A MULTIMODAL LOGIT MODEL OF MODAL SPLIT FOR A SHORT JOURNEY

A MULTIMODAL LOGIT MODEL OF MODAL SPLIT FOR A SHORT JOURNEY

BY

PAUL F. INGLIS, B. ENG.

A THESIS
SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF ENGINEERING

McMASTER UNIVERSITY

DECEMBER 1971

MASTER OF ENGINEERING (1971) McMASTER UNIVERSITY

(Civil Engineering) Hamilton, Ontario.

TITLE: A MULTIMODAL LOGIT MODEL OF MODAL SPLIT FOR A SHORT JOURNEY

AUTHOR: P. F. Inglis, B. Eng., McMaster University

SUPERVISOR: Dr. P. R. Stopher

NUMBER OF PAGES: vi, 122

SCOPE AND CONTENTS:

The logit format for a modal split model, which has previously been used for only binary cases, is used to build a new set of behavioural, probabilistic, multi-mode models. The models and the testing were carried out on a CDC 6400 Computer.

A program developed at Chicago was used to construct the models while a separate program was developed to analyze the results. The type and number of variables to be used in the different sections of the model were investigated and an attempt was made to find the best method of aggregation. An inferred 'value of time' was also calculated and statistical testing of the individual and aggregate models was made.

It is shown that this method of modelling is indeed feasible in terms of the significance of the models and the accuracy of the predictions on a separate data set.

ACKNOWLEDGMENTS

I would like to acknowledge the help, support and guidance given to me by my supervisor, Dr. Peter Stopher for the following research project.

I would also like to show appreciation to my friends and family for the encouragement which they freely gave.

I would also like to thank Mrs. Anne MacIntosh for typing the final copy of the thesis.

The thesis work was carried out with research monies from the National Research Council, and a Teaching Assistantship from the Civil Engineering Department of McMaster University.

TABLE OF CONTENTS

LIST OF TABLES

NOTATION

SYMBOL	MEANING
X_i	The ith value of an independent variable
X_{ki}	The ith value of the kth independent variable
$\sum_{i=1}^{n}$	Summation over all i
M	Number of alternatives
n	Sample size
N1, N2	Total number of variables
R	Total correlation
t	Student's t statistic
F	F statistic
α_k	Coefficient of the kth independent variable X_k
G( )	Function of
i,j,k,ℓ	Subscripts to denote typical observations
p, q	Probabilities
p_ℓ	Probabilitv of the ℓth alternative
e	Exponential base
A, B	Subscripts denoting particular alternatives

CHAPTER I

INTRODUCTION

I-1 Background

As the standard of living in the world, and particularly in North America, has increased, an awareness of problems within our environment has developed. Since the large proportion of the population are city-dwellers, this has been reflected in a desire for improved conditions in our large urban centres.

Cities are still the most efficient form of habitation and must remain in basically the same form for many years to come. Many cultural and economic aspects of modern living are dependent upon a certain size of community for their sustenance. The post-war years have seen a higher reliance upon personal transportation which has afforded greater mobility and has led to the resultant urban sprawl.

It is becoming more necessary that the growth of the automobile population of our city centres must be checked before the transportation arteries either strangle the city by clogging the streets or demand a too large share of the core land. A trend towards more efficient mass transit systems is evident.

In Canada, both the federal and provincial governments have just recently become aware of these trends. The Ottawa government has set up the Canadian Transportation Commission to control all facets of transportation. The Ontario government has halted work on a major freeway in Toronto by withholding financial support. At the same time the former Department of Highways has been reorganized as the Department of Transportation and Communication and a provincial subsidy of fifty per cent on all costs for rapid transit systems is now available.

New technology will provide our society with still further concepts which are feasible for travel within our large urban centres. These will most likely be compromises between the absolute privacy and convenience of the private auto and the less convenient but more efficient mass transit systems.

We must not forget also that man evolved with a built-in means of mobility. In the "New Towns" in both North American and in Europe, a greater emphasis has been placed on pedestrian accessibility. See Buchanan, 1963.

It may be seen then that several different means of travel will be available to the urban population in the future. Even now with present technology, a city like London, England has at least nine distinct means of transport including two types of rail service and two types of bus service. The smallest city can easily count on taxi, bus or private auto for transportation.

I-2 Present Transportation Planning

Transportation planning has developed to a point at which travel demand is conventionally predicted using a sequence of four steps: trip generation, trip distribution, modal split and network assignment. These four steps are each taken using a separate model for each process. The first step, trip generation, predicts the absolute number of trip-starts and trip-ends which will occur in given urban areas. The second model, trip distribution, apportions the given trips to specific generating and absorbing areas creating a corridor flow pattern. Mode split models determine the share of trips which will be made on each separate mode of travel available. The last model, trip assignment, then uses the limitations of the plant within the corridor to allot the vehicles to specific thoroughfares. See Davis, 1969, or Martin, Memmott, Bone, 1961.

Since the building of new models for each separate city is an inefficient use of time, a definite and standard "Urban Transportation Planning" (UTP) Package is desirable. None of the existing models are entirely satisfactory. To remedy this situation we may start by taking one of the present models and reformulating it to make it more satisfactory. This may be used as a starting point to develop a whole new UTP Package.

The modal split model has the best defined bounds with a most definite and measureable result. It then provides a reasonable starting point upon which other models can be based and to which they may be co-ordinated. Most models have so far used an aggregate approach. The present work is an attempt to follow a path only recently opened towards disaggregate, behavioural models.

The idea is to use behavioural theory and the individual's traits as well as the characteristics of the systems involved, as an indication of his probability of choice of mode. The disaggregate approach is a change from the initial theory in modal split which involved aggregate statistics. Using the same type of zonal division barriers the aggregate models produced absolute numbers or percentages of total travel for each mode between the zonal pairs. This method was highly inaccurate because of the variation in trip lengths and characteristics of the individuals within the zone. The results were limited by zoning and geographical considerations of the urban geometry, to each unique study area. A new model was required for each city. The disaggregate approach should have a more universal application, since it is based on non-local variables.

The identification of the correct variables to be employed is still very much in the embryo stage. The disaggregate technique tries to pattern the choice through theoretical behavioural prejudices of the individual trip-maker evolving from most socio-economic conditions such as his sex, income, age and stage in the family life-cycle.

The form which the 'ultimate model' will eventually have is also an unsolved problem. As pointed out above, this choice has become more varied, as new facets of transportation have been introduced. One school of thought says that the individual has a binary choice no matter how many choices he has in the absolute sense. This creates a problem for the modeller who must determine the two most appropriate choices for the users. This thesis tries to show that a multi-dimensional format can be used to predict the probabilistic choice. It hopes to show that a simplification can be realistically accomplished by building a multi-mode dis-aggregate stochastic type model.

The actual model form is an extension of an existing binary form which has been previously proven viable. See Stopher, 1969. The multinomial extension is easily achieved mathematically, Theil, 1969. The set of variables to be used is also of the same pattern as that used in the previous binary models. The linear relationship which is a part of the technique also allows us to develop an understanding of the individuals' comparitive attitudes towards the separate variables and choices involved.

The thesis tries to discover whether a multinomial logit formulation of the modal split model is feasible. Chapter II reviews past work in modelling modal split. Chapter III outlines the theory and strategy. Chapter IV gives the results of the empirical testing of the model and

Chapter V points out the conclusions to be reached and the future directions of research. In the Appendix there are listings of the computer programs used and the data available.

CHAPTER II

IN REVIEW OF THE STATE OF THE ART

Studies to improve the flow and direction of traffic volumes have been made for many years, starting before the turn of the century. As early as 1844, traffic counts were being made in France. Yet it was not until Federal legislation in the United States in 1944, that transportation planning in the form of Origin-Destination Studies evolved to a recognizable form. Before 1955, these studies concentrated on an extrapolation of present trends. Modern analytic predictive planning, then, has only a relatively short life of less than 20 years up to this point. Even so, institutional research into this field and dollar volume to consultants has mushroomed. See Oi and Shuldiner, 1969.

This thesis concerns itself with only one part of the 'Urban Transportation Planning' (UTP) Package; the choice of mode for a relatively short journey. The UTP package comprises four models, trip generation, trip distribution, modal split and network assignment. The relative placing of trip distribution and modal split in the sequence are not rigidly set. Both Davis, 1969, and Martin, Memmott, Bone, 1961, discuss the UTP package in full. One sequence prescribes modal split as the second factor and trip distribution as the following model, while the other reverses this alignment.

In the late fifties major transportation studies were first commissioned. Each of these studies was required to produce its own modal split model. Early research used zonal aggregate values in multiple linear regression formats. A large emphasis was placed on urban land use and zoning as reflected in early work by Wynn, 1955, Carroll, 1954 and Adams, 1959. This was followed by Chicago Area Transportation Study Reports by Howe, 1958, Biciunnas, 1964, and Sharkey, 1958, 1959, and a Milwaukee Study Report by Hadden, 1962. These models used only socio-economic measures and activity levels derived from urban zonal theory. Since these measures - age, income, car ownership, residential density, etc. - were the only variables, any changes in the systems were not reflected in the model. Also, since these measures all suffered an inflationary trend as the standard of living increased, an ever increasing share was predicted for the car over transit. No significant change in mathematical form occurred in any of these works.

By using high levels of aggregation there was a very high variance within the zones, especially when short trips were being considered. A trip from zone A to adjoining zone B could vary from several blocks to more than a mile. A generalized zonal activity could over-ride small pockets of different types. Since the techniques of these studies also used socio-economic factors such as income, car ownership, level of education, the within zone variation was further generalized and thus distorted. This was perhaps the largest source of error, for even on a given street in an urban area, car ownership could vary from 0-3 cars per family or education from public school to post-graduate level. These high variance levels, together with errors in generating the co-efficients and errors inherent within random sampling surveys, were enough to make these models unreliable for the present, let alone for use in predictions of the future, for which there are additional errors generated.

In the late fifties, a large amount of data was collected in conjunction with the development of a new series of models. The pendulum swung away from the socio-economic variable to the system variable, although the level of aggregation still remained on a zonal basis. Large studies in Washington, Chicago, San Francisco, Toronto, and Philadelphia resulted in the definition of a new set of diversion curves for modal split prediction. These models worked on either time saved or time ratio vs. per cent of total trips diverted from one mode to the other. The techniques used are well documented in papers by Quinby, 1961, Hamburg and Guinn, 1966, and Hill and Von Cube, 1963. Quinby recognized that the multiple regression techniques used up to this time might not be the best-fit solution.

He proposed that Pearl-Reed logistic curves be fitted, but eventually settled on a Gompertz exponential curve formulation. His curves, however, took into account only travel time ratio. Hamburg and Guinn extended this into another dimension with some research into 'transit response surfaces'. Hill and Von Cube did a more extensive study on the effects of many different variables, both system and user types. This work led to a large set of diversion curves developed by Traffic Research Corporation which set up relationships between travel time ratios and per cent on transit for different cost ratios and income levels, which was used in Washington, Toronto, and Philadelphia and duly documented by Deen, Mertz, and Irwin, 1963. Similar studies were made by Moskowitz, 1956, in California using diversion curves for an analogous route-choice problem involving a free-way-toll situation. One gross problem involved projections into the future. The portions of the curves of highest importance in the predictive sense were also the areas of greatest uncertainty. The diversion curves also profess a geographical constancy, that is, the use of results from one city upon diversion curves for prediction in another. This probably has only limited regional validity. Most of the problems of accuracy and large confidence limits were still not solved by this second generation of models. The aggregate deterministic technique was suspect as a predictive tool.

The first attempts at solving the many problems of the early deterministic approach were such as McClynn and Watkins, 1965, who decided that feedback and interchange of results were required between the separate steps of the UTP package. They tried to combine trip generation and mode choice, feeling that the latter step had some definite effect on the decision to make a trip. Reichman and Stopher, 1971, point out one major flaw, saying that in order to operationalize this model, trip distribution also must be included, so that the specific system characteristics operating in a given direction can be described. This of course adds to the complexity of the model. Charles Rivers Associates, 1968, further feel that statistical validity for these models is very low. They feel that the number of residual errors arising from oversimplifying this decision into one model combined with all other sampling and predictive errors give a too high uncertainty for the model to be of practical use.

The third wave of opinion is still building. Research is not complete nor has any major study found an optimum way of using the newest techniques. Errors in earlier models were as much as 300 per cent. Much of the error in prediction was attributable to the level of aggregation at which the models were built. By working at the zonal level, the large variance within the zonal populations could not be accounted for. By reducing to the basic component, the individual user, this problem can be eradicated. The resulting model could then be aggregated to any level desired with only the relatively small variance error of the individual carried into the final prediction. Early results have tended to indicate that there is some future for these types of models. Modal split provides an excellent starting point to redevelop the entire UTP package. This model is most easily developed and data are very easily gathered. The range of variables can be more easily defined and a definite result can be expected within definable bounds. From this model a complete set of UTP models can be integrated using the same mix of user and system characteristics.

This concept borrows from individual behaviour theory. It takes the relative desires and preferences of the individual user towards the attributes of each mode in order to predict a probability for his use of each individual mode. This then, is an attempt to present the modal choice as the market-place decision process that it truly is, and indeed many of the statistical techniques so far employed have had a basis in economic theory previous to being adopted into the transportation field. See McClynn, Goldman, Meyer, Watkins, 1967.

The resulting experimentation was involved with studies into the form that the model should take. The first stochastic models developed were by Warner, 1962, in Chicago.

Quarmby, 1967, in Management Science, and Lisco, 1967, an economist, followed with further studies. Currently three separate classes of models exist, depending upon the form of the mathematics; discriminant analysis; probit analysis; logit analysis. The use of linear regression has largely been discarded because of the invalidity once it passes probabilities of one or has negative probabilities.

The discriminant function was used first to tell the difference between different strains of plant life and in work on taxonomic problems. See Fisher, 1936. The theory is based on the existence of overlapping normal sub-populations which are distinct in the decision sense - either by strain or in this case, choice of mode. The analysis tries to pinpoint attributes of both populations which can account for the difference in choice and develop a function which 'discriminates' between the two populations, by minimizing the number which are misclassified by the model. In the binary case, it sets a limit for the discriminant function below which the member is classified in group I. It sets a second limit above which the member is classified in group II. The area between these two limits is a probabilistic area for which a secondary logit-type exponential description is provided. Mongini, 1965, was the first to apply this theory to modal split for intercity trips in the "Northeast Corridor Project" of the Atlantic seaboard. Quarmby, 1967, also uses discriminant functions employing differences for his time and cost variables, while

McGillivray, 1969, uses ratios for systems characteristics. Both models incorporate user characteristics as well but the set of variables and their form is still a subject for much debate and research amongst all the current model developers.

Several researchers have tried to put the problem of significant descriptive variables into a concise form, Paine, 1962, at the University of Maryland, Bock, 1968, for the Highway Research Board, and John and Claudia Betak, 1969, at Northwestern University. The Betak paper presents an overview of all thought on both systems and user variables as well as an exhaustive bibliography of the modal split literature. The papers do not agree on any concrete conclusions for the form which any of the variables should take, or indeed the exact number which should enter. This is a problem which is common to all forms of the behavioural disaggregate models.

The form of the model which incorporates Probit analysis was first suggested by Warner, 1967, for use in modal split, who rejected it as computationally too complex. Lisco, 1967, was the first to use this method successfully in his studies on the 'value of time', a result derived from the cost and time coefficients of the model. Since he was most interested in this secondary result, it was left up to Lave, 1968, at Stanford to build the first true modal split models using this mathematical form.

Probit analysis assumes that the frequency of choice is normally distributed with respect to the function of user and system characteristics. Thus by using a normal probability distribution function and different weights for the characteristics in the form of coefficients, the actual split can be determined. Any user with probability of .5 or greater for a given mode accepts that choice. This is a technique which also has derivative roots outside the Transportation field, having been developed for use in toxicology by Finney, 1964.

