at the most around 10 people. Therefore, a comparison of the expected and actual number of users for each or any of the stations would not bear much significance.

Evaluation of the Estimated Weighted Inclusive Price Station Model

This model will be evaluated in the same way as the previous station model: the evaluation of the statistical indicators, the evaluation of the coefficients and elasticities, and the evaluation of the model application to control data.

The coefficients of the weighted inclusive price station model and other relevant information are given in Table 9.

The statistical tests indicate that the weighted inclusive price variable and the parking dummy variable are significant with better than .97 level of confidence. The linehaul time variable is significant with almost .90 level of confidence. The model itself is statistically significant with .99 level of confidence.

TABLE 9

Coefficients and Relevant Information of the
Weighted Inclusive Price Station Model
P(s)

Variable	Coefficient	Standard Error
Weighted inclusive price	.585	.107
On-train line-haul time difference	.230	.194
Parking dummy	1.189	.563

χ^2 = 100.4353 with degrees of freedom = 3

The direct elasticities with respect to these variables are computed at the mean for the weighted inclusive price variable, at 4 minutes for the linehaul time variable, and at 1 for parking dummy. They are listed in Table 10.

Table 10 shows that the direct elasticity of the weighted inclusive price variable computed at the mean is greater than one at the given probability values. The linehaul time difference and the parking dummy variable are both inelastic at the given probability values. It appears that in both this model and the weighted price station model that the modal level-of-service, which describes the accessibility to the station, has much greater influence over the travellers' station selection decisions than the linehaul time and the parking dummy attributes. Also, the parking dummy attribute seems to affect the station choices more than the line haul time attribute.

TABLE 10

Direct Elasticities of the Weighted Inclusive Price Station Model

Variable	Variable Value	Direct Elasticity (E)		
		P = .30	P = .50	P = .70
Weighted inclusive price	8.07	3.31	2.36	1.42
On-train line-haul time difference	4.00	.64	.46	.28
Parking dummy	1	.83	.59	.36

Graphs of the direct elasticities plotted between the upper and lower limits of WIP and LHT are shown in Figs. 9-11. These graphs show that, as the values of the variables increase, the direct elasticities of the weighted inclusive price increase more rapidly than those of the linehaul time difference variable. This again indicates that the weighted inclusive price is the most important attribute when selecting the stations.

It has been discussed previously that the coefficient of the weighted inclusive price variable should have a value of 1 according to the assumption that the travellers assign the same "weights" to each of the modal level-of-service variables when making the mode and station choice decisions. The coefficient of the weighted inclusive price variable in this model is .5850. It is tested to be significantly different from 1.0000. This indicates that the above mentioned assumption is invalid. In other

