SIMPLIFIED METHODS IN TRANSPORTATION ANALYSIS

by

Serge Tsygalnitzky

Submitted in partial fulfillment

of the requirements for the degree of

Master of Science

at the

Massachusetts Institute of Technology

May 1977

Signature redacted

Signature of Author..
Department of Civil Engineering (May 11, 1977)

Signature redacted

Certified by...
Thesis Supervisor

Signature redacted

Accepted by..
Chairman, Departmental Committee on Graduate Students
of the Department of Civil Engineering

ABSTRACT

SIMPLIFIED METHODS IN TRANSPORTATION ANALYSIS

by

Serge Tsygalnitzky

Submitted to the Department of Civil Engineering
on May 11, 1977 in partial fulfillment of the
requirements for the degree of Master of Science

This thesis is concerned with the development of simplified methods of analysis in transportation systems. The objective is to find ways to predict the impacts and estimate a demand model, which are not based on large scale data or computer-based analyses. Instead, the methods are designed to be used on programmable pocket calculators for ease of analyst use.

The methods are intended to be useful to analyse decisions from the perspective of a transportation operator. Among the specific methods developed, the most important are:

- Design of a simplified way of forecasting using a logit model;
- Estimation of a logit model with least squares method;
- Inferences about the context when a lot of data is missing, as obtaining of O-D matrix for all modes when the characteristics of only one mode are known.

In conclusion, directions for future work are also indicated. The results of this thesis demonstrate that useful simplified methods can be developed, and thus open up an important new field of research.

Thesis Supervisor: Marvin L. Manheim

Title: Professor

ACKNOWLEDGMENTS

I would like to take this opportunity to thank all those who assisted in this effort.

Professor Marvin L. Manheim, my thesis supervisor for his guidance and patience with my efforts and for his attempts to make this document readable.

My friends, Jesse Jacobson, who helped me in the discussion of the ideas presented in this thesis, and Peter Furth, who gave freely of his time in writing calculator programs.

Finally, I would like to thank Gilbert High and Jane Roth for their wonderful typing.

Cette these est dediée a mes parents.

TABLE OF CONTENTS

FIGURES AND TABLES

CHAPTER I

Summary

In transportation systems analysis the concern is to try to predict the effect of any change in the transportation system. These changes are short range as well as long range. In this thesis we will focus on the first type of prediction. To do this prediction an analyst needs a demand function. During many years this demand function has been the fundamental subject of many studies in transportation. This demand function has first been an aggregate one giving prediction in an aggregate way, and taking aggregate data as inputs. Unfortunately, this model was more correlative than causal. Lately the behavioral models have been introduced trying to put heavier emphasis on human psychology for prediction than the previous ones. In both cases it has appeared that the accuracy of the prediction has increased with years. On the other hand, the amount of data required to estimate a model, or to make a forecast has also increased. The availability of computers to process all this data has become nowadays a decisive point. However, if there is a need for such accurate demand functions, it is not the ideal solution for all kinds of problems. In many cases there is a need for short range approximate predictions, and the use of developed models can be too expansive for the use of the results.

Therefore, there is a need for simplified methods of prediction. This thesis will try to find some of them. In the second chapter, we will define the goal of simplified methods. Mostly they are useful for transportation operators. We will also develop a basic methodology. We will choose a generic demand function (logit model) and try to apply some of these methods. In particular we will develop a simplified method of forecast (pivot point method) and estimation (least squares method).

In Chapter 3 we will focus primarily on the decision process of a transportation operator (b us operator). We will develop the methods presented in the previous chapter and study them for the different implementations the operator can take. Basically, these implementations are going to be changes in the level of service of the transportation system.

We will see how the pivot point method can be applied for different purposes, and how an operator can be able to predict the new ridership on the transportation system after a change in the service.

We will also develop the model estimation giving some examples for the two basic cases: modal split model and multiple choice model. We will then study a simplified way of running multiple linear regressions in order to be able to program them on a pocket calculator.

In Chapter 4 we will be concerned with the inferences about the context. Every stage of the prediction process is based on data, and the data is expensive to get. Therefore, we will try to develop some methods in order to infer the basic data for a model estimation or a prediction from a minimum data. We will not study all the cases but will give some interesting examples, such as: how to get an O-D matrix for transit when only boarding and alighting counts are available. How is it possible to derive an O-D matrix for all modes when only a Transit O-D matrix is available? Finally, we will introduce an alternative way of making simplified predictions through the bias of elasticities.

In Chapter 5 we will conclude this thesis trying to show the imperfections of such methods and their limits.

This thesis is by no means a guide for simplified methods. Its goal is primarily to show that there exist some methods that can be applied to get relevant results. Yet they are only examples of what can be done. Mainly in Chapter IV we will try to develop some alternative ways to get some data, but they are only a few examples in an unlimited field of research. Furthermore, we will study some of these methods applied to a specific case of buas operations. The next step would be to extend these methods to other purposes. Also we will make some implicit assumptions that can certainly be argued (as the split of the trips between three purposes: work, shopping and social-recreation).

Lastly, we will develop some methods that are within the capacity of our pocket calculator. (Several programs and worksheets are included in the appendix.) As time passes the capacity of thes pocket calculators increases, and these methods can very well be outdated in a few years. Also, it is important to remember that many of these methods are empirical and have no mathematical argument. Their only quality is that they give fair results according to few experiments.

A caution must then be given to the reader that some of these methods are still subject to controversy and our hope is that one day they will be fully justified.

CHAPTER II

II-1. The Nature of the Approach

In transportation systems analysis the centrl concern is to try to predict the consequences of any proposed changes in the transportation system. Two types of predictions are possible: long run and short run prediction. For long run prediction changes in the social and economic activity are important. This kind of prediction is essential for Urban Planning.

In short run prediction one is interested primarily in changes in travel patterns: how people will react to a change in the level of service of the system; what kinds of shifts from one mode of transportation to another might result? What changes in frequency or destination of trips will occur?

This thesis is concerned with short run prediction. To make a short run prediction a demand function is an essential component. Therefore the first task of the analyst will be to get such a function. Two cases are possible: the demand function already exists and can be used for prediction or it does not exist and a model must then be estimated.

In the past the demand models used were based on aggregate data (i.e. zonal averages), were more correlative than causal and utilized data obtained by standard survey, especially home interviews. In the present the emphasis has been put on disaggregate behavioral models because they represented more accurately the situation. But in both cases, a lot of data are required for the calibration of the model.

And this is expensive. A home interview survey costs $50, and at least 200 home interviews are needed to calibrate a disaggregate model, i.e. $10,000. In fact, 1,000 to 2,000 are more often needed. For aggregate models 3% of the entire metropolitan area households are required. Furthermore, this information must be processed and it appears that the only practical way of processing it is to write a program on a computer. As an example the coefficients of the logit demand model are computed as solutions of a non-linear equation (maximum likelihood).

As we can see, the estimation of such models is based on the availability of computers and on a large budget.

If the demand function is available the problem of prediction becomes in a sense simpler. If the model is an aggregate one, the anlayst will need only data fitted to the model to make a forecast and apply the model directly. If the model is disaggregate a problem of aggregation arises and the results are obtained with more computations. At that level the required data will be sometimes less numerous and the computer will not always be necessary.

But of course the accuracy of the results is an increasing function of the data, computer availability, and consequently of the price (for both prediction and estimation). As the budget of a study is a constraint, every analyst should adapt a methodology to this constraint. However, it is not clear that the accuracy is a linear function of the price or that a budget twice bigger than another one will in case lead to results of which the accuracy will be two times greater (if this ever has a meaning!).

Another general comment is the great uncertainty to which some prediction results are subject. A prediction can be performed if the economic environment remains steady but often some unpredictable changes in economy occur (ex. the oil crisis of 1973) and deprives some results of significance.

Then it appears that the very expensive methodologies are not the best way of making precictions all the time.

If the goal of the analysis is to get rough estimates in a prediction, a detailed analysis might be too expensive for the use of the results. Therefore, there is a need for a new direction of analysis, where the methodology developed will consist of simple methods, easily applicable and giving results still significant. These methods should be based on "paper and pencil" or even on pocket programmable calculator and consequently avoid to require any computer programs. It will be a challenge to find some results, but the interests are obvious.

The goal of this thesis is to try to deal with this challenge, develop some simplified methods, try if possible to recognize their limits, and apply them to some examples in order to show their efficiency.

II-2. To Whom Is This Methodology Addressed?

The role of an analyst is to do an analysis which is valid, practical, and relevant to the issues. It means that the relevance of any analysis must be defined in terms of the goals of the decision makers involved. If the goals are comprehensive it may be important to be accurate and very detailed. On the contrary, if the goals are primarily short range and concerned with specific operational decisions the analysis can be short and more approximate. In many situations a decision maker needs only approximate results to support a decision, because that decision does not require a lot of resources or imply drastic changes in the economics of the situation. This last is the case we are concerned with. This case is especially relevant when the decision makers are operators of a transportation system. Generally operators are mainly concerned with the financial impacts of decisions on their company; as an ultimate reflection of the consequences of changes in service and in market responses they try to implement changes which will increase their profit (or more often decrease their loss). These changes are mostly changes in level of service attributes such as frequency or fare or changes in coverage (for bus lines) such as new routes or extensions of an existing route.

To reach a decision in this context the process is an incremental one. First, the operator makes initial judgements about the effect of a possible change, then if the change seems desirable he will supplement his initial judgements with analyses, to develop more information. Simplified methods of transportation analysis are important in this process. Detailed methods are too expensive relative to the importance of the results; but on the other hand, human judgements unsupported by mathematical help is insufficient.

Therefore, this thesis will focus on operator decisions, seek to develop some useful tools to assist them in their first decision. We will focus primarily on bus operators, but the results are readily transferred to other problems.

II-3. Basic Stpes of Decision for a Bus Operator*

The role of a bus operator (or any transportation system manager) can be visualized in 8 steps which are:

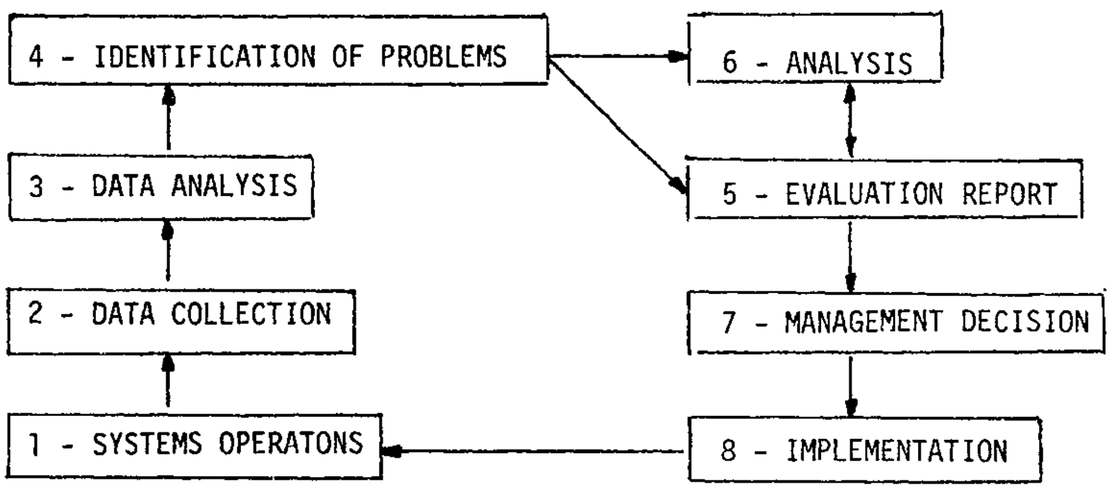
1. The system operation summarizing all the activities which are involved in producing transportation
2. The data collection involving data from supply (fares collected, vehicle hours, etc.) as well as demand (ridership, O-D matrix, etc.)
3. "Data analysis" - this is explained by itself
4. "Identification of the problems"
5. "Evaluation report," which summarizes the previous steps and presents some management actions
6. "Analysis." The actions presented in step 5 are analysed to reach a decision.
7. "Management decision"
8. "Implementation." The decision of step 7 is implemented gradually and the results are studied.

This can be summarized by Figure 1.

---

*(Cf. Marvin L. Manheim, Fundamentals of Transportation Systems Analysis, 1976, in print.

Figure 1.



4 - IDENTIFICATION OF PROBLEMS
6 - ANALYSIS
3 - DATA ANALYSIS
5 - EVALUATION REPORT
2 - DATA COLLECTION
7 - MANAGEMENT DECISION
1 - SYSTEMS OPERATONS
8 - IMPLEMENTATION

As can be seen the two key steps are step 2 and step 6: data collection and analysis. Of course, everything depends on step 6, which itself depends on step 2.

The level of importance of these different steps can vary according to the importance of the change. But as we are here studying only simple incrementations where the environment is stable economically and where the different alternatives are known by the consumers, simplified methods are necessary for prediction.

Therefore, some simplifications must be done in these eight steps and mainly in steps 2 and 6.

The simplification in step 2 (data collection) will be to classify the data by order of availability and then choose the types that are the most easily obtained. The simplification in step 6 will be, as we said previously, to use easily applicable models, i.e. models which do not require computers or unlimited data to give some results.

These two sorts of simplification are greatly interrelated. It is obvious that in most of the cases it will be impossible to use the simplest model with the most easily available data just because there will be no compatibility with the two concepts. The challenge of the analyst will then be to adjust the data and the model in order to minimize the cost of the study under the constraint to get still relevant results for prediction.

II-4. Specification of Demand Function

As we said previously an operator needs a demand function to make some predictions. This one may take all sorts of possible forms including the aggregate or disaggregate type. But the operator needs also some data as inputs to get the results. Then we can see that the problems will arise when the demand model or the data will be missing. We will study the different cases and the inferences that can be made in Chapter IV. However, we can already assume that the prediction will be straightforward when the data and the model will be available. As we cannot possibly study all sorts of models we will try in this chapter to choose one of them for the rest of the study. This demand model should have two qualities: it should be estimable by hand or pocket calculator, and readily applicable for a forecast.

The choice exists mostly between the aggregate type and the disaggregate type of model. It appears that for a forecast the aggregate models are easier to handle than the disaggregate ones where an aggregation is needed. Furthermore, the calibration of aggregate models is often based on linear regression (that can be programmed on a pocket calculator) while the disaggregate ones are based on maximum likelihood because the data are discrete (the data are based on a binary Yes/No answer from some surveys). On the other hand, the disaggregate models are definitely the most up-to-date. The introduction of the logit model by Mc Fadden (1968) and studies by Ben Akiva (Phd. 1973) have shown a lot of qualities. First, it is more accurate and more policy-sensitive than the aggregate models. Second, it allows a simultaneous procedure for the alternative choices of decision. And this is very realistic. Mainly the stages of demand are divided into generation, distribution and modal split. With the exception of work trips where the generation and the distribtuion are supposed to reamin constant, it has appeared that the level of service was interfering on the 3 stages of demand for the other purposes, and the logit model takes into account this fact.

The major problem comes precisely from the disaggregate nature of this model; thus, it will need an aggregation to make some forecasts, and requires the resolution of a non-linear equation to be calibrated. The way to overcome these difficulties will be to aggregate this model for both estimation and prediction. Of course, this model will lose some of its qualities (as being very accurate) but on the other hand it will gain the qualities of the aggregate models, and this is what we seek. That is why we will choose this alternative even if we have some inconveniences coming from the aggregation.

The logit model can be presented as:

$$P_r(i,A_t) = \frac{e^{U_{it}}}{\sum_{j \in A_t} e^{U_{jt}}} \qquad \text{where}$$

t = behavioral unit ($t = 1, \ldots, T$)

A_t = set of relevant alternatives for behavioral unit t.

$P_r(i;A_t)$ = probability that behavioral unit t will choose alternative i out of the set A_t.

U_{it} = the utility of alternative i to behavioral unit t. V_{it} is a function of the characteristics of alternative i and the socio-economic characteristics of behavioral unit t.

$$V_{it} = X_{it}' \theta = \sum_{k=1}^{K} X_{itk} \theta_k$$

X_{it} = vector of function depending on alternative i and behavioral unit;

θ = vector of coefficients.

To aggregate this model we will use the classification given by Koppelman.* Yet the aggregation is not easy and many errors can result from it. The approach of aggregation can be summarized as in Figure 2.

The difficulty consists mainly in the aggregation procedure. There are five types of procedures; the expected share of people choosing an alternative is given:

--- Procedure of enumeration

by the average over all the individuals knowing their characteristics. It is the perfect way of aggregating a model but it requires all the data possible.

--- Procedure of summation

by the sum over all the individuals assuming there is a distribution of the characteristics over the consumers. It requires the knowledge of the distributions.

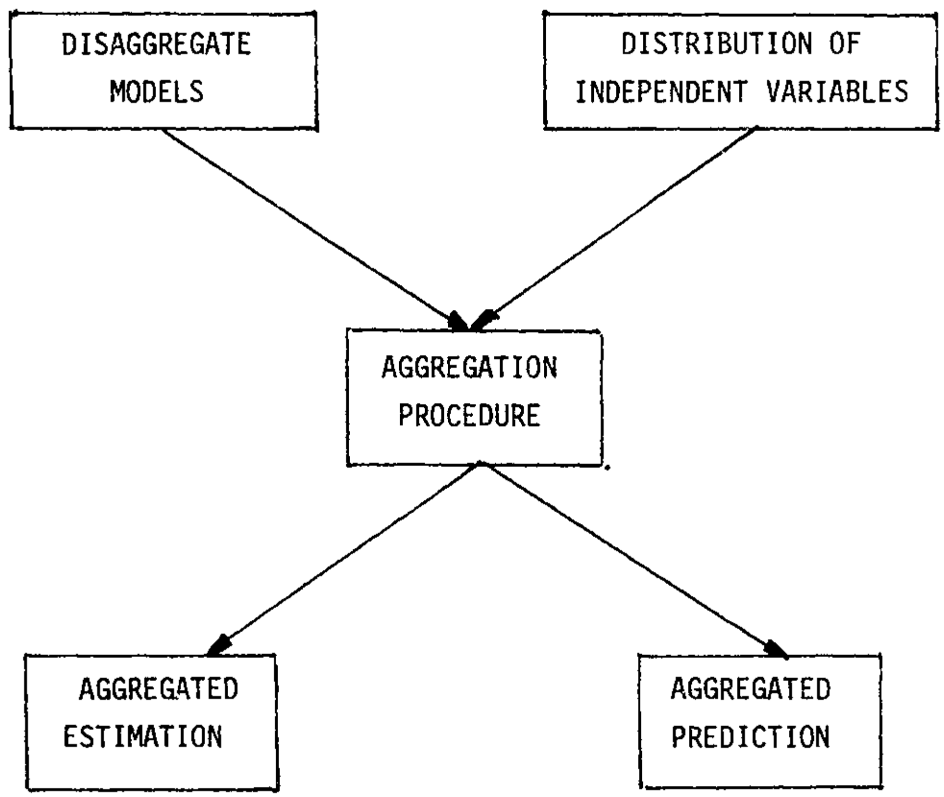
--- Procedure of statistical differentials

by linearizing the disaggregate choice function by a Taylor series expansion and taking the expectation across the aggregate prediction groups.

---

*Frank Koppelman, Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models (MIT, PhD. Thesis 1975 unpublished).

Figure 2.


DISAGGREGATE MODELS
DISTRIBUTION OF INDEPENDENT VARIABLES
AGGREGATION PROCEDURE
AGGREGATED ESTIMATION
AGGREGATED PREDICTION

--- Procedure of classification

by summing over weighted homogeneous groups. But it requires the knowledge of all these groups.

--- Naive procedure

by assuming that all the consumers form an homogeneous group and taking then the average over them.

Of course the more sophisticated the aggregation method, the more accurate the results will be. But once again we want simplified methods based on data which are frequently limited. Therefore, the type of aggregation will be directly related to the data available. However, two types of aggregation should be more important: the procedure of classification and the naive procedure. If our data is explicit enough for the operator to be able to classify the consumers into groups roughly homogeneous, the procedure of classification should be used. This will happen if a home interview has been done, or using the census data, or even if we just classify the consumers by auto availability and income.

In most cases the naive procedure will be the one used; often the aggregation errors are substantial, but compared to the errors coming from other sources (as errors coming from the regressions) they are not so important. Furthermore, the bias can be diminished if the auto availability and bus availability variables can be included in the aggregation, because they are the major sources of the bias. So now we will assume that in most of our analyses we will use a logit model with naive aggregation procedures, i.e. an aggregated logit model.