Logit analysis was developed into its present form by H. Theil, 1969. Stopher, 1969, has built his models using this technique at Northwestern. Theil has also developed the logit theory to the multinomial case. Rassam, Ellis and Bennett, 1970, have been the only ones to use this concept in modal split so far, but have confined themselves to an aggregate form, rather than the disaggregate form, which seems to hold more promise, and which is the subject of this thesis.

CHAPTER III

THEORY AND DATA

In almost all cases of model building, the builder has restricted the model form to two dimensions. He has assumed that the choice has been made between only two modes for the model. This has been found to be inadequate, and so attempts have been made to build a series of models which break down the choice into binary choice steps. This is conceptually and behaviourally inadequate. This method also tends to accumulate high error terms. A multinomial approach would seem to be an improvement. This Chapter explores the background for the models in terms of:

(1) Theory of multinomial models

(2) Data base and variables choice

(3) Statistics used to measure the value of the models

III-1 Theory

Binary models have been successfully built using the method explained below. The question is, "Is the multinomial extension viable?"

Henri Theil, 1967, first developed the mathematical theory for the model. What follows is a simple explanation of the mathematical workings, as modified by Stopher, 1969.

A complete development of the model mathematics is found in Appendix II. The following is a brief outline.

For a binary choice, the model takes the form III-1-1

$$p = \frac{e^{G(x)}}{1 + e^{G(x)}}, \quad q = 1 - p = \frac{1}{1 + e^{G(x)}} \qquad \text{III-1-1}$$

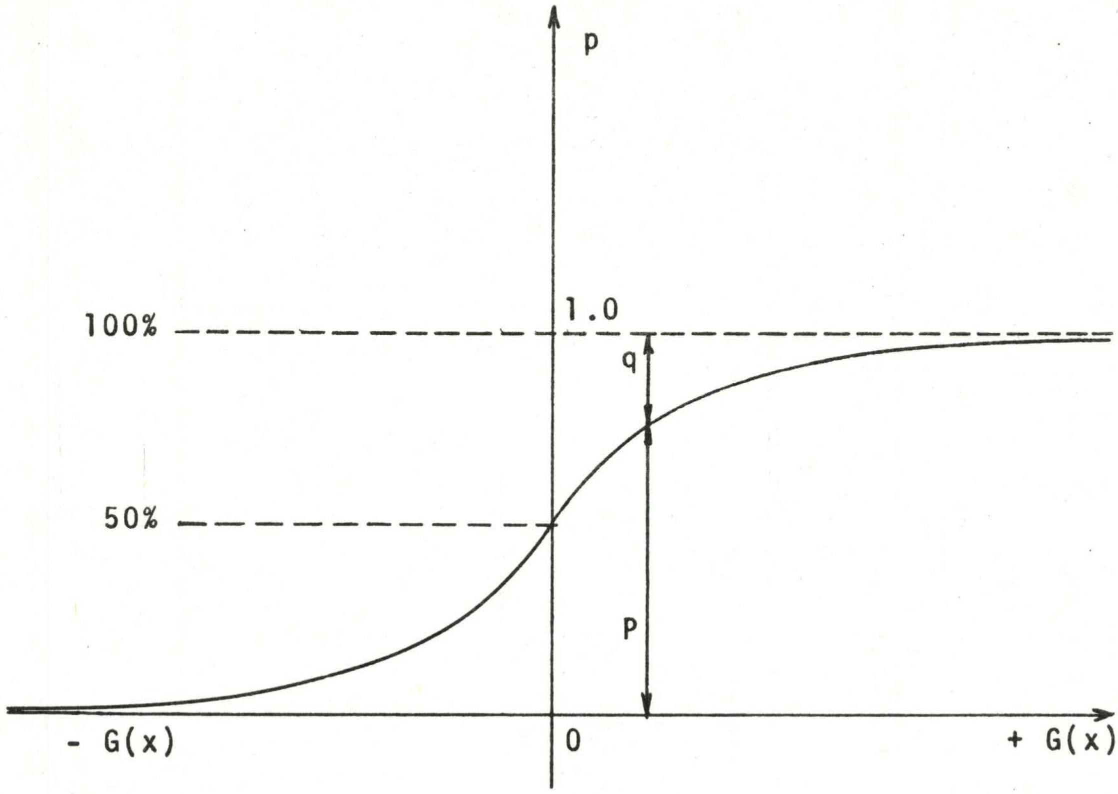
A linear relation is hypothesized for G(x), so that $G(x) = \text{constant} + \Sigma\, \alpha_k X_k$. The X_k are system and user variables of the individual.

From the binary, it is possible to propose the multi-choice form of the model as in III-1-2.

$$p_\ell = \frac{e^{G_\ell(x)}}{\sum_{\ell=1}^{M} e^{G_\ell(x)}} \qquad \text{III-1-2}$$

A pictorial representation is given in Figure III-1 below of a binary choice situation.

The major problem given the theory above is the estimation of the coefficients in each equation. To do this a maximum likelihood estimator program developed by John Cragg was used. Previous to this application, the program had been used only sparingly in the field of economics.



NOTE: Given a G(x) for a particular observation, p and q are the resulting probabilities.

TYPICAL 2-DIMENSIONAL LOGIT CURVE

FIGURE III-1

III-2 Data Base and Variables

The data used to derive and test the models have been titled the "Suburban Station Access" data. They were collected from people using the Chicago Transit Authority suburban routes from the Northwest corridor out of Chicago. The trips that were modelled were not to the Centre of the city but rather the shorter trip between home and the commuter station.

Ten different stations were involved as destination points, but only one area (211th Street) had a situation in which all four modes were not available to each user. The four modes involved were: a) Walking, b) Drive and park, c) Driven (Kiss and Ride) and, d) Bus. Bus was not available for these people. A final set of 117 observations was used to build the models from the area of Skokie, Illinois, since it had a good balance of each mode-type user. A second set was used in evaluating the models. This set had 400 observations and was labelled the Northwest Corridor Study Area. This second set did not have this same balance but this can be interpreted as a desirable feature for, if we hope that the models will have some validity in being transferred from one area to another, then obviously it is better to test the models with a set that has different characteristics than the original data set.

The information obtained from this study was placed on computer cards according to the card format also found in the appendices. In addition, a number of these absolute terms were regenerated into dummy variables on a third card. Those involved were distance - 1 variable taking the value 0 if less than .5 miles, 1 otherwise-, income- four dummy variables-, and age- three dummy variables. All the data used were complete except for the Pavement variable, which tried to evaluate the condition of the road surface for the trip (wet or dry). This variable was collected for the

Northwest Corridor only and was not used in the model at all.

Probably the most important task for the model builder is the choice of variables to be used in the models. Below, is a discussion of each of the variables which were tried and whether they eventually proved significant or not.

1. Cost - This variable has surprisingly not often proven very important in explaining the choice process. It is the measure of almost every other good in our society, and as such, it is logical that it be the measure of this commodity as well. The form of cost that is used is another question, however. In the choice process, the real cost of transportation is seldom the cost which the user considers for his decision. Rather, he uses the cost which he perceives. In the case of bus and walking his out of pocket costs are going to be quite accurate, not taking into account the hidden costs involved in subsidies through taxes. On the other hand however, the real costs of operating a personal means of transport is probably underestimated to a high degree, since usually only out of pocket gas, parking and toll costs are taken into account. It is not a universal opinion that this variable is of great import. Betak feels that because there is no true substitutability between transportation modes, costs do not substantially affect mode choice. This may be true of the higher income bracket user but should not be true of the low income user. This correlation between income and costs has been noted by others before, but the scope of this work does not allow for experimentation

into this matter.

2. Time - This is another of the important system characteristics. Again it is not a clear cut variable since the effects of different segments of the total time seem to be more important in the choice process than others. The decision criterion does not seem to be based on the absolute time itself but rather on the activity connected with that time. Thus by segmenting the time, this difference in attitude between separate activities will be reflected as a different coefficient. This would not necessarily reflect a change in the value of time saved but rather would be a measure of the inconvenience associated with that activity. The type of activity inferred would be such as waiting time, transfer time, or time spent in getting to a bus stop.

A corollary to these first two variables is the inference of a "Value of Time". That is, since time and cost are entered into the equation linearly with respect to each other, it is possible to calculate a 'value of time' by finding the ratio of the coefficients of time to that of cost. This has been done previously by Lisco with binary probit models, but has not been attempted in the multi-mode case. This could possibly raise some interesting points concerning the change from mode to mode of this value, as well as in comparison with the values found in the binary case.

3. Rush - This is another of the system characteristics. (Rush is a dummy variable which has a value 1 if the trip is taken between 6:30 and 8:30 a.m. or 0 if at any other time.) It was entered linearly, but there would seem to be a good argument for using this dummy-type variable to stratify the model into two separate entities. Within this program however, this would add too many variables (double the number) to keep the calculation within the bounds of the program. For a stratification, two separate models would be built with this variable determining which would be used. Used as it was, it served to change the effect of each variable by the same ratio, instead of the change being calculated uniquely for all variables. That is, there is an assumption that all of the variables change in the same way between the two conditions. Although the effect entered was not optimal, it proved to be significant.

4. Other System Characteristics - The data source reported only those system variables mentioned above, so that the only way to gain an insight into their effect is collectively through the difference in time and cost coefficients as mentioned above.

5. Age - Age was entered in two forms. The first way was as a linear variable, where the actual value was used. This is conceptually, as well as computationally, less accurate, since it can not be expected that the effect upon choice at say age 50 would be twice as much as at age 25. As a partial solution, this variable was transgenerated into three dummy

variables. The dummy variables were 0,1 variables with a maximum of one variable having the value 1. This allowed for division into four groups; those younger than 25; those between 26 and 45; those betwen 46 and 65; those older than 65. These seem to be reasonable divisions for age as far as life-cycle is concerned; the rebellious youth; the young family; the middle age; the retirement age. The first group would be expected to walk and take the bus, since they would be less able to afford the personal transportation and quite capable of walking to save the bus fare. The second group would probably be more able to afford the costs of a car and therefore be less likely to take the bus or walk. The third group would be most able to afford the luxury of an auto since costs would be of least importance to these people. The people of retirement age would be most likely to take the bus since it is inexpensive, and least likely to take to the sidewalks, due to their advanced age.

6. <u>Income</u> - Income was treated in the same way as age. A direct correlation should not be expected between a variable such as this and choice, so that a set of four dummy variables were put forward. The groups were: less than $5 thousand; $5 - $8 thousand; $8 - 12 thousand; $12 - $17 thousand; $17 - $25 thousand; greater than 25 thousand. Obviously then, this survey does not reflect the cross-section of an ordinary city but the upper-middle class fringe regions. It is not then going to be satisfactory in the centre-city region, where the distribution of income will be more in the lower ranges

than this data set. This, of course, does not affect the validity of the model, only its application to a different type of trip-maker.

Many researchers have used income as a combining variable, especially with items of cost, for example, De Donnea, 1970. Again, time limitations prevent this idea from being tested with respect to the multi-mode case.

7. Auto Ownership - This variable should be of extreme importance for obviously without a vehicle available it is impossible to drive and park at the station. Likewise, after having made the investment in a personal vehicle, there is a desire to have a high utilization and hence the greater probability of driving or being driven. This variable could likely be more descriptive as a dummy-type variable, since the degree of utilization decreases as the number of cars available increases. This avenue was not explored however.

8. Sex - This variable is in the form of a dummy 0,1 variable -0 for male -1 for female. The difference in attitudes between the two sexes will be reflected by the coefficient of this variable. It would seem that the female should be less likely to regard the car as a status symbol and hence would have a greater probability of taking the bus or of walking.

9. Stage in the Family Life-cycle - This variable is in the form of three dummy variables. The first has a value of 1 if the user is unmarried and living at home of parent, otherwise it is 0. The second has a value of 1 if he is unmarried and not living at his parent's home or is married with a spouse who does not drive or with a spouse who drives to work with him. Otherwise the value is 0. The final dummy is 1 if the user is married and has a spouse who drives to work independently. All other categories have all three dummy variables with a value of 0.

The difference between these categories would seem to be the availability of the auto in the family. A second criterion of separation would be the achievement of status. These groupings also reflect distinct income levels as well and may be highly correlated to the income variable. At any rate the first group would be expected to use the bus more often because of a lower car availability and a lower desire for status. The second group would be more likely to have a car available, hence would be most likely to use it in the journey to work. If the user falls into the third category there is going to be competition for an available means of transport and hence one of the competitors is likely to take the bus or to walk. This is valid if only one car is available or the ratio of cars/drivers is less than 1. Hence there should be some inter-relation between this variable and the ownership variable.

10. Trip Purpose - This variable could be used to describe the different economic demand generated by the separate trip purposes. This would involve stratifying and thus complicating the model, making it computationally impossible for the program. Thus we assumed a constant demand function and trip purpose was rejected as a variable.

11. Other User Characteristics - No other variables were available from the data. Other relevant indicators might be race, housing type, occupation, education or religion.

III-3 Statistics

In order to evaluate the models, there must be some criterion of their worth with respect to other models which have been built. To fulfill this requirement, statistical theory has developed many tests of significance which can be used as a measure of comparison and also as a measure of worth in their own right. Many of these values were incorporated in the original program by Cragg. These statistics have been supplemented by the use of a second program, STAT3, developed by the author for this research. A listing of the program is to be found in the appendices.

The Theil program developed by J. Cragg, uses the maximum likelihood estimates of the coefficients and combines these with standard errors which it also calculates and uses to calculate student's t-values for each of the estimates. These figures can be used to assess the significance of the individual coefficients in the equation. They are the sole basis for rejection of - or acceptance of - the variables in the model.

Besides these values, there are two aggregate statistics for the model, the first is the value for the "Likelihood Ratio Test" with its accompanying degrees of freedom. The second and probably most useful for comparison purposes within the research is the "Proportionate Pseudo R-square". It is not clear however whether or not it is a valid approximation of the R squared values commonly found in statistical literature.

In addition to tests of the model there is also a "Variance-Co-variance Matrix" given to allow some analysis of the relation between variables. The diagonal of the matrix is the variance and the off-diagonal entries are the covariances. This allows an investigation of a possible duplication of worth for a variable. Any non-zero value of the covariance is evidence of non-independence between variables. Thus a large covariance between two variables would indicate that one of them is superfluous to the equation. The approximations made in the program have resulted in imperfections in the accuracy of these values as indicated by differences in covariances between the same two variables, when they have been entered into all three equations of a four choice model. The order of the magnitude of those values is, however, extremely useful.

To supplement the above, the STAT3 program was developed. It was designed to assess the models with respect to a second data set. This set was considerably larger than the set used in deriving the models to show the effects of aggregation upon the accuracy.

This program also has statistics related to the individual variables. There is a complete "Correlation Ratio Matrix" and a complete "Simple F-scores Matrix" for this purpose. The former is used to determine the degree to which a relationship exists between variables. This is useful since the variables are combined linearly in the formulation. This is not of course any guarantee of a causal relationship since some third common factor could be involved in a causal relation with both. Rounding errors in the program computations cause a correlation of income with itself in two different equations to give the result .9972 instead of 1.000. The relative values and their order of magnitude can still be considered as highly indicative of relations between the variables. The latter matrix contains values for the simple F-scores. This statistic is in common use, and significance tables are easily located.

The program also has two aggregate measures of the model. The "Multiple Correlation Ratio" and "Multiple F-Value" both test the overall model significance in describing the choice process. The program further determines the number individually predicted from the model by assuming the maximum probability mode to be the one chosen. It also sums the total probabilities for each mode and determines the average probability given by the aggregate figures. All of these are then compared to the values from the data source. It is to be assumed that the summing of the probabilities will result in the superior prediction since the data focuses on only one day, and probably the situation would change from day-to-day for the individual while maintaining some overall consistency. To illustrate, assuming that an individual had probabilities of .2, .3, .4, and .1 for a four mode choice, it is likely that on the survey date he chose the third mode yet he 'probably' chooses some other mode 60 per cent of the time. Thus by summing the probabilities we attain a more illustrative explanation of the causal relationships between the variables and the choice.