Weighted Inclusive Price Station Model

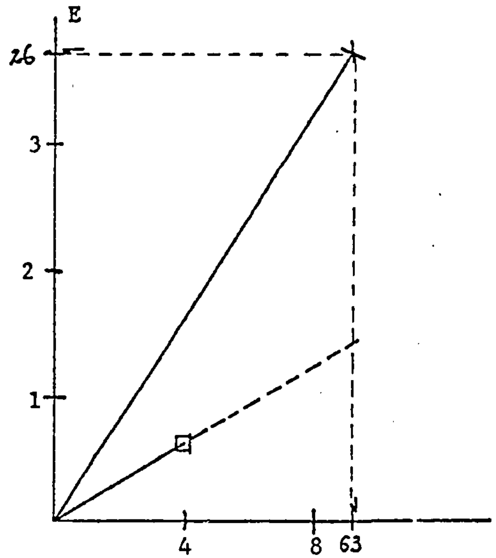

FIGURE 9. Direct Elasticity at P = .30

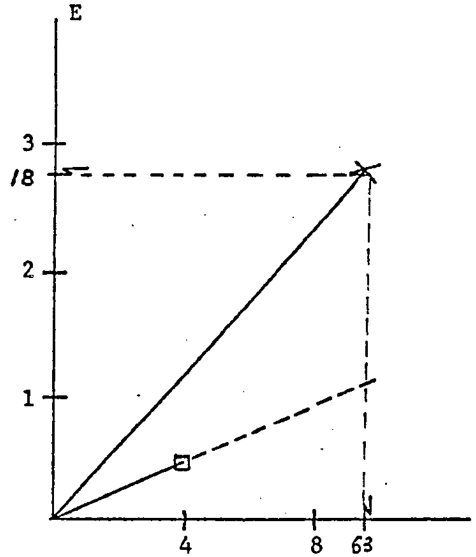

FIGURE 10. Direct Elasticity at P = .50

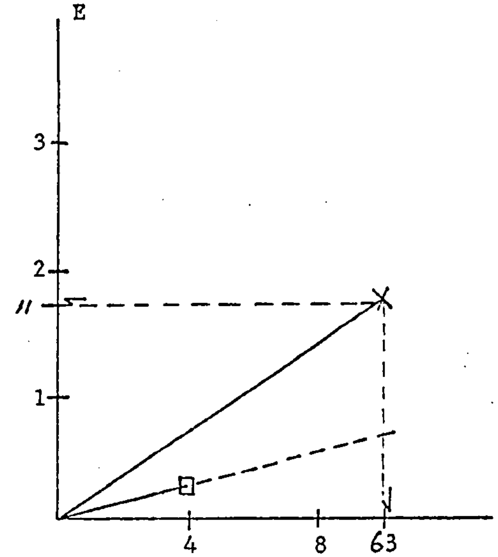

FIGURE 11. Direct Elasticity at P = .50

X-- WIP   ⊡ -- LHT

words, the traveller does assign different weights to the set of transportation system attributes when making his access mode and station choice decisions.

The model is further applied to both the base data and the control data. The results of the evaluation in terms of the numbers of misclassifications and the predictive accuracy rates of the model are given in Table 11

TABLE 11

Model Application Results of the Weighted Inclusive Price Station Model

Data	N	M	$\alpha = 1 - \frac{M}{N}$
I.C.	150	27	82%
R.I./S.S.	50	2	96%

N, M and α have the same meanings as in Table 9.

The comparison of the actual and the expected number of travellers choosing each of the stations are not presented, also because of the relatively few number of observed travellers in the base and control data as compared to the many alternative stations available on the three railroads.

The two access station models, one developed by using the weighted price method and the other developed by using the weighted inclusive price method to aggregate the modal level-of-service information, are evaluated to have good predictive performances.

Both of them indicate that the accessibility to the station have predominant influence over the travellers' station selection decisions. In addition, the weighted inclusive price model also indicated that the travellers do not value the same transportation system attributes equally when making decisions on the access mode and station choices. Therefore, it can not be recommended for use. Thus, the access mode model developed earlier, has to be complemented with its "weighted price" counterpart for station selection situations.

CHAPTER V

DEVELOPMENT OF ACCESS TRIP CHOICE MODELS DECISION SEQUENCES: MODE-STATION AND SIMULTANEOUS

Introduction

Two other model structures are used to develop the access mode and station selection models. The first is based on the assumption that the traveller chooses the access mode before he chooses the station. The second is based on the assumption that decisions of the access mode and station choices are made simultaneously. In the former sequential model structure, two of the estimated access station selection models appeared to be reasonable as well as statistically satisfactory. However, the attempt to estimate the access mode choice model, which is estimated after the station choice model in this sequence, failed to produce any valid models. The simultaneous method also resulted in models with incorrect coefficients. The development of these models along with their model coefficients and other general characteristics are presented in this chapter.

THE MODE-STATION SEQUENCE

The Access Station Selection Models

The same set of base data which had been used for the station-mode sequence model structure were used for the estimation of the access station models. These models predict the conditional probability of choosing the station for predetermined modes $P(s|m)$. However, for some of the observed travellers, there is only one station that is accessible by the given mode. To explain this, consider the case of an individual traveller who is observed to use the bus mode to go to the chosen station. In the previous sequence, the station-mode sequence, the station is fixed (predetermined) and there are often more than one way such as auto or walk mode by which the person can also reach his chosen station. Thus, the auto or walk mode or both of them become the relevant alternative(s) of the chosen mode, the bus mode. In the present sequence, the mode-station sequence; the chosen mode, in this case the bus mode, is predetermined. Often times, there is bus service going to only one particular station and not to any other station adjacent to the chosen station. Therefore, there is no relevant station alternative for this bus traveller. Consequently, those observations for which a relevant station alternative does not exist were taken out of the set of the base data. 110 observations still remain in the base data set with which the access station models were estimated.