II-4-1. Procedure of Forecast

The next point is to have general methods of forecast using this model. We have:

$$\Pr\text{ (choose alternative i)} = \frac{e^{U_{it}}}{\sum\limits_{j \varepsilon A_i} e^{U_{jt}}}$$

where this time the behavioral units t will be the group defined by the aggregation procedure. From the point of view of a bus operator, the important thing to know is the probability of choosing transit. (For illustration we consider only mode split. But these methods can be generalized to other choices dimension as will be seen later). We can represent the logit model by:

$$\Pr\text{ (choose transit)} = \frac{e^{U_T}}{\sum\limits_{j \varepsilon \text{ all modes}} e^{U_J}}$$

U_J = utility of mode J (T = transit)

If some changes occur in the level of service of transit, the utility of transit will change:

$$U_T \longrightarrow U'_T = U_T + \Delta U_T$$

We can assume after that the utilities of the other modes of transportation will not depend on the utility of transit (i.e. no cross-elasticities). It is a strong assumption, but it is often done. Therefore, it means that U_J with $J \neq T$ will remain the same after the changes on the bus line.

Then we can predict the new probability in two ways:*

A - Synthetic

There we just replace U_T by U'_T.

and we get the new probability of transit by:

$$\Pr(\text{choose transit}) = \frac{e^{U'_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{U'_T}}$$

but we will need to know all the characteristics of each utility in order to compute that ratio.

B - Incremental

This method is also called Pivot point method. We have:

$$P' = \frac{e^{U'_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{U'_T}} = \frac{e^{\Delta U_T} \times e^{U_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{\Delta U_T} e^{U_T}} = e^{\Delta U_T} \times \frac{e^{U_T}}{\sum\limits_{J} e^{U_J}} \times \frac{\sum e^{U_J}}{\sum\limits_{J \neq T} e^{U_J} + e^{\Delta U_T} \times e^{U_T}}$$

$$P' = P \times e^{\Delta U_T} \times \frac{1}{1 + P(e^{\Delta U_T} - 1)} \qquad (1)$$

We see then that the new probability depends only on the old one and on the change in transit utility. Therefore we do not need to know the utilities of all the modes nor all the utility function of transit, but only the present probability or share.

As we are looking for simplified methods, this method seems to be particularly suitable to our concern. However, we will demonstrate other methods, mainly when there will be a change in the coverage of the line.

---

*cf Marvin Manheim, Fundamentals of Transportation Systems Analysis, 1977, in press.

Then we can predict the new probability in two ways:*

A - Synthetic

There we just replace U_T by U'_T.

and we get the new probability of transit by:

$$\text{Pr (choose transit)} = \frac{e^{U'_T}}{\sum_{J \neq T} e^{U_J} + e^{U'_T}}$$

but we will need to know all the characteristics of each utility in order to compute that ratio.

B - Incremental

This method is also called Pivot point method. We have:

$$P' = \frac{e^{U'_T}}{\sum_{J \neq T} e^{U_J} + e^{U'_T}} = \frac{e^{\Delta U_T} \times e^{U_T}}{\sum_{J \neq T} e^{U_J} + e^{\Delta U_T} e^{U_T}} = e^{\Delta U_T} \times \frac{e^{U_T}}{\sum_{J} e^{U_J}} \times \frac{\sum e^{U_J}}{\sum_{J \neq T} e^{U_J} + e^{\Delta U_T} \times e^{U_T}}$$

$$P' = P \times e^{\Delta U_T} \times \frac{1}{1 + P(e^{\Delta U_T} - 1)} \qquad (1)$$

We see then that the new probability depends only on the old one and on the change in transit utility. Therefore we do not need to know the utilities of all the modes nor all the utility function of transit, but only the present probability or share.

As we are looking for simplified methods, this method seems to be particularly suitable to our concern. However, we will demonstrate other methods, mainly when there will be a change in the coverage of the line.

*cf Marvin Manheim, Fundamentals of Transportation Systems Analysis, 1977, in press.

Then we can predict the new probability in two ways:*

A - Synthetic

There we just replace U_T by U'_T.

and we get the new probability of transit by:

$$\text{Pr (choose transit)} = \frac{e^{U'_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{U'_T}}$$

but we will need to know all the characteristics of each utility in order to compute that ratio.

B - Incremental

This method is also called Pivot point method. We have:

$$P' = \frac{e^{U'_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{U'_T}} = \frac{e^{\Delta U_T} \times e^{U_T}}{\sum\limits_{J \neq T} e^{U_J} + e^{\Delta U_T} e^{U_T}} = e^{\Delta U_T} \times \frac{e^{U_T}}{\sum\limits_{J} e^{U_J}} \times \frac{\sum e^{U_J}}{\sum\limits_{J \neq T} e^{U_J} + e^{\Delta U_T} \times e^{U_T}}$$

$$P' = P \times e^{\Delta U_T} \times \frac{1}{1 + P(e^{\Delta U_T} - 1)} \qquad (1)$$

We see then that the new probability depends only on the old one and on the change in transit utility. Therefore we do not need to know the utilities of all the modes nor all the utility function of transit, but only the present probability or share.

As we are looking for simplified methods, this method seems to be particularly suitable to our concern. However, we will demonstrate other methods, mainly when there will be a change in the coverage of the line.

---

*cf Marvin Manheim, Fundamentals of Transportation Systems Analysis, 1977, in press.

II-4-2. Procedures of Estimation

There are cases where the model does not exist, and there is a need to calibrate it. We will set here a general way of estimating this model. The basic assumption will be that we have all the required data. The different cases where this will not be true will be discussed in Chapter IV.

Again we have:

$$Pr\ (i;\ A_t) = \frac{e^{U_{it}}}{\sum_{j \varepsilon A_t} e^{U_{jt}}}$$

where $U_{it} = X'_{it}\theta = \sum_{k=1}^{K} X_{itk}\theta_k$

To calibrate a model means to compute this latter vector θ.

There are two ways to compute this vector:

-- The maximum likelihood method

-- The least squares method

A - Maximum likelihood*

In this case we do not aggregate this model. The data required will be disaggregate (i.e. individual answers). The likelihood function is:

$$L = \prod_{t=1}^{T} \prod_{i \varepsilon At} P(i;\ A_t)^{g_{it}}$$

where $g_{it} = 1$ is chosen, 0 otherwise.

Taking the natural log and then taking derivatives with respect to the coefficient we want to compute, we get:

---

*cf Moshe Ben Akiva; The Logit Model and Functional Specifications of Alternative Models (Ph.D., 1975).

$$\frac{\partial L^*}{\partial \theta_k} = \sum_{t=1}^{T} \sum_{i \varepsilon A_t} (g_{it} - P(i;\, A_t)) \cdot X_{itk} = 0$$

and θ_k is the solution of this non linear equation. As we can see this equation is rather complicated, and unless the number of variables is small (1 or 2) this cannot be solved by a pocket calculator. (In the case of 1 Or 2 varaibles it might be possible to solve this equation, but unfortunately no program has been written yet.)

Then it appears that the maximum likelihood method is not the most appropriate for simplified methods of estimation.

B- Least Squares

This time we will use the aggregation procedure. We have:

$$P_{it} = P_r \text{ (choose alternative i)} = \frac{e^{U_{it}}}{\sum_{j \varepsilon A_t} e^{U_{jt}}}$$

$$P_{nt} = Pr \text{ (choose alternative n)} = \frac{e^{U_{nt}}}{\sum_{j \varepsilon At} e^{U_{jt}}}$$

We can then derive

$$\frac{P_{it}}{P_{nt}} = \frac{e^{U_{it}}}{e^{U_{nt}}} = e^{U_{it} - U_{nt}}$$

If we take the natrual log on both sides of the equation, we get:

$$\ln \frac{P_i}{P_n} = U_{it} - U_{nt} = \sum_{k=1}^{K} (X_{itk} - X_{ntk})\theta_k \qquad (2)$$

As we can see here we have a linear regression of $\ln \frac{P_i}{P_n}$ on θ, within the same behavioral unit. We can then run linear regression using the least-squares method.

The major difficulty will consist of computing P_i and P_n within the behavioral unit. But this kind of problem will be overcome by the different types of aggregation, and mostly the naive aggregation (this will be discussed in the next chapter).

As we are looking for a simplified method, this method seems more suitable for our concern, because it can be programmed on a pocket calculator. Therefore we will use it in the rest of the study.

CHAPTER III

Our concern is now to study a given bus line of a network. A lot of these lines are located in a corridor CBD-outside town oriented. As we cannot possibly study all sorts of bus lines we will study one of this type because it allows only one degree of freedom for the destination of the trips for the people living in this corridor, and thus, is simpler to study.

The first point is to determine the level of detail of the study, i.e., we can study the line as a whole, or try to split it into non-overlapping sections with varying sizes. The smallest size of a section would be a bus stop attraction zone. These attraction zones will be delimited by the walking distance.

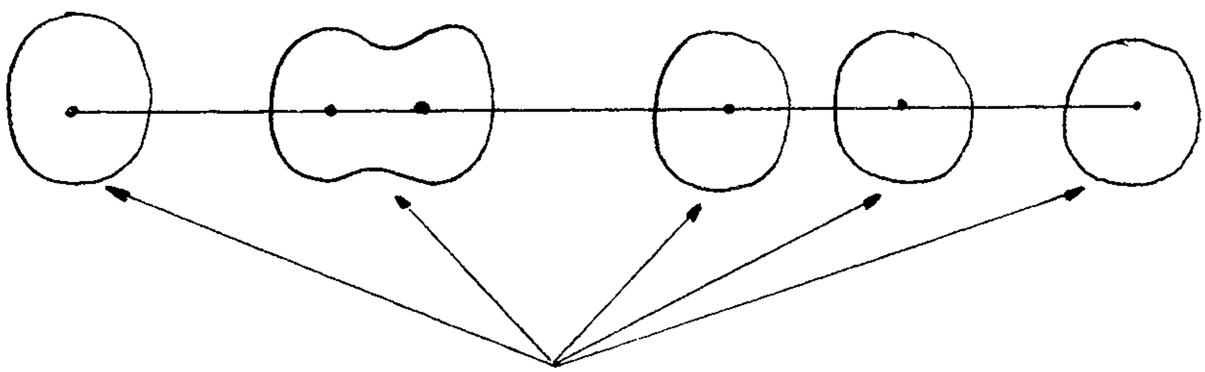
It is commonly admitted that a person will not walk more than 1/4 mile to go to a bus station. Therefore the sections will be geographically drawn according to that criteria. However, this may vary from one city to another. (figure 3).

III-1 Data Collection

The second point is to gather data related to these sections. The relationship is bilateral, i.e., we can adjust the sections to the data we have or collect the data according to the split of the line. These data are of four different types:

--- The socioeconomic characteristic variables, i.e., the income distribution, the auto ownership, the number of people in the household, the number of workers, etc...characterizing the generation at origin.

FIGURE 3.



SECTIONS

--- The attraction variables, i.e., the retail employment, the ratio of vacancies, the total employment etc..., i.e., characterizing the destinations.

--- The distribution of trips per zones, for the three purposes we will consider, i.e., work, shopping and social recreation. This may or may not be represented by an O.D. matrix per mode between the different zones. (This would include a fortiori the generation figures and the modal split.)

--- The level of service of the available modes between the zones i.e., the in-vehicle time, the waiting time, the out of pocket cost, etc... per mode. We will consider here that there are 3 different modes available (as is the case in France): auto, transit and 2-wheels. This latter category includes bicycles, mopeds and motorcycles.

A complete set of data would consist of all of these described here. However, most of the time only a portion of this data is available, and so we are concerned with exploring different degrees of data availability.

Typically in U.S. cities the following data will usually be available:

--- The socioeconomic data are given by the census which covers the city in blocks (equivalent to a city block) and tracts (areas with approximately 4,000 inhabitants). Block data include population, racial composition, number of rented and owned units. The tract data gives information such as family income, age, household composition, and occupation. As we can see the level of accuracy can be good.

--- The attraction variables can often be obtained from land use, planning data or employment sources. Therefore it can be possible to derive the retail employment, the land use etc.

--- The level of service variables are obtained with surveys. To know the waiting time some surveys can be done at the stations or if we assume that the buses are exactly on time, a look at the schedules might be enough. (We assume that the waiting time for auto and 2-wheels is insignificant). To know the walking and parking time some surveys are needed. The out of pocket cost will be the fare for bus trips and operating cost plus parking cost for auto and 2-wheels. The fare can easily be known as usually a fixed fare exists whatever the destination is. For the operating and parking costs some surveys are necessary or some studies previously made on auto and 2-wheels can be used (i.e., studies on comsumption, etc.) For in-vehicle time we just use the data available to the bus operator for busses and use some studies on the speed of the autos or 2-wheels to know if for these modes.

--- The distribution variables are the most difficult to get and, of course, the most important. Nearly all these data can be obtained only through surveys and the level of confidence of the results is proportional to the size of the sample. Yet some data can be gathered in another way. For example the census tracts (again) will give some journey-to-work data. This is an easy way of getting it.

As we can see some of thse data are sometimes hard to get and inferences about the context will be necessary. This will be discussed in Chapter IV. As we saw before we need mainly data aggregated in a naive way, to do so we need to know how to aggregate them.

The basic assumption is that all the variables are uniformly distributed between the zones or within the zones. It is a strong assumption which will be false in most cases because the zones have been delimited geographically. But it will help for the first approximation we need.

- The socioeconomic variables will be obtained taking the averages within the zones.

- The level of service between the zones will be obtained in the same way.

For the out-of-pocket cost the average fare will be taken for the bus (in most cases this fare will be constant) and the average operating cost (obtained by the computation of the average distance between the zones) will be considered. The same will apply for in-vehicle-time. As a matter of fact the average variables are much easier to get and often they are the only data that exist.

III-2 Application of the Logit Model

We have assumed previously that we could use the naive aggregation to make an estimation or a forecast. The first point will be to determine the homogeneous groups. As the aggregation is naive we will assume that the people in the sections will form these groups. Therefore within a certain zone t the logit model can be written as:

$$P_r \text{ (choose alternative n)} = \frac{e^{U_{nt}}}{\sum_{p \in A_t} e^{U_{pt}}} \quad \text{where}$$

A_t = set of possible alternative in zone t

U_{nt} = utility of alternative n given by:

$$U_{nt} = f(T_n, S_t) = \sum_{k=1}^{n} a_k X_{nk}$$

where

T_n = transportation system variable (level of service)

S_t = socioeconomic characteristic of zone t

X_{nk} = generic variable representing T_n or S_t

a_k = coefficient.

Finally,

$\overline{U}_{nt}$ = average utility of alternative i taken over the zone t

$$= \sum_{k=1}^{k} a_k \overline{X}_{nk}$$

Then instead of having a behavioral unit submitted to a set of alternatives A_t we will have a group V_t submitted to the same set and V_{nt} will be the sub-group of V_t having chosen alternative n. As we supposed that the zone t was homogeneous we have:

$$P_r \text{ (choose alternative n)} = \frac{V_{nt}}{V_t}$$

Therefore we can derive:

$$V_{nt} = V_t \times \frac{e^{U_{nt}}}{\sum_{P \epsilon A_t} e^{U_{pt}}} \qquad \text{(B)}$$

At this point we can see that this general formula gives the volumes of zone t having chosen the alternative n as a function of the volume V_t submitted to the set of alternatives A_t and the average utility of each alternative U_{pt}.

But obviously as the set of alternative A_t varies for each different purpose, the volume of people V_t submitted to this set will also vary, and we have to study them differently.

The three basic purposes for people's trips we are considering are: work, shopping and social-recreational.

III-2-1 Work Trips

For work trips the generation and the distribution are supposed to remain constant. The workers are a definite number. Therefore the only thing that might happen if a change in the line occurs will be a shift from one mode to another. The logit model must take into account this fact, and then be only a mode split model for work trips.

As the bus line is in a corridor, outside town-CBD, oriented, the workers have only the choice to go to one of the bus sections of the line. (The CBD section would include all destinations in the CBD.) Therefore:

--- the set A_t will be a set of choices between the auto, 2 wheels, or transit mode (indices A, W, T) to go to work;

--- the volume V_t will be a sub-group of workers going from a random section i to a random section j, $(V_t = V^{ij})$;

--- the sub-group V_{nt} of V_t will be the sub-group of V^{ij} having chosen the mode N to go to work $(V_{nt} = V_N^{ij})$;

--- the utility U_{nt} of the alternative n will be the utility U_N^{ij} of the mode N to go from i to j. This utility should include the level of service of mode N from i to j;

--- the average utility U_{nt} will be the average utility $\overline{U}_N^{ij}$ taken over the zones i and j and over the link (i,j).

The formula (3) can be rewritten as:

$$P_N^{ij} = \frac{V_N^{ij}}{V^{ij}} = \frac{e^{U_N^{ij}}}{\sum\limits_{M=\{A,T,W\}} e^{U_N^{ij}}} \qquad (4)$$

III-2-2 Shopping & Social-recreational Trips

As we said in the previous chapter, one of the qualities of the logit model was that it could represent a simultaneous choice logic especially useful for shopping and social-recreational trips. Consequently, in this case, if a change occurs in the service of the bus line the modal split as well as the generation and the distribution of the trips can change. Still, as we are in a corridor, the change of distribution will remain within that corridor. Therefore:

--- the volume V_t described previously will be the entire population of a random section i (POP_i)

--- The set of alternative A_t will be:

- going to a zone j by one of the three modes,
- not making a trip;

--- the volume V_{nt} will be the volume having chosen to go to a section j by the mode N ($V_{nt} = V_N^{ij}$);

--- the utility U_{nt} will be the utility of the trip to j by mode N. This utility should include the level of service of the mode N and the attraction variables of destination j. ($U_{nt} = U_N^{ij}$; the utility of not making a trip will be U_0);

--- the average utility U_{nt} will be the average utility $\overline{U}_N^{ij}$ taken over the zones i and j and the link (i,j).

The formula (3) becomes

$$P_N^{ij} = \frac{V_N^{ij}}{P_o P_i} \times \frac{e^{\overline{U}_N^{ij}}}{\sum_{k \in A_i} \sum_{M \in \{A,T,W\}} e^{\overline{U}_M^{ij}} + e^{\overline{U}_o}} \qquad (5)$$

where A_i is the set of possible destinations k

III-3 Forecast

Given that the bus operator has all the context data available and that he has also an estimated demand model, his task will consist basically in knowing the future ridership of the bus line when a change in the servcie will occur, in order to derive the gross revenues. On the other hand, he will have to estimate the cost of such changes in the service. The changes considered here will be only changes in the level of service (changes in fare, frequency or speed) and changes in the coverage of the line

We are mainly interested in the first part of the forecast. How is the ridership going to evolve? What will be the new volumes on transit for peak hours and off peak hours? Basically the forecast processing will be the following:

--- Analyse the changes in the service and translate them into changes in the demand model. As we said before the changes will be essentially changes in frequency, fare, speed, or coverage.

--- Study the changes for each of the three pruposes, work (and school), shopping and social-recreational trips, and compute the new ridership.

--- Split this new ridership between peak and off peak hours.

--- Sum up the new riderships for peak and off peak hours per purpose.

III-3-1 Analysis of the changes

Let's consider the bus line split into several sections. At this point of the analysis we are supposed to have the necessary data (i.e., distribution, attraction, level of service, socio-economic variables) for all the sections. If the logit model is well calibrated, the most important characteristics according to the purposes are included in the utility function. Among others the level of service of all modes will always be included. Therefore when changes in the level of service of transit occur (as changes in fare, speed, or frequency) it will be translated into changes in the utility functions. But on the contrary, as the coverage is not explicitly expressed in the utility functions, we will have to argue differently. However, as we saw the difference between work and non work logit models we will treat them separately.

A-Changes in frequency

The frequency is essentially felt by the consumer as the waiting time. Therefore it will be captured by all the variables that include waiting time as: off-vehicle time, total travel time, or even only waiting time, which are the most common.

Usually the waiting time is supposed to be half the headway (though it is only an approximation). In fact the waiting time is greater than 1/2 headway for question of regularity. If Q is the frequency (i.e.. number of vehicle-trips per hour) the headway is given by:

$$h = \frac{1}{Q} \rightarrow \text{the waiting time } W = \frac{1}{2Q}$$

If we change the frequency $Q \rightarrow Q' = Q + \Delta Q$, the waiting time becomes:

$$W' = \frac{1}{2(Q + \Delta Q)} = \frac{1}{2Q} \times \frac{2Q}{2(Q + \Delta Q)} = W(1 - \frac{\Delta Q}{Q'})$$

Therefore the change in waiting time will be

$$\Delta W = - W \frac{\Delta Q}{Q'} \qquad (6)$$

For example, if n is the number of vehicles of the fleet of the line, v the average speed of those vehicles, including the stops and 2L the le length of a round trip, the frequency will be given by:

$$Q = \frac{nv}{2L}$$

Then if we want to add a new vehicle on the fleet the frequency will become:

$$Q' = \frac{(n + 1)v}{2L} \rightarrow \Delta Q = \frac{v}{2L}$$

and we get:

$$\Delta W = -W \frac{\Delta Q}{Q'} = \frac{-W}{n + 1} \qquad (7)$$

But, in fact, many other factors will interfere in a change in frequency. The regularity, for example, should also increase with an increase of frequency. But these factors cannot readily be evaluated and for a first approximation, they are not taken into account.