The use of the second data set in the analysis allowed for an examination of the models for a larger data set. Since there are several means of aggregating as discussed above, it allows for an analysis of the methods on an unbiased set. The use of the larger set can perhaps give an indication of whether the method of choosing the highest probability for the individual is superior or inferior to the method of summing probabilities over the whole range. With a larger set, there is a greater opportunity for a sure evaluation. It can serve as a definite indicator of the effects of aggregation. One of the deficiencies inherent in these data is the salary range within which the people interviewed fall. The Skokie set has an average wage of approximately $8 thousand, while the Northwest Corridor data has an average wage of $10 thousand. These values are quite above the national average if all classes are taken into account. The former group could be classed as lower middle class while the latter could be classed as upper middle class. This is to be expected of people living in the suburbs of a large city, since a considerable income would be required to allow for such a considerable expenditure on transportation.

Other deficiencies in the data set are readily apparent. The trip under study is itself only a part of a much longer trip, so that attitudes with respect to this trip may differ drastically from attitudes for a trip to a more local destination of the same distance. It may be more appropriate to consider the whole trip instead of this one fragment. This is especially true since the length of the trip on the CTA is not constant across the data. The sociological information available here is not adequate. Other factors which would be perhaps significant have been mentioned above. An attempt should also be made to derive measures of system characteristics which were also not available. Attributes which could be studied would include safety, convenience, personal privacy and comfort, flexibility in destination and timing, status, perhaps in the form of age and appearance of the vehicles, or reliability. Not every member of the original data was complete. As a result a great many of the observations were useless in the calculations. Of course in all cases not all modes were available. It should be possible to incorporate this problem into the model since the probabilities are calculated separately. For this reason models were developed using a varying number of modes. In this way the change in coefficients could be observed as the mode-availability changed. One drawback to taking this into account is the increase in complexity which is introduced into the model by this secondary structuring.

Several advantages are attributable to this technique, particularly in terms of theoretical assumptions. There is no assumption of normality to be met in the G(x) function resulting in a more general model. Also the proposed use of a probabilistic sum for aggregation gives a better conceptual idea of the true process that occurs in this mechanism. That is, instead of summing individual absolute choices, which may vary from day to day, the individual probabilities are summed to give totals. This better describes the behavioural process since we are dealing with human beings who can and will change their minds in this non-exact manner.

However, in order to maintain the conceptual basis of the model, it is necessary to constrain the coefficients of system dependent variables which appear in more than one sector. That is not to say that a freeing of the values cannot be a good fitting technique, rather, on the contrary, it may prove to give a better fit since it is less restrictive. All user characteristics are mode independent and may be left free if appearing in more than one sector.

CHAPTER IV

RESULTS

This Chapter presents the most important results of the research work done for the project. There are two view-points from which to work, both of which are discussed. As explained in Appendix II, the conceptual basis of the model depends upon the use of a constant coefficient for each separate system dependent variable in all sectors of the model, each mode defining one sector. The coefficients of time and cost must then be the same in each sector of the model. It may be assumed that there is some change in attitude from mode to mode. This would be reflected in different coefficients in each sector. There is no simple solution for this mathematically. Thus, if we allow the coefficients to take different values independently, it must be considered only as fitting technique and can thus have little analytic value.

The Chapter will show best-fit models of both of these two positions. From this basis, a comparison can be made of the two different techniques using the STAT3 program. Another set of models are presented involving time and cost only in the functions. In addition, there are three choice models using the best-fit variables. All of these models have STAT3 analyses accompanying them.

A secondary result of keeping the time and cost coefficients constant, is the possibility of deriving the 'value of time' calculations of Lisco and Stopher and also it is possible then to compare the models derived here with those of these authors.

In order to understand the effect of a change in the variable magnitude, a sensitivity test is done on the best-fit model. This also affords a better idea of how the model will work when it is operationalized.

IV-I <u>Best-Fit Models</u>

Table 1 and Table 2 below contain the best-fit model using the restriction for the time and cost coefficients. In the tables there are only three choice equations. This is because the fourth choice probability is defined by the limitation that all the probabilities sum to unity. The effect of the time and the cost of the base mode are included by the use of difference formulations for these variables in the other sections of the model.

In Table 2, "The counts according to the model" are found by finding for each individual, which mode has the highest probability of being chosen and assigning that individual to that mode. "The true counts" represent the sums of the choices actually made by the members. "The counts by summing probabilities" are found by summing all of the individual probabilities for each mode as calculated from the model.

"The model mean probabilities" are the previous values divided by the number in the population. Similarily, "The mean true probabilities" are the true counts divided by the number of observations. "The number correctly predicted" counts the number for each mode in which the model had the highest probability for the mode which was actually chosen.

To help understand both the signs and the magnitudes of the coefficients, a sensitivity test of this model follows. The test presents four different individuals each of which takes one of the modes as his choice in the sampling. This should also show and explain more completely how the model works.

The degrees of freedom are calculated for the multiple statistics and quoted on the table. The first value quoted is the number of arrays minus one. The number of arrays determine the intervals in the technique used. In the case of STAT3 this value is constant at 10. The second degree of freedom is the number of observations minus the number of arrays, again for the testing of this data set a constant at 399 - 10 = 389.

TABLE 1

BEST-FIT MODEL FOR FOUR CHOICE WITH RESTRICTED COEFFICIENTS

VARIABLE	COEFFICIENT	t-VALUE	SIGNIFICANCE LEVEL
<u>Drive Mode</u>			
Cost	-.11397	5.0254	0.9995
Time Difference	-.00421337	5.1352	0.9995
Frict	.23824	3.2905	0.995
Walking Time PL-Stn.	-.0123129	2.1434	0.975
Car Ownership	1.4924	3.3222	0.995
Age	.0107903	1.5720	0.900
Income Dummy 1	-5.5707	2.2152	0.975
Income Dummy 2	-7.4174	3.2807	0.990
Income Dummy 3	-7.9819	3.8004	0.999
Income Dummy 4	-7.0880	3.5293	0.995
<u>Driven Mode</u>			
Cost	- .11397	5.0254	0.9995
Time Difference	- .00421337	5.1352	0.9995
Rush Dummy	-2.2600	3.4537	0.995
Income Dummy 4	1.1685	1.9737	0.950
<u>Bus Mode</u>			
Cost	- .11397	5.0254	0.9995
Time Difference	- .00421337	5.1352	0.9995
Rush Dummy	-1.3456	1.8551	0.950
Constant	2.0248		

Proportionate Pseudo R-squared .6862

Likelihood Ratio Test 120.062 with 15 degrees of freedom

TABLE 2

BEST-FIT MODEL FOR FOUR CHOICE WITH RESTRICTED COEFFICIENTS

MULTIPLE F-VALUE = 6.954123 (3.38 at .999 LEVEL)

MULTIPLE CORRELATION RATIO = .138594

WITH 9 AND 389 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	DRIVEN	BUS
31	111	181	76

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	DRIVEN	BUS
54	117	187	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
44.9	86.5	158.9	108.7

THE MEAN TRUE PROBABILITIES ARE

OBS = 399

WALK	DRIVE	DRIVEN	BUS
.135338	.293233	.468672	.102757

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
.112412	.216785	.398361	.272442

THE NUMBER CORRECTLY PREDICTED IS	16	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	28	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	75	IN MODE	DRIVEN
THE NUMBER CORRECTLY PREDICTED IS	19	IN MODE	BUS
TOTAL	138		

The first individual chose to walk to the commuting station. Below are the values of the variables found on Table 1 for this person, along with the contributions to the value of G(x), the model function as explained in Chapter III.

COST	17 cents	-1.94
TIME	-410 seconds	1.72
FRICT	19	4.52
WALK	70 seconds	-0.86
CARS	1 car	1.49
AGE	50 years	0.54
INCOME DUMMY 4 - 1		-7.09

Total G(x) for the drive mode -1.62 $e^{G(x)}$ = .1980

Probability of this mode = .085

COST	10 cents	-1.14
TIME	-480 seconds	2.02
RUSH DUMMY-1		-2.26
INCOME DUMMY-1		1.17

Total G(x) for the driven mode -.23 $e^{G(x)}$ = .7945

Probability of this mode = .341

COST	20 cents	-2.28
TIME	-120 seconds	0.51
RUSH DUMMY-1		-1.35
CONSTANT		2.02

Total G(X) for the bus mode -1.10 $e^{G(X)}$ = .3329

Probability of this mode = .143

Probability of the walk mode = 1.00 -.085 -.341 - .143

= .431

The choice of the highest probability is the walk mode which this member of the population did indeed choose. If the proper method of aggregation is to be the summing of probabilities, then we would assume that he follows the probabilities in his choice of transport. That is, he would choose to walk 43% of the time, to drive 8 1/2% of the time, to be driven 34% of the time and to ride the bus for the remaining 14% of the time. Below are three other individuals of the population, one for each of the other three modes.

	VARIABLE	TO DRIVE Value	TO DRIVE G(x)	TO BE DRIVEN Value	TO BE DRIVEN G(x)	TO TAKE THE BUS Value	TO TAKE THE BUS G(x)
		ACTUAL CHOICE					
DRIVE MODE	Cost	44	-5.02	20	-2.28	42	-4.78
	Time	-3940	16.60	-2470	10.40	-3.70	13.34
	Frict	21	5.00	9	2.16	21	5.00
	Walk Time PL-Stn	30	-.37	90	-1.08	30	-.35
	Car Ownership	1	1.49	1	1.49	1	1.49
	Age	60	.65	50	.54	60	.65
	Income Dummy	3	-7.98	3	-7.98	4	-7.98
	G(x)		+10.37		+3.25		+8.27
	$e^{G(x)}$		31,900		25.79		3,910
	Probability		.709		.02		.103
DRIVEN MODE	Cost	30	-3.42	20	-2.28	48	-5.46
	Time	-3970	16.71	-2560	11.57	-3200	13.41
	Rush Dummy	1	-2.26	1	-2.26	1	-2.26
	Income Dummy	3	-	3	-	4	1.17
	G(x)		+8.52		+7.03		+6.86
	$e^{G(x)}$		5,020		1,130		951
	Probability		.111		.880		.025
BUS MODE	Cost	30	-3.42	20	-2.28	20	-2.28
	Time	-2790	11.75	-1530	6.46	-2850	12.00
	Rush	1	-2.26	1	-2.26	1	-2.26
	Constant		2.02		2.02		2.02
	G(x)		+9.00		+4.85		+10.40
	$e^{G(x)}$		8,103		127.8		32,900
	Probability		.180		.099		.871
	Walk Mode Probability		.000		.000		.001

In each case the choice and the maximum probability were the same. As in the first example a sum of the individual probabilities is possible. For these four data the prediction, a sum of their four probabilities separately, is, drive .085 + .709 + .021 + .103 = .918, driven = 1.357, bus = 1.293 and walk = .432. The sum of these four values is of course 4.000. This method of aggregation changes the picture, if we were using each of these as a proxy for 10 users for example. The first method predicts 10 on each mode. This method predicts 9 drive, 14 are driven 13 are on the bus while the remaining 4 walk. Some further work is required to satisfactorily determine the superior method.

The counts according to the model seem to give a superior prediction in Table 2. Only 35 of the population are misclassified as to mode using this method while 76 are wrongly assigned by summing the probabilities. This is the first indication that the model formulation is more suited to an absolute choice on the individual basis than to a summing approach as hoped. This would mean that for best results every user would have to be characterized. For absolute testing, the model could only agree with the actual choice on 138 of the 399 sample members. The attractiveness of the bus was less for the second data group, even though they were making similar trips, as far as purpose and length are concerned. The model also expected that fewer people would walk, but was very good on the other two mode splits.

For the sensitivity test itself, the above datum, the one who chose the drive mode, was selected arbitrarily. This test is one means of illustrating the effect of a change in a variable on the probability of the mode choice.

To start suppose that there is a ten per cent decrease in the cost of operation of a car, perhaps because the user has purchased a car which is more economical to run. His cost is now 40 cents instead of 44. This results in an increase in G(x) for the drive mode. G(x) now has the value 10.87 and all of the probabilities change. The drive probability goes up to .800, while all others decrease proportionately according to their magnitude. Of course, for the advantage which has been mentioned, there is also the same cost saving for the driven mode, and hence, an accompanying change in the cost value for that mode. G(x) for the third choice now becomes 8.86. Now the final probabilities are walk .00, drive .777, driven .104 and bus .118. This interdependence seems to confirm that cost, and indeed time, should be spoken of in terms of the type of time or cost that is being considered. To illustrate, taking another kind of cost, the parking fee, it is noted that its effect is only on the drive mode. But, it is entered in two difference variables, Cost and Frict. This time, suppose a ten per cent increase in cost occurs in the form of increased parking. For this datum the parking cost is 35¢, half of which is charged to the inbound trip and half of which is charged to the outbound trip. A ten per cent change in cost means a 4¢ change. Therefore, a change of 8¢ in the parking charge results. The effect on G(x) is to 10.82 almost the same as previously calculated for a decrease in the running cost. At any rate a change in costs is not going to effect the result as far as absolute choice is concerned at this level. In the case of sums of probabilities, there is some change of ten per cent for the drive mode, in both of the above examples. Change in cost is not confined to the drive mode. One oft discussed change to increase the transit patronage is the abolition of fares. In this model it would mean a change in G(x) for the bus sector to 12.42, since the cost difference is now nil. This is highly significant in this particular case, since it means that this mode now becomes the user's most probable. The new probability is now .873, while the driven is .017 and drive is .110, walk being almost completely eliminated. A twelve cent decrease in the bus cost is required to make the drive and bus modes equally attractive. This is a percentage decrease of 40%. The resulting probability is .465 for both choices. The driven mode can also have an independent reduction of cost if for example the driver had a destination closer to the station after the transit rider was dropped. Here a 10% reduction means G(x) = 8.86 and probability .153 up .042 from the previous value. A unique change can also be achieved for the drive mode if a car pool is set up amongst a number of users who live in the same area. In this way it is possible to halve or even quarter the cost without seriously changing the values of any of the other variables. If we assume a two user car pool, this means that the G(x) value is now 12.88 and the probability is .966. By adding any more passengers, this mode becomes a virtual certainty. The high cost elasticity may reflect the nul requirement for mobility during the day, hence reducing the factor involved in choosing the auto.

The other system characteristic which can easily be changed by the transportation planner is the time difference. Savings in time can be achieved by many traffic methods such as traffic light synchronization, or by reduction of congestion by removing parking from one or both sides of the street. In the same way, an increase in the travel time is also quite possible as the number of cars on the road increases. A ten per cent rise in the time difference, meaning that the motorized mode is now quicker, means a change to 12.03 for the drive mode. This means an increase of its probability to .926. If the change is in the driven mode alone, the probability rises to .536 for this mode and it becomes the most probable choice. If by setting up separate right-of-ways or if by some other method the bus time alone is decreased, its share is elevated to .378, not far below the drive value of .488.

As in the case of cost, there are secondary changes depending upon the type of time change. A change in the line haul portion of the trip proceeds as above, but a change in the walking time from the parking lot to the station is different. It is reflected in three of the variables entered into the first sector, time difference walking time to the station from the parking lot directly and in the Frict variable. Assuming a ten second change in this segment of the trip, the new G(x) for the second choice is 11.10, the new probability .840. This is only a 2.5 per cent change in the time difference.

The user variables are not so easily changed in the aggregate sense with which we must be ultimately concerned. Even in the behavioural sense of the individual, they are variables which are not susceptible to radical and large change, with the possible exception of the dummies where a change in group can have a large effect on the overall probability.