The variables entered into the models were out-of-vehicle time (OVT), auto time (AT), bus time (BT), in-vehicle time (IVT), operating cost (OC), on-train line-haul time difference (LHT), and parking dummy (PD).

The coefficients of the two estimated access station selection models and other relevant statistical information are given in Table 12.

TABLE 12

Coefficients and Relevant Information of the Access Station Model P(s|m)

Variable	Model 1		Model 2	
	Coefficient	Standard Error	Coefficient	Standard Error
Out-of-vehicle time	-.664	.180	-.653	.177
Bus time	-.310	.208	-	-
In-vehicle time	-	-	-.184	.173
Operating cost	-.136	.066	-.062	.094
Parking dummy	.859	.480	.833	.480

The value of time implied by the coefficients of Model 1 are approximately \$3.00 (S.D. 1.73) for the out-of-vehicle time and \$1.40 (S.D. 1.12) for the bus time. The value of time implied by the coefficients of Model 2 are approximately \$6.00 (S.D. 9.94) for the out-of-vehicle time and \$1.80 (S.D. 4.20) for the in-vehicle-time.

The misclassification of these two models when applied to the the base and the control data are listed in Table 13.

TABLE 13

Accuracy of the Access Station Models

	Data	N	M	$\alpha = 1 - \frac{M}{N}$
MODEL 1	I.C.	110	20	82%
	R.I./S.S.	23	3	87%
MODEL 2	I.C.	110	20	82%
	R.I./S.S.	23	3	87%

N, total number of observations
M, number of misclassifications
α, predictive accuracy

These two access station models appear to be fairly good models. Nevertheless, it must be noted that the station models only constitute part of the sequential modeling processes. The access mode models also have to be examined before the validity of the assumption of this particular access mode and station decision-making sequence can be determined.

The Access Mode Choice Models

In order to estimate the access mode models, the modal level-of-service variables which describe the access to a station by the various modes were aggregated by the "weighted price" method. Furthermore, dummy variables were also used to indicate whether the in-vehicle time information was associated with the autommode or the bus mode. However, all the estimated models involve incorrect coefficient signs. The various estimated models and their corresponding coefficients are shown in Table 14.

TABLE 14

The Access Mode Choice Models

P(m)

Variable	Model 1	Model 2	Model 3
Out-of-vehicle time	-.083	-.082	-.347
Bus time	.137*	-	-
In-vehicle time	-	.142*	.210*
Operating cost	-.189	-.248	-.093
Dummy variable 1	-	-	2.837
Dummy variable 2	-	-	3.638

D1 and D2 are the dummy variables such that:

D1 = 0, D2 = 0 for auto mode
D1 = 0, D2 = 1 for walk mode
D1 = 1, D2 = 0 for bus mode

*Incorrect coefficient sign.

It has been shown in this chapter that models for the station choice with predetermined access mode, if judged by themselves, are reasonably good models and have satisfactory predictive performances. However, the station model is only part of the composite access trip choice model system. The fact that no valid access mode model could be estimated indicates that the entire assumption of the traveller's making his mode choice prior to his station choice appears to be an invalid assumption. Therefore, the access mode models as well as the access station models in this mode-station decision-making sequence can not be applied with confidence to planning problems.