B-Changes in Speed

The speed of the vehicles is essentially felt as the in-vehicle time by the consumers. But as we saw previously the speed interferes with the waiting time (the headway is an inverse function of the generalized speed). Therefore, changes in the speed of the vehicle (i.e., if some newer vehicles were bought) should be expressed as changes in in-vehicle-time, off-vehicle-time if the time spent at the bus stops remains the same, and consequently in total travel time.

If v'is the new speed (including the stops) of the vehicle, l the distance between 2 random stations, the old in-vehicle time will be

$$\mathrm{IVT} = \frac{l}{V}$$

and the new one

$$\mathrm{IVT}' = \frac{l}{V'} = \frac{v}{v'}\,\mathrm{IVT}$$

Therefore the change will be:

$$\Delta\mathrm{IVT} = \frac{v-v'}{v}\,\mathrm{IVT} \qquad (8)$$

If the time spent at the bus stops remains the same we have the old waiting time $W = \frac{l}{nv'}$

the new waiting time $W' = \frac{l}{nv'}$

→ the change in waiting time will be

$$\Delta W = \frac{v - v'}{v}\,W \qquad (9)$$

Again, some other changes would result from an increasing speed, as regularity, but it is difficult to take them into account unless there is an explicit judgement about the variance of the waiting time.

C - Changes in Fare

Obviously the fares are perceived as out-of-pocket cost and a change in fare will only result in a change in this variable.

speed). Therefore, changes in the speed of the vehicle (i.e., if some newer vehicles were bought) should be expressed as changes in in-vehicle-time, off-vehicle-time if the time spent at the bus stops remains the same, and consequently in total travel time.

If v'is the new speed (including the stops) of the vehicle, l the distance between 2 random stations, the old in-vehicle time will be

$$IVT = \frac{l}{v}$$

and the new one

$$IVT' = \frac{l}{v'} = \frac{v}{v'}\, IVT$$

Therefore the change will be:

$$\Delta IVT = \frac{v-v'}{v}\, IVT \qquad (8)$$

If the time spent at the bus stops remains the same we have the old waiting time $W = \frac{l}{nv'}$

the new waiting time $W' = \frac{l}{nv'}$

$\rightarrow$ the change in waiting time will be

$$\Delta W = \frac{v - v'}{v}\, W \qquad (9)$$

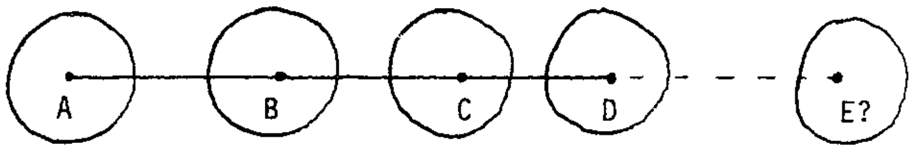
Again, some other changes would result from an increasing speed, as regularity, but it is difficult to take them into account unless there is an explicit judgement about the variance of the waiting time.

C - Changes in Fare

Obviously the fares are perceived as out-of-pocket cost and a change in fare will only result in a change in this variable.

D - Changes in Coverage

This is the most complicated change. It cannot reasonably be translated into a change in the utility function. The basic change will be the following: (figure 4)



We have a line going to the stops A B C D, and we want to add a new station E, which will still remain in the corridor. How is the ridership going to change?

In fact we have to treat this line as a new one. A new attraction zone has been created. As long as the other sections are not overlapping with this new one, they will remain the same. For this new section the 4 types of variables should be computed: the socioeconomic characteristicc, the level of service variables, as well as the distribution for work trips which is very important.

As we are dealing here with minor changes, i.e., the new station remains in a constant environment, we can assume that the demand function will be the same i.e., that the coefficients of the utility functions will keep the same coefficients.

III-3-2 Calculation of the New Ridership

We have at this point to split the study for work trips and social-recreational and shopping trips. And within these purposes forecast the new ridership for the different changes in service.

A-Work Trips

A-1. Changes in the level of service

As we said before if we have a change in the level of service, and if we know the demand function, the simplest way to forecast the new transit ridership is to use the incremental or "pivot point" method.

As the generation between the zones is supposed to remain constant, the logit model is just a modal split and we have:

$$P_T^{ij} = \frac{V_T^{ij}}{V^{ij}} \qquad \text{where}$$

P_T^{ij} = Probability of taking transit between i and j

V_T^{ij} = volume of transit between i and j

V^{ij} = total volume between i and j

(it is the generation between i and j)

The pivot point can then be written as:

$$P'^{ij}_T = \frac{e^{\Delta\bar{U}_T^{ij}} \times P_T^{ij}}{P_T^{ij} \times (e^{\Delta\bar{U}_T^{ij}} - 1) + 1}$$

where the ' index indicates the new variables after the changes

Then

$$V'^{ij}_T = V_T^{ij} \frac{e^{\Delta\bar{U}_T^{ij}} \times V^{ij}}{V_T^{ij} \times (e^{\Delta\bar{U}_T^{ij}} - 1) + V^{ij}} \tag{10}$$

We can see that the new ridership V'^{ij}_T is only a function of the old one V_T^{ij}, the distribution of work trips V^{ij} and the change in utility

$\Delta\bar{U}_T^{ij}$. This latter variable is important because it shows that sometimes, we will need only one coefficient of the utility function to make a forecast. For example, if we make a change in the fare, and if the out-of-pocket cost (OPC) is included in the utility function we have:

$$\Delta\bar{U}_T^{ij} = \alpha_{OPC}\overline{\Delta OPC}_{ij}$$

where OPC_{ij} is the out-of-pocket cost (fare) between i and j. We see that we need here only to know α_{OPC} to make a forecast and that in any case we do not need to know anything about the other modes.

A-2. Change in Coverage

If we add a new station to the line, the change in the ridership will come from the flows between this new station and the already existing ones. As we have assumed that the generation and distribution were constant for work trips, we can suppose that no change will occur in the ridership between the existing stations.

Therefore, if we had n stations (i = 1...n) and we add and $(n+1)^{th}$ station, the new ridership will be expressed as:

$$V_T' = \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq 1}}^{n} V_T^{ik} + \sum_{i=1} V_T^{i(n+1)} + \sum_{k=1}^{n} V^{(n+1)k}$$

We know how to compute V_T^{ik}. To compute $V_T^{i(n+1)}$ or $V_T^{(n+1)k}$ we will use the same model, as we supposed that the utility coefficients remained the same.

$$V_T^{i(n+1)} = V^{i(n+1)} \frac{e^{\bar{U}_T^{i(n+1)}}}{\sum\limits_{M=\{A,T,W\}} e^{U_M^{i(n+1)}}} \qquad \text{(same for } V_T^{(n+1)k}\text{)}$$

Therefore to compute the new ridership to and from (n+1) we need:

- The coefficients of all the utility functions, and all the characterictics of the new zone.
- The distribution of work trips to and from the new zone.

But V_T' can be written as:

$$V_T' = V_T + \sum_{i=1}^{n} V_T^{i(n+1)} + \sum_{k=1}^{n} V_T^{(n+1)k} \qquad (11)$$

and we see here that we do not need the distribution of the work trips between the existing zones, but only the existing total transit ridership. Furthermore if the origin zone characteristics are not included in the utility function, we will need only the characteristics of the new zone. This will happen when the utility functions will be written as:

$$V_{it} = f(T_n; S_t) = f(T_n) \text{ (i.e. only function of the level of service)}$$

<u>B-Shopping and Social-Recreational Trips</u>

We know that the operator has at his disposal a simultaneous logit model (for both purposes) where

$$\Pr(\text{go to } j \text{ by transit from zone } i) = \frac{e^{\overline{U}_T^{ij}}}{\sum_M \sum_{j \epsilon A_i} e^{\overline{U}_M^{ij}} + e^{\overline{U}_0}}$$

where M is the mode and A_i the set of possible stops.

<u>B-1-Change in level of service</u>

If we change the level of service of the bus line, all the utilities of transit $\overline{U}_T^{ij}$ are going to change:

$$\overline{U}_T^{ij} \rightarrow \overline{U}'^{ij}_T = \overline{U}_T^{ij} + \Delta\overline{U}_T^{ij}$$

Then the new probability can be expressed as:

$$P'^{ik}_T = \frac{e^{\bar{U}_T^{ij}}}{\sum\limits_{j \varepsilon A_i} \left(e^{\bar{U}_T^{ij}} + \sum\limits_M e^{\bar{U}_M^{ij}}\right) + e^{\bar{U}_0}} \times \frac{e^{\bar{U}_T^{ik}} e^{\Delta \bar{U}_T^{ik}}}{\sum\limits_{j\, E_i} \left(e^{\bar{U}_T^{ij}} e^{\Delta \bar{U}_T^{ij}} + \sum\limits_M e^{\bar{U}_M^{ij}}\right) + e^{\bar{U}_0}}$$

$$= \frac{e^{\bar{U}_T^{ik}}}{\sum\limits_{j \varepsilon A_i} \sum\limits_M e^{\bar{U}_M^{ij}} + e^{\bar{U}_0}} \times e^{\Delta U_T^{ik}} \times \frac{\sum\limits_J \sum\limits_M e^{\bar{U}_M^{ij}} + e^{\bar{U}_0}}{\sum\limits_{j \varepsilon A_i} \left(e^{\bar{U}_T^{ij}} \times e^{\Delta \bar{U}_T^{ij}} + \sum\limits_M e^{\bar{U}_T^{ij}}\right) + e^{\bar{U}_0}}$$

$$= P_T^{ik} \times \frac{e^{\bar{U}_T^{ik}}}{1 + \sum\limits_{j \varepsilon A_i} P_T^{ik} \left(e^{\Delta \bar{U}_T^{ij}} - 1\right)}$$

But as we saw before: $P_T^{ik} = \dfrac{V_T^{ik}}{POP_i}$

Then we can write:

$$V'^{ik}_T = V_T^{ik} \frac{e^{\Delta \bar{U}_T^{ik}}}{1 + \sum\limits_{j \varepsilon A_i} \dfrac{V_T^{ij}}{POPi} \left(e^{\Delta \bar{U}_T^{ij}} - 1\right)} \qquad (12)$$

This formula is very important because it shows that the new transit ridership is a function of the old ones, of the changes in the transit utility functions and of the population in the zones.

We do not need here to know anything about the other modes. If we have a survey giving the volumes of transit V_T^{ij} (and this is frequent), if we know the population of the zones (and this is relatively easy) and some coefficients in the utility function (as we saw previously) we can very well predict the new ridership.

B-2 Changes in Coverage

We still want to add a new station (n+1) to the line. The problem here will be less easy. The set of possible choices has increased by 3 alternatives, for the already existing sections: go to (n+1) by 3 different modes.

Therefore the new probability to go from a section i to a section k by transit will be written as:

$$P'^{ik}_{T} = \frac{e^{\overline{U}^{ik}_{T}}}{\sum\limits_{\text{Modes}} \sum\limits_{j \epsilon Ai} e^{\overline{U}^{ij}_{M}} + \sum\limits_{\text{Modes}} e^{\overline{U}^{i(n+1)}_{M}} + e^{\overline{U}_{O}}}$$

We can derive

$$P'^{ik}_{T} = \frac{P^{ik}_{T}}{1 + P^{ik}_{T} \left(\sum\limits_{\text{Modes}} e^{\overline{U}^{i(n+1)}_{M} - \overline{U}^{ik}_{T}} \right)}$$

From where we can derive

$$V'^{ik}_{T} = \frac{V^{ik}_{T}}{1 + \frac{V^{ik}_{T}}{POP_{i}} \left[\sum\limits_{M} e^{\overline{U}^{i(n+1)}_{M} - \overline{U}^{ik}_{T}} \right]} \qquad (13)$$

Unfortunately we se here that we will need the characteristics of all the zones as well as all the utility functions.

This situation is even worse when we want to forecast the volume going from (n+1):

$$P^{(n+1)k}_{T} = \frac{e^{\overline{U}^{(n+1)k}_{T}}}{\sum\limits_{M} \sum\limits_{j \epsilon An+1} e^{\overline{U}^{nj}_{M}} + e^{\overline{U}_{O}}}$$

We need to know everything perfectly in order to compute this new probability and then the new ridership.

It appears here that this kind of problem might be a little too complex to get efficient results with limited data. A way to solve this would be to keep the number of zones small.

III-3-3 Time Split

Now that we know how to compute the new riderships we must be able to split them between peak and off peak hour, because it is the point the operator is interested in.

Some studies have been done about the time of day choice models (as CRA Pittsburgh) but they cannot easily be applied. Then the methods presented here are only empirical.

The basic assumption is that the trips split during peak and off-peak hours is in fixed proportion. Therefore if we know the time split by purpose of the old ridership, we can derive the new one. If we know:

$$\left.\begin{array}{ll}\text{During peak hours we have} & V_T^P \\ \text{During off-peak hours we have} & V_T^O\end{array}\right\} \text{ for a given purpose}$$

With $V_T^P + V_T^O = V_T$ (V_T = old ridership)

Then the new split will be given by:

$$V'^P_T = \frac{V_T^P}{V_T} \quad V'_T \quad \text{and } V'^O_T = \frac{V_T^O}{V_T} \times V'_T \qquad (14)$$

If we do not have these riderships V_P^T, V_T^O, we can try to apply some coefficients derived form some studies. It is known for example, that 85% of the work trips occur during the morning peak in Washington.* That 18% of shopping trips and social-recreational trips occur during the same period.

III-4 Methods of Getting a Demand Model

There are many cases where the operator does not have any demand model available, and needs one to make a forecast. At this point different alternatives are offered to him.

1. Use a model developed elsewhere
2. Use some elasticities
3. Update or transfer an existing model
4. Estimate a completely new model.

Of course the choice of any of these alternative will depend essentially on the data the operator will have. The more detailed the data, the more accurate the method will be.

- Case 1 is straightforward but very inaccurate.
- Case 2 will be studied in Chapter IV.
- Case 3 and 4 will be studied here. Of course the real challenge is Case 4 and this is the one we will study first.

III-4-1 Estimation of Demand Models

We will assume all along the arguments that we have the necessary data available. According to Chapter II, we have seen that the least. squares method was the most appropriate to our concern. But we need

*Jesse Jacobson - Studies on Washington, home interviews, surveys, 1976, unpublished an aggregation procedure. We will use the naive one.

Therefore the equation (2) becomes

$$\ln \frac{P_{nt}}{P_{pt}} = \ln \frac{V_{nt}}{V_t} \times \frac{V_t}{V_{pt}} = \overline{U}_{nt} - \overline{U}_{pt}$$

$$\rightarrow \ln \frac{V_{nt}}{V_{pt}} = \sum_{k=1}^{K} a_k[\overline{X}_{nk} - \overline{X}_{pk}] \qquad (15)$$

Again we have to make a distinction between the purposes.

<u>A-Calibration of a work mode split model</u>

As we saw previously the logit model was only a modal split for work trips between three different modes: auto, 2 wheels, and transit. Therefore the equation (15) can be summarized by 2 independent equations:

$$\begin{aligned} \ln \frac{V_A^{ij}}{V_T^{ij}} &= \sum_{k=1} a_k[X_{Ak}^{ij} - X_{Tk}^{ij}] \\ \ln \frac{V_W^{ij}}{V_T^{ij}} &= \sum_{k=1} a_k[\overline{X}_{Wk}^{ij} - \overline{X}_{Tk}^{ij}] \end{aligned} \qquad (16)$$

Where i and j are two random sections taken on the line. As it can be seen the significance of this regression is directly related to the number of zones (i.e. the number of pairs (i,j)). Thus, if we divide the line in into N zones, we will have N x (N-1) zone pairs. If we have the counts V_M^{ij} and the values of $\overline{X}_M^{ij}$, we can estimate the values of the K parameters a_k if k does not exceed 2 x N x (N-1) - 1.

However the greater the value of $\frac{2 \times N \times (N-1)}{K}$, the more significant the results will be.

Among all possible variables the major ones are:

OVT/DST	(out-of-vehicle time/distance)
AAC	(auto ownership)
IVT	(in-vehicle-time)
TWO	(two wheels ownership)
CBD	(dummy CBD variable)

Some of these variables are alternative specific (as auto ownership and 2 wheel ownership) and some are generic. The type of variable to choose depends a lot on the data we have. As we are using the least-square method to run our regression the R^2 test is significant.

Here an example which demonstrates this type of calibration follows. At the same time we will see some simple assumptions to generate input data (this will be discussed in Chapter IV).

The example is a bus line in Toulouse (France), which goes through three basic zones, the CBD, the 1st ring, and the 2nd ring.(figure5).

CBD	1st Ring	2nd Ring
Zone 1	Zone 2	Zone 3

For simplification sake we will consider only two modes. The 2 major variables taken into account are in-vehicle-time and off-vehicle-time/ distance.

The data we have are expressed as modal shares (in percent). Therefore we have the regression:

$$\ln \frac{P_A^{ij}}{P_T^{ij}} = \bar{U}_A^{ij} - \bar{U}_T^{ij} = a_0 + a_1(\overline{IVT}_A^{ij} - \overline{IVT}_T^{ij}) + a_2\frac{(\overline{OVT}_A^{ij} - \overline{OVT}_T^{ij})}{\overline{DST}^{ij}}$$

We can easily compute IVT_T and OVT_T examining the schedules of the bus line.

If we have some data available we can get IVT_A and OVT_A. Otherwise we must make some assumptions. Eg: first we can assume that the speed of a bus is roughly .85 the speed of a car.

$$IVT_A \simeq IVT_T \times .85$$

Second, we can assume that $OVT_A \simeq$ 2mn (this represents mostly the time to go from the parking to the work location). In this case the utility function can be reduced to:

$$\overline{U}_A^{ij} - \overline{U}_T^{ij} = b_o + b_1 \overline{IVT}_T^{ij} + b_2 \overline{OVT}_T^{ij} / \overline{DST}^{ij}$$

Thus we see that only the bus characteristics are taken into account in this reduced model.

The distance will be expressed in minutes (in vehicle time). The CBD will be represented by the index 1, the First ring by 2, and the second ring by 3.

By estimates we have: DST(1,1) = 4 DST(1,2) = 12

DST(1,3) = 16 DST(2,2) = 5.5

DST(2,3) = 4

To compute the OVT_T we assume the relation:

$$OVT_T = \text{headway}/2 + \text{walking time}$$

To estimate the walking time we can assume that the ridership of the bus is uniformly distributed within 1/4 mile of the bus stops. It means that an average person will have to walk 1/8 mile to get the bus and the same distance to go to the office. So in total he will have to walk 1/4 mile. Knowing that an average person walks at the speed of 3 miles per hour the walking time will be equal to:

$$1/4 \div 1/3 = 1/12 \text{ hour} = 5 \text{ mn}$$

In Toulouse during the peak hour the headway is equal to 6mn. So in total $OVT_T = 5 + 6/2 = 8mn$.

Finally, from analysis of available data we have:

$P^1_{A1} = .60 \quad P^1_{T1} = .22 \qquad\qquad P^2_{A1} = .55 \quad P^2_{A1} = .22$

$P^2_{A2} = .64 \quad P^2_{T2} = .14 \qquad\qquad P^3_{A1} = .65 \quad P^3_{T1} = .21$

$P^3_{A3} = .69 \quad P^3_{T2} = .12$

These then lead to the following relationships to be estimated:

$\ln(\frac{60}{22}) = b_0 + b_1 4 = b_2 6/4 \qquad \ln(\frac{55}{22}) = b_0 + b_1 x12 + b_2 x6/12$

$\ln(\frac{64}{14}) = b_0 + b_1 x5.5 + b_2 x6/5.5 \quad \ln(\frac{65}{21}) = b_0 + b_1 x16 + b_2 x6/16$

$\ln(\frac{69}{12}) = b_0 + b_1 x4 + b_2 x6/4$

Using a standard regression routine available for a programmed pocket calculator, we get the values for the coefficients:

$b_0 = 1.28$

$b_1 = .02$

$b_2 = .15$

These results are not satisfactory with respect to signs; we see that the utility of transit is going to increase with the increase of in-vehicle-time. This can be due to the fact that the in-vechicle-time interferes also in the second variable. Anyway, if we check the correlation we get a R^2 equal to .27 which is very bad for the small sample we had. So instead of using IVT_T as a characteristic variable we will use a dummy variable "CBD" which will be equal to 1 if one end of the trip is the CBD and 0 otherwise.