A change of ten years in a person's age means only a change of .107 for the value of G(x). From 10.37 to 10.48 means a change in share from .709 to .728 in the drive case. Only when the distance to the station is smaller will it have a large effect, as the proclivity to walk is lessened with age increase.

A change to two cars from one would definitely cause a large change in the value of the G(x) for choice two, but in the data sets used it must be considered as a dummy type variable since all values were either one or zero. If this user did not own a car his revised probabilities would be walk .001, drive .351, driven .248, bus .400. As might be expected the bus becomes the choice. The high value for drive seems to indicate that some control should be put on car ownership, that the present method does not describe the effect well.

The user chosen was in income dummy 3. If this were to change to dummy 4, changes would occur in the first two of the three sectors of the model. $G_1(x)$ becomes 11.27 and $G_2(x)$ becomes 9.68. The four probabilities are then walk - .001, drive - .762, driven - .157, bus - .080. If the change was to income dummy 2, then only the first sector is changed to 10.94 and the probability to .791. If the income is further decreased to dummy 1, then drive becomes the choice at .927 probability. No member of the data set has a level income in the lowest range so that at least one of the four dummies entered had a value of one. This makes the use of the model in the lowest range invalid.

The Rush dummy is the only other variable which has not been dealt with in the sensitivity testing. In its form as a dummy it too can have a drastic effect upon the results. If it were not the rush hour for this user, his new probabilities would be walk - .001, drive - .287, driven - .431, bus - .282 and the absolute choice is to be driven.

The sensitivity test gives an insight into the workings of the model, in particular into the magnitudes and signs of the variable coefficients in the model. It is also noteworthy to point out the variables which failed to enter significantly into the equation. These variables were rejected on the basis of low t-value scores for their coefficients. That is, the t-value scores were not significant at the .90 level.

One of these variables is sex, a variable which Lisco, 1967, found to be significant in his work. There was a 21 per cent representation for the female in the data source for deriving the equation, so that there are enough data available to allow a differentiation to be made if one exists. However, even in preliminary testing of the data it was noted that the correlation of this variable to the dependent variable is very low. The t-values that it achieved were amongst the lowest so that it was one of the first to be rejected.

An attempt was made to enter a distance dummy especially for the relation with the walk mode. Perhaps if a different base mode were used and a user characteristic dependence investigated for this mode it would prove significant, but in the other three segments and using the walk mode as a base, no significant t-value was achieved.

Age, as a linear variable was found to be insignificant in two of the modes. The 'a priori' assumption that a dummy variable use would be superior, was tested, but complications arose since their use generated an unsolvable matrix for the maximum likelihood estimator. Unfortunately, it is felt that the present form of the variable is much less than satisfactory for describing the effect upon the choice process. It was not then surprising that the coefficients had low t-values in two cases, and even in the first sector the magnitude is only marginally acceptable at the .90 level.

Another big disappointment was that the life-cycle dummy variables proved to be insignificant in the model. Again the 'a priori' assumption was that they would prove to be very helpful in describing the choice process. It was thought that they would help to determine the availability of the car by indicating the number of drivers competing for the use of the car in the household.

An attempt was made to try to break up the time difference into line haul and waiting times to uncover some relation between the inconvenience and the value of time which is allotted to the different activities. Again the attempt failed due to low t-values being achieved for the coefficients.

The next thing which can be discussed about this model is the 'value of time' which can be derived using the coefficients of the time difference and the cost difference. By dividing the former coefficient by the latter and multiplying this ratio by 36 a 'value of time' in terms of $/hour can be obtained. For this model this figure is $1.33 per hour. Based on a 2000 hour work year and the average yearly wage of the data set ($11,000) this value works out at 24.1% of the wage rate. This can now be postulated as a 'value of time' for a short trip to a commuter station. This is not an absolute value for the total time as such but rather it is a measure of the worth of saved time. Whether or not the magnitude would be the same is a matter for conjecture. This 'value of time' is only about half the value found by Lisco, 1967, in his probit model of modal split. His data, however, was for a longer line haul trip and as such the characteristics would be different. The user is less concerned about saving a small amount of time when the trip is of considerable length. Therefore, for the user, this small segment of the total trip is not regarded as a separate entity, but rather in the context of the whole voyage. A trip which is in itself complete, of the same length as the trips with which we are concerned, is likely to have a 'value of time' more in line with the values found by Lisco.

Values are also available from other authors. Stopher, 1968, found values of time as a proportion of the wage rate to vary from .33 to .14 depending upon the salary range. This compares very favourably with the values above. This is the only valid comparitive measure since the monetary values in England where the studies were made are different.

Thomas and Thompson, 1970, give a large range of values depending upon the trip purpose and income level. This is also set up for savings by a rural toll freeway. The values which appear vary greatly. If we may assume that 10 minutes is a fair time savings level, applicable to this study, then for our income levels his average value is $2.23/hour, higher than that above. As a per cent of income the average is 33% again above the values previously discussed. The values, it may be argued, are for a totally different type of trip and thus the values are not validly comparable.

Assuming that it is based on valid premises, we can make comparisons with the R-squared values of other models. The size of the R-squared statistic for this model is small when compared to those of Stopher, 1969, but the dimension of this model is that much greater. To model a complete set of four choices, a binary system requires some method of stratification. A minimum of three separate models is required to find the ultimate split. To achieve results as good as those for the given model R-squared values greater than .89 are needed. This assumes that R-square represents the % of variance explained and that the cumulative effect is multiplicative. Stopher's models achieved .90 so that they are possibly as good in their predictive ability. There is a great space for improvement of the variables entered in the multi-mode models. As far as computational simplicity, the multi-mode model also holds an edge over the binary models.

The model developed by Lisco, 1967, used a probit format. Using the method of elimination for variables used here, his models would not enter the sex value. Therefore his statistics cannot be compared to those of this model if the criterion for significance is to be maintained. The Likelihood Ratio does confirm that the model is highly significant well above the .99 level.

In using the fit technique for the data in the same formulation but not restricting the coefficients of time difference and cost difference as in Tables 3 and 4 the same variables do not enter.

In the first segment of the model, the first income dummy variable was found to be insignificant. Since there were no members of the data set which had the lowest grouping this does not change the power of explanation in any way since those in the 1st group now have a zero value for all the remaining dummies, this simply means that the ceiling wage for the first group has been raised to include those of the previously second group. Other than this variable all others were still included in both of the models. In the second segment, one more variable has been added. Age proved to enter significantly in both the second and the third segments of the model. The same number of variables result because the age replaces the Rush dummy of the first model.

The multiple statistics for the two models are not significantly different so that we can not assume any advantage to using this technique simply as a fitting method, over the conceptually more accurate first model.

In the STAT3 counts however the prediction of the second are superior in every way to those of the first model. By absolute count, 35 are misclassified in the first model, 30 in the second, by summing probabilities, 76 are misclassified for the first, 48 by the second. In the matter of correct individual predictions the latter also did better, 146 to 138. Especially in the case of summing of probabilities, the fit technique is better.

The computation required for the second model is slightly more than that for the first model making it more expensive to run. The size of the central memory required for the calculation is not effected in the Theil program, and in STAT3 there is no difference in complexity for the two. So only a little advantage is gained for the first model.

The coefficients in the fit model can be compared not on any conceptual basis but as a matter of interest. The cost and time coefficients are widely varying over the range of the model and only in the bus mode is there any similarity to the magnitude of the first model. In this case, the difference is only in the second significant figure. All of the common variables have the same sign showing that there is no large change in correlation. On the other hand the magnitudes are all significantly changed from the one to the other.

TABLE 3

LARGE SCALE BEST-FIT MODEL

VARIABLE	COEFFICIENT	t-VALUE
DRIVE MODE		
Car Ownership	1.2245	2.8845
Age	-.0442	1.9550
Walking P-L to Stn.	-.02805	1.4257
Frict (see below)	.34127	1.5227
Cost	-.25785	2.0226
Time difference	-.005378	2.9277
Income Dummy 2	-3.2799	2.1844
Income Dummy 3	-3.8929	3.2269
Income Dummy 4	-2.9545	2.7130
DRIVEN MODE		
Age	-.042356	2.1026
Cost difference	-.092659	3.3403
Time difference	-.004333	4.7149
Income Dummy	1.3042	2.1104
Rush Dummy	-.93175	1.4898
BUS MODE		
Age	-.037945	1.3624
Time difference	-.004602	4.8414
Cost difference	-.13720	1.6476
Constant	2.5821	42.4919

Likelihood Ratio Test 119.161 with 15 degrees of freedom
Proportionate Pseudo R-square .6883

TABLE 4

LARGE SCALE BEST-FIT MODEL - STAT3

MULTIPLE CORRELATION RATIO = .13898

MULTIPLE F-VALUE = 6.96521 (3.38 AT .999 LEVEL)

WITH 9 AND 389 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	DRIVEN	BUS
39	102	202	56

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	DRIVEN	BUS
54	117	187	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
45.6	91.0	173.2	89.3

THE MEAN TRUE PROBABILITIES ARE

OBS = 399	WALK	DRIVE	DRIVEN	BUS
	.135338	.293233	.468672	.102757

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
.114194	.228022	.433968	.223816

THE NUMBER CORRECTLY PREDICTED IS	11	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	33	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	84	IN MODE	DRIVEN
THE NUMBER CORRECTLY PREDICTED IS	18	IN MODE	BUS
TOTAL	146		

IV-2 Secondary Testing

The following tests merely show the effects upon the model of changing the variables to strictly system characteristics, or by changing the number of dimensions of the model. There is a further desire to find an accurate 'value of time' through possibly eliminating the interference of other variables.

Table 5 and Table 6 contain the testing for time and cost only models. It may be noted that three constants appear, one in each of the sectors of the model. This was not the case in the previous model formulations. The reason for this is that many iterations are required to find a solution for the coefficients if all three constants are used, thus generating both high time costs on the computer and also causing the no solution result, since there is a limit put on the number of iterations which can be performed in the program as it now stands. This can be rectified in the future to allow better models to be built, since there is convergence.

The time difference coefficient changed only 5% from the first model, but the cost difference coefficient decreased by 33%. The t-values are still very highly significant and the multiple statistics are slightly lower than the first model. This shows that the behavioural models are superior to simple time and cost models.

TABLE 5

RESTRICTED COEFFICIENTS - TIME AND COST ONLY

VARIABLE	COEFFICIENT	t-VALUE
DRIVE MODE		
Cost	-.0765749	4.8761
Time Difference	-.00445492	5.2761
Constant	-1.4143	
DRIVEN MODE		
Cost	-.0765749	4.8761
Time Difference	-.00445492	5.2761
Constant	-2.1116	
BUS MODE		
Cost	- .0765749	4.8761
Time Difference	- .00445492	5.2761
Constant	.14068	

Likelihood Ratio Test 91.4085 with 6 degrees of freedom

Proportionate Pseudo R-square .5799

TABLE 6

RESTRICTED COEFFICIENTS - TIME AND COST ONLY - STAT3

MULTIPLE CORRELATION RATIO = .132339

MULTIPLE F-VALUE = 6.592448 (3.38 at .999 LEVEL)

WITH 9 AND 389 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	DRIVEN	BUS
51	105	178	65

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	DRIVEN	BUS
54	117	187	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
55.2	111.3	124.2	108.3

THE MEAN TRUE PROBABILITIES ARE

OBS = 399	WALK	DRIVE	DRIVEN	BUS
	.35338	.293238	.468672	.102757

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
.138357	.279027	.311281	.271335

THE NUMBER CORRECTLY PREDICTED IS	26	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	41	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	97	IN MODE	DRIVEN
THE NUMBER CORRECTLY PREDICTED IS	15	IN MODE	BUS
TOTAL	179		

In contrast, this model does the best job of prediction in the STAT3 results of Table 6. Only 24 members are misclassified in the counts of the model. 59 are misclassified in the summing of probabilities. A total of 179 correct individual choices were achieved. This may indicate that the individuals made their decision on the basis of time and cost more than the Skokie area people did.

The 'value of time' for this model is $2.09 per hour. This is much closer to those of Lisco, 1967, and Thomas and Thompson, 1970, than the previous values, but the model itself is far inferior making this less trust-worthy.

Table 7 and Table 8 set forth the fitting technique time and cost difference model. Several marked changes occur. Firstly, the sign of the cost difference changes in the drive sector. As there is no conceptual basis for this model there is no particular way in which to interpret this change. Since there is a sign change there is no reason to compare the magnitudes to the first case. In the other two sectors the time difference as before did not change by a great amount. Only the values of the cost coefficients varied drastically. The constants change since the nature of the constant is to account for the variation which is not directly attributable to the entered variables.

The R-squared value of the last model and the Likelihood Ratio Test are both up over the restricted coefficient model. The aggregate values of the second test program however, counter this result. There is misclassification of 48 in the absolute count, 69 by summing probabilities. Strangely, the exact prediction number is the highest yet at 188.

Tables 9, 10, 11, 12 below contain the analysis of the model split when there are only three dimensions. Those members of the population who chose the fourth alternative were eliminated in the derivation and testing of these models. This is not totally realistic in the real world situation since the users that remain still do have the fourth alternative in reality. The purpose is to show the changes which occur with respect to the multiple statistics and the forecasting part of STAT3.

In this case the time difference and cost difference coefficients have been restricted to the same value for each mode. The constant is placed on the last sector of the mode, either the bus mode or the driven mode. The variables used are those which were developed from the four choice model, and they are adapted to the three choice situation only by the movement of the constant.

TABLE 7

FREED COEFFICIENTS - TIME AND COST ONLY

VARIABLE	COEFFICIENT	t-VALUE
DRIVE MODE		
Cost	.0091026	.22734
Time Difference	-.0036333	4.1208
Constant	-2.4589	
DRIVEN MODE		
Cost	-.0818613	3.0855
Time Difference	-.0041409	4.9989
Constant	-1.7568	
BUS MODE		
Cost	-.10935	1.8974
Time Difference	-.0044824	5.3511
Constant	.34819	

PROPORTIONATE PSEUDO R-square .6082

LIKELIHOOD RATIO TEST 98.3839 with 6 degrees of freedom

TABLE 8

FOUR CHOICE - TIME AND COST FREED COEFFICIENTS

MULTIPLE CORRELATION RATIO = .132048

MULTIPLE F-VALUE = 6.575738 (3.38 AT .999 LEVEL)

WITH 9 AND 38 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	DRIVEN	BUS
51	143	142	63

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	DRIVEN	BUS
54	117	187	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
57.9	130.2	117.4	93.6

THE MEAN TRUE PROBABILITIES ARE

OBS = 399	WALK	DRIVE	DRIVEN	BUS
	.135338	.293233	.468672	.102757

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	DRIVEN	BUS
.144989	.326328	.294201	.234482

THE NUMBER CORRECTLY PREDICTED IS	25	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	66	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	81	IN MODE	DRIVEN
THE NUMBER CORRECTLY PREDICTED IS	16	IN MODE	BUS
TOTAL	188		

TABLE 9

THREE CHOICE MODELS RESTRICTED COEFFICIENTS

VARIABLE	MODEL 1	MODEL 2	MODEL 3
DRIVE MODE			
Cost	-.11177	-.15552	
Time Difference	-.0086448	-.0053491	
Frict	.29327	.35606	
Walking PL-S	-.010014	-.018493	
Car Ownership	1.3798	1.4232	
Age	.0072984	.0089530	
Income Dummy 1	-10.470	-7.7027	
Income Dummy 2	-9.3919	-9.1352	
Income Dummy 3	-13.400	-9.5666	
Income Dummy 4	-9.7496	-8.6719	
DRIVEN MODE			
Cost	-.11177		-.11852
Time Difference	-.0086448		-.0043701
Rush Dummy	-2.1604		-2.6090
Income Dummy 4	3.0037		1.4559
Constant	-3.9025		
BUS MODE			
Cost		-.15552	-.11852
Time Difference		-.0053491	-.0043701
Rush Dummy		-.55339	-2.3490
Constant		2.0212	2.9254
Proportionate Pseudo R-squared	.7381	.7463	.7607
Likelihood Ratio	93.90	92.51	89.67
Degrees of Freedom	13	12	6