THE DEVELOPMENT OF THE SIMULTANEOUS MODEL

The simultaneous models were estimated by entering those attributes that describe the access modes, out-of-vehicle time, auto time, bus time, etc.; and those attributes that directly describe the train station, on-train line-haul time difference and parking dummy. However, no valid model resulted from using the simultaneous model structure. In fact, each of the estimated models involves at least one incorrect coefficient sign. The coefficients of the most acceptable simultaneous models are given in Table 15

TABLE 15

The Coefficients of the Incorrect Simultaneous Model

Variable	Coefficient
Out-of-vehicle time	-.0119
Auto time	-.0571
Bus time	.1974*
Parking Dummy	1.340

*Incorrect coefficient sign

The results of this study tend to suggest that the simultaneous model structure is also an invalid traveller decision-making assumption.

CHAPTER VI

SUMMARY AND CONCLUSION

Summary

The main purpose of this study was to develop disaggregate choice models of the access mode and access station for those individual travellers who make their work-journeys by rail transit.

The mathematical model employed is a multinomial logit model based on the "independence of irrelevant alternatives" assumption, which is capable to deal with different number of choice alternatives for each of the individual behavioral unit and which is considered to be the most suitable model for the situation under investigation.

The data used for the estimation and evaluation of the various probability models were obtained from the Chicago area.

It is assumed that a person may make the access mode and access station choice decisions either simultaneously or in one of two sequences, the station-mode sequence or the mode-station sequence. In the case of the sequential assumption, the joint probability of the access mode and station is separated into a conditional probability of one choice given the other choice and a marginal probability of the other choice, depending on the particular choice sequence assumed. The investigation in this study using the simultaneous model structure and the mode-station sequence structure failed to produce valid choice models.

The results of the research based on the other sequential structure, the station-mode sequence, revealed some interesting behavioral characteristics of the individual travellers when making their access trips. It has been discovered from the previous travel demand studies[1] that the private auto mode is always preferred to the public transit (bus) mode, e.g. the auto demand is insensitive to auto travel time or cost. In this study, however, it has been found from the access mode model $P(m|s)$ that the elasticity of auto time is much higher than that of bus time. This indicates that spending travel time inside an automobile is very much disliked by the travellers as compared to spending travel time inside a bus. Furthermore, the relatively low elasticity of the out-of-vehicle time variable also suggests that the travellers even prefer to choose those access modes, such as the walk mode and the bus mode for which the out-of-vehicle time may constitute a large part of the "total price". The above findings suggest that the auto mode is the least preferred mode among the access modes considered. This discovery may initially seem contradictory to the information obtained from the previous travel demand studies. Nevertheless, it must be realized that the type of trips considered in those studies are concerned with the entire trip whereas the type of trips considered in this study is the access part of the entire journey. These two types of trips are different in nature and consequently the behavioral responses of the

---

[1]Ben-Akiva. *Disaggregate Direct Demand Model;* CRA, *Disaggregate Direct Demand Models;* Talvitie. *Comparison of Probabilistic Modal-Choice Models.*

travellers when confronted with these two different situations should not be expected to be the same. Unfortunately, no other access trip studies are available at this time for comparison with the results of this research.

Still, the different behaviors may be justified from the intuitive point of view. For the entire journey, the travel distance between the trip origin (home) and the trip destination (job site) is generally quite large. Therefore, the comfort, the privacy and other advantages offered by the auto mode become rather important to the travellers and thus make the use of a private auto a more desirable mode of travel. Of course, the walk mode is usually not considered as one of the available alternative modes for such a trip. On the other hand for the access trip, the travel distance between the trip origin and the trip destination (train station) is very short. For example the average access trip travel distance for the 150 observed travellers in the set of base data is only 1.5 miles. Clearly, not much comfort or privacy can be derived by using a car for a short trip of this magnitude. On the contrary, the various inconveniences of using an auto, such as finding parking spaces for the car, cost of owning it or having someone else to drive the traveller to the station become predominant disadvantages.

The travellers' implied values of time (bus time and auto time) obtained from this model ($.74/hr for AT and $.41/hr for BT) compare closely to the value of the in-vehicle time of the previous studies.[1] The value of the out-of-vehicle time implied in this study

---

[1]Ben-Akiva. *Disaggregate Direct Demand Model;* Talvitie, *Comparison of Probabilistic Modal-Choice Models.*

($.48/hr) is much smaller than that from other studies. Again, it must be reminded that the comparisons are made between two different types of trips.