We have the new regression:

$$\ln \frac{P_A^{ij}}{P_T^{ij}} = b_0 + b_1\ OVT_T^{ij}/DST^{ij} + b_2\ CBD$$

$\ln \frac{60}{22} = b_0 + b_1 \times 6/4 + b_2$ $\qquad$ $\ln \frac{55}{22} = b_0 + b_1 \times 6/12 + b_2$

$\ln \frac{64}{14} = b_0 + b_1 \times 6/5.5$ $\qquad$ $\ln \frac{55}{21} = b_0 + b_1 \times 6/16 + b_2$

$\ln \frac{69}{12} = b_0 + b_1 \times 6/4$

We get the values for the coefficients:

$b_0 = 1.60$

$b_1 = .02$

$b_2 = -.61$

This time these coefficients have the expected signs; if the OVT_T/DST increases the utility of transit will decrease and if the destination of the trip is the CBD, it will increase.

Furthermore the correlation test is acceptable because we get a R^2 equal to .90 (and an adjusted $\bar{R}^2$ equal to .80). This shows that with few data and a pocket calculator some results can be found, but the significance of the results are a function of the data.

B-Calibration of Multiple Choice Shopping Model

We saw previously that the logit represented a simultaneous logic for shopping model. Thus, using naive aggregation we have:

$$P_N^{ij} = P_0 \text{ (going to j by mode N)} = \frac{e^{\bar{U}_N^{ij}}}{\sum\limits_{k \epsilon A_i} \sum\limits_{M=\text{modes}} e^{\bar{U}_M^{ik}} + e^{\bar{U}_o}} = \frac{V_n^{ij}}{POP_i}$$

As we have here a model where the generation, the distribution and the modal split are represented at the same time, the regression (15) must represent these properties. Therefore, we can derive from (15)

$$\ln \frac{P_N^{ij}}{P_N^{ij}} = \ln \frac{V_N^{ij}}{V_N^{ij}} = \sum_{p=1}^{P} a_p [\overline{X}_{Np}^{ij} - \overline{X}_{Mp}^{ik}] \tag{17}$$

This equation reveals two steps:

--- If the two alternatives compared are "making atrip by a certain mode" we are estimating the parameters of destination and mode.

--- On the other hand if one of the alternatives is the no trip alternative (alternative 0) we will derive the coefficient for the generation also (i.e., no trip – one trip alternative)

In this case the equation would become:

$$\ln \frac{V_N^{ij}}{V_0} = \sum_{p=1}^{P} a_p (X_{Np}^{ij} - X_{0p})$$

where $V_0 = POP_i - \sum_{j \varepsilon Ai} \sum_{M} V_M^{ij}$

But this last equation (18) is very important, because it makes a lot of simplifications as the X_{op}'s will be zero for most of the P variables.

Among the possible variables, the most characteristic appeared to be

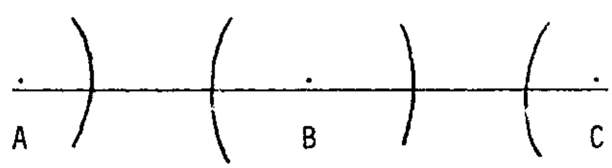
Total travel time/Distance	(TTT/DST)
Out-of-pocket cost/income	(OPC/INC)
Retail employment at destination	($REMP_j$)
Auto ownership	(AAC)
Two wheels ownership	(TWO)

Retail employment at origin $(REMP_i)$

CBD constant (CBD)

Of course, the number of possible variables depends essentially on the data available. At most if we have all the wanted data, the number of variables could not exceed $3N(N-1) - 1$ where N represents the number of zones. The number $3N(N-1)$ represents the number of aggregate equations ($N(N-1)$ for each mode) and -1 represents the constant in the regression.

Example: Let's take a line with 3 zones, A and B and C and 2 modes: auto and transit. The data available is (hypothetical data):



Population: A = 10,000

B = 1,000

C = 2,000

Retail employment: A = 800

B = 50

C = 100

The in-vehicle times:

$TT_T^{BA} = 20mn \qquad TT_A^{BA} = 17mn$

$TT_T^{BC} = 20mn \qquad TT_A^{BC} = 8mn$

$TT_T^{CA} = 25mn \qquad TT_A^{CA} = 21mn$

Volumes by all modes:

$V^{AB} = V^{AC} = 0$ $\quad V_T^{BC} = 150 \quad V_A^{BC} = 200$

$V_T^{BA} = 50 \quad V_A^{BA} = 100$ $\quad V_A^{CB} = V_T^{CB} = 0$

$V_T^{CA} = 100 \quad V_A^{CA} = 200$

Then if we include in the utility function 2 variables - retail employment and travel time, we will have

$$\ln \frac{V_M^{ij}}{V_o^i} = a_0 + a_1 \, TT_M^{ij} + a_2(REMP_j - REMP_i)$$

Then we just replace the varaibles by their value (we have six equations) knowing that:

$$V_o^i = POP_i - \sum_j V^{ij}$$

and we get

$a_o = .16$

$a_1 = -.139$

$a_2 = .001$

and $R^2 = .975$

But here we had all the wanted data. Usually it is not so, and inferences must be done about the context.

<u>C-Social-recreational trips</u>

The pattern of the demand model is the same as in shopping trip. Only the variables are going to change. Among the set of possible variables the most characteristic will be:

total travel time/distance
out-of-pocket cost/income
household size

employment density
race of household
ratio of vacant to total area, etc.

Again the number of variables should at any time not exceed $3N(N-1) - 1$.

III-4-2 Resolution of the Regression Problem

We are going to introduce here two separate issues.

A-Resolution of Simultaneous Regressions

As we saw previously we had to run simultaneous regressions in order to get some results. The general technique is rather complicated and should not be applied here. The only difference between these simultaneous equations are some alternative specific variables. The way this difficulty will then be overcome will be to introduce some dummy variables (unconstrained data pooling) and run the regression as a single one.

The best way to explain it is to show an example. Let's take the case of work trip where we would have two alternative specific variables: AAC (auto ownership) and TWO (two-wheels ownership). We would have to run the regressions:

$$\ln \frac{V_A}{V_T} = a_o + a_2 AAC$$

$$\ln \frac{V_W}{V_T} = a_1 + a_3 TWO$$

We introduce here some dummy variables (it is valid), X_1, X_2, X_3 and we get the single regression:

$$\ln \frac{V_X}{V_T} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3$$

where $\begin{cases} X_1 = 1 & \text{when } X = W \\ = 0 & \text{otherwise} \end{cases}$ $\begin{cases} X_2 = AAC & \text{when } X = A \\ = 0 & \text{otherwise} \end{cases}$

$\begin{cases} X_3 = TWO & \text{when } X = W \\ = 0 & \text{otherwise} \end{cases}$

and this can be solved more easily.

B-Regression with more than two variables*

To solve regressions with the help of a pocket calculator, we have to find a way to overcome its limitations especially limited memories. There is a program running directly a double regression, but not more.

If X is a nxk matrix of observations on k independent variables, and if y is the column vector of observations on these variables, to run a regression is to find a vector b as $\hat{y} = Xb$ where $\hat{y}$ is as close a possible to y. Using the least-squares criteria we want to minimize the function fo z:

$(y-Xz)'\ (y-Xz)$ this leads to the equations:

$X'Xb = X'y$ or $b = (X'X)^{-1}X'y$

As it can be seen the important point is the inversion of the matrix $(X'X)$. If we have n variables we will have a $(n+1,\ n+1)$ matrix to invert and beyond a certain value of n, this inversion will be impossible because of the limited memories, even if the calculator routines do not invert matrices directly.

An alternative is as following. Let's suppose we are considering a least squares fit of y y on k + 1 independent variable as (X x). That is we seek a vector b and a scalar c such that:

$\hat{y} = Xb + xc$ $\qquad$ $S = (n,k)$ matrix
$\qquad\qquad\qquad\qquad$ $x = n$ - vector

*T. Rothenberg; Linear Model (1976 course notes)

This time we want to minimize the function of z and m:

$$(y - Xz - xm)'(y - Xz - xm)$$

and taking the derivatives with respect to z and m we get:

$$X'Xb + X'Xc = X'y$$

$$x'Xb + x'xc = x'y \qquad \text{as } (X'X)^{-1} \text{ exists we get}$$

$$b = (X'X)^{-1}X'y - (X'X)^{-1}Xxc$$

and substituting in the second equation we get

$$c = \frac{x'(I - X(X'X)^{-1}X')y}{x'(I - X(X'X)^{-1}X')x} = \frac{x'My}{x'Mx}$$

but

$$x'M = (Mx)' \quad \text{as M is symetric}$$

$$= x - S(X'X)^{-1}X'x = (x-\hat{x})' = \overset{o}{x}'$$

where $\overset{o}{x}$ is the residual vector obtained from a least squares regression of x on X.

This is a very important result coming form some properties of the least squares which say that the residual vector $y - \hat{y}$ of a regression of y on X is perpendicular to each independent variable vector included in the regression.

Therefore this is the solution to use. If the number of variables exceeds the capacity of the calculator, we first choose a variable and run a regression on the other variables. As an example let's imagine we want to run a regression on:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

but we can only run a regression with 2 variables at a time. Then we have the following procedure:

Step 1:

We run a regression

$$X_3 = a_o + a_1X_1 + a_2X_2$$

and we get: $\hat{a}_0, \hat{a}_1, \hat{a}_2$

Then we have

$$X_3 = \hat{a}_0 + \hat{a}_1X_1 + \hat{a}_2X_x + \hat{u} = \hat{X}_3 + \hat{u}$$

We compute this residual vector $\hat{u}$.

Step 2:

Our initial regression can now be written as:

$$y = b_o + b_1X_1 + b_2X_2 + b_3\hat{X}_3 + b_3\hat{u}$$

As $\hat{u}$ is perpendicular to the other variables we can run a simple regression of y on $\hat{u}$:

$$y = c_0 + c_1\hat{u}$$

and according to the previous properties we will have $\hat{b}_3 = \hat{c}_1$.

Step 3

Now we know $\hat{b}_3$ we can run a new regression derived from the initial one:

$$y - \hat{b}_3X_3 = b_0 + b_1X_1 + b_2X_2$$

As it is a double regression we can do it this time.

As we can see this method can be very useful if we want to compute all the coefficients of the regression. But more often we will be interested in only one coefficient of the regression to make some forecast (as we saw in the Pivot point method). Therefore this method is extremely useful because it will allow to compute this coefficient very easily.

III-4-3 Updating Models

Sometimes it is hard to calibrate a model because the data are missing. However the bus operator can try to use an already existing model used for another place or another time and try to transfer or update it.

There are many ways to update a model and the more sophisticated the method the better the results are.* But once again we are limited with our pocket calculator and we cannot use some complicated methods as the Bayesian updating which is often based on maximum likelihood results. To update a logit model means to update the coefficients of the utility function. This utility function is written as:

$$U_i = a_0 + \sum_{k=1}^{K} a_k X_{ik}$$

Three alternatives are offered to us:

Case I - use the utility function as it already exists from a model estimated elsewhere

Case II - try to update the constant a_0 only $\rightarrow a_0'$

Case III - try to update the constant a_0 and the coefficients $a_k \rightarrow a_0'$ and α

$$U'i = a_0' + \alpha\left(\sum_{k=1}^{k} a_k X_{ik}\right)$$

In this case we will have to find a_0' and α considering that $\sum_{k=1}^{K} a_k X_{ik}$ is a single one variable.

The two latter cases can be shown by an example. Let's suppose we have an existing model giving:

---

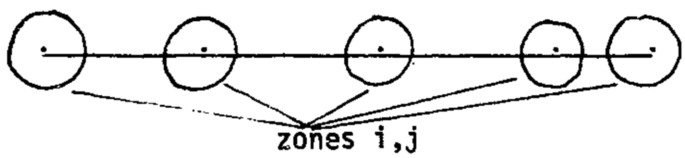
*Terry J. Atherton and Moshe Ben Akiva; Transferability and Updating of Disaggregate Travel Demand Models - M.I.T., 1975.

$$U_A - U_T = a_0 + \sum_{k=1}^{K} a_k (X_{Ak} - X_{Tk})$$

Using the naive aggregation procedure, we get:

$$\ln \frac{V_A}{V_T} = \bar{U}_A - \bar{U}_T = a'_0 + \sum_{k=1}^{k} a_k \, (\bar{X}_{Ak} - \bar{X}_{Tk})$$

The variables V_A, V_T are a generic way of representing all the flows between the zones i and j of the bus lines



Case II

If we want just to update the constant we will use:

$$\frac{1}{n} \left[\ln \frac{V_A}{V_T} - \sum_{k=1}^{K} (\bar{X}_{Ak} - \bar{X}_{Tk}) \cdot a_k \right] = \hat{\bar{a}}'_0$$

taking the mean on all the O.D. pairs (i,j).

Case III

If we want to update the constant and the variable coefficients we just run a regression of:

$$\sum_{k=1}^{K} a_k \, (X_{Ak} - X_{Tk}) \text{ on } \ln \frac{V_A}{V_T}$$

and we will get:

a_0' and λ

$$\ln\frac{V_A}{V_T} = a_0 + \lambda[\sum_{k=1}^{N} a_k (X_{Ak}-X_{Tk})]$$

Sometimes it appears that to update a model is even more efficient than to calibrate a model when the data are rather limited. As a matter of fact it has appeared throughout all the studies that some constant were existing in all the models. For example it is a reasonable assumption that $a_{OVT} = 2.5a_{IVT}$ and $a_{OPC} = .167a_{IVT}$* where OVT, IVT are expressed in minutes and OPC is in ¢. These results are derived from many models, and they are surprisingly exact for first approximation.

We could then try to use them for the work trips example we had. We will run a regression on:

$$\ln\frac{V_A}{V_T} = a_o + a_1 \text{ (IVT + 2.5OVT)/DST}$$

and we get:

$a_0 = .62$

$a_1 = 2.05$

$R^2 = .27$

which is not very good but sitll more significant than a two variables regression. (first regression we made).

---

*Marvin Manheim, Fundamentals of Transportation Systems Analysis.

CHAPTER IV

4-1. INTRODUCTION

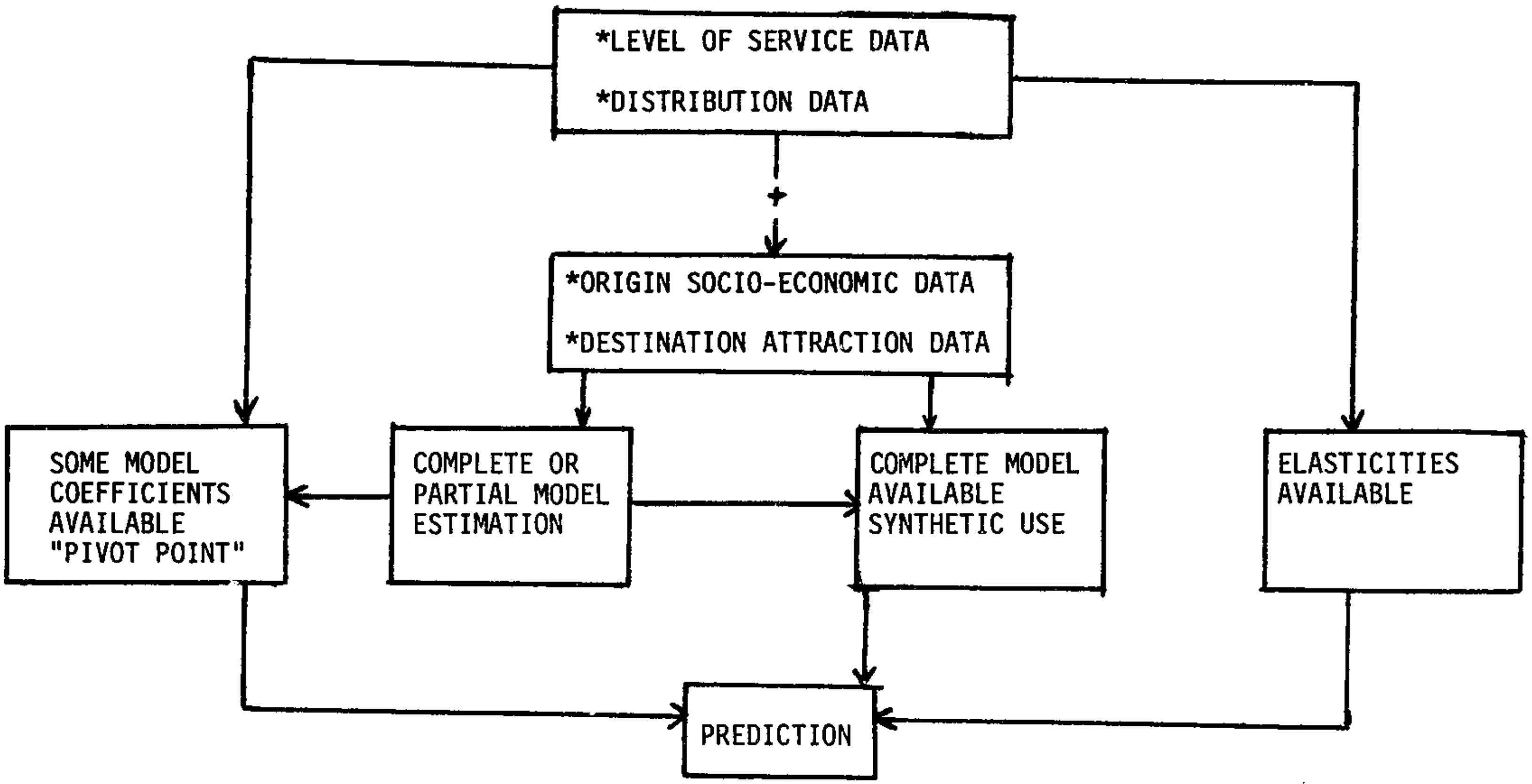
As we saw in the previous chapters the possibility to make a prediction is limited by the existence of some relevant data. Unfortunately in most cases these data do not exist or exist but not in the expected form, and cannot be used as they are. Still some results are needed--consequently inferences about the context must be done. The basic examples of such inferences would be: How is it possible to get reliable data when poor and few surveys are made? How is it possible to make a forecast when the model is missing? It is a challenge to try to find some results but it is worth doing it. Of course the reliability of these results is directly related to the information gathered, but the goal of the study is to determine this relationship.

We saw previously that the data were of 4 types:

- The socio-economic data--important for the origin zones
- The attraction data--i.e. the socio-economic data of the destination zones
- The level of service data
- The distribution data, i.e. riderships, generation distribution, etc.

According to the type of data, and a model availability, the operator has different ways of making a forecast. This can be summarized by Figure 6.

FIGURE 6


*LEVEL OF SERVICE DATA
*DISTRIBUTION DATA
+
*ORIGIN SOCIO-ECONOMIC DATA
*DESTINATION ATTRACTION DATA
SOME MODEL COEFFICIENTS AVAILABLE "PIVOT POINT"
COMPLETE OR PARTIAL MODEL ESTIMATION
COMPLETE MODEL AVAILABLE SYNTHETIC USE
ELASTICITIES AVAILABLE
PREDICTION

As we can see there are 4 major ways to make a forecast according to the type of model available. These four cases can be described as:

a) Some model coefficients are available:

If we are concerned with changes in the level of service of the bus line, we can use the pivot point method to make a forecast. The coefficients we will need to know at that moment will be the coefficients related to the change in the level of service (for example, coefficients of out-of-pocket-cost if the change is a change in fare). We saw also that we needed the distribtuion data and the existing level of service to use that pivot point method. If we are now concerned with changes in coverage, this method cannot be applied.

b) Some elasticities are available:

With some elasticities (with respect to fare or frequency, for example) we are in a position of doing a forecast. This method will be discussed in this chapter. To make this forecast we will need the existing ridership (i.e. distribution data) and the change in the level of service.

c) A complete model is available:

We can of course use the "pivot point" method for changes in the level of service. If a change in coverage has to be studied, we can predict the new ridership if in addition to the distribution data and level of service data we know the origin socio-economic and the destination attraction data.

d) No model is available:

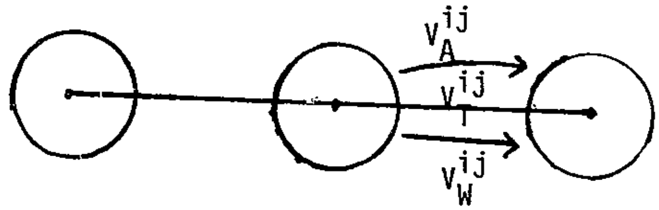
We saw in the previous chapter how it was possible to estimate a model. In order to do so, we need the four types of data in the corridor. After having calibrated the model we can use either method (a) or (c).