TABLE 10

THREE CHOICE - RESTRICTED COEFFICIENTS - MODEL 1

MULTIPLE CORRELATION RATIO = .085312

MULTIPLE F-VALUE = 3.606396 (3.38 AT .999 LEVEL)

WITH 9 AND 348 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	DRIVEN
67	91	200

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	DRIVEN
54	117	187

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	DRIVEN
68.7	80.2	209.1

THE MEAN TRUE PROBABILITIES ARE

OBS = 358	WALK	DRIVE	DRIVEN
	.150838	.326816	.522346

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	DRIVEN
.191791	.224119	.584090

THE NUMBER CORRECTLY PREDICTED IS	21	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	24	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	97	IN MODE	DRIVEN
TOTAL	142		

TABLE 11

THREE CHOICE - RESTRICTED COEFFICIENTS - MODEL 2

MULTIPLE CORRELATION RATIO = .054633

MULTIPLE F-VALUE = 1.290643 (3.38 AT .999 LEVEL)

WITH 9 AND 201 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVE	BUS
5	184	22

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVE	BUS
54	116	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVE	BUS
6.3	185.1	19.7

THE MEAN TRUE PROBABILITIES ARE

OBS = 211

WALK	DRIVE	BUS
.255924	.549763	.194313

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVE	BUS
.029665	.877089	.093246

THE NUMBER CORRECTLY PREDICTED IS	3	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	102	IN MODE	DRIVE
THE NUMBER CORRECTLY PREDICTED IS	7	IN MODE	BUS
TOTAL	112		

TABLE 12

THREE CHOICE - RESTRICTED COEFFICIENTS - MODEL 3

MULTIPLE CORRELATION RATIO = .222619

MULTIPLE F-VALUE = 8.654763 (3.38 AT .999 LEVEL)

WITH 9 AND 272 DEGREES OF FREEDOM

THE COUNTS ACCORDING TO THE MODEL ARE

WALK	DRIVEN	BUS
57	128	97

THE TRUE COUNTS ARE GIVEN AS

WALK	DRIVEN	BUS
54	187	41

THE COUNTS BY SUMMING PROBABILITIES ARE

WALK	DRIVEN	BUS
61.5	119.4	101.0

THE MEAN TRUE PROBABILITIES ARE

OBS = 282

WALK	DRIVEN	BUS
.191489	.663121	.45390

THE MODEL MEAN PROBABILITIES ARE

WALK	DRIVEN	BUS
.218249	.423576	.358175

THE NUMBER CORRECTLY PREDICTED IS	22	IN MODE	WALK
THE NUMBER CORRECTLY PREDICTED IS	82	IN MODE	DRIVEN
THE NUMBER CORRECTLY PREDICTED IS	22	IN MODE	BUS
TOTAL	126		

The first thing to note is that the value of the R-squared statistic is increased even though there is no change in the variables entered. This means that by increasing the number of choices we are going to have to expect a smaller explanatory power for the model. We would expect that an increase to five different modes would reduce the R-squared value to below .600 or in that area at least. Therefore the variable investigation should point towards a good explanation of the behavioural phenomenon so that larger mode choices can be accommodated with sufficient accuracy.

It would seem that, although the drive mode has the most variables for explanation, it is not as good a model for this sector as for the other two. The R-squared figures for Models 1 and 2 are less than for Model 3, in which the drive choice does not enter.

Previously, the time difference coefficient has remained fairly consistent while the cost coefficient varied widely. This is not the case in this transition. The first varied from the third model's close figure to Model 1's figure which is twice the magnitude of the four choice value. Both the values of Model 1 and Model 3 cost coefficients are close to the original four choice value, and Model 2 has only increased this by 25%. These changes are reflected in the new 'values of time' which can be found from these models. They are $2.78 per hour for Model 1, $1.24 per hour for Model 2 and $1.33 per hour for Model 3. The first value is right in the range that both Lisco, 1967, and Thomas and Thompson, 1970 found while the other two values are similar to the previous findings of this paper in the four choice model. Certainly, the fact that the constant has been moved, for the one case where the value changes, may have some bearing on that change.

In this light, it may be noted that the coefficients in general changed quite highly from the four choice model to Model 1. Again this may be an effect of the moved constant. Although the magnitudes of the income dummies changed in Model 2, they maintained their relative positions with respect to each other. This is not true in Model 1.

There seems to be little correlation between the R-squared statistic and the results of testing in aggregation. The best results in this regard come from Model 1, which has the smallest value for R-squared. Only 26 were misclassified in the count method and 31 in the summing of probabilities. This contrasts to 68 and 69 for Model 2 and 59 and 67 for Model 3, both on smaller totals. As before the Likelihood Ratio Test shows that the models are highly significant even at the .99 level.

A comparison of trends in prediction can be made. The prediction for walk is low in both large scale models, while bus is high. Walk is accurate with time and cost only, while Bus is quite high. It is only in Table 11 that

Bus is underestimated. This model is dominated by the drive mode. Driven is usually underestimated. Table 4 once and Table 10 twice being the exceptions. Drive is only underestimated in Table 11.

The "multiple statistics" shown in the earlier tables still require some validation. It is not yet clear whether they represent the fit to one mode or to all. They have been included only as a comparison between the models presented.

CHAPTER V

CONCLUSIONS

This chapter sets forth the conclusions drawn from the experimental work. Following that is a discussion of the direction and content of future work in this field.

The main interest of the research was to determine whether or not the multi-mode, multinomial logit formulation is significant as a descriptive and analytic model of mode split. The foremost conclusion then is that this type of model does produce significant models. In every case the Likelihood ratio statistic proved to be highly significant at the .99 level.

In addition to the technique, which is to fit a mathematical form to an observed pattern, the type of model which is most efficient is another matter. The thesis shows that statistically the behavioural concept, which enters user as well as system characteristics, is superior in explanatory power to the simple market place model, which enters only time and cost differences. As far as the use of the two models for prediction is concerned, the thesis work has not been able to prove any advantage to either method. The predictions on the second data set were equally as accurate for both of the models. There also seems to be no advantage to a non-conceptually based model.

Another of the questions raised, was the problem of aggregation. The method of predicting peoples' absolute choice and then summing in this way has proved to give more accurate analyses in all cases investigated. No real conclusion can be drawn here I feel because this was simply an exercise to test a model on a larger set of data and not to predict the reaction of grouped populations. That is, in an operational model, the total population would be represented by a number of different characteristics and thus divided into sub-groups represented by one member. To aggregate then, his probabilities would be extended to the group. When each individual is entered into the model, no such extension is required and so his maximum probability is his choice. In the operational model the idea is to model the choices of a group of people at an instant in time, whereas in the thesis form of aggregation there is an analysis of the individual user over a period of time. As the data are formulated for an instant in time, the use of an absolute sums method should prove to be better, as it does.

Although it is possible to state that this sort of model will be significant, from this limited study it is not possible to make any absolute statement on the accuracy compared to the binary stratification technique. The statistics of these models are lower than those for the binary models, but the problems of stratification and aggregation are greater, so that any real value judgement is impossible at this time. Several points can be made in this vein however. The multi-mode structure is most definitely more conceptually satisfying in that there is no arbitrary decision pattern set up. In the binary stratification model, there is a two step decision process set up so that, for example, in our four mode case, the first decision would be whether or not to use the family auto for the trip. Having decided to use that mode, the second decision is then to determine whether to drive or be driven. This seems to be quite plausible. On the other hand, if the decision is not to use the car, then the secondary decision is to walk or to take the bus. It is not likely that these two alternatives are considered separately from the use of the auto. Another way of dealing with this problem of stratification may be more conceptually satisfying, but must then be more structurally complex. The problem becomes more unsatisfactory as the number of modes increases while the structural problems do not change for the multi-mode method. The problems of unavailable modes is quite easily handled in the multi-mode case simply by setting the cost and time differences to arbitrarily high values. For example, if the bus mode of the modes is not available, then a large and positive time difference is sufficient to cause that mode to have a very small finite probability. In theory, if the availability of modes varied widely across the population, it would still be possible to use this model without any structural change, whereas the use of a binary structure would cause much confusion.

One of the most vexing problems for the modeller is the choice of variables which are to be included in the model and once the choice has been made, the form which the variables are to take. The conclusions which are forthcoming from this research are not highly sophisticated since the main purpose was to test the model not to investigate the variables. No detailed study of form for the variables has been done in this work. It is safe to say that these variables which have been accepted as valid in the explanatory sense should have some place in all behavioural models. These main elements are time, cost, for the system and income, age and car ownership for the user. This is a study of a particular group, making a particular type of trip, so that there is little basis for making any general statement without a more comprehensive study of a wider data base on this matter. The one variable which is unlikely to hold any hope for inclusion is the sex of the user, which this thesis found so insignificant.

It is immediately obvious that this is not an operational model. Problems that must still be resolved include many of those mentioned above. Also there is a problem of directionality. It is fallacious to assume a mirror of the trend for the return trip. There are limitations of course since anyone who drives to the station is obliged to return by the same mode. However, in other cases, this is not true. It is quite likely that the 'value of time' may change for the return trip. It is also likely that the driven mode, having a different structure as it must, will have a completely changed G(x) form. It is obvious that it has a completely different waiting time structure, for example.

The 'value of time' which is found from the first model is $1.33 per hour. This should be the most trustworthy value, since it comes from the most accurate model, with the least interference. This seems to indicate that the value of time is less for this short intermediate trip than for the longer line haul trip.

The data collected for the testing of these models were inferior to what would be desired for an adequate building base upon which to develop the model. The information available was incomplete for many members of the population, leading to a small final population for model design and calibration. Future work in the behavioural field must then include the collection of adequate data. Not only was the size of the set too small, but also the range of information which was forthcoming was too small. A greater number of variables must be investigated. The type of socio-economic variables which must be evaluated should be such as the information which is available through the census, so that aggregation is easily accomplished. It is easy to see that given the effect of housing, for example upon the choice of mode and also given the housing breakdown for a particular area from the census data, a definite prediction of the modal split for the whole area is conveniently achieved. Deficiencies also occur in the variables which were presented in the data set. The way in which the income was recorded posed difficulties in use because the information was coded for different sections in different ways. The difference was not easily reconciled between the groups, and, in fact, the income variables for the two separate data sets were slightly different in the ranges that were covered for each variable. Data should also be collected for both directions of the trip. A more accurate picture of the true probability distribution might be derived if the dependent variable were structured so that it reflects the percentage of the time that the user takes each separate mode.

Given a good data source and a good mathematical and conceptual form for modal split, the next step for research is the extension of the behavioural approach to the other steps in the UTP package. The ultimate goal then is an integrated universally applicable set of models which use this approach taking the results of the first to supply the input for the second and so on.

APPENDIX I

The following pages show the format for the data source which was used for building the models. The total sample was about 2000, but when it was assessed for completeness, it was found to be lacking, so that the sample amounted to only about one third of the original. Often the data had the information for a complete mode missing indicating that the alternative was not available to that member.

The text has indicated that this data set was deficient in many ways. Yet without this source of information the time required to make any such study would have more than doubled. Therefore the author wishes to thank the Chicago Area Transportation Study for making this source available.

SUBURBAN STATION ACCESS

CARD FORMAT

(CARD # 1)

ITEM NO.	COLUMN NO.		DESCRIPTION
1	1 - 6	Sample Number:	Number assigned to commuter.
		01XXXX 51XXXX 02XXXX 52XXXX	N.W. Corridor - Outer Study Area
		03XXXX 53XXXX 04XXXX 54XXXX	N.W. Corridor - Inner Study Area
		14XXXX	211th Street
		200XXX	Skokie
2	8	Access Mode: Mode used to Station: 1----- Walk 2----- Drive & Park 3----- Driven 4----- Bus	
3	10 - 11	Station used: 01---- Palatine 02---- Arlington Heights 03---- Mt. Prospect 04---- Dee Road 05---- Park Ridge 06---- Edison Park 07---- Norwood Park 08---- Dempster 09---- Evanston 10---- Howard 11---- 211th Street	

ITEM NO.	COLUMN NO.	DESCRIPTION
4	13	Trip Purpose: N.W. Corridor & Skokie 2 ----- Work 3 ----- Shop 4 ----- School 5 ----- Social Recreation 6 ----- Eat Meal 7 ----- Personal Business 8 ----- Serve a passenger 211th Street 1 ----- Work 2 ----- School 3 ----- Shopping 4 ----- Personal Business 5 ----- Social Recreation 6 ----- Other
5	15 - 16	Street Distance: Residence to Station; Coded to nearest tenth of mile; Example: 24 - 2.4 miles
6	18	Rush or Non-Rush Hour Trip: Departure from Suburban Station N.W. Corridor 6:15 - 8:30 a.m. -----1 Other Times --------0 Skokie 6:30 - 8:30 a.m. -----1 Other Times --------0 211th Street 6:15 - 8:05 a.m. -----1 Other Times --------0
7	20	Pavement Condition: N.W. Corridor Only 0 ----- Dry 1 ----- Wet
8	22	Household Income: N.W. Corridor

ITEM NO.	COLUMN NO.	DESCRIPTION
8	22	1 ----- Under $5,000 2 ----- $5,000 - $7,999 3 ----- $8,000 - $11,999 4 ----- $12,000 - $16,999 5 ----- $17,0000 - $24,999 6 ----- $25,000 + Skokie 1 ----- Under $5,000 2 ----- $5,000 - $6,999 3 ----- $7,000 - $8,999 4 ----- $9,000 -$11,999 5 -----$12,000 -$15,000 6 -----$15,000 +
9	24	Auto Ownership: Number of cars at household. (exact number)
10	26 - 27	Age: Actual age given in original interview (N.W.C.) Average of age range given in original interview --- (Skokie)
11	29	Sex: 0 ----- Male 1 ----- Female
12	31	Dummy Variable I: (N.W. Corridor Only) If unmarried, living at home of parents ----- 1 Otherwise: -- 0
13	33	Dummy Variable II: (N.W. Corridor & Skokie) If unmarried <u>not</u> living at parents home <u>or</u> married and spouse cannot drive <u>or</u> drives to work with spouse. ----- 1 Otherwise:--- 0
14	35	Dummy Variable III: (N.W. Corridor & Skokie) If married and has a spouse who drives to work or school independently of trip taker --------1 Otherwise: ---------- 0
15	37 - 40	Walk Time: Residence to Station Coded in seconds: Walking speed 3 MPH or 1200 seconds per mile.

ITEM NO.	COLUMN NO.	DESCRIPTION
16	42 - 44	Platform Wait Time: Coded in in Seconds: One-half headway time between assigned train and next later train up to a maximum of 4 minutes. (8 minutes or more between trains).
17	46 - 49	Total Time: Walk time + Platform Wait Time. (Items 15 + 16)
18	51 - 54	Driving Time: Residence to Parking Lot - Coded in Seconds - Driving speed approximately 20 MPH or 1800 seconds per mile.
19	56 - 58	Walk Time: Parking Lot to Station - Coded in seconds - Walking speed 3 MPH Varies with time of day according to how full parking lot is.
20	60 - 62	Platform Wait Time: Same as Item # 16
21	64 - 67	Total Time: Driving Time + Walk Time + Platform Wait Time - (Items # 18 + 19 + 20)
22	69 - 70	Driving Cost: 7.5¢ per mile
23	72 - 73	Parking Cost: One-half daily parking cost. (One-half the daily parking cost is assigned to the inbound trip, one-half to the out-bound trip.)
24	75 - 77	Total Cost: Driving Cost + Parking Cost. (Items # 22 + 23)
25	79 - 80	Drive and Park Friction: This is a measure of the disutility of parking expressed in cents at a suburban station; it is comprised of: 1. one-half the daily parking cost plus 2. average walk time from parked auto to station entrance assessed at 6¢ per minute.