The socio-economic variable was also considered in estimating the access mode model. A very rough income figure, the median income of the traffic zone which is the only income information available, was used to derive the value of the individual socio-economic variable. It is not known whether this is the reason that the model with the socio-economic variable yielded the wrong value of time. However, more specific income information of each individual would have been desirable.

The access station models indicate that the travellers' choice decisions are sensitive to the modal level-of-service which describe the access to the station. The weighted price station model in particular indicates that the weighted auto time is an attribute to which the travellers' station choice decisions are most sensitive. This is because the weighted auto time has a much greater coefficient than any other variable in the model.

In the weighted inclusive price station model, the coefficient of the inclusive price variable is statistically significantly different from the value of 1.0, which suggests that the travellers do not value or assign the same weights to the set of transportation system attributes when making decisions of the access mode and station choices.

Finally, the access mode and station models are applied to different situations and they produce good predictive results.

Regarding the simultaneous models and the mode-station sequential models, no conclusive explanation can be given at this point to the failure to produce valid choice models. However, it may be speculated from the results of this research that the travellers' decision making process for the mode and station choices of the access trip is behaviorally of a sequential nature, the sequence of the station choice followed by the mode choice.

Extension of the Research

It has been learned through this research that more detailed information concerning the level of service of transportation system as well as the individual samples is required in order to estimate the disaggregate choice models. Within the extent of this study, for example, the exact location of the trip origin, the individual income, the specific information of the parking facility at the stations, and most importantly the 'relevant alternatives' of the choices actually considered by each individual behavioral unit - the traveller, should be specificly asked in taking surveys for disaggregate choice modeling. Of specific concern is the auto mode, which in this study is considered a relevant mode for every traveller.[1]

Secondly, the disaggregate models have to be aggregated for implementation in transportation and urban planning.[2] Up to date, little work has been done on analyzing the travel demand by means of aggregating the disaggregate models. Of the few existing aggregation procedures, more empirical studies are warranted.

---

[1]See Appendix III.

[2]Talvitie, A. P. *Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models*, Unpublished paper, University of Oklahoma (1973).

Finally, although some disaggregate travel demand studies have been made concerning the frequency, destination, mode choice, time of day and routes for the various trip purposes, these studies pertain to the entire trip. It is felt that more emphasis should also be placed on the access, egress part of the main trip, especially when mass transit will obviously become more and more widely used in the future, for both work-trips and non-work trips.

In conclusion, the disaggregate modeling technique is found to be highly plausible in terms of the relatively small data sample required and the savings in time and cost as compared to the aggregated demand models. Valid access mode and station selection models are developed in this study and they have accurate predictive ability. More research in the application and eventually the implementation of disaggregate demand models in the area of the main trip as well as the access trip are necessary.

BIBLIOGRAPHY

Deen, T. B., Irwin, N. A. and Mertz, L. "Application of A Modal Split Model to Travel Estimated for the Washington Area," *Highway Research Record 38*, pp. 97-123 (1963).

Ben-Akiva, M. *Structure of Passenger Travel Demand Models*, Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology (1973).

Charles River Associates, Inc. (CRA), *A Disaggregate Behavioral Model of Urban Travel Demand*, Federal Highway Administration, U.S. Department of Transportation, Washington, D. C. (1972).

Lave, C. A. *Modal Choice in Urban Transportation: A Behavioral Approach*, Ph.D. Dissertation, Department of Economics, Stanford University (1968).

Lisco, T. E. *The Value of Commuter's Travel Time: A Study in Urban Transportation*, Ph.D. Dissertation, Department of Economics, University of Chicago (1967).

McFadden, D. *The Revealed Preference of a Government Bureaucracy*, University of California, Berkeley (1968).

Rassam, P. R. *et al.* "The N-Dimensional Logit Model: Development and Application," *Highway Research Record 369*, pp. 135-147 (1971).

Stopher, P. R. "A Probability Model of Travel Mode Choice for the Work Journey, *Highway Research Record 283*, pp. 57-65 (1969).