The conclusion to these different cases is the absolute need of the distribution and level of service data for any case. Therefore, the emphasis should be put on inferences about these kinds of data. However, as the level of service of transit is usually easily available for the bus operator, it appears that the inferences about distribution data are by far the most important. As a first priority we will study them. Inferences about the two other types must not be neglected, but as they are both available from the census tracts, in most cases we can assume that they will be known.

4-2. INFERENCES ABOUT DISTRIBUTION DATA

When we have a bus line with several zones we are mainly interested in the volumes going between two zones i and j, for different purposes, and different modes.

Figure 7



In other words we are interested in V_M^{ij} for work trip, shopping trip and social-recreational trip. In each of these volumes we have three different stages of the transportation demand process for a given purpose:

- The generation:

  From zone i are generated V^{ij} trips given by:

  $$V^{ij} = V_A^{ij} + V_T^{ij} + V_W^{ij}$$

- The distribution:

  As we have the volumes by O-D pairs the distribution is included.

- The modal split:

  If we have the total volumes and the volumes by mode the modal split is immediately given by:

$$P_A^{ij} = \frac{V_A^{ij}}{V^{ij}} \;;\; P_T^{ij} = \frac{V_T^{ij}}{V^{ij}} \;;\; P_W^{ij} = \frac{V_W^{ij}}{V^{ij}}$$

Then if we have all these different characteristics for each of these volumes, we can try to work on the reverse way--i.e., knowing the characteristics of transportation demand try to get the volumes V_M^{ij}.

The different possibilities to get these volumes are numerous and depend only on the data available. But as we are supposed to work for a bus operator, this method will be heavily based on the transit volumes because these are the most easily available for him.

A. Transit Data

Our first task will be to explain the transit data. As we are playing the role of a bus operator we are mainly interested in:

- The distribution of the trips by purpose.
- The time split (i.e. the distribution between Peak and off-Peak).

The way to obtain this data is:

- Home interview surveys
- Surveys aboard bus lines; i.e., ask the people waht their origin and destination are and what is the purpose of their trip.
- Peak load counts--i.e., experienced people will estimate the ridership during peak and off-peak hours on the buses.
- boarding and alighting counts: we do not know here what will be the destination of the people boarding or the origin of the people egressing.

It is from this data that inferences can be made. Three major problems, however, will be studied:

- How to get an O-D matrix for all purposes
- How to split the trips between the purposes
- How to split the trips by purposes between peak and off-peak hours.

For each of these problems there are different solutions according to the existing data. We cannot study all these solutions, but we will study some cases of inferences.

A.1 Obtaining an O-D matrix

CASE 1

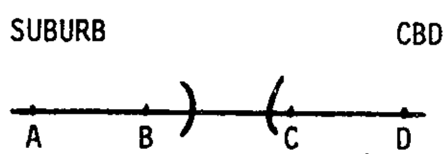
We will assume here that the collected data are some boarding and alighting counts. We must then make some assumptions on the line itself. We decided in the previous chapters to choose an outside town-CBD-oriented line among the whole network because they were the most characteristic. They are also the easiest to study.

1st method

As a matter of fact if a line is suburb-CBD-oriented, most of the people in the suburb will board in the different stations at the origin of the line, and alight at the last stations at the end of the line. Very few people will alight in the suburb or board in the CBD. This can be verified by some surveys on such bus lines (see the following example).

The second assumption we will make is that the desination if uniformally distributed over the origin. Clearly it means that if a certain percent of the total ridership alights at a certain station j, the same percent of people boarding at a prior station i will alight at that station j.

To explain this we can use an example. We have a bus line with two stops in the suburb, and 2 stops in the CBD:



Our assumption says that the people will board at A and B and alight at C and D. No one (or only an insignificant percent) will go from A to B or from C to D. It is a strong assumption but it is verified by some surveys.

If we have some boarding counts b_A and b_B at A and B, and some alighting counts a_C, a_D at C and D our assumption can be written as:

V^{ij} being the volumes between i and j:

$$V^{AB} = V^{CD} = 0$$

$$V^{AC} = b_A \times \frac{a_C}{a_C + a_D}; \; V^{AD} = b_A \times \frac{a_D}{a_C + a_D}$$

$$V^{BC} = b_B \times \frac{a_C}{a_C + a_D} \; ; \; V^{BC} = b_B \times \frac{a_D}{a_C + a_D} \qquad (19)$$

As we have assumed that nobody alighted at B, we have:

$$a_C + a_D = b_A + b_B = V^{AC} + V^{AD} + V^{BC} + V^{BD}$$

<u>2nd method</u>

Unfortunately, all the lines are not of that kind; however, we must find a way to get an O-D matrix.

Other assumptions must then be made. Let's figure we have a line with n stations i = 1, . . . , n

1	2	3	i	i+1	n-1	n
(a_1,b_1)	(a_2,b_2)	(a_3,b_3)	(a_i,b_i)		(a_{n-1},b_{n-1})	(a_n,b_n)

a) our first assumption will be that the people boarding at station i will not alight at the station (i + 1). It is a reasonable assumption that can be explained in words of utility: within a certain distance, the people will prefer the walking mode to the transit. It depends of course on the distance between the zones, but here the distance is usually less than 1/4 mile.

b) The second assumption will be the same as in the previous case with a slight modification. We will assume that the volume alighting at a station will be proportionally distributed over the prior station according to the volumes boarding at these stations diminished by the volumes already alighted. This seems rather complicated but in fact is not--
let b_1, b_2, b_3 . . . b_n be the boarding counts on the stations and a_1, a_2m a_3 . . . a_n be the alighting counts. Of course $a_1 = 0$ and $b_n = 0$, but it will be simpler to use them for the rest of the argument. According to the first hypothesis, we should have also $a_2 = 0$ (if it is not so, we will know that $v^{12} = a_2$ because the people can come only from station 1). Still, according to the first hypothesis, we will have, for $n \geq 2$,

$$a_n = V^{1n} + V^{2n} + - - - - + V^{(n-2)n}$$

(Nobody will go from station (n-1) to station n).

According to the second hypothesis we will have:

$$\frac{V^{1n}}{b_1 - \sum\limits_{i=1}^{n-1} V^{1i}} = \frac{V^{2n}}{b_2 - \sum\limits_{i=2}^{n-1} V^{2i}} = \frac{V^{jn}}{b_j - \sum\limits_{i=j}^{n-1} V^{ji}} = --- = \frac{V^{n-2\ n}}{b_{n-2}} = k$$

This represents the volumes already alighted.

Using a simple rule of arithemetic we get:

$$k = \frac{V^{1n} + V^{2n} + --- + V^{n-2n}}{\sum_{j=1}^{n-2} (b_j - \sum_{i=j}^{n-1} V^{ji})}$$

but $\sum_{j=1}^{n-2} (bj - \sum_{i=j}^{n-1} V^{ji}) = \sum_{j=1}^{n-2} bj - \sum_{j=1}^{n-2} (\sum_{i=j}^{n-1} V^{ji})$

$$= \sum_{j=1}^{n-2} bj - \sum_{i=1}^{n-1} (\sum_{j=1}^{n-2} V^{ji}) = \sum_{j-1}^{n-2} b_j - \sum_{i=1}^{n-1} a_i$$

Therefore, we get the formula:

$$V^{pn} = \frac{b_p - \sum_{i=1}^{n-1} V^{pi}}{\sum_{j=1}^{n-2} b_j - \sum_{i=1}^{n-1} a_i} \times a_n \qquad (20)$$

This is of course an empirical formula, which cannot be proven in any case. But it gives sometimes results which are quite sufficient for first approximations. The two following examples given are taken from surveys in the city of Toulouse (France) made by the IRT.* The 2 lines are CBD--suburb oriented--but one is typically the illustration of the first case, while the other one has a more general type and illustrates the second case.

FIRST TABLE

As can be seen this type of line corresponds exactly to the description made previously. As it is CBD-suburb oriented, nobody alights in the CBD and nobody boards in the suburb. The pair of figures in the matrix represent: - for the top ones the synthetic volumes, and the bottom ones the real volumes.

As can be seen the results are extremely satisfactory. If we take into account only the relatively important volumes ($V^{ij} \geq 20$) we do not have any relative error greater than 18% which is a very low rate. And the greatest error for the first station where most people have boarded does not exceed 13%, which is excellent.

SECOND TABLE

This line is suburb-CBD oriented, but this time the suburb is a close one. That is why the line does not correspond to the first criteria. The results here are less accurate. However, the accuracy increases with the volumes. For volumes greater than 35 we get a maximum error of 23%. Furthermore, the average error decreases for the alighting at the CBD station.

These two tables have shown that some efficient inferences could be made when some data were missing. But we can realize also that there is some limit to the accuracy of these results, and we can only use these synthetic volumes for a first approximation.

---

*AVAT - Surveys on line 64 and Q (1976).

TABLE 1

BUS DIRECTION: SUBURB → CBD

	ESQUIROL	PONT NEUF	COURS-DILLON	OLIVIER	ST. CYPRIEN	ROGUET	PATTE D'OIE	MASSONIE	MURILLO	ST. MARTIN (3)	SNIAS	FONTAINE	SIDOBRE	COLOMIERS C	VERCORS	COLOMIER S	COLOMIERS N	COMINGES	BOARDING COUNTS
ESQUIROL		-	-	-	-	-	-	-	-	-	(13) 13	(124) 110	(88) 88	(115) 117	(57) 66	(101) 106	(96) 94	(424) 427	1021
PONT-NEUF		-	-	-	-	-	-	-	-	-	-	(5) 7	(4) 5	(5) 7	(2) 2	(4) 1	(4) 1	(18) 21	46
COURS-DILLON			-	-	-	-	-	-	-	-	-	(2) 2	(1) -	(1) 2	(1) -	(1) 2	(1) 2	(5) 5	13
OLIVIER					-	-	-	-	-	-	-	(2) 5	(2) 4	(2) 2	(1) -	(2) 3	(2) 2	(8) 3	19
ST. CYPRIEN						-	-	-	-	-	(1) 1	(5) 5	(4) 6	(5) 3	(2) 3	(4) 5	(4) 3	(18) 18	44
ROGUET							-	-	-	-	(–) 4	(2) 4	(2) –	(2) 1	(1) –	(2) 1	(2) 3	(8) 6	19
PATTE D'OIE								-	-	-	-	(8) 16	(6) 3	(7) 4	(4) 1	(6) 3	(6) 5	(27) 33	65
MASSONIE									-	-	-	(2) 7	(1) –	(2) 4	(1) 1	(2) 1	(2) –	(7) 3	16
MURILLO									-	-	-	(2) 2	(1) -	(1) 3	(1) –	(1) –	(1) 3	(5) 5	13
PURPAN (3)										-	-	(9) 5	(7) 11	(8) 6	(4) 2	(8) 9	(7) 11	(32) 32	76
SAINT-MARTIN											-	-	-	(2) 2	(1) 1	(2) 2	(2) 2	(7) 7	16
SNIAS												1	-	1	-	-	-	-	2
FONTAINE													-	-	-	-	-	-	0
SIDOBRE														-	-	-	-	1	0
COLOMIERS C															-	-	-	-	1
COLOMIERS S																	-	-	0
COLOMIERS N																		-	0
COMINGES																			
ALIGHTING COUNTS		0	0	0	0	0	0	0	0	0	18	166	117	152	76	133	126	561	

TABLE 2

	J. D'ARC	WILSON	CAFARELLI	BON REPOS	COLOMBETTE	PROVIDENCE	SOUPETARD	Ch. PELLEPORT	CHAUBET	LOUIS PLANA	PT. de BALMA	ARENES	BALMA C	MONS	CROISEMENT	Rte. de LAB.	LASBORDES	
CAFARELLI	(1) 1																	1
BON REPOS	(5) 6	(3) 2																8
COLOMBETTE	(11) 10	(5) 6	(1) 3															17
PROVIDENCE	(15) 16	(7) 8	(2) -	(4) 6														28
SOUPETARD	(37) 44	(17) 15	(4) 3	(9) 7	(5) 3													72
CH. PELLEPORT	(37) 43	(18) 18	(4) 4	(9) 4	(6) 4	(1) 1												76
CHAUBET	(32) 30	(15) 19	(3) 5	(8) 9	(5) 3	(1) -	(2) 1											67
LOUIS PLANA	(38) 36	(18) 24	(4) 2	(9) 10	(6) 4	(1) -	(2) 3	(1) 1										80
PT. de BALMA	(32) 20	(15) 9	(3) 16	(8) 18	(5) 2	(1) 2	(2) 1	(1) -	(1) 1									67
ARENES	(19) 14	(9) 13	(2) 2	(5) 2	(3) 5	(1) -	(1) 3	- -	- 1	1 1								41
BALMA C	(134) 139	(65) 63	(14) 8	(34) 34	(20) 31	(5) 8	(9) 9	(4) 1	(2) -	(6) 3	(2) 1							299
MONS	(27) 26	(13) 14	(2) 2	(7) 10	(4) 5	(1) -	(2) -	(1) -	(1) -	(1) -	- 1	- -						58
CROISEMENT	(33) 43	(15) 21	(3) 3	(2) 4	(5) 2	(1) 1	(2) -	(1) -	(1) 1	(1) 1	- -	(1) -	(5) 0					76
Rte. de LAB.	(17) 8	(8) 3	(2) -	(4) 2	(2) 4	(1) 2	(1) 6	(-) 4	(-) 1	(1) 6	(-) -	(-) -	(3) 4	(1) -				40
LASBORDES	(37) 43	(18) 13	(4) 1	(9) 10	(5) 3	(1) -	(2) -	(1) 3	(1) 2	(1) -	(1) 1	(1) -	(6) 8	(1) 2	(6) 6	(5) 5	y	99
	479	226	67	116	66	16	23	9	6	11	3	2	16	2	6	5		

CASE 2

This time we will assume that we have already an O-D matrix but it is outdated and we want to update it with some boarding and alighting counts.*

We suppose that we have an O-D matrix (V^{ij}) where

$$\sum_{j=1}^{N} V^{ij} = B^{i} \text{ (total boarding at station i)}$$

$$\sum_{i=1}^{j-1} V^{ij} = A^{j} \text{ (total alightin at station j)}$$

We want to update the matrix using some boarding and alighting counts $B^{\prime i}$ and $A^{\prime j}$. The first point is to transform the O-D matrix into a contingency table, with marginals.

a marginal of V^{ij} will be given by: $v^{ij} = \frac{V^{ij}}{V}$

where V is the total volume.

$$V = \sum_j \sum_i V^{ij} = \sum_j A^{j} = \sum_i B^{i}$$

In the same way, the marginal of a boarding and alighting count will be:

$$a^{j} = \frac{A^{j}}{V} ; \quad b^{i} = \frac{B^{i}}{V}$$

$$a^{\prime j} = \frac{A^{\prime j}}{V^{\prime}} ; \quad b^{\prime}_{i} = \frac{B^{\prime}_{i}}{V^{\prime}}$$

Then we use an algorithm which converges most of the time (it is not guaranteed to converge with tricky initial values).

---

*John Nordin, "Iterative Proportional Fit and Extrapolation," (MIT, 1977, unpublished.)

We have

$$v^{ij}_{(n+1)} = v^{ij}_{(n)} \times \frac{a'^{i}}{a^{i}_{(n)}} \quad ; \quad v^{ij}_{(n+2)} = v^{ij}_{(n+1)} \times b'^{j}_{(n+1)} \times \frac{b'^{j}}{b^{j}_{(n+1)}} \qquad (21)$$

where

$v^{ij}_{(n)}$ is the result of the n^{th} iteration

and

$$a^{i}_{(n)} = \sum_{j} v^{ij}_{(n)} \quad ; \quad b^{j}_{(n)} = \sum_{i} v^{ij}_{(n)}$$

We can show the efficiency of such an algorithm with a simple example. The numbers chosen are absolutely random (Table 3).

ORIGINAL TABLE

.02	.04	.07	.13
.03	.13	.10	.26
.05	.27	.29	.61
.10	.44	.46	1.0

NEW MARGINALS

			.10
			.32
			.58
.20	.35	.45	1.0

ITERATION 1:

.0154	.0308	.0538	.1000
.0369	.1600	.1231	.3200
.0475	.2567	.2757	.5800
.0998	.4475	.4526	

.0309	.0241	.0535	.1085
.0739	.1251	.1224	.3214
.0952	.2008	.2741	.5701
.2000	.3500	.4500	

ITERATION 2:

.0285	.0222	.0493	.1000
.0736	.1246	.1219	.3200
.0969	.2043	.2789	.5800
.1990	.3511	.4501	

.0286	.0221	.0493	.1000
.0470	.1242	.1219	.3201
.0974	.2037	.2788	.5799
.2000	.3500	.4500	

TABLE 3.

A-2 Split of the Transit Trips by Purpose

This time we want to know how many people are going to use the transit system for shopping, to work or for social-recreational reasons. To derive these volumes we will use the peak load count.

Our assumption will be that the trips for a given purpose are constnatly distributed overtime, i.e. a given proportion of people are going to work, to shop or for other purposes during the morning peak, the day off-peak and the evening peak.

As we can assume that the time in the day of the trip has really no effect on the mode chosen, this proportion will remain the same for the three different modes.

Then if:

α_1 is the percentage of work trips during the morning peak

β_1 " " " shopping " " " " "

γ_2

α_2 " " " work " " " " "

β_2 " " " shopping " " " " "

γ_2 social-recreational " " "

We have:

	MORNING PEAK	DAY OFF PEAK	EVENING PEAK
WORK	$\alpha_1 \times W$	$(1-\alpha_1-\alpha_2)W$	α_2 W
SHOPPING	$\beta_1 \times S$	$(1-\beta_1-\beta_2)S$	β_2 S
SOC.-RECREATIONAL	$\gamma_p \times R$	$(1-\gamma_1-\gamma_2)R$	γ_2 R

where W, S, and R are the volumes for the different purposes.

Then, Y P_M is the morning peak load count.

" " P_E " evening " " " .

" " P_0 " off peak " " " .

We have:

$$P_M = \alpha_1 W + \beta_1 S + \gamma_1 R$$
$$P_0 = (1-\alpha_1-\alpha_2) W + (1-\beta_1-\beta_2) S + (1-\gamma_1-\gamma_2) R \qquad (22)$$
$$P_E = \alpha_2 W + \beta_2 S + \gamma_2 R$$

As we can see here, we have a system of three equations with three variables W, S, R that in most of the cases can be solved. Therefore, if we have the peak load counts for the three periods of the day, we can easily compute the three volumes W, S, R. In fact, we do not need to make these three counts. Only two will suffice. If T is the total daily volume we have:

$$T = W + S + R$$

Consequently, with two of the previous equations we have a new system equivalent.

The coefficients α_i, β_i, $_i$ are variable according to the towns. In a survey made in Washington, it has appeared that some of these coefficients were:

$$\alpha_1 = .85 \qquad \beta_1 = .18 \qquad \gamma_1 = .18$$

A-3 Time Distribution

This time we want to know how the transit volumes are going to be distributed between peak and off peak hour, assuming that we know them for each purpose.

The method has already been discussed in the previous chapter. It is nothing but the converse of the previous argument.

We know, as a first approximation, that some coefficients α_i, β_i, γ_i exist which are the peak rates for each purpose. Consequently, if we know W, S and R:

$\alpha_1 W$,	$\beta_1 S$,	$\gamma_1 R$	will be	the	volumes	during	morning	peak
$\alpha_2 W$,	$\beta_1 S$,	$\gamma_2 R$	"	"	"	"	evening	"
$(1-\alpha_1-\alpha_2)W$,	$(1-\beta_1-\beta_2)S$,	$(1-\gamma_1-\gamma_2)R$		"	"	"	off	"

B. Distribution Variables for All Modes

We want this time to derive the OD matrix for all purposes and all modes. As we have heavily put the emphasis on transit data, we will assume that we already know the O-D matrix for transit (obtained synthetically, or by survey). But this is not enough, and we will assume furthermore that we know also the generation figures and the modal shares at a high scale (i.e. very general data) for the three purposes. For example, we can know the generation figure for the whole town and the modal shares between the 3 basic zones, CBD, 1st Ring, 2nd Ring, for the 3 purposes. Then we will limit our study to a bus line which is located in a corridor. In this case the people living in that corridor will have only one degree of freedom for the destination of their trips. If we split the line again into sections i (i = 1 . . . N) and assume that the generation figure, for a first approximation, will be the same for the corridor as for the town, the total number of trips originated at a zone i will be:

$$T_{P_i} = \alpha_P \times PoP_i$$

where T_{P_i} is the number of trips originated in i for purpose P

α_P is the generation rate for purpose P

$(POP)_i$ is the population of zone i

Then, for a purpose P (we will drop the index to make the formulas easier) we will have:

$$T_i \underset{j \neq i}{=} V^{ij}$$

where $V^{ij} = V_A^{ij} + V_T^{ij} + V_W^{ij}$

As we know V_T^{ij} the challenge will be to derive V_A^{ij} and V_W^{ij}.