ITEM NO.	COLUMN NO.	DESCRIPTION
1	1 - 6	Sample Number: Number assigned to commuter (Same as Item # 1 - Card 1)
26	8 - 11	Driving Time: Residence to Station Coded in Seconds -
27	13 - 15	Platform Wait Time: (Same as Item # 16 - Card 1)
28	17 - 20	Total Time: Driving Time + Platform Wait (Items # 26 and 27)
29	22 - 23	Driving Cost - 15¢ per mile
30	25 - 27	Walk Time: Residence to Bus Stop Coded in Seconds - Walking speed 3 MPH or 1200 seconds per mile.
31	29 - 31	Bus Wait Time: Coded in Seconds - One-half of headway time between assigned bus and next bus up to a maximum wait time of 4 minutes. (8 minutes headway).
32	33 - 36	Bus Travel Time: Coded in Seconds - Scheduled departure time of bus from commuter residence bus stop less scheduled arrival time of bus at suburban station.
33	38 - 40	Platform Wait Time: (Same as Item # 16 - Card 1)
34	42 - 45	Total Time: Walk Time + Bus Wait + Bus Travel Time + Platform Wait Time. (Item # 6 + 7 + 8 + 9)
35	47 - 48	Bus Fare
36 - 45		Dummy Variables as explained in the Text

APPENDIX II

The following program listing is that used to calculate the maximum likelihood estimates for the model coefficients. The program was developed by an economist, John Cragg, and modified first by Dr. Peter Stopher for the CDC 6400 computer at Northwestern University. Testing for the multi-mode case was limited and it was not until the author adapted it to the CDC 6400 facility at McMaster University that it became totally operational.

The primary basis for trying to model the modal split is based upon the premise that the probability of using a particular mode is a continuous function whose dependent variable p_ℓ varies in the range from 0 - 1 according to some function of his sociological traits and the characteristics of the mode. Thus, as any of these variables change, so does the function and hence the probability. The use of a simple linear relationship is rejected because of the bounds imposed by the 0 - 1 range. The function should be assymptotic to both of these limits. This can be done using a logit formulation as in AII-1 for a binary case

$$p = \frac{e^{G(x)}}{1 + e^{G(x)}} \qquad \text{AII-1}$$

The choice probability for the second mode is then AII-2

$$q = 1 - p = 1 - \frac{e^{G(x)}}{1+e^{G(x)}} = \frac{1}{1 + e^{G(x)}} \qquad \text{AII-2}$$

where $G(x) = \text{constant} + \sum_{k=1}^{N1+N2} \alpha_k X_k$ where N1 is the number of system dependent variables (such as time, cost) and N2 is the number of system independent variables. In order to take into account the system dependent variables of the second mode while only dealing with one function we can enter this type of variable as a difference $X_{k_2} - X_{k_1}$

therefore $$G(x) = \text{constant} + \sum_{k=1}^{N1} \alpha_k(X_{k_2} - X_{k_1}) + \sum_{k=N1+1}^{N1+N2} \alpha_k X_k \qquad \text{AII-3}$$

By substituting $\alpha_k = \alpha_{k_2} - \alpha_{k_1}$ for the system independent variables and simplifying AII-4 results

$$G(x) = (\text{constant} + \sum_{k=1}^{N1} \alpha_k X_{k_2} + \sum_{k=N1+1}^{N1+N2} \alpha_{k_2} X_k)$$

$$- (\sum_{k=1}^{N1} \alpha_k X_{k_1} + \sum_{k=N1+1}^{N1+N2} \alpha_{k_1} X_k) \qquad \text{AII-4}$$

We may represent this by a difference of two functions as in AII-5

$$G(x) = G_2(x) - G_1(x) \qquad \text{AII-5}$$

Substituting in AII-1 we have

$$p = \frac{e^{G_2(x)-G_1(x)}}{1 + e^{G_2(x)-G_1(x)}} = \frac{e^{G_2(x)} / e^{G_1(x)}}{1 + e^{G_2(x)} / e^{G_1(x)}}$$

$$= \frac{e^{G_2(x)}}{e^{G_1(x)} + e^{G_2(x)}}$$

and $$q = \frac{e^{G_1(x)}}{e^{G_1(x)} + e^{G_2(x)}} \qquad \text{AII-6}$$

An obvious symmetry exists which may be extended to the multi dimensional form by proposing AII-7.

$$p_\ell = \frac{e^{G_\ell(x)}}{\sum_{\ell=1}^{M} e^{G_\ell(x)}} \qquad \text{AII-7}$$

A similar derivation can bring us to the same result using ratios instead of differences for the system variables. It is important to note the first derivation depends upon a constant value for α_k, k = 1, N1 although α_k, k = N1+1, N2 can and will vary greatly. The use of the model without this restriction has no conceptual base but has been used in this thesis simply as a fitting technique.

To illustrate a binary case where

$$G(x) = 1.0 + .23\ \Delta t + .067\ \Delta c + .05\ S$$

where Δt is minutes and Δc is cost difference in cents, and S is the sex of the user. The two choices are Bus and Drive, p is the probability of using the car, while q is the probability of using the bus. If we hypothesize a time saving of 3 minutes by car (i.e. $\Delta t = + 3.0$) and a loss of 15 cents by driving ($\Delta c = -15$) for a male driver ($S = 1$) the value of G(x) is now $1.0 + .23 \times 3.0 + .067 \times (-15) + .05(1)$

$$= + .74$$

$$p = \frac{e^{.74}}{1 + e^{.74}} = \frac{2.096}{3.096} = .677$$

$$q = .323.$$

Therefore the probability of taking the bus is only 32% while the probability of driving is 68%.

The above is a contrived model, and while quite realistic is not derived from any empirical data.

In aggregation we may use this method by having one of the sample act as a proxy for n like members of the population. Using the above example we would conclude that 32% of those n would use the bus on any given trip and that 68% would drive. Alternatively the whole n could be assigned to the drive mode. The merits of the two methods have yet to be investigated.

The program uses a maximum likelihood estimator of the coefficients. This value then replaces the 0 value initially assumed. From this point the program refines the values of the coefficients in an iterative loop using the Newton method until an acceptably small difference in value is reached. This convergence value is entered by the user. The program will terminate if this value is not satisfied after 25 iterations.

The program does not derive the model exactly as constituted by AII-7, however. To achieve the form which is used, divide the top and bottom of each probability by $e^{G_M(x)}$. A similar derivation to that for the two-dimensional case simplifies to the form of AII-8.

$$P_{(\ell)} = \frac{e^{G_\ell(x)}}{1 + \sum_{\ell=1}^{M-1} e^{G_\ell(x)}} \,, \quad P_{(M)} = \frac{1}{1 + \sum_{\ell=1}^{M-1} e^{G_\ell(x)}} \qquad \text{AII-8}$$

This is the final form of the model to be tested by the program. There is a limitation to AII-8, however, in that the coefficients of the system variables for which differences are used must be the same in each of the sections of the model. User characteristics are the same in each section so that the magnitude of the coefficients can vary.

One advantage that should be built into this program is the ability to build the model in a stepwise fashion. As it is now conceived, complete new runs must be made each time a variable is found to be insignificant. Beyond this the program proved to be extremely useful and quite easy to use and understand.