Stopher, P. R. *A Multinomial Extension of the Binary Logit Model for Choice of Mode of Travel*, Unpublished paper, Northwestern University (1969).

Talvitie, A. P. "Comparison of Probabilistic Modal Choice Models: Estimation Methods and System Inputs," *Highway Research Record 392*, pp. 111-120 (1972).

Talvitie, A. P. *Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models*, Unpublished paper, University of Oklahoma (1973).

Theil, H. A. "A Multinomial Extension of the Binary Logit Model," *Rept. 6631* (1966).

Warner, S. L. "Stochastic Choice of Mode in Urban Travel," *Northwestern University Press* (1962).

APPENDIX I

Derivation of McFadden's Multilogit Model

Using the same notation in Chapter II, the independence of irrelevant alternatives assumption can be expressed mathematically as:

$$P_{ij}P(j:A) = (P_{ji})P(i:A) \quad . \qquad \text{(AI-1)}$$

From one of the basic conditions of a probability model, $P_{ij} + P_{ji} = 1$ or $P_{ji} = 1 - P_{ij}$ for the subset $A' \varepsilon A$ which contains elements i and j only, Equation (AI-1) may be written as

$$P_{ij}P(j:A) = (1 - P_{ij})P(i:A) \qquad \text{(AI-2)}$$

and

$$P(j:A) = (\frac{1 - P_{ij}}{P_{ij}})P(i:A) \quad . \qquad \text{(AI-3)}$$

For any element, k, of set A, the above equation may be expressed as,

$$P(k:A) = (\frac{1 - P_{ik}}{P_{ik}})P(i:A) \quad . \qquad \text{(AI-4)}$$

From the other basic assumption of a probability model, $\sum_{k \varepsilon A} P(k:A) = 1$, the following may be obtained by summing P(k:A) in Eq. (AI-4) over all $k \varepsilon \{A\}$.

$$1 = \sum_{k\varepsilon A} P(k:A) = \sum_{k\varepsilon A} [\frac{(1 - P_{ik})}{P_{ik}} P(i:A)] \qquad \text{(AI-5)}$$

thus,

$$1 = P(i:A) \sum_{k\varepsilon A} \frac{(1 - P_{ik})}{P_{ik}} \qquad \text{(AI-6)}$$

and

$$P(i:A) = \frac{1}{\sum_{k\varepsilon A} \frac{(1 - P_{ik})}{P_{ik}}} = \frac{1}{\sum_{k\varepsilon A} \frac{P_{ki}}{P_{ik}}} \; . \qquad \text{(AI-7)}$$

Assume the ratio of P_{ki} and P_{ik} is the exponential of a certain index function, f(i,k), which is a function pertaining to elements i and k, e.g.,

$$\frac{P_{ki}}{P_{ik}} = \exp[f(i,k)] \qquad \text{(AI-8)}$$

then, f(i,k) has the following properties:

$$f(i,i) = 0 \qquad \text{(AI-9)}$$

$$f(i,j) = - f(j,i) \qquad \text{(AI-10)}$$

$$f(i,j) + f(j,k) + f(k,i) = 0^1 \qquad \text{(AI-11)}$$

The above three properties which the index function must satisfy are satisfied if f(i,k) = B(k) - B(i), where B(k) and B(i) are linear functions describing characteristics of elements k and i, respectively.

---

[1]Equation (AI-11) can be proved by permuting subscripts i, j and k in expression (AI-1) and multiplying to obtain $P_{ij}P_{jk}P_{ki} = P_{ji}P_{ik}P_{ki}$. Therefore $(P_{ij}/P_{ji})(P_{jk}/P_{kj})(P_{ki}/P_{ik}) = 1$, and $\{\exp[f(j,i)]\}\{\exp[f(k,j)]\}\{\exp[f(i,k)]\} = 1$. Combine the exponents of the last expression, one obtains, $\exp[f(j,i) + f(k,j) + f(i,k)] = 1$ and finally, $f(j,i) + f(k,j) + f(i,k) = 0$.