The first step of the argument will be to build a Transit matrix at the same scale as the modal shares we have. For example, if we have the modal shares between or within the CBD, the first ring, and the second one, we must derive the Transit O-D matrix for these zones.

Then we can assume that the modal shares between these zones in the corridor will be the same as the global ones. At this point we can compute the intermediate volumes V_A^{ij} and V_W^{ij}. We have:

$$V_A{}^{ij} = \frac{P_A{}^{ij}}{P_T{}^{ij}} V_T{}^{ij} \quad , \quad V_W{}^{ij} = \frac{P_W{}^{ij}}{P_T{}^{ij}} V_T{}^{ij}$$

where $P_A{}^{ij}$, $P_T{}^{ij}$ and $P_W{}^{ij}$ are the modal shares for the zones.

If the volumes were the real ones we should have:

$$\sum_{j=1}^{N} (V_A{}^{ij} + V_T{}^{ij} + V_W^{ij}) \overset{?}{=} \alpha \, (PoP)i \qquad (23)$$

But this will not be the case most of the time. Therefore, we have to adjust the results. We will make strong hypotheses which once again are only an approximation of the reality.

We will assume that

a) the modal share between auto and two wheels will remain the same for all the zones. By the way we are isolating the transit; of course the total modal share will vary.

b) the change of the volumes V_A^{ij} and V_W^{ij} will be proportional to the value of these volumes.

This can be summarized by the two following formulas:

$$\text{a)}\quad \frac{V_A^{ij} + \Delta V_A^{ij}}{V_W^{ij} + \Delta V_W^{ij}} = \frac{V_A^{ij}}{V_W^{ij}} = \frac{P_A^{ij}}{P_W^{ij}}$$

and

$$\text{b)}\quad \frac{V_M^{ij} + \Delta V_M^{ij}}{V_M^{ik} + \Delta V_M^{ik}} = \frac{V_M^{ij}}{V_M^{ik}} \qquad (k \neq i)\ \ (M = A \text{ or } W)$$

where ΔV_M^{ij} represents the change in volume between i and j using mode M.

These two formulas can be rewritten as:

$$\frac{\Delta V_A^{ij}}{\Delta V_W^{ij}} = \frac{V_A^{ij}}{V_W^{ij}} \qquad ; \qquad \frac{\Delta V_M^{ij}}{\Delta V_M^{ik}} = \frac{V_M^{ij}}{V_M^{ik}}$$

As it can be seen this is equivalent to:

$$\frac{\Delta V_M^{ij}}{\Delta V_{M'}^{ik}} = \frac{V_M^{ij}}{V_{M'}^{ik}} \iff \Delta V_M^{ij} = \frac{V_M^{ij}}{V_{M'}^{ik}}\, \Delta V_{M'}^{ik} \quad (M_1 M' = A \text{ or } W)$$

So we have 2N unknown ΔV_M^{ij} (M = A, W; j = 1 . . . N) and 2N-1 equations coming from (2N-1) independent rates $\dfrac{V_M^{ij}}{V_{M'}^{ij}}$. But we have also the formula (23) that can be written as:

$$\sum_{j=1}^{N} (V_A^{ij} + \Delta V_A^{ij} + V_T^{ij} + V_W^{ij} + \Delta V_W^{ij}) = \alpha(PoP_i)$$

Therefore:

$$\sum_{j=1}^{N} (\Delta V_A^{ij} + \Delta V_W^{ij}) = \alpha\ PoP_i - \sum_{j=1}^{N} (V_A^{ij} + V_T^{ij} + V_W^{ij})$$

So we have a system of 2N linear equations with 2N unknown which will be solvable all the time. The solution is:

$$\Delta V_M^{ij} = \frac{V_M^{ij}}{\sum_{j=1}^{N} (V_A^{ij} + V_W^{ij})} \times \left[\alpha\ PoP_i - \sum_{j=1}^{N} (V_A^{ij} + V_T^{ij} + V_W^{ij})\right] \quad (24)$$

To show the efficiency of such a method we will take the example of a bus line (line Q) in Toulouse. This line goes from the CBD to the east suburb of Toulouse, and it is the only line available on a large length creating a corridor (first and second ring).

We are given:

- a transit O-D matrix (station by station) coming from a survey aboard the buses in May 1976.
- global generation rates for the three purposes (for the whole town).
- modal shares for work trips between and within the CBD (zone 0), the first ring (zone 1), the second ring (zone 2).
- the approximate population served by the line, in these three zones

With these data we are able to apply the method described previously, for work trips.

The first point is to get an aggregate transit O-D matrix for work trips between the basic zones. Using the O-D matrix (shown previously in the study) by station, it is easy to aggregate it. Then we know that only 44% of the people on that survey were making work trips. Applying therefore the coefficient .44 we get the following matrix:

	CBD	1st RING	2nd RING
CBD (Zone 0)	95	149	272
1st Ring (Zone 1)	105	2	12
2nd Ring (Zone 2)	276	12	22

But there again inferences must be made. This O-D matrix represents the distribution of work trips for the WHOLE DAY. In other words it means that the volumes shown can be as well trips to go to work or return trips. But according to the existing employment in the different zones, it seems reasonable that all the work trips are going towards the CBD. Furthermore, the matrix is rather symmetric and would show that the trips going from the CBD to the 1st ring or 2nd ring are return trips.

As we have assumed that the part of the first and second ring served by the line are in a corridor, we can apply the method.

The population of the zone 2 served by the line is 7250.

" " " " zone 1 " " " 6000.

The generation rate for work trips at the city scale is a .32. The modal shares are:

within 2nd ring $P_A^{22} = .71$ $P_W^{22} = .23$ $P_T^{22} = .06$

within 1st ring $P_A^{11} = .66$ $P_W^{11} = .26$ $P_T^{11} = .12$

Between 1st and 2nd ring $P_A^{21} = .69$ $P_W^{21} = .19$ $P_T^{21} = .12$

Between 1st ring and CBD $P_A^{10} = .55$ $P_W^{10} = .23$ $P_T^{10} = .22$

Between 2nd ring and CBD $P_A^{20} = .65$ $P_T^{20} = .16$ $P_T^{20} = .21$

Then, according to these data we should have for the trips originated in zone 2:

$$T_2 = \alpha \times PoP_2 = .32 \times 7250 = 2320$$

As $V_M^{2i} = \frac{P_M^{2i}}{P_T^{2i}} \times V_T^{2i}$, we get the tableau:

	TRANSIT	AUTO	2 WHEELS	TOTAL
Zone 2 ⟶ CBD	274	848	183	1305
Zone 2 ⟶ Zone 1	12	69	19	100
Zone 2 ⟶ Zone 2	22	260	84	366
TOTAL	308	1177	286	1771

As we see we do not reach the 2320 predicted trips. Therefore, we can use the formula found previously:

$$\Delta V_M^{2i} = \frac{V_M^{2i}}{\sum_{j=o}^{2} (V_A^{2j} + V_T^{2j})} \times (\alpha \ PoP_i - \sum_{j=o}^{2} (V_A^{2j} + V_T^{2j} + V_W^{2j})$$

here $\sum_{j=o}^{2} (V_A^{2j} + V_T^{2j}) = 1463$

$$\alpha PoP_i - \sum_{j=o}^{2} (V_A^{2j} + V_W^{2j} + V_T^{2j}) = 2320 - 1771 = 549$$

So we get the new table, giving the computed volumes and the modal shares:

	TRANSIT	AUTO	2 WHEELS	TOTAL
Zone 2 → CBD	274 (16%)	1156 (70%)	252 (14%)	1692
Zone 2 → Zone 1	12 (9%)	95 (71%)	26 (20%)	133
Zone 2 → Zone 2	22 (5%)	358 (72%)	115 (23%)	495
TOTAL	308	1619	393	2320

As it can be seen here the rates of transit are slightly lower than the average. However, this could be explained by the fact that there is only one bus line in this area. In other areas of Toulouse the network is much denser and the ridership is greater. Anyway, we see that we can find some O-D matrix for any mode knowing the global modal shares and the generation rate. Yet the accuracy of the results is greatly related to the exact generation figure, the exact global modal shares and overall the exact population served by the line. The example of the results for the zone 1 can illustrate perfectly this fact. The population has approximately been estimated to 6000 inhabitants in Zone 1. Therefore, the total number of trips originated in Zone 1 will be:

$$T_1 = .32 \times 6000 = 1920$$

If we compute the volumes according to the global modal shares we get the table:

	TRANSIT	AUTO	2 WHEELS	TOTAL
Zone 1 ⟶ CBD	105	263	110	478
Zone 1 ⟶ Zone 1	2	11	4	17
Zone 1 ⟶ Zone 2	0	0	0 (we have assumed that the trips were return trips from 2→1)	
TOTAL	107	276	114	495

As we see there is a tremendous gap between the 2 volumes. And at this point we have not made any important assumption; only that the global shares remained the same in the corridor. The mistake obviously comes from the population that has been overestimated. This shows then that the results will be consistent only if the data are. The inferences about the context are useful if the preliminary studies give some relevant results.

4-3. INFERENCES ABOUT LEVEL OF SERVICE

In many calibrations of models the level of service of all the modes is needed. If the level of service of transit is easily available for the bus operator it is not the case for the level of service of the other modes, and we must then find a way to derive them.

Mainly the level of service variables are:

- the out of pocket cost
- the in vehicle time
- the waiting time
- the walking time

$\Big\}$ which can be replaced by the out of vehicle time

We can assume that the out po pocket cost of the waiting time and the in vehicle time are known for the transit mode. Therefore, we will have to derive from these results and from data the other variables.

A. In Vehicle Time

Two cases are possible:

Case 1 - No data at all exists and inferences must be done only from transit data.

Case 2 - Some data are available and the inferences can be done about them.

Case I

The in-vehicle time is essentially based on the generalized speed of the vehicles, i.e. the speed including all the necessary stops (at the stop lights, for example). We are supposed to know the speed of the buses and of course the in-vehicle time. It appears that there exist some rough coefficients which give the speed of the automobile (and two wheel) as a function of the speed of transit. These coefficients come from some general observations on the traffic.

We have then:

$S_A = \alpha_A S_T$

$S_W = \alpha_W S_T$ where S_M represents the speed of mode M, and α_i is a constant.

As the in vehicle time can be represented by:

$$IVT = \frac{DST}{S} \begin{array}{l}\text{(distance)}\\ \text{(generalized speed)}\end{array}$$

we have:

$$IVT_A = \frac{DST}{S_A} = \frac{DST}{\alpha_A ST} = \frac{1}{\alpha_A} IVT_T$$

$$IVT_W = \frac{1}{\alpha_W} IVT_T$$

The generally admitted values of α_W are $\frac{1}{.85}$ and 1 on an urban network. But of course these values are inaccurate and represent only roughly the reality. Furthermore, if we apply a coefficient of proportionality between the modes, we will lose an essential property of these variables for an eventual regression: the variance of their difference

Case 2

If we want to keep some variance in the difference of in vehicle time between the modes we need some other results. Many surveys have been done on the urban traffic, and often these surveys can be transferred from one city to another. Usually they give the average speed as a function of the nature of the road. Here follows an example of such data (source Hall and George, Highway Research Board 1959).

TYPE OF FACILITY	SPEED FOR SAN DIEGO DURING PEAK HOUR FOR AUTO
Major St. - Normal	30 MPH
Major St. - Highly Developed	20 "
Collector Street	25 "
Local and Business District	15 "

This kind of data exists very frequently and can easily be used. Knowing the distance between the zones it is easy to compute the in-vehicle time. We first study the kind of link joining the two zones and then a first approximation of the in vehicle time will be:

$$IVT_A = \frac{DST}{S_A} \qquad NT_W = \frac{DST}{S_W}$$

Another kind of data gives the average speed of the cars as a function of the existing volume and the capacity of the facility in urban areas. Here follows an example.

Ratio $\frac{\text{VOLUME}}{\text{CAPACITY}}$	AVERAGE OVERALL TRAVEL SPEED
less than .70	> 30 MPH
less than .90	> 25 MPH
less than .85	> 20 MPH
less than .90	> 15 MPH
less than .95	$\simeq$ 15 MPH
more than .95 (traffic jam)	< 15 MPH

(source: Highway Research Board; Highway Capacity Manual 1965)

Therefore, the task of the analyst is to determine the average distance between the zones, the average spped of the car according to the kind of traffic between the zones and compute the in-vehicle time.

The same kind of inferences can be done about the 2-wheels but unfortunately the data is missing in the U.S. about that mode of transportation.

B. Waiting (Excess) Time

If the waiting time commonly accepted for the transit is half the headway (as a first approximation), the excess time for the auto mode is not so neatly defined. The excess time for the auto consists of the parking and unparking time, and of course it will depends essentially on the location of the parking space. The time to park will certainly vary from on city to another. But if no data is available one can try to transfer some data from another city. Here are given rough estimates of the parking time in Pittsburgh.

Type of Parking Area	Parking time (min.)	Unparking time (min.)
Industrial	2	2
High density shopping area	2	2
Low density shopping area	2	2
High density resident	1	1
Low density resident	0	0

(Source: Domencich McFadden; Urban travel demand) (1975)

The excess time for the two-wheels can be assumed to be zero. These vehicles can be parked everywhere, and consequently there is not waiting time.

C. Walking Time

According to some surveys most auto drivers declared that they did not walk more than one block to get their car. Therefore, the waiting time for the auto mode can be assumed to be negligible (i.e. zero). This will be even more true for the 2-wheels mode, where most of these vehicles are parked in the very building where the people work or live. Therefore, the walking time will be important mainly for the transit mode. We have delimited the emission and attraction zones within 1/4 mile of the bus stations on the line. As we are dealing with aggregate data, the first point will be to compute the average distance walked by an average passenger. Many assumptions can be made about the distribution of the location of the people taking the bus within the emission zone. The simplest assumption that can be used if nothing else is available is a uniform distribution giving 1/8 mile as the average walking distance. But some studies have been done and showed that this distance can be a function of the income, and that in general more people were taking the bus within 1/8 mile of the station than between 1/8 and 1/4 miles.*

The walking distance will not suffice. Some estimates of the walking speed are also required. Some surveys have been done, and this time the results can be transferred because the walking conditions are similar in all cities. Here follows a survey in Pittsburgh giving the walking speed as a function of the walking location.

Walk Location	Average Speed
Downtown (congested streets and sidewalks)	2 mph
High density residential and commercial suburban districts	3 mph
medium density residential and commerical suburban districts	3 mph

(Source: Domencich MacFadden, Urban Travel Demand)

Thus, with the average walking distance and the average speed, the operator will be able to determine the average walking time.

*John Shortreed, Transit and Pricing Policy.

D. Out of Pocket Cost

The transit out of pocket cost is the fare. The auto out of pocket c s will be the operating and parking costs, and the 2 wheels out of pocket cost will be the operating costs only.

a) Operating Costs

Many studies have been made about the auto operating cost or more exactly about the mileage of a car according to the type of the road. Using these studies, and having the average gas price will give a rough first estimate of the operating cost. In fact, to get a more accurate estimate of these costs we need the costs of repais, tires, etc. Here are two tables giving some of these estimates. They are old but can give an idea of these kinds of data.

Roadway Condition	Miles/Gallon for passenger vehicles
Expressway	22.3
Arterial coordinated signal	18.4
Arterial ordinary	14.9
Congested businees street	8.7

(Source: Gibbons and Proctor, Economic Cost of Traffic Congestion.)

Average travel speed MPH	Cost, cents per vehicle mile: Estimated # of stops per miles	Cost of Fuel	Cost of tires, repairs, etc.	Cost of stops	TOTAL COSTS
10	6.26	3.40	.96	.28	4.62
15	2.71	2.71	.95	.18	3.86
20	1.29	2.30	.98	.11	3.39
30	.26	1.82	1.05	.06	2.91
40	0	1.55	1.19	0	2.76

(Source: CATS Research News 1959)

With this kind of data, and the description of the links between the zones it is possible to compute the operating cost for the auto.

The same kind of data might exist for the 2 wheels (in Europe). Unfortunately this data is missing in the U.S.

b) Parking Cost

There is no rule to compute the average parking cost in the different zones. Some surveys are needed in each zone to know the average price of a parking or we can only use the average price for the whole town.

4-4. Inferences about the Socio-Economic and Attraction Data

As we said previously, the socio-economic and attraction variables are usually given (in the U.S.) by the census tracts. Unfortunately, there are cases when even this data is not available. The challenge of the analyst will then become greater, and the inferences harder to make.

I have to admit that unfortunately I could not find any inferences about these variables. But of course it does not mean that it is impossible. The field of research is wide open and it is probably one of the most interesting problems to explore.

4-5. The Concept of Elasticity*

A. Definition

The elasticity is a concept which is often used inaccurately. Its definition is: the elasticity of y with respect to n $E_n(y)$ is the precent change of y for 1 percent change in n.

Thus the elasticity of the volume V with respect to the level of service S will be

$$E_S(V) = \frac{So}{Vo} \cdot \frac{\partial V}{\partial S}$$

where Vo, So are the existing volume and the existing level of service. From this exact concept of elasticity must be sometimes derived an approximate concept, the arc elasticity. The arc elasticity will be given by:

$$E'_S(V) = \frac{So}{Vo} \times \frac{\Delta V}{\Delta S} \qquad (25)$$

wehre in this case we are examining the changes at a level which is not marginal anymore. The difference between the two concepts is given by Figure 8.

This arc elasticity can in some ways be improved by the introduction of an average arc elasticity given by:

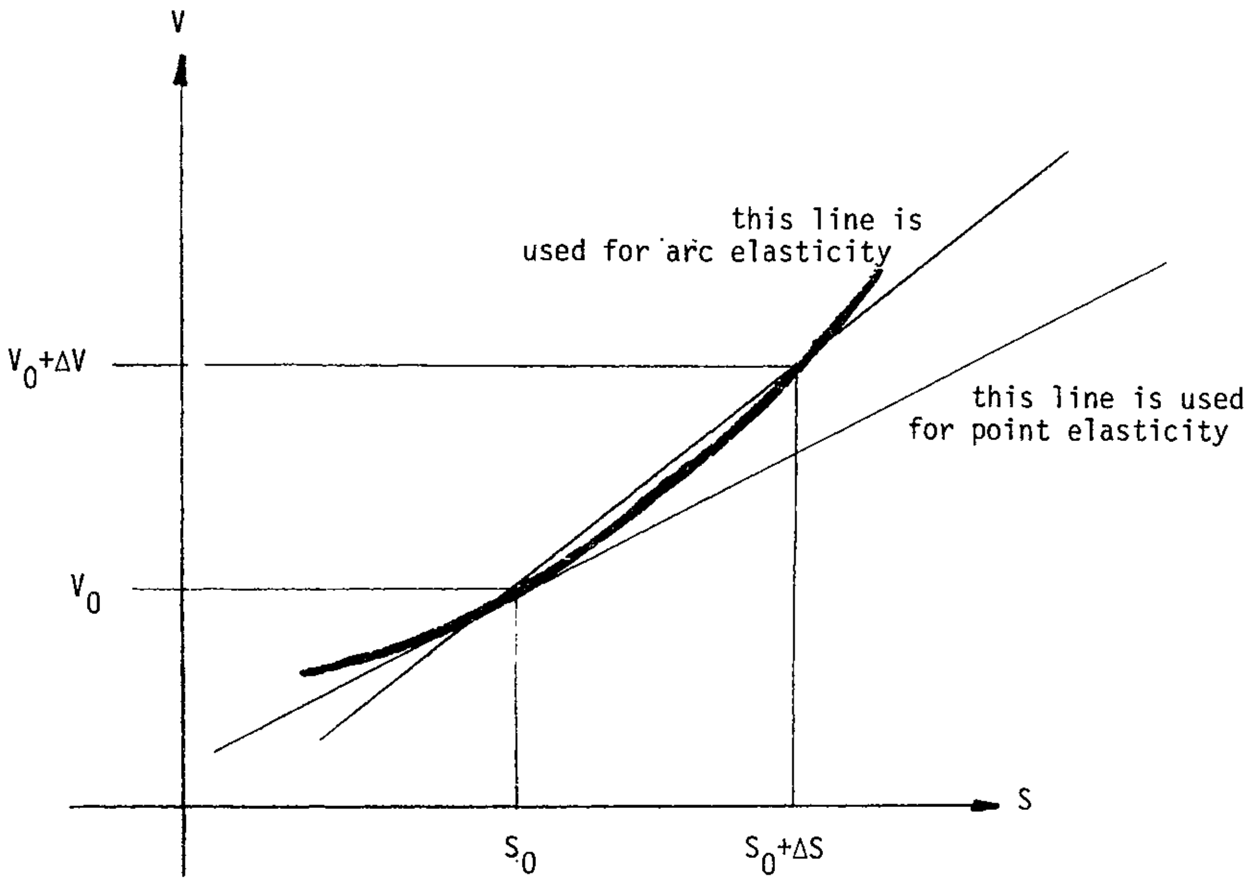
$$E''_S(V) = \frac{\bar{S}}{\bar{V}} \frac{\Delta V}{\Delta S} \qquad (26)$$

where $\bar{S}$ and $\bar{V}$ represent the average S and V before and after the change. To compute these last two elasticities, when no demand function is

---

*Marvin L. Manheim, Fundamentals of Transportation Systems Analysis, MIT, 1976, in print.