The program uses a central memory of 60K and the time required for solving a problem is of the order of 30 decimal seconds for the size of model which has been dealt with in this work.

```
      PROGRAM THEIL(INPUT,OUTPUT,TAPE1,TAPE2,TAPE3,TAPE4,TAPE5=INPUT,TAPPRST0010
     $E6=OUTPUT)
C
C J.G.CRAGG PROGRAM FOR MULTINOMIAL MULTIVARIATE LOGIT ANALYSIS, ADAPTEDPRST0040
C TO CDC6400 AT NORTHWESTERN UNIVERSITY BY PETER STOPHER, JULY 1970     PRST0050
C
      DIMENSION XDAT(70),AIR(10),PLIM(10),AV(10)
      DIMENSION DATE(2)                                                 PRST0080
      DIMENSION VARNAM(70),NAME(12)                                     PRST0090
      DIMENSION CV(70),KVR(70,5)
      DIMENSION        DER(70),TRIL(70),XX(70,70)
      DIMENSION TRAT(70)    ,PAMES(70)
      DIMENSION SUBNAM(8)                                               PRST0013
      DIMENSION RKEEP(65)
       COMMON TRAT
      DATA  PROG/6HTHEIL /                                              PRST0016
      REWIND  1
      REWIND  2
      READ(5,1111)KPROB
 1111 FORMAT(10X,I4)                                                    PRSP0090
      ICOUNT=0
    5 ICOUNT=ICOUNT+1
      IF(ICOUNT-KPROB)5555,5555,9989
 9989 STOP
 5555 READ(5,1)(KRUN,KSUBS,(DATE(I),I=1,2))                             PRST0370
    1 FORMAT(10X,2I4,A6,A2)                                             PRST0390
      KSUB = 0                                                          THEI0400
      KPAGE = 0                                                         THEI0410
      READ(5,2)(NAME(I),I=1,12)                                         PRST0420
    2 FORMAT(6X,12A6)                                                   PRST0460
      CALL HEAD(PROG,DATE,KPAGE,KRUN,KSUB,NAME)
      CALL TSAT (IOBA,KTAPE,NBV,NFIR,LTO)
    4 KSUB = KSUB + 1                                                   THEI0470
      KPAGE = 0                                                         THEI0480
      IF(KSUB-KSUBS) 7,7,5                                              THEI0490
 7    READ(5,11) KOPT,KVAR,TOLA,(SUBNAM(I),I=1,8)                       PRST0500
   11 FORMAT(6X,2I4,F8,8A6)                                             PRST0510
  654 KPAGE = 0                                                         THEI0550
      CALL HEAD(PROG,DATE,KPAGE,KRUN,KSUB,NAME)                         THEI0560
      CALL SUBHED(KSUB,SUBNAM,KOPT)                                     THEI0570
      CALL SETDR(IOBA,IOBS,NFIR,LTO,MBORT,NY,M2,VARNAM,NBV,DER,
     1XX,AIR,KVR,CV,NPAR,KTAPE,KVAR ,PAMES )
      IF(MBORT) 5000,5000,4
5000  CALL HEAD(PROG,DATE,KPAGE,KRUN,KSUB,NAME)
      CALL SUBHED(KSUB,SUBNAM,KOPT)
      STEP=1.0
      M1=M2-1                                                           THEI0630
      M=M2-2                                                            THEI0640
      OBS =  FLOAT(IOBS)                                                THEI0660
      CLIK = 0.0                                                        THEI0670
      WRITE (6   ,813)(AIR(I),I=1,NY)                                   THEI0680
813   FORMAT(//6X,11HFREQUENCIES //6X,8F7.0)
      DO 107 I=1,NY                                                     THEI0700
      PLIM(I) = AIR(I) /  OBS                                           THEI0710
      IF(PLIM(I)) 800,800,801
801   IF(PLIM(I)-1.0)803,800,800
  800 WRITE (6   ,802)I                                                 THEI0740
  802 FORMAT(//6X,12HLIMIT VALUE   ,  I6)                               THEI0750
      KNUMQ=777                                                         THEI0760
```

```
      GO TO 107                                                         THEI0770
  803 CLIK = CLIK + AIR(I)*ALOG(PLIM(I))                                THEI0780
  107 CONTINUE                                                          THEI0790
      IF(KNUMQ-777) 810,811,810                                         THEI0800
  811 GO TO 4                                                           THEI0810
  810 KITS = 0                                                          THEI0820
  414 WRITE (6   ,416)CLIK,(I,PLIM(I),I=1,NY)                           THEI0830
      WRITE (6   ,600)TOLA                                              THEI0840
  600 FORMAT(  /6X,23HITERATIONS TO TOLERANCE     , G 12.4)             THEI0850
  418 FORMAT(6X,A6,10X,G15.5,21X,G15.5)                                 THEI0860
  416 FORMAT(/6X,43HLOG OF LIKELIHOOD FOR MULTINOMIAL MODEL =   ,G13.5/ THEI0870
     1 (6X,5HVALUE,I2,10H ESTIMATE ,F5.3))                              THEI0880
      MQ=M2
      MD=NPAR
      IF(KOPT-3)7889,7890,7890
 7890 READ(5,7888)(TRIL(I),I=1,MD)
      READ(5,7876)KITS,QLIM
 7876 FORMAT(I4,E20.4)
      GO TO 437
 7889 DO 150 I=1,MQ
      KF=KVR(I,1)
      KC=KVR(I,3)
      KQ=KVR(I,2)
      DER(KC)=DER(KC)-PLIM(KQ)*CV(KF)
  150 TRIL(I)=0.0
      DO 152 I=1,MQ                                                     PI
      DO 152 J=1,MQ                                                     PI
      K=KVR(I,2)
       L=KVR(J,2)
      D=PLIM(K)*PLIM(L)
      IF(K-L) 152,515,152
  515 D=D-PLIM(K)
  152 XX(I,J)=-1.0*D*XX(I,J)
      DO 252 I=1,NPAR
      DO 252 J=I,NPAR
  252 XX(I,J+1)=0.0
      DO 253 I=1,MQ                                                     PI
      DO 253 J=1,MQ                                                     PI
      KC=KVR(I,3)
      KQ=KVR(J,3)
      IF(KQ-KC)253,255,255
  255 XX(KC,KQ+1)=XX(KC,KQ+1)+XX(J,I)
  253 CONTINUE
      DO 254 I=1,NPAR
      DO 254 J=I,NPAR
      XX(J,I)=XX(I,J+1)
  254 XX(I,J)=XX(J,I)
      DO 155 I=1,MQ                                                     PI
      IF(KVR(I,1)-KVAR+1)155,159,155
  159   L=KVR(I,2)
      K=KVR(I,3)
      TRIL(K)=ALOG(PLIM(L)/PLIM(1 ))
  155 CONTINUE
      MD=NPAR
  158 GLIK = CLIK                                                       THEI1130
       QLIM=GLIK-100.0
```

```
      GO TO 480                                                         THEI1140
  437 DO 419 I=1,MD                                                     THEI1150
      DER(I) = 0.0                                                      THEI1160
      DO 419 J=1,MD
  419 XX(I,J) = 0.0                                                     THEI1180
      KJAB=0
      KLUG=0
      GLIK = 0.0                                                        THEI1190
      DO 181 LL= 1,IOBS
      READ ( 1 )(XDAT(I),I=1,KVAR)
      DO 2111 I=1,NY
2111  AV(I)=0.0
      DO 171 I=1,MQ                                                     THEI1240
      K=KVR(I,2)
      L=KVR(I,3)
      KF=KVR(I,1)
171   AV(K)=AV(K)+XDAT(KF)*TRIL(L)
      TEMP = 0.0                                                        THEI1350
      DO 173 I=1,NY
      IF(AV(I)-320.)5004,5004,5001
 5004 IF(AV(I)+320.)5002,5003,5003
 5002 AV(I)=-320.
      GO TO 5003
 5001 AV(I)=320.
 5003 AV(I)=  EXP(AV(I))
  173 TEMP = TEMP + AV(I)                                               THEI1370
      DO 174 I=1,NY                                                     THEI1380
  174 PLIM(I)=AV(I)/TEMP                                                THEI1390
      KL=IFIX(XDAT(KVAR))
      DO 6743  I=2,NY
       IF(I-KL)6744,6745,6744
6744   PAMP=AV(I)
      GO TO 6746
6745  PAMP=-AV(I)
6746   IF(PAMP)6743,6743,6747
6747  KJAB=KJAB+1
6743   CONTINUE
      IF(PLIM(KL)) 710,710,711                                          THEI1410
 710  GLIK=GLIK-670.
      KLUG=KLUG+1
      GO TO 175
  711 GLIK = GLIK+ALOG(PLIM(KL))                                        THEI1440
175   DO 181    I=1,MQ
      KF=KVR(I,1)
      KC=KVR(I,2)
      KQ=KVR(I,3)
      IF(KC-KL)1177,1178,1177
1178   DER(KQ)=DER(KQ)+XDAT(KF)
1177  CONTINUE
      DER(KQ)=DER(KQ)-XDAT(KF)*PLIM(KC)
      DO 181   J=1,MQ
      KG=KVR(J,1)
      K=KVR(J,2)
      L=KVR(J,3)
      D=PLIM(KC)*PLIM(K)
      IF(KC-K) 1181,1180,1181
1180   D=D-PLIM(K)
```

```
1181    XX(KQ,L)=XX(KQ,L)-D*XDAT(KF)*XDAT(KG)
181     CONTINUE
        IF(KJAB)6750,6750,6751
6750    WRITE(6,6752)
6752     FORMAT(6X,12HPERFECT FIT     )
         GO TO 4
6751    IF(KLUG-4)480,480,7891
7891     WRITE(6,7892)   KLUG
        GO TO 4
7892     FORMAT(/6X,8HTROUBLE  ,I6)
  480 REWIND  1
C        DER IS FIRST DERIVATIVES                                        THEI1660
C        XX IS MATRIX OF SECOND PARTIAL DERIVATIVES                      THEI1670
C        GLIK IS LOG OF LIKELIHOOD ( PREVIOUS ITERATION)                 THEI1680
      WRITE ( 6 ,427)KITS, GLIK                                          THEI1690
  427 FORMAT( 6X,11HITERATION  ,I3,6X,31HLOG OF LIKELIHOOD FUNCTION =    THEI1700
     1,G13.5)                                                            THEI1710
      IF((GLIK-QLIM).GE.0.0)GO TO 1496
      STEP=STEP*.6
      TOLA=TOLA*.6
      QLIM=GLIK-100.
      DO 1497 I=1,MD
1497   TRIL(I)=RKEEP(I)
      GO TO 437
 1496 CALL INVERT(XX,DET,MD,70)
      IF(DET)4503,4503,4505
4503   WRITE(6,4504)
4504  FORMAT(//6X,16HSINGULAR MATRIX        )
      GO TO 4
4505  DO 4506 I=1,MD
      RKEEP(I)=TRIL(I)
      DO 4506 J=1,MD
4506  TRIL(I)=TRIL(I)+XX(I,J)*DER(J)*STEP
      KITS = KITS + 1
7808   FORMAT(/(2X,9E12.4))
7888  FORMAT(5E16.6)
C      CHECK FOR CONVERGENCE OF ITERATION                                THEI1900
      IF(KOPT-2) 349,348,348
  348 WRITE(7,7888)(TRIL(I),I=1,MD)
      WRITE(7,7876) KITS,GLIK
349   IF(KITS-2) 436,436,667
  667 IF( KITS - 25) 666,666,668                                         THEI1920
  668 WRITE ( 6 ,669)                                                    THEI1930
  669 FORMAT(/40H TWENTY-FIVE ITERATIONS  - DISCONTINUED     )           THEI1940
      GO TO 435
  666 IF( ABS((GLIK-QLIM)/GLIK)- TOLA) 435,435,436
  436 QLIM =  GLIK                                                       THEI1970
      GO TO 437                                                          THEI1980
435   DO 442 I=1,MD
      DER(I) =  SQRT(XX(I,I))
442   TRAT(I)=TRIL(I)/DER(I)
      WRITE ( 6 ,443)                                                    THEI2030
 443  FORMAT(/6X,30HMAXIMUM LIKELIHOOD ESTIMATES  ,//20X,9HESTIMATES,
     19X,15HSTANDARD ERRORS,7X,8HT-VALUES/)
      WRITE(6,160)(PAMES(I),I,TRIL(I),DER(I),TRAT(I),I=1,MD)
160   FORMAT(2X,A6,I4,2X,3G19.5)
```

```
      ALIK =-2.0  *(CLIK - GLIK)                                        THEI2150
      MM=MQ-NY +1
      WRITE (6  ,452)ALIK,MM                                            THEI2170
  452 FORMAT(// 6X,24HLIKELIHOOD RATIO  TEST      ,G14.6,6H WITH  ,I6,6H THEI2180
     1 D.F.    )                                                        THEI2190
      ALIK=1.0-EXP(-ALIK/FLOAT(IOBS))
      WRITE(6,467)ALIK
467   FORMAT(/6X,18HPSEUDO R-SQUARE -     ,G18.4)
      ALIK=ALIK/(1.0-EXP(2.0*CLIK/FLOAT(IOBS)))
      WRITE(6,468)ALIK
468   FORMAT(/6X,31HPROPORTIONATE PSEUDO R-SQUARE = ,G18.4)
      ALIK=GLIK+FLOAT(MD)*ALOG(2.0*3.14159)/2.0-ALOG(DET)/2.0
      WRITE(6,469)ALIK
469   FORMAT(/6X,44HLOG OF POSTERIOR PROBABILITY = CONSTANT +    ,G18.5)
      DO 900 I=2,MD
      K=I-1
      DO 900 L=1,K
900   XX(L,I)=XX(L,I)/SQRT  (XX(L,L)*XX(I,I))
      WRITE (6  ,445)                                                   THEI2200
  445 FORMAT(// 6X,39HVARIANCE COVARIANCE MATRIX (ASYMPTOTIC)      )    THEI2210
      DO 450 I=1,MD,8                                                   THEI2220
      NB = I+7                                                          THEI2230
      IF(MD-NB) 447,447,448                                             THEI2240
447   NB=MD
448   WRITE(6,449)(PAMES(J),J=I,NB)
449   FORMAT(/20X,8(A6,6X))
      DO 450 K=1,MD                                                     THEI2280
450   WRITE(6,451)PAMES(K),(XX(K,J),J=I,NB)
451   FORMAT(6X,A6,4X,8G12.4)                                           PFI
  499 GO TO 4                                                           THEI2310
1000  WRITE(6,1001)                                                        PROLO
1001  FORMAT(60X,'END OF RUN')                                             PROLO
      GO TO 9989                                                        PRST2314
      END                                                               THEI2320
      SUBROUTINE SETDR(IOBA,IOBS,NFIR,LTO,MBORT,KA,NY,VARNAM,NBV,DER,
     1XX,AIR,KVR,AV,NPAR,KTAPE,KVAR,PAMES)
      DIMENSION VARNAM(70),PAMES(70)                                    PRSD0030
      DIMENSION NVVP(10),NLQP(10)
      DATA  CONS/6HCONST./                                              PRSD0050
      DIMENSION DER(70),XX(70,70),AIR(10),SPIN(20,120),SPO(70)
      DIMENSION KNUM(7,10),KBB(7),MODE(7),KD(10),           VD(8)
      DIMENSION AV(70),KVR(70,5),KLIST(7,10)
      DIMENSION KVVR(40),LAG(40),KSVE(40),LLAG(40)
      COMMON KVVR,LAG,KSVE,LLAG
      COMMON SPIN,SPO,KNUM,KLIST,VD,KBB,MODE,KD
      REWIND 3
      REWIND  2
      REWIND 1
      MBORT=0
      IOBS = 0                                                          SETD0230
      DO 400 I=1,70
400   AV(I)=0.0
    6 FORMAT(20I4)                                                      SETD0290
      READ(5,6)NST,NSP,KA,NY,MA
      READ(5,6)(KVVR(I),I=1,KVAR)
      READ(5,6)(LAG(I),I=1,KVAR)
```

```
      READ(5,6)(KVR(I,1),I=1,NY)
      READ(5,6)(KVR(I,3),I=1,NY)
      READ(5,6)(NLQP(I),I=2,KA)
      KQ=1
      DO 840 I=2,KA
      KM=NLQP(I)
      DO 841 J=KQ,KM
841   KVR(J,2)=I
840   KQ=KM+1
      DO 120 I=1,KA
      READ(5,6)KB,MODE(I),(KLIST(I,J),KNUM(I,J),J=1,KB)
120   KBB(I)=KB
      IF(NY-70)460,460,5000
5000  MBORT=1
      WRITE(6,5001)
      RETURN
5001  FORMAT(//6X,20HTOO MANY VARIABLES       )
 460  DO 901 I=1,NY                                                     PFI
      DER(I) = 0.0                                                      SETD0390
      DO 901 J=1,NY                                                     PFI
  901 XX (I,J) = 0.0                                                    SETD0420
350   DO 512 I=1,KA                                                     SETD0430
  512 AIR(I) = 0.0                                                      SETD0440
      IF(LTO) 4000,4000,4001
4001  REWIND 3
      READ ( 3 )(SPIN(1,I),I= 1,KTAPE)
      GO TO 4002
4000  READ (  2)(SPIN(1,I),I= 1,KTAPE)
4002  WRITE(6,1116)
1116  FORMAT(///6X,27HMULTIPLE THEILIT  ANALYSIS  //6X,12HVARIABLES    )
      DO 1107 I=1,KVAR
      KSST=KVVR(I)
      IF(KSST)1105,1105,1106
1105  VARNAM(I)=CONS
      GO TO 1107
1106  VARNAM(I)=SPIN(1,KSST)
1107  CONTINUE
1104  FORMAT(3(6X,I4))
      WRITE(6,2000)(VARNAM(I),LAG(I),I=1,KVAR)
2000  FORMAT(8(2X,A8,I2,2X))
      DO 3036 I=1,NY
      L=KVR(I,1)
      K=KVR(I,3)
3036  PAMES(K)=VARNAM(L)
      WRITE(6,2001)
2001  FORMAT(6X,15HPARAMETERS       )
      WRITE(6,1104)((KVR(I,J),J=1,3),I=1,NY)
      WRITE (6  ,101)                                                   SETD0610
  101 FORMAT(/6X,25HDEPENDENT VARIABLE VALUES  )                        SETD0620
      DO 103 I=1,KA
      KB=KBB(I)
      DO 102 J=1,KB
      L=KLIST(I,J)
102   VD(J)=SPIN(1,L)
103   WRITE(6,100)I,MODE(I),(VD(J),KNUM(I,J),J=1,KB)
100   FORMAT(6X,2H= ,I3,6H MODE ,I3,4H IF ,6(A8,3H = ,I2,2X))
```

```
      DO 10 II = 1,NFIR
      IF (LTO) 4010,4010,4011
4011   READ(3)((SPIN(I,J), J=1,KTAPE),I=1,IOBA)
      GO TO 4012
4010   READ(2)((SPIN(I,J), J=1,KTAPE),I=1,IOBA)
4012  DO 10 JPP=NST,NSP
      DO 12 I = 1, KVAR
      L=KVVR(I)
      IF(L)1117,1117,1118
1117  SPO(I)=1.0
      GO TO 12
1118   KSL=JPP+LAG(I)
1123  IF(KSL)1124,1124,1126
1126  IF(KSL-IOBA)1125,1125,1124
1124  SPO(I)=-.09
      GO TO 12
1125  SPO(I)=SPIN(KSL,L)
12    CONTINUE
      JX=KVAR+1
      JP=JX
14    KLAG=0
      DO 110 I=1,KA
      KB=KBB(I)
      DO 90 L=1,KB
      LPZ=KLIST(I,L)
      KD(L)=IFIX(SPIN(JPP,LPZ))
      IF(MODE(I))91,91,92
92    IF(KD(L))91,91,1110
1110  KD(L)=1
91    CONTINUE
90    CONTINUE
       KKKK=0
      DO 1111 L=1,KB
      IF(KD(L)-KNUM(I,L))1112,1111,1112