Thus, from Eq. (AI-7), the probability of choosing element $i\varepsilon\{A\}$ from set A is,

$$P(i:A) = \frac{1}{\sum_{k\varepsilon A} \exp[B(k) - B(i)]} \quad . \qquad \text{(AI-12)}$$

Multiplying both the numerator and the denominator of Eq. (AI-12) by, $\exp[B(i)]$, and

$$P(i:A) = \frac{\exp[B(i)]}{\sum_{k\varepsilon A} \exp[B(k)]} \quad . \qquad \text{(AI-13)}$$

Equation (AI-13) is a symmetrical, generalized mathematical expression which describes the probability that an alternative is chosen is a function of the characteristics associated with all the alternatives in the choice set.

APPENDIX II

Derivation of the Direct and Indirect Elasticities for the Multilogit Model

The elasticity of the choice probability of alternative i, P_i, with respect to the lth variable of alternative j, X_{lj}, is

$$E_{X_{lj}}(P_i) = \frac{\partial P_i}{\partial X_{lj}} \cdot \frac{X_{lj}}{P_i} \quad , \tag{AII-1}$$

where alternative j may be any alternative in the choice set.

For the multilogit model,

$$P_i = \frac{e^{U_i}}{\sum_{all\ k} e^{U_k}} \quad , \tag{AII-2}$$

where U is the utility function while subscripts i and k indicate the alternatives; and $U_k = U_k(X_k)$ is a linear function of the characteristics of alternative k.

Since P_i is a function of U_j, and U_j is a function of X_{lj}.

$$E_{X_{lj}}(P_i) = E_{Uj}(P_i) \cdot E_{X_{lj}}(Uj) \tag{AII-3}$$

In Eq. (AII-3), if j = i, then the elasticity is called the direct elasticity, and

$$E_{X_{1i}}(P_i) = E_{U_i}(P_i) \cdot E_{X_{1i}}(U_i) \ . \qquad \text{(AII-4)}$$

Assuming independence among the utilities of the various alternatives,

$$E_{U_i}(P_i) = \frac{\partial P_i}{\partial U_i} \frac{U_i}{P_i} \qquad \text{(AII-5)}$$

thus,

$$E_{U_i}(P_i) = \frac{[\,(\sum_k e^{U_k}) - (e^{U_i})] \cdot U_i}{\sum_k e^{U_k}} \ . \qquad \text{(AII-6)}$$

It follows that,

$$E_{U_i}(P_i) = U_i(1 - P_i) \ . \qquad \text{(AII-7)}$$

On the other hand,

$$E_{X_{1i}}(U_i) = \frac{\partial U_i}{\partial X_{1i}} \frac{X_{1i}}{U_i} = b_1 \cdot \frac{X_{1i}}{U_i} \qquad \text{(AII-8)}$$

b_ℓ is the coefficient of the lth variable. Conbining Eqs. (AII-7) and (AII-8),

$$E_{X_{1i}}(P_i) = b_1 X_{1i}(1 - P_i) \ . \qquad \text{(AII-9)}$$

Cross Elasticity

Back to Eq. (AII-3), if $J \neq i$, then the elasticity is called the cross elasticity and,

$$E_{X_{1j}}(P_i) = E_{U_j}(P_i) \cdot E_{X_{1j}}(U_j) \ . \qquad \text{(AII-10)}$$

Again, assuming the independency among the various utilities,

$$E_{U_j}(P_i) = - U_j P_j \qquad \text{(AII-11)}$$

also,

$$E_{X_{1j}}(U_j) = b_1 \frac{X_{1j}}{U_j} \qquad \text{(AII-12)}$$

therefore,

$$E_{X_{1j}}(P_i) = - b_1 X_{1j} P_j \ . \qquad \text{(AII-13)}$$

Therefore generalized formula for the direct and cross elasticities of a multilogit model is

$$E_{X_{1j}}(P_i) = b_1 X_{1j} (\delta - P_j) \qquad \text{(AII-14)}$$

$$\delta = 1 \text{ for } j = i$$

$$\delta = 0 \text{ for } j \neq i$$

APPENDIX III

Test of Relevant Alternatives

One of the key questions concerning the multi-logit model employed in this research is whether or not the alternatives in each choice assigned to the individual traveller are indeed the "relevant alternatives" actually considered by the traveller. Specifically, the auto mode was assumed in this research as a relevant access mode alternative to every traveller. On the other hand, the observed data suggest that the auto mode may not always be a relevant access mode alternative, because there are only few travellers observed to use the auto mode when the distance between the trip origin and the access train station is less than one half of a mile.