FIGURE 8.


V
S
$V_0+\Delta V$
V_0
S_0
$S_0+\Delta S$
this line is
used for arc elasticity
this line is used
for point elasticity

available, the operator has to implement a change in the level of service, analyse the change in ridership provoked by this change, and apply one of the two formulas.

Use of the concept of elasticity:

To make a forecast using this concept, the operator will need to know three parameters:

- The existing ridership
- The existing level of service
- The change in level of service he will try to implement

In this case we can derive from the formula (25)

$$\Delta V = \frac{Vo}{So} \times E'_S(V) \times \Delta S$$

where we see that the change in volume is proportional to the change in level of service and the rate $\frac{Vo}{So}$. The new ridership is given then by the formula

$$V_1 = V_o + \Delta V = V_o \left[1 + E'_S(V) \times \frac{\Delta S}{So}\right] \qquad (27)$$

From the second formula (26) we can derive:

$$\frac{\Delta V}{\overline{V}} = \frac{\Delta S}{\overline{S}} \times E''_S(V)$$

but $\frac{\Delta V}{\overline{V}} = 2 \times \frac{V_1 - V_o}{V_1 + V_o}$ therefore we get:

$$2(V_1 - V_o) = \frac{\Delta S}{\overline{S}} \times E''_S(V) \times (V_1 + V_o)$$

if we note $K = \frac{\Delta S}{\overline{S}} \times E''_S(V)$, we get

$$V_1 = Vo \times \frac{2 + K}{2 - K} \qquad (28)$$

As we see this is a method to predict a new ridership when the three parameters described previously are known. But this is an inaccurate way of computing it. The elasticity is an approximate concept which varies a lot from one place to another. It also varies along the demand curve (e.g. with logit). Furthermore, it has appeared that the elasticity could be oriented, i.e. the elasticity with respect to an increase of level of service might be different from the elasticity with respect to a decrease in the level of service. In conclusion, it is not a reliable concept (because of the approximation and should be used as a last resort. In France (Paris), the elasticity considered are:

- with respect to frequency: .07*
- with respect to fare: 1.6

---

*Christian Bourgin, "L'utilisation du concept d'elasticite dans les evaluations de clientele T.C. (IRT 1976).

CHAPTER V

CONCLUSION

When I first started this sutdy, months ago, I was very doubtful about its conclusion. Most of the transportation analysis methods I had learnt were so much based on great collections of data or on computer time that I did not really think it would be possible to find some simplified results. Obviously I was wrong. The methods developed here are simple and give some relevant results. Of course they are far from being perfect and they certainly need a to be improved. Nevertheless, they can help as examples on a new direction or research.

Some preliminary conclusions have come up. This methodology, though simple, is sound. The methods employed are mathematically justified, but incomplete. Therefore, some further research should be done for:

- prediction

As we have used for the prediction a naive procedure of aggregation, a bias has been introduced. This bias should be estimated (cf. Koppelman estimation of bias in aggregation procedure). At this point there is a way to diminish that bias. It can be done with the use of sample *. The pattern of the process is the following:

(1) Establish the basic population units of the corridor

(2) Identify and assemble data for the population of units

(3) Decide on major classification variables and classes

(4) Classify population of units into above categories. This is

---

*Marvin L. Manheim, "Demand Prediction with Geographic Classification," M.I.T., 1977, unpublished.

then the sampling frame.

(5) Decide on the desired sample size, both total and number in each category of the sampling frame

(6) Randomly select a sample of unit from the sample frame

(7) Summarize the relevant socio-economic characteristics from these units.

(8) Calculate the "base case" service level for each unit in the sample

(10) For each unit using the appropriate models, calculate the demand volumes at the base levels of service; expand

(11) If any data is available on actual usage in the corridor, develop adjustment factors to match predicted and observed volumes.

Most of our results have been based on the fact that the people using the bus line were in a corridor and had only one degree of freedom for the destination of their trips. This is an approximation and the error of such an assumption needs to be computed. Our methods of prediction were based on the fact that we were using a logit model but other models can very well be used and some new simplified way, perhaps leading to better results could be found.

- estimation

The method of estimation used here for the logit model was based on the least squares estimators, and needed an aggregation of the data. We surely introduced an error in aggregating these data. An estimation of that error is needed in order to give some limits to the results.

Again, we used the logit model, but all sorts of models could be used, and it would be interesting to compare the results coming from different models estimated in a simplified way..

- inferences about the context

We have here studied only a few specific cases of inferences, and this field of search is nearly unlimited. Many more methods should be developed to try to get the maximum information from the minimum data. Mainly the emphasis should be put on inferences on the socio-economic and the attraction variables. Then the next point would be, once again, to estimate the errors made with such inferences. As a matter of fact, all the results given here have not been bound between some upper and lower limits, and for that reason are not completely valid.

We must not forget that all the methods were developed because they could be programmed on a pocket calculator and did not need any computer hardware. But as the time passes these pocket calculators are improving and it is obvious that the limits we were bound to will be expanded in the future. At this time the field of research will be enlarged and new methods will be possibly applied. For example, it may be possible to calibrate a logit model on a pocket calculator with the maximum likelihood method.

But this is future and we are now using these methods. That's why we must recognize their limits. These methods are useful as a first help for people who just want a rough feeling on a given situation. It may help to predict if a change in the service of the line is desirable but it will by no means determine with accuracy the consequences of such a change. It means that these methods cannot replace in any case a detailed analysis using the maximum data. The results found by our methods should be limited between upper and lower estimates (unfortunately this part of the study has not been done). Therefore, these methods will be more useful for the extreme ends of the range of decisions. If the criteria to implementa change is a minimum limit for the future ridership and if our estimates have a lower bound greater than this minimum limit we can assume that we can implement the change. The same kind of inferences will occur with maximum limits. The problem will arise when our methods will give upper and lower estimates which will bracket these maximum or minimum limits for an implementation decision. It is at that point that a simplified method is insufficient and some further studies are required.

Also, we have seen that our methods were heavily based on the few available data. As all sorts of inferences are made from these few data it is essentail for these data to be accurate. By our methods we introduced many errors. If by chance our data are already biased our results will not actually meany anything.

The last conclusion will be a caution for the future. If later these kinds of methods are developed, and they surely will be, they will have to include, all the time, the behavioral characterisitcs of the consumer. The simplification of the methods does not mean the complete abandon of the psychological aspect of the transportation analysis to the profit of the mathematical one. We have simplified or found some mathematical formulas in this thesis, but we tried when it possible to interpret these simplifications by a behavioral aspect of the passengers. It has to be so if the analyst wants to continue to give a meaning to the transportation analysis.

APPENDIX

A. Logit Model Program

This program allows a series of different computations with the logit model. Among others it allows the computation of the existing shares between various alternatives, and the prediction of the new shares, if some changes occur in the utilities of the alternative, or if an extra laternative is add to the set. The number of possible alternatives is equal to 5. For all the alternatives, we have a vector of coefficients that can be up to 16 dimensions. We will give 3 different worksheets for each of the different use of the program.

1. BASE CASE

	θ: Vector of coefficients
INPUTS:	X: Vector of variable of the utility i
	n: number of alternatives

θ	(θ_o) (θ_1) (θ_2) - - - - - - - - $-(\theta_p)$
X_1	
- -	- - - - - - - - - - - - - - - - - -
X_n	

OUTPUTS
V_i: Utility of alternative i
S: Share of alternative i

i	1	2	-	-	n
V_i					
S_i					

Notice: If the utilities of the different alternatives are known they can be entered as inputs directly.

2. CHANGE IN UTILITIES (PIVOT POINT)

INPUTS ΔV_i = change in utility i

θ	(θ_o)	(θ_1)	(θ_2) - - - - - - - -	(θ_p)
ΔX_i				
$\theta_o \Delta X_{ik}$				

$\Delta V_i = \Sigma\theta_k \Delta X_{ik} =$ $\boxed{}$

OUTPUTS

V'_i: new utility of alternative i

S'_i: new share of alternative i

i (key)	1	2	-	-	n
V'_i					
S'_i					

3. ADDITION OF NEW ALTERNATIVES

INPUTS

n': new # of alternatives

X_j: vector of variables of the new utility j

θ	θ_o	-	-	θ_n
X_j				

OUTPUTS

U_j: utility of new alternatives

S'_i: new shares

i	1	-	-	-	-	n'
U_i						
S'_i						

CAUTION: As we had to use some arithmetic stack memories to s⁻ore certain values, never press hte key CLR unless indicated. Press CE or = and then 0.

SR-52 User Instructions

TITLE LOGIT MODEL BASE CASE PAGE 1 OF 4

◄A►		LOGIT MODEL		
n	θ	chg Vi	Del Xik	INIT Si
$X_1 \rightarrow S_1$	$X_2 \rightarrow S_2$	$X_3 \rightarrow S_3$	$X_4 \rightarrow S_4$	$X_5 \rightarrow S_5$

◄B►		LOGIT MODEL		

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
1	READ PROGRAM A and B		2ND	READ		
2	INITIALIZE		CLR			
			2ND	CMS		
			2ND	rset		
3	ENTER # ALTERNATIVES	n	2ND	A´		n
4	ENTER VECTOR θ	θ_o	2ND	B´		-13
	(The # of coefficients	θ_1	2ND	B´		-12
	must not exceed 14. The					
	display shows the #	θ_p	2ND	B´		-13+P
	still allowed).					
5	IF NECESSARY RECALL θ_p		RCL	P		θ_p
6	IF NECESSARY CHANGE θ_p	θ'_p	STO	P		θ'_p
7	LOAD UTILITY ALT 1	X_{11}	A			P-1
	(The display shows the	X_{12}	A			P-2
	# of variables necessary					
	to fill up the utility					
	function)	X_{1p}	A			1 Flashing
8	IF NECESSARY DELETE	X_{ik}	2ND	D´		P-k
	LAST X_{ik} AND REPLACE	X'_{ik}	+/-	2ND	D´	P-k
	BY X'_{ik}					
9	LOAD UTILITY ALT 2		CE			
	REPEAT STEP 7	X_{21}	B			P-1
		X_{22}	B			P-2

 Part No. 1220479

SR-52 User Instructions

TITLE LOGIT MODEL BASE CASE PAGE 2 OF 4

◄A►		LOGIT MODEL		
n	θ	Chg Ui	Del Xik	INIT Si
$X_1 \rightarrow S_1$	$X_2 \rightarrow S_2$	$X_3 \rightarrow X_3$	$X_4 \rightarrow X_4$	$X_5 \rightarrow X_5$

◄B►		LOGIT MODEL		

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
		X_{2p}	B			2 FLASHING
	REPEAT STEP 7 9 FOR					
	ALL ALTERNATIVES UNTIL		CE			n-1
	ALTERNATIVE n	X_{n1}	B,C	D, or	E	p-1
			(according to n)			
		X_{np}	B,C,D, or E			0 flashing
			CE			0
10	COMPUTE SHARES		2nd	E´		
	ALT. 1		A			S_1
	ALT. 2		B			S_2
	ALT. 3		C			S_3
	etc....					
11	RECALL IF NECESSARY		RCL	(14+i)		e^{Ui}
	UTILITY ALTERNATIVE i		ln			Ui
12	IF POSSIBLE THE	1	STO	(14+i)		1
	ALTERNATIVE UTILITIES		SUM	68		1
	CAN BE ENTERED DIRECTLY	i	2nd	C´		i+14
	GO THEN DIRECTLY	Ui	2nd	C´		i flashing
	TO STEP 10					

 Part No. 1220479

SR-52 User Instructions

TITLE LOGIT MODEL PIVOT POINT PAGE 3 OF 4

	A				
	LOGIT MODEL				
n	θ	CHG UiDEL	Xik	INITSi	
$X_1 \to S_1$	$X_2 \to S_2$	$X_3 \to S_3$	$X_4 \to S_4$	$X_5 \to S_5$	

	B			
	LOGIT MODEL			

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
14	INTRODUCE CHANGES	i	2nd	C´		14+i
	IN THE UTILITY ALTi	ΔUi	2nd	C´		i flashing
15	REPEAT 13 IF NECESSARY		CE			o flashing alternative n
11	RECALL IF NECESSARY		RCL	(14+i)		e^{Ui}
	UTILITY Ui			ln		Ui
10	COMPUTE NEW SHARES		2nd	E´		
	NEW SHARE ALT 1		A			S_1
	ALT 2		B			S_2
	etc......					
17	IF ALL THE X_i's		CLR			0
	ARE CHANGED BUT IF		STO	1	5	0
	θ REMAIN THE SAME		STO	1	6	0
	(Different problem)		STO	1	7	0
	RETURN TO STEP 7		STO	1	8	0
			STO	1	9	0

 Part No 1220479

SR-52 User Instructions

TITLE LOGIT MODEL ADDITION OF A NEW ALTERNATIVE PAGE 4 OF 4

	A				
	LOGIT MODEL				
n	θ	CHG UiDEL	Xik	INIT Si	
$X_1 \to S_1$	$X_2 \to S_2$	$X_3 \to S_3$	$X_4 \to S_4$	$X_5 \to S_5$	

	B			
	LOGIT MODEL			

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
19	INITIALIZE		2nd	reset		
20	ENTER NEW # OF	n´	2nd	A´		n´
	ALTERNATIVES					
8	FILL UP VECTOR OF	$X_{(n+1)1}$	B, C,	D, E		p-1
	NEW UTILITY					
	(PROCEDURE RESUMES					
	WHERE STEP 8 HAD STOPPED)	$X_{(n+1)p}$	B, C,	D, E		(n+1) flashing
10	COMPUTE NEW SHARES		2nd	E´		
	NEW SHARE ALT 1		A			S_1
	ALT 2		B			S_2
11	RECALL IF NECESSARY		RCL	(14+i)		e^{Ui}
	NEW UTILITIES		ln			Ui

 Part No 1220479

000	46
001	16
002	42
003	01
004	04
005	81
006	46
007	17
008	36
009	42
010	06
011	09
012	01
013	44
014	06
015	09
016	43
017	06
018	09
019	75
020	01
021	04
022	95
023	80
024	01
025	05
026	05
027	81
028	46
029	23
030	42
031	06
032	06
033	60
034	00
035	01
036	06
037	03
038	43
039	09
040	08
041	65
042	36
043	43
044	06
045	07
046	95
047	36
048	44
049	06
050	06
051	01
052	44
053	06
054	07
055	43
056	06
057	07
058	75
059	43
060	06
061	09
062	95
063	80
064	01
065	01
066	09
067	56
068	46
069	11
070	51
071	42
072	01
073	05
074	51
075	23
076	81
077	46
078	12
079	51
080	42
081	01
082	06
083	51
084	23
085	81
086	46
087	13
088	51
089	42
090	01
091	07
092	51
093	23
094	81
095	46
096	14
097	51
098	42
099	01
100	08
101	51
102	23
103	81
104	46
105	15
106	51
107	42
108	01
109	09
110	51
111	23
112	81
113	46
114	42
115	42
116	09
117	08
118	56
119	42
120	06
121	07
122	36
123	43
124	06
125	06
126	22
127	23
128	36
129	42
130	06
131	06
132	44
133	06
134	08
135	43
136	06
137	06
138	75
139	01
140	04
141	75
142	43
143	01
144	04
145	95
146	80
147	01
148	05
149	05
150	85
151	43
152	01
153	04
154	95
155	45
156	95
157	56
158	46
159	10
160	50
161	00
162	81
163	36
164	43
165	06
166	06
167	55
168	43
169	06
170	08
171	95
172	56
173	46
174	19
175	94
176	42
177	09
178	08
179	01
180	22
181	44
182	06
183	07
184	41
185	00
186	03
187	08
188	46
189	18
190	60
191	01
192	02
193	00
194	05
195	85
196	01
197	04
198	95
199	42
200	06
201	06
202	50
203	01
204	81
205	22
206	23
207	65
208	36
209	43
210	06
211	06
212	22
213	44
214	06
215	08
216	95
217	22
218	50
219	01
220	41
221	01
222	02
223	08

B. O-D Matrix Program

This program has been developed according to the formula (20). The inputs will be the boarding and alighting counts and the outputs will be the volumes V^{ij} between the stations.

Because of the limited memories we can consider only the lines with less than 11 sections.

1	2	3	4	5		n	n+1	m+2
o	o	o	o	o	- - - -	o	o	o
b_1	b_2	b_3, a_1	b_4, a_2	b_5, a_3		b_n, a_{n-2}	a_{n-1}	a_n

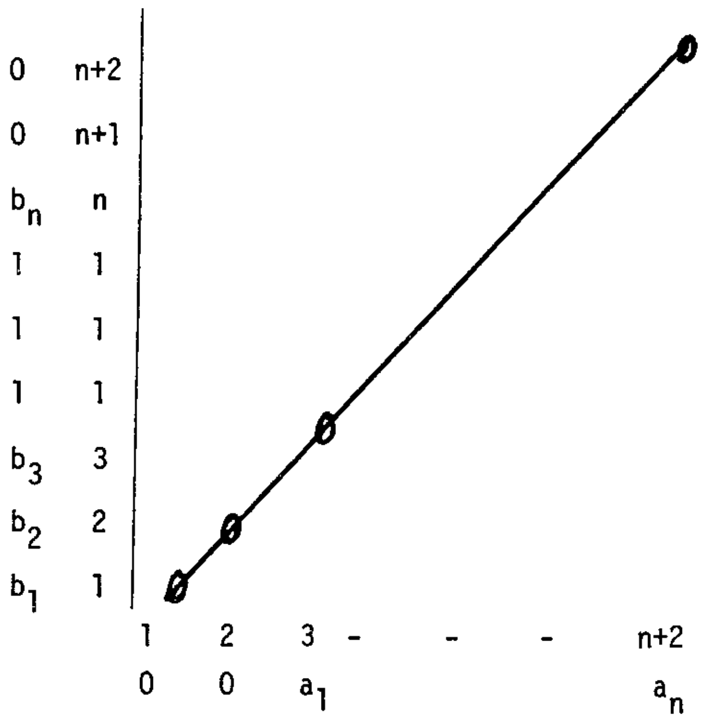
According to the first hypothesis we should have no alighting at station 2, and no boarding at station (n-1). If it was not so, and if we had a people alighting at station 2 and b people boarding at station n, we would first have to transform b, $b_1 - a$ and a_n $a_n - b$, because we will know that the people can only come from station 1 if they alight at station 2, and go to station n+2 if they board at station (n+1). Therefore, we would know V^{12} and $V^{(n+1)(n+2)}$, and we can treat the problem without these volumes.

We have then as

INPUTS: a_j and b_i (defined as on the drawing)

OUTPUTS: V^{ij}

The worksheet will just be the O-D matrix.



<u>CAUTION</u>: We have used some arithmetic unit memories, consequenlty never press the key CLR. Use the key CE or = and then 0.

SR-52 User Instructions

TITLE O-D MATRIX PAGE 1 OF 1

A O-D MATRIX				
del ai	del bi			
ai	bi	Vij		INIT

B O-D MATRIX				
del ai	del bi			
ai	bi	Vij		INIT

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
1	ENTER PROGRAM(A and B)					
2	INITIALIZE		E			
3	ENTER BOARDING COUNTS	bi	B			-(9-i)
4	ENTER ALIGHTING COUNTS	aj	A			-(9-j)
	(Step 3 can be processed					
	after step 4 if wanted)					
	(Only 9 counts can be					
	entered. The display					
	shows how many left are					
	allowed)					
5	DELETE LAST bi,if necessary	bi	2nd	B´		-(9-i+1)
6	DELETE LAST ai,if necessary	aj	2nd	A´		-(9-j+1)
7	COMPUTE V^{ij}		C			V^{ij}´
	REPEAT FOR EACH POINT					
	WHEN ROW i IS FINISHED					(i+1)flashing
8	COMPUTE $V^{(i+1)j}$		C			
	REPEAT STEP 7					
	WHEN MATRIX IS FINISHED					-1 flashing

Part No. [illegible]

000 46	045 53	090 46	135 81	180 43
001 15	046 36	091 32	136 01	181 01
002 25	047 43	092 43	137 94	182 09
003 47	048 09	093 06	138 45	183 75
004 09	049 09	094 07	139 95	184 01
005 42	050 75	095 75	140 81	185 08
006 06	051 43	096 43	141 46	186 95
007 08	052 06	097 09	142 12	187 81
008 42	053 06	098 09	143 42	188 46
009 01	054 54	099 95	144 09	189 17
010 09	055 55	100 90	145 08	190 24
011 01	056 43	101 01	146 01	191 01
012 42	057 06	102 02	147 44	192 22
013 09	058 09	103 06	148 00	193 44
014 09	059 65	104 01	149 00	194 00
015 81	060 36	105 44	150 43	195 00
016 46	061 43	106 06	151 09	196 43
017 13	062 06	107 07	152 08	197 00
018 24	063 08	108 44	153 36	198 00
019 43	064 22	109 06	154 42	199 75
020 06	065 44	110 08	155 00	200 09
021 08	066 06	111 36	156 00	201 95
022 75	067 09	112 43	157 43	202 81
023 43	068 95	113 06	158 00	203 46
024 01	069 44	114 07	159 00	204 16
025 09	070 06	115 75	160 75	205 24
026 95	071 06	116 36	161 09	206 01
027 90	072 81	117 43	162 95	207 22
028 00	073 43	118 06	163 81	208 44
029 07	074 09	119 08	164 46	209 01
030 03	075 09	120 95	165 11	210 09
031 01	076 75	121 44	166 42	211 43
032 44	077 43	122 06	167 09	212 01
033 06	078 00	123 09	168 08	213 09
034 07	079 00	124 41	169 01	214 75
035 44	080 95	125 32	170 44	215 01
036 06	081 80	126 01	171 01	216 08
037 08	082 01	127 44	172 09	217 95
038 36	083 03	128 09	173 43	218 81
039 43	084 06	129 09	174 09	219 00
040 06	085 25	130 43	175 08	220 00
041 07	086 09	131 09	176 36	221 00
042 44	087 42	132 09	177 42	222 00
043 06	088 06	133 45	178 01	223 00
044 09	089 08	134 95	179 09	

4 MULTIPLE REGRESSION

This program has been developped to be able to run regressions up to 4 variables. It can run indifferently regressions for 2,3 or 4 variables. The method used has been studied in chapter III (resolution of the regression problem).

We can represent the program by the following flow chart:

ENTER CARD 1 — INPUT DATA → COMPUTE MOMENTS

ENTER CARD 2 — REGRESS $X_3 = F(X_1, X_2)$ FIND c_0, c_1, c_2

2- variables regression ends here.

READ CARD 3 — REGRESS $X_4 = F(X_3 - X_3)$ FIND b_3

READ CARD 4 — REGRESS $X_4 - b_3X_3 = F(X_1, X_2)$ FIND b_0, b_1, b_2

3- variables regression ends here.

READ CARD 5 — REGRESS $Y = F(X_4 - X_4)$ FIND a_4

READ CARD 6 — REGRESS $Y - a_4X_4 = F(X_3 - X_3)$ FIND a_3

READ CARD 7 — REGRESS $Y - a_4X_4 - a_3X_3 = F(X_1, X_2)$ FIND a_0, a_1, a_2

end

SR-52 User Instructions

TITLE LINEAR REGRESSION PAGE 1 OF 2

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
1	ENTER CARD 1 (A and B)					
2	CLEAR MEMORIES		2ND	A'		0
3	ENTER DATA	X_{1i}	A			2
	(REPEAT FOR ALL i)	X_{2i}	B			3
		X_{3i}	C			4
		X_{4i}	D			5
		Y_i	E			i (n)
4	IF NECESSARY CORRECT X_{1i},	X_{1i}	+/-	A		2
	X_{2i}, X_{3i}, and/or X_{4i}	X'_{1i}	A			2
	BEFORE COMPLETION OBS. i	X_{2i}	+/-	B		3
		X'_{2i}	B			3
		.	.			.
		.	.			.
		.	.			.
5	DELETE AN OBSERVATION	X_{1i}	2ND	D'		2
	(ANY TIME BEFORE CARD 2)	X_{2i}	B			3
		X_{3i}	C			4
		X_{4i}	D			5
		Y_i	E			n-1
6	ENTER CARD 2 (A and B)		A			c_0
			B (optional)			c_1
			C (optional)			c_2
7	ENTER CARD 3 (A and B)		D			b_3
8	ENTER CARD 4 (A and B)		A			b_0
			B (optional)			b_1

SR-52 User Instructions

TITLE LINEAR REGRESSION PAGE 2 OF 2

STEP	PROCEDURE	ENTER	PRESS			DISPLAY
			C (optional)			b_2
9	ENTER CARD 5 (A and B)		E			a_4 (printed)
10	ENTER CARD 6 (A and B)		D			a_3 (printed)
11	ENTER CARD 7 (A and B)		C			a_2 (printed)
			RUN			a_1 (printed)
			RUN			a_0 (printed)
12	IF NECESSARY RECALL		RCL	0	0	a_0
	PARAMETERS		RCL	0	1	a_1
			RCL	0	2	a_2
			RCL	0	3	a_3
			RCL	0	4	a_4
13	ENTER CARD 8 (A and B)					
	COMPUTE THE VARIANCE OF		A			var(a_0)
	THE COEFFICIENTS		B			var(a_1)
			C			var(a_2)
			D			var(a_3)
			E			var(a_4)
			2ND	A'		R^2

PROGRAM 1

```
000 46
001 16
002 25
003 47
004 00
005 42
006 09
007 08
008 42
009 09
010 09
011 56
012 46
013 11
014 42
015 00
016 00
017 10
018 44
019 06
020 06
021 02
022 81
023 46
024 12
025 42
026 00
027 01
028 10
029 44
030 06
031 07
032 03
033 81
034 46
035 13
036 42
037 00
038 02
039 10
040 44
041 06
042 08
043 04
044 81
045 46
046 14
047 42
048 00
049 03
050 10
051 44
052 09
053 08
054 05
055 81
056 46
057 15
058 42
059 00
060 04
061 10
062 44
063 09
064 09
065 01
066 94
067 42
068 06
069 03
070 05
071 42
072 06
073 05
074 18
075 18
076 18
077 18
078 18
079 01
080 10
081 44
082 06
083 09
084 22
085 50
086 04
087 43
088 06
089 09
090 81
091 46
092 18
093 01
094 44
095 06
096 03
097 00
098 42
099 06
100 04
101 17
102 56
103 46
104 17
105 36
106 43
107 06
108 03
109 65
110 36
111 43
112 06
113 04
114 95
115 22
116 60
117 04
118 41
119 94
120 46
121 41
122 36
123 44
124 06
125 05
126 01
127 44
128 06
129 04
130 44
131 06
132 05
133 43
134 06
135 03
136 75
137 43
138 06
139 04
140 95
141 22
142 80
143 86
144 17
145 46
146 86
147 56
148 46
149 19
150 50
151 04
152 41
153 11
154 46
155 10
156 22
157 60
158 04
159 86
160 94
161 56
```

PROGRAM 2

000	46
001	11
002	43
003	06
004	09
005	65
006	43
007	00
008	05
009	75
010	43
011	06
012	06
013	40
014	95
015	42
016	09
017	07
018	65
019	53
020	43
021	06
022	09
023	65
024	43
025	00
026	07
027	75
028	43
029	06
030	07
031	40
032	54
033	75
034	53
035	43
036	06
037	09
038	65
039	43
040	00
041	06
042	75
043	43
044	06
045	06
046	65
047	43
048	06
049	07
050	54
051	42
052	09
053	06
054	40
055	95
056	20
057	70
058	97
059	65
060	53
061	43
062	09
063	07
064	65
065	53
066	43
067	06
068	09
069	65
070	43
071	00
072	09
073	75
074	43
075	06
076	07
077	65
078	43
079	06
080	08
081	54
082	75
083	53
084	43
085	09
086	06
087	65
088	53
089	43
090	06
091	09
092	65
093	43
094	00
095	08
096	75
097	43
098	06
099	06
100	65
101	43
102	06
103	08
104	54
105	42
106	09
107	05
108	95
109	42
110	00
111	02
112	65
113	43
114	09
115	06
116	94
117	85
118	43
119	09
120	05
121	95
122	55
123	43
124	09
125	07
126	95
127	42
128	00
129	01
130	65
131	43
132	06
133	06
134	94
135	85
136	43
137	06
138	08
139	75
140	43
141	00
142	02
143	65
144	43
145	06
146	07
147	95
148	55
149	43
150	06
151	09
152	95
153	42
154	00
155	00
156	56
157	46
158	97
159	00
160	56
161	46
162	12
163	43
164	00
165	01
166	56
167	46
168	13
169	43
170	00
171	02
172	56

PROGRAM 2

000	46
001	11
002	43
003	06
004	09
005	65
006	43
007	00
008	05
009	75
010	43
011	06
012	06
013	40
014	95
015	42
016	09
017	07
018	65
019	53
020	43
021	06
022	09
023	65
024	43
025	00
026	07
027	75
028	43
029	06
030	07
031	40
032	54
033	75
034	53
035	43
036	06
037	09
038	65
039	43
040	00
041	06
042	75
043	43
044	06
045	06
046	65
047	43
048	06
049	07
050	54
051	42
052	09
053	06
054	40
055	95
056	20
057	70
058	97
059	65
060	53
061	43
062	09
063	07
064	65
065	53
066	43
067	06
068	09
069	65
070	43
071	00
072	09
073	75
074	43
075	06
076	07
077	65
078	43
079	06
080	08
081	54
082	75
083	53
084	43
085	09
086	06
087	65
088	53
089	43
090	06
091	09
092	65
093	43
094	00
095	08
096	75
097	43
098	06
099	06
100	65
101	43
102	06
103	08
104	54
105	42
106	09
107	05
108	95
109	42
110	00
111	02
112	65
113	43
114	09
115	06
116	94
117	85
118	43
119	09
120	05
121	95
122	55
123	43
124	09
125	07
126	95
127	42
128	00
129	01
130	65
131	43
132	06
133	06
134	94
135	85
136	43
137	06
138	08
139	75
140	43
141	00
142	02
143	65
144	43
145	06
146	07
147	95
148	55
149	43
150	06
151	09
152	95
153	42
154	00
155	00
156	56
157	46
158	97
159	00
160	56
161	46
162	12
163	43
164	00
165	01
166	56
167	46
168	13
169	43
170	00
171	02
172	56

PROGRAM 3

```
000 46
001 14
002 43
003 01
004 03
005 75
006 43
007 00
008 00
009 65
010 43
011 09
012 08
013 75
014 43
015 00
016 01
017 65
018 43
019 01
020 01
021 75
022 43
023 00
024 02
025 65
026 43
027 01
028 02
029 95
030 55
031 53
032 43
033 01
034 00
035 85
036 43
037 00
038 00
039 40
040 65
041 43
042 06
043 09
044 85
045 43
046 00
047 01
048 40
049 65
050 43
051 00
052 05
053 85
054 43
055 00
056 02
057 40
058 65
059 43
060 00
061 07
062 85
063 02
064 65
065 53
066 43
067 00
068 00
069 65
070 53
071 43
072 00
073 01
074 65
075 43
076 06
077 06
078 85
079 43
080 00
081 02
082 65
083 43
084 06
085 07
086 75
087 43
088 06
089 08
090 54
091 85
092 43
093 00
094 01
095 65
096 53
097 43
098 00
099 02
100 65
101 43
102 00
103 06
104 75
105 43
106 00
107 08
108 54
109 75
110 43
111 00
112 02
113 65
114 43
115 00
116 09
117 54
118 54
119 95
120 42
121 00
122 03
123 81
```

PROGRAM 4

000	46	050	53	100	00	150	18
001	11	051	43	101	06	151	01
002	18	052	01	102	54	152	95
003	43	053	02	103	75	153	42
004	00	054	75	104	16	154	06
005	05	055	43	105	65	155	04
006	75	056	00	106	43	156	81
007	16	057	03	107	09	157	46
008	40	058	65	108	04	158	97
009	95	059	43	109	54	159	00
010	42	060	00	110	42	160	56
011	09	061	09	111	09	161	46
012	07	062	54	112	05	162	16
013	65	063	75	113	95	163	43
014	53	064	17	114	42	164	06
015	18	065	65	115	00	165	06
016	43	066	53	116	04	166	56
017	00	067	43	117	65	167	46
018	07	068	09	118	43	168	17
019	75	069	08	119	09	169	43
020	17	070	75	120	06	170	06
021	40	071	43	121	94	171	07
022	54	072	00	122	85	172	56
023	75	073	03	123	43	173	46
024	53	074	65	124	09	174	18
025	18	075	43	125	05	175	43
026	43	076	06	126	95	176	06
027	00	077	08	127	55	177	09
028	06	078	54	128	43	178	65
029	75	079	42	129	09	179	56
030	16	080	09	130	07	180	46
031	65	081	04	131	95	181	12
032	17	082	54	132	42	182	43
033	54	083	75	133	06	183	06
034	42	084	43	134	05	184	05
035	09	085	09	135	65	185	56
036	06	086	06	136	16	186	46
037	40	087	65	137	94	187	13
038	95	088	53	138	85	188	43
039	20	089	18	139	43	189	00
040	70	090	53	140	09	190	04
041	97	091	43	141	04	191	56
042	65	092	01	142	75		
043	53	093	01	143	43		
044	43	094	75	144	00		
045	09	095	43	145	04		
046	07	096	00	146	65		
047	65	097	03	147	17		
048	53	098	65	148	95		
049	18	099	43	149	55		

PROGRAM 5

```
000 46
001 15
002 43
003 06
004 04
005 42
006 09
007 06
008 43
009 06
010 05
011 42
012 09
013 07
014 43
015 01
016 08
017 75
018 16
019 65
020 43
021 09
022 09
023 75
024 17
025 65
026 43
027 01
028 05
029 75
030 18
031 65
032 43
033 01
034 06
035 75
036 19
037 65
038 43
039 01
040 07
041 95
042 55
043 53
044 43
045 01
046 04
047 85
048 43
049 06
050 09
051 65
052 16
053 40
054 85
055 17
056 40
057 65
058 43
059 00
060 05
061 85
062 18
063 40
064 65
065 43
066 00
067 07
068 85
069 19
070 40
071 65
072 43
073 01
074 00
075 75
076 53
077 16
078 65
079 53
080 43
081 09
082 08
083 75
084 17
085 65
086 43
087 06
088 06
089 42
090 09
091 05
092 75
093 18
094 65
095 43
096 06
097 07
098 75
099 19
100 65
101 43
102 06
103 08
104 54
105 85
106 17
107 65
108 53
109 43
110 01
111 01
112 75
113 18
114 65
115 43
116 00
117 06
118 75
119 19
120 65
121 43
122 00
123 08
124 54
125 85
126 18
127 65
128 53
129 43
130 01
131 02
132 75
133 19
134 65
135 43
136 00
137 09
138 54
139 85
140 19
141 65
142 43
143 01
144 03
145 54
146 65
147 02
148 95
149 42
150 00
151 04
152 43
153 09
154 05
155 42
156 06
157 06
158 43
159 00
160 04
161 98
162 81
163 46
164 16
165 43
166 09
167 06
168 56
169 46
170 17
171 43
172 09
173 07
174 56
175 46
176 18
177 43
178 00
179 04
180 56
181 46
182 19
183 43
184 00
185 03
186 56
```

PROGRAM 6

000	46
001	14
002	43
003	01
004	07
005	75
006	16
007	65
008	43
009	09
010	09
011	75
012	17
013	65
014	43
015	01
016	05
017	75
018	18
019	65
020	43
021	01
022	06
023	75
024	43
025	00
026	04
027	65
028	53
029	43
030	01
031	03
032	75
033	16
034	65
035	43
036	09
037	08
038	75
039	17
040	65
041	43
042	01
043	01
044	75
045	18
046	65
047	43
048	01
049	02
050	95
051	55
052	53
053	43
054	01
055	00
056	85
057	43
058	06
059	09
060	65
061	16
062	40
063	85
064	17
065	40
066	65
067	43
068	00
069	05
070	85
071	18
072	40
073	65
074	43
075	00
076	07
077	75
078	53
079	16
080	65
081	53
082	43
083	06
084	08
085	75
086	17
087	65
088	43
089	06
090	06
091	75
092	18
093	65
094	43
095	06
096	07
097	54
098	85
099	17
100	65
101	53
102	43
103	00
104	08
105	75
106	18
107	65
108	43
109	00
110	06
111	54
112	85
113	18
114	65
115	43
116	00
117	09
118	54
119	65
120	02
121	95
122	42
123	00
124	03
125	98
126	81
127	46
128	16
129	43
130	00
131	00
132	56
133	46
134	17
135	43
136	00
137	01
138	56
139	46
140	18
141	43
142	00
143	02
144	56

<u>PROGRAM 7</u>

000	46	041	97	082	54	123	98	164	00
001	13	042	65	083	42	124	81	165	00
002	18	043	53	084	00	125	65	166	98
003	43	044	43	085	00	126	43	167	56
004	00	045	09	086	54	127	09	168	46
005	05	046	07	087	75	128	06	169	97
006	75	047	65	088	43	129	94	170	00
007	16	048	53	089	09	130	85	171	56
008	40	049	18	090	06	131	43	172	46
009	95	050	53	091	65	132	00	173	16
010	42	051	43	092	53	133	01	174	43
011	09	052	01	093	18	134	95	175	06
012	07	053	06	094	53	135	55	176	06
013	65	054	75	095	43	136	43	177	56
014	53	055	10	096	01	137	09	178	46
015	18	056	43	097	05	138	07	179	17
016	43	057	01	098	75	139	95	180	43
017	00	058	02	099	10	140	42	181	06
018	07	059	75	100	43	141	00	182	07
019	75	060	19	101	01	142	01	183	56
020	17	061	43	102	01	143	98	184	46
021	40	062	00	103	75	144	81	185	18
022	54	063	09	104	19	145	65	186	43
023	75	064	54	105	43	146	16	187	06
024	53	065	75	106	00	147	94	188	09
025	18	066	17	107	08	148	85	189	65
026	43	067	65	108	54	149	43	190	56
027	00	068	53	109	75	150	00	191	46
028	06	069	43	110	16	151	00	192	19
029	75	070	09	111	65	152	75	193	43
030	16	071	09	112	43	153	43	194	00
031	65	072	75	113	00	154	00	195	03
032	17	073	10	114	00	155	02	196	65
033	54	074	43	115	54	156	65	197	56
034	42	075	09	116	42	157	17	198	46
035	09	076	08	117	00	158	95	199	10
036	06	077	75	118	01	159	55	200	43
037	40	078	19	119	95	160	18	201	00
038	95	079	43	120	42	161	01	202	04
039	20	080	06	121	00	162	95	203	65
040	70	081	08	122	02	163	42	204	56

BIBLIOGRAPHY

1- Atherton,T.J and Moshe Ben Akiva," Transferability and Updating of Disaggregate Travel Demand Models" TRB,January 1976.

2- Ben Akiva Moshe," The Logit Model And Functional Specifications of Alternative Models". MIT Ph.D thesis, June 1973.

3- Bourgin Christian,"L'Utilisation du Concept d'Elasticite dans les Evaluations de Clientele T.C." Note de Travail IRT September 1976.

4- Jacobson,Jesse ," A Case Study Comparison of Alternative Urban Travel Forecasting Methodologies" MIT M.S. thesis June 1977.

5- Koppelman,Franck,"Guidelines for Aggregate Travel Prediction Using Disaggregate Models" MIT Ph.D thesis, June 1975.

6- Highway Research Board,"special Report # 85 " 1965

7- Manheim, Marvin L,"Fudamentals of Transportation Analysis 1977 in Press.

8- Mc Fadden,Domencich,"Urban Travel Demand;A Behavioral Analysis" Amsterdam,The Netherlands,North Holland Press 1976.