As an extension of this research, a small test was conducted to investigate the sensitivity of the choice model with respect to how the "relevant alternatives" are defined. This is done by assuming that the auto mode is not a relevant alternative for those travellers who live very close to their chosen stations.

Consequently, unless the individual traveller was observed to use the auto mode, the auto mode was not considered as an access mode alternative for a traveller if the distance between his trip origin and his chosen station is less than 0.5 mile. The access mode models $P(m|s)$ were then estimated and evaluated in the same way as the access mode model described in Chapter IV. One of the estimated access mode models appears to have reasonable coefficient signs and statistical indicators. The coefficients and the statistical information of this model are given in Table 15.

TABLE 15

Coefficients and Relevant Information of the Access Mode Model P(m|s) with "Relevant" Auto Mode Alternative

Variable	Coefficient	Standard Error
Out-of-vehicle time	-.266	.061
Auto time	-.316	.258
Bus time	-.189	.097
Operating cost	-.350	.118

χ^2 = 80.6170 with degrees of freedom = 4

The statistical tests indicate that all the variables in this model as well as the model as a whole are statistically significant with at least .90 level of confidence.

Among the time variables in this model, the coefficient of auto time has the greatest value followed by the coefficient values of out-of-vehicle time and bus time, in that order. The value of time implied by the time and cost variables are \$.44 (S.D. .30) for out-of-vehicle time, \$.54 (S.D. .54) for auto time and .32 (S.D. .24) for bus time.

The numbers of misclassification and the prediction rates are given in Table 16.

TABLE 16

Accuracy of the Access Mode Model with "Relevant" Auto Mode Alternative

Data	N	M	$\alpha = 1 - \frac{M}{N}$
I.C.	150	15	90%
R.I./S.S.	50	8	84%

N, total number of observations

M, number of misclassifications

α, predictive accuracy

The comparison of the expected and the actual number of mode users when the model is applied to the base data and the control data are given in Table 17.

TABLE 17

Comparison of Number of Mode Users

Data	Mode	N_E	N_S	Rate
Base (I.C.)	Auto	41	47	87%
	Walk	54	50	92%
	Bus	54	53	98%
Control (R.I./S.S.)	Auto	9	16	56%
	Walk	20	18	89%
	Bus	21	16	69%

N_E, expected number of travellers

N_S, actual number of travellers

The expected and actual number of users for auto mode and bus mode does not compare too closely. For auto mode, especially, the accuracy rate of the comparison is only 56%, which is less than desirable.

Comparison between this access mode model (with "relevant" auto mode alternative) and the access mode model in the main text reveal several interesting points. Superficially, it appears that the variables in the two models have similar coefficient values. The coefficient of the two models also yielded similar implied values of time, although the values of time implied by the present model are somewhat lower than by the model of the main text. This model, as the other model, also indicates that travellers are most sensitive to changes in auto time, not too sensitive to changes in out-of-vehicle time and least sensitive to changes in bus time.

Statistical tests between the coefficients of the variables indicate that the variable coefficients of the two models are significantly different. The value of time implied by the coefficients of the two models are also statistically different with the exception of the value of out-of-vehicle time. This suggests that different models are resulted depending on how the "relevant alternatives" are defined.

Nevertheless, the decision was made to conduct the entire research by assuming that auto mode was a relevant alternative for every traveller, because some individuals were in fact observed to use auto mode even though the distances between their trip origins and their chosen stations are less than 0.5 mile,[1] and because the chosen model was judged to have a slightly better predictive ability.

[1]This has to be qualified by the statement that the trip origin was observed only by ¼ sq. mile and the distance to station was reported by the "number of blocks" to the station in the survey.

74

6969