1112  KKKK=1
1111  CONTINUE
      IF(KKKK)110,1113,110
1113  KLAG=I
110   CONTINUE
      IF(KLAG) 95,95,96                                              SETD0930
95    SPO(JX)=-.09                                                   SETD0940
      GO TO 97                                                       SETD0950
96    SPO(JX)= FLOAT(KLAG)
97    CONTINUE
  311 KLAG=0                                                         SETD1030
      DO 312 J=1,JP                                                  SETD1040
      IF(SPO(J) +.09) 312,313,312                                    SETD1050
  313 KLAG=1                                                         SETD1060
  312 CONTINUE                                                       SETD1070
353   IF(KLAG) 315,315,10                                            SETD1080
  315 WRITE ( 1 )(SPO(J) ,J=1,JP)
      IOBS = IOBS +1                                                 SETD1100
      DO 410 JL=1,JP                                                 SETD1110
410   AV(JL) = AV(JL) + SPO(JL)                                      SETD1120
      L=IFIX(SPO (JX) )
      AIR(L) = AIR(L) + 1.0                                          SETD1190
      DO 1138 I=1,NY                                                  PFI
```

```
      KCD=KVR(I,1)
      IF(KVR(I,2)-L)1136,1137,1136
 1137 KC=KVR(I,3)
      DER(KC)=DER(KC)+SPO(KCD)
 1136 CONTINUE
      DO 1138 J=1,NY                                              PFI
      KCP=KVR(J,1)
 1138 XX(I,J)=XX(I,J)+SPO(KCD)*SPO(KCP)
   10 CONTINUE                                                    SETD1270
      REWIND 1
      REWIND  2
      IF(LTO.GT.0)REWIND 3
      KVAR=JX
      NPAR=MA
   31 RETURN
      END
      SUBROUTINE TSAT   (N,J,NBV,NFIR,LIO)
C     PARAMETERS                                                  TSAT0030
C     1  -  TAPE NUMBER                                           TSAT0040
C     2 NO. OF FINAL VARIABLES                                    TSAT0050
C     3 NO. VARIABLES FROM TAPE                                   TSAT0060
C    4 NO OF TRANSGENERATIONS                                     TSAT0070
C     5 NO OF ROUTINE BAD VALUES                                  TSAT0080
C     6 NUMBER OF SPECIAL BAD VALUES                              TSAT0090
C     7  START NUMBER                                             TSAT0100
C     8  -  STOP NUMBER                                           TSAT0110
C     9 NUMBER OF FIRMS                                           TSAT0120
C     10  -  NUMBER OF LINES OF DATA PRINTED.                     TSAT0130
C     11 - NUMBER OF FORMAT CARDS
C     2 -  VARIABLE  NUMBERS (18I4)                               TSAT0140
C     VARIABLE NAMES  8(A6,4X)                                    TSAT0150
C     3 - BAD VALUE CARDS - FORMAT (8F10.5)                       TSAT0160
C     SPECIAL BAD VALUE CARD(S)         (I4,F10. )                TSAT0170
C     4 - TRANSGENERATION  CARDS -  FORMAT 3I4,F10.0              TSAT0180
C     PLACE, OPERATION, FIRST CONTROL , SECOND CONTROL            TSAT0190
C     REPLACES ALL ILLEGAL TRANSGENERATIONS BY -.09               TSAT0200
C     REPLACES ALL BAD VALVES  BY -.09                            TSAT0210
C     TRANSGENERATION CODES                                       TSAT0220
C       1 =   A+K                  7 = LOGF(A)        14 = A/B    TSAT0230
C       2 =   A-K                  8 = 1/A            15 =A**B    TSAT0240
C       3 =   A*K                  9 = EXPF(A)        16=X=1 IF A GREATER  TSAT0250
C       4 =   A/K                 10 = ABSF(A)           OR EQUAL B        TSAT0260
C       5 =   A**K               11 = A+B                X=0 IF A LESS B   TSAT0270
C       6 = X=1 IF A ABOVE K     12 = A-B             17=ARITH REG GR      TSAT0280
C           X=0 IF A BELOW K     13 = A*B             18=LOG  REG  GR      TSAT0290
C           X=1 IF A =K                               19=RATIO GR          TSAT0300
C     20  -  LAG ARITHMETIC GROWTH RATE
C     21  -  LAG LOG REG. GROWTH RATE
C     22  -  LAG RATIO GROWTH RATE
C     23  -  AVERAGE OVER PERIOD
C     24  -  STANDARD DEVIATION OVER PERIOD
C     25  -  LOWEST TO AVERAGE OVER PERIOD
C     26  -  CURRENT TO AVERAGE OVER N YEARS
C     27   -   LAG 1
C     28   -  LAG  2
C     29 -- SQRT(A)
C     30  --   MINIMUM/AVERAGE
```

```
C       31  --  RECENT/AVERAGE
C       32   -- TREND VALUE
        COMMON TITLE
        DIMENSION FMT(80),TITLE(11)
        DIMENSION BVSS(20),KSBVS(20)
        DIMENSION BV(104) ,SPIN(20,59),VVV(20), ILT(200,3)
        DIMENSION  TLF(200),LVAR(100),LTAPE(100)
        DIMENSION ILG( 80,8)
        COMMON  ILT,VVV
        COMMON  BV      , SPIN    , SPO     , TLF     , LVAR    , BVSS
        COMMON  KSBVS
        COMMON FMT,LTAPE
        DIMENSION SPO(20,120),MODE(200)
        EQUIVALENCE(ILG(1),ILT(1))
        NIT=5
        REWIND  2
        READ(5,1        )NBT,J,NTAPE,NTG,NBV,NBVS,NSTRT,NSTOP,NFIR,NRT,NFMT
      1,LTO,MLIN
        WRITE(6,781    )NBT,J,NTAPE,NTG,NBV,NBVS,NSTRT,NSTOP,NFIR,NRT,NFMT
      1,LTO   ,MLIN
  781   FORMAT(//6X,10HPARAMETERS  //(6X,18I4))
      1 FORMAT(20I4)
         IF(LTO-2)4023,4024,4024
 4024    WRITE(6,4025)
         REWIND 3
 4025   FORMAT(///6X,25HDATA FROM PREVIOUS RUN          )
        READ(3)(SPO(1,MM),MM=1,J)
        WRITE(6,4030)(SPO(1,MM),MM=1,J)
 4030   FORMAT(//6X,9HVARIABLES/9(7X,A6))
        IF(NRT) 4068,4068,4069
 4069   WRITE(6,23)NRT
        DO 4031 IK=1,NRT
         READ(3)((SPO (I,JJ),JJ=1,J),I=NSTRT,NSTOP)
        DO 4031 I=NSTRT,NSTOP
 4031    WRITE(6,42)(SPO(I,MM),MM=1,J)
 4068   REWIND 3
        GO TO 788
 4023   CONTINUE
        IF(NFMT)767,767,768
  768   KFMT=NFMT*10
         READ(5,770)(FMT(I),I=1,KFMT)
        WRITE(6,780  )(FMT(I),I=1,KFMT)
  770   FORMAT(10A8)
  780   FORMAT(///6X,17HFORMAT FOR DATA   ///(6X,10A8))
        GO TO 769
  767   REWIND 4
  769   WRITE (6,10)
        WRITE (6,19)
        DO 455  L = 1,J
        READ (5,456)  KC ,PNMES,  (TITLE(MM),MM = 1,11)
  457   FORMAT (6X,I4,2X,A6,2X,11A6)
  456   FORMAT(I4,2X,12A6)
        SPO(1,KC) = PNMES
  455   WRITE (6,457) KC,PNMES,  (TITLE(MM),MM = 1,11)
        IF(LTO)4000,4000,4001
 4001   REWIND 3
        WRITE(3)(SPO(1,MM),MM=1,J)
```

```
      GO TO 4002
4000  WRITE( 2)(SPO(1,MM),MM=1,J)
4002   IF(NTAPE.EQ.0) GO TO 4987
      WRITE (6, 17)
      DO 458 L=1,NTAPE
      READ(5,1       )LVAR(L),LTAPE(L)
458   WRITE(6,18    )LVAR(L),LTAPE(L)
18    FORMAT (6X,I4,6X,I4)
    3 FORMAT(8(A6,4X))                                                  TSAT0540
4987  IF (NBV) 6,6,4                                                    TSAT0550
 4      READ(5,5)(BV(I),I=1,NBV)
    5 FORMAT(8F10.5)                                                    TSAT0570
    6 IF(NBVS) 43,43,44                                                 TSAT0580
   44 READ (5   ,131)(KSBVS(I),BVSS(I),I=1,NBVS)                        TSAT0590
  131 FORMAT(I4,F10.5)                                                  TSAT0600
   43 IF (NTG) 9,9,7                                                    TSAT0610
    7 READ (5   ,8)((ILT(I,JK),JK=1,3),TLF(I),MODE(I),I=1,NTG)          TSAT0620
    8 FORMAT(3I4,F10.0,3X,I1)                                           TSAT0630
   10 FORMAT(/6X,12HDATA FOR RUN )                                      TSAT0650
    9 IF(NBV) 613,613,612                                               TSAT0660
  612 WRITE (6   ,13)                                                   TSAT0670
   13 FORMAT( /6X, 51HTHE FOLLOWING VALUES FLAG UNAVAILABLE OBSERVATIONSTSAT0680
     1   )                                                              TSAT0690
      WRITE ( 6 ,14)(BV(I),I=1,NBV)                                     TSAT0700
   14 FORMAT(40X, F12.6 )                                               TSAT0710
  613 IF(NBVS) 119,119,12                                               TSAT0720
   12 WRITE (6   ,45)(KSBVS(I),BVSS(I),I=1,NBVS)                        TSAT0730
   45 FORMAT(//6X,19HSPECIAL BAD VALUES   /(6X,I4,2X,G18.7)    )        TSAT0740
  119 IF(NTG) 614,614,615                                               TSAT0750
  615 WRITE (6,15)                                                      TSAT0760
   15 FORMAT ( /6X,17HTRANSGENERATIONS ,12X,5HPLACE,6X,3HOP.,5X,6HA VAR.TSAT0770
     1,9X,6HB VAR. )
      WRITE(6,16)((ILT(I,JK),JK=1,3),TLF(I),MODE(I),I=1,NTG)
16    FORMAT(36X,I3,7X,I3,7X,I3,7X,F11.3,4X,I2)
   17 FORMAT( /6X, 27HVARIABLES LIFTED FROM TAPE   )                    TSAT0820
   19 FORMAT( /6X , 15HVARIABLE NAMES   )                               TSAT0860
   20 FORMAT(   9(7X,A6)      )                                         TSAT0890
614   IF (NRT) 21,21,22                                                 TSAT0900
   21 NRT = 5                                                           TSAT0910
   22 WRITE ( 6 ,23)NRT                                                 TSAT0920
   23 FORMAT (//6X, 6HFIRST ,I3,3X, 12HOBSERVATIONS )                   TSAT0930
      KLOK=0                                                            TSAT0940
C                                                                       PRSP0250
C CHECK FOR CARD INPUT. IF CARD INPUT SKIP END OF FILE CHECK            PRSP0260
C                                                                       PRSP0270
      IF(NBT-5)6001,1003,6001                                           PRSP0280
 6001 ASSIGN 8777 TO KKK                                                PRSP0290
C                                                                       PRSP0281
C END OF FILE CHECK ALTERED FOR CDC 6400                                PRSP0291
C                                                                       PRSP0300
      IF(EOF,4)1003,8777
 1003 DO 24  IK= 1,NFIR                                                 TSAT0950
      IF(NBT-5)121,777,121
777   READ(5,FMT)((SPIN(I,KKK),KKK=1,NTAPE),I=NSTRT,NSTOP)
      GO TO 778
121   IF(MLIN)8888,8888,8889
```

```
8889   DO 8989 I=NSTRT,NSTOP
 8989 READ(4)(SPIN(I,KKK),KKK=1,MLIN)                                   PRSA0970
      GO TO 778
 8888 READ(4)SPIN                                                       PRSA0972
 778  DO 3000 JX=NSTRT,NSTOP
      DO 26 I=1,NTAPE                                                   TSAT0980
      K= LVAR(I)                                                        TSAT0990
      KLM=LTAPE(I)
      SPO(JX,K)=SPIN(JX,KLM)
  123 IF(NBV) 26,26,25                                                  TSAT1010
 25   DO 789 KR=1,NBV
      IF(SPO(JX,K)-BV(KR))789,27,789
   27 SPO(JX,K)= -.09                                                   TSAT1040
 789  CONTINUE
 26   CONTINUE
       IF(NBVS.EQ.0) GO TO 5000
      DO 5001  I=1,NBVS
      K=KSBVS(I)
 5001  IF(SPO(JX,K).EQ.BVSS(I)) SPO(JX,K)=-.09
C
C     TRANSFORMATION LOOP
C
5000   IF(NTG)3000,3000,29
   29 DO 30 I= 1,NTG                                                    TSAT1070
      KA= ILT(I,1)                                                      TSAT1080
      KB= ILT(I,2)                                                      TSAT1090
      KC= ILT(I,3)                                                      TSAT1100
      IF(MODE(I)) 801,801,800
 800   C = SPIN(JX,KC)
      GO TO 803
 801  C =  SPO (JX,KC)
 803  IF (KB -10) 32,32,31
   31 KD=IFIX(TLF(I))                                                   TSAT1130
      IF(KB.GE.17) GO TO 32
      IF(MODE(I)) 805,805,804
 804  D = SPIN (JX,KD)
      GO TO  806
 805  D = SPO  (JX,KD)
 806  IF (  D     +.09) 32,34,32                                         TSAT1140
 34    SPO(JX,KA)=-.09
      GO TO 30                                                          TSAT1160
 32   IF(KB.GT.16.AND.KB.LT.27)GO TO 36
      IF(      C   +.09) 36,34,36
   36 GO TO(110,120,130,140,150,155,160,170,180,190,210,220,230,         TSAT1180
     1240,250,260,270,280,290,270,280,290,400,400,400,400,420,421,422,  TSAT1190
     21200,1200,1200),KB
 1200   KXX=JX-KD
      KLAG=0
      DO 1201 K=KXX,JX
      IF(MODE(I))1202,1202,1203
 1202  VVV(K)=SPO(K,KC)
      GO TO 1204
 1203  VVV(K)=SPIN(K,KC)
 1204  IF(VVV(K)+.09)1201,1205,1201
 1205 KLAG=1
 1201 CONTINUE
```

```
      IF(KLAG)1206,1206,34
1206  TVAR=0.0
      BETA=0.0
      YT=0.0
       YVAR=0.0
      COSSS=0.0
        DO 1207 K=KXX,JX
      TVAR=TVAR+FLOAT(K)
      BETA=BETA+FLOAT(K*K)
      YBAR=YBAR+VVV(K)
1207  YT=YT+VVV(K)*FLOAT(K)
      YVAR=FLOAT(JX-KXX+1)
      YT=YT-TVAR*YBAR/YVAR
       BETA=BETA-TVAR*TVAR/YVAR
      BETA=YT/BETA
      TVAR=TVAR/YVAR
      YBAR=YBAR/YVAR
      IF(KB-32)1208,1208,1209
1208  DO 1210 K=KXX,JX
1210        VVV(K)=VVV(K)-YBAR-BETA*(FLOAT(K)-TVAR)
      IF(KB-31)1211,1212,1212
1212  SPO(JX,KA)=VVV(JX)/YBAR
      GO TO 30
1211  YT=VVV(KXX)
      DO 1213 K=KXX,JX
      IF(VVV(K)-YT)1214,1213,1213
1214  YT=VVV(K)
1213  CONTINUE
      SPO(JX,KA)=YT/YVAR
       GO TO 30
1209  SPO(JX,KA)=YBAR+BETA*(FLOAT(K)-TBAR)
      GO TO 30
420   IF(JX-1)34,34,423
423   KLP=JX-1
      GO TO 424
421   IF(JX-2)34,34,425
425   KLP=JX-2
424   SPO(JX,KA)=SPO(KLP,KC)
      IF(MODE(I).EQ.1)SPO(JX,KA)=SPIN(KLP,KC)
      GO TO 30
422   IF(C)34,426,426
426   SPO(JX,KA)=SQRT(C)
      GO TO 30
428   SPO(JX,KA)=SPIN(KLP,KC)
      GO TO 30
400    KX=JX-KD+1
      KLLL=0
      IF(KX) 34,34,401
401   AVV = 0.0
      SDD = 0.0
      DO 402 K = KX,JX
      IF(MODE(I))403,403,404
404   DCC=SPIN(K,KC)
      GO TO 405
403   DCC=SPO(K,KC)
405   IF(K-KX)406,406,407
```

```
406    PIN = DCC
407    AVV = AVV+DCC
       SDD = SDD + DCC*DCC
       IF(NBV.EQ.0) GO TO 4687
       DO 4688 LLF=1,NBV
4688    IF(DCC.EQ.BV(LLF)) KLLL=1
4687   IF(DCC-PIN)408,408,402
408    PIN = DCC
402    CONTINUE
       IF(KLLL.EQ.1) GO TO 34
       AVV = AVV/FLOAT(JX-KX+1)
       KP = KB-22
       GO TO (409,410,411,412),KP
409    SPO(JX,KA)=AVV
       GO TO 30
410    SPO(JX,KA) = SQRT((SDD-AVV*AVV*FLOAT(JX-KX+1))/FLOAT(JX-KX))
       GO TO 30
411    SPO(JX,KA) = PIN/AVV
       GO TO 30
412    SPO(JX,KA) = DCC/AVV
       GO TO 30
   110 SPO(JX,KA) = C       + TLF(I)
       GO TO 30
   120 SPO(JX,KA) = C -TLF(I)
       GO TO 30
   130 SPO(JX,KA) = C* TLF(I)
       GO TO 30
   140 IF(TLF(I) ) 141, 34, 141
   141 SPO(JX,KA) = C/ TLF(I)
       GO TO 30
   150 IF(C)        34, 151, 151
   151 SPO(JX,KA) = C     ** TLF(I)
       GO TO 30
   160 IF      (C) 34,34   ,  161
161    SPO(JX,KA)=ALOG(C)
       GO TO 30
   170 IF      (C)  171,34  , 171
   171 SPO(JX,KA) = 1.0/C
       GO TO 30
   180 SPO(JX,KA) = EXP(C)
       GO TO 30
   155 IF(C -TLF(I)) 156,157,157
   156 SPO(JX,KA) = 0.0
       GO TO 30
   157 SPO(JX,KA) = 1.0
       GO TO 30
   190 SPO(JX,KA) = ABS(C)
       GO TO 30
   210 SPO(JX,KA) = C+D
       IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
       GO TO 30
   220 SPO(JX,KA) = C-D
       IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
       GO TO 30
   230 SPO(JX,KA) = C*D
       IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
```

```
      GO TO 30
  240 IF(D)241,34,241
  241 SPO(JX,KA) = C / D
      IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
      GO TO 30
  250 IF(C)34,251,251
  251 SPO(JX,KA) = C**D
      GO TO 30
  260 IF(C-D)261,262,262
  261 SPO(JX,KA) = 0.0
      GO TO 30
  262 SPO(JX,KA) = 1.0
      GO TO 30
  270 KXX=JX
      IF(KB-20) 310,311,310
  311 KXX=KXX-1
  310 KX=KXX-KD
      IF(KX) 274,274,271
  271 TVAR= 0.0
      BETA= 0.0
      YBAR= 0.0
      YT  = 0.0
      YVAR= 0.0
      DO 272 K=KX,KXX
      BETA= BETA +1.0
      TVAR= TVAR + BETA* BETA
      YBAR=YBAR+SPIN(K,KC)
      YT=YT+SPIN(K,KC)*BETA
  272 YVAR=YVAR+SPIN(K,KC)*SPIN(K,KC)
      TBAR= (BETA+1.0)/2.0
      YBAR= YBAR/BETA
      YVAR= YVAR/BETA - YBAR * YBAR
      TVAR= TVAR/BETA - TBAR * TBAR
      YT  = YT/BETA -YBAR*TBAR
      IF(YBAR)274,274,273
  274 SPO(JX,KA) = -.09
      KA=KA+1
      SPO(JX,KA) = -.09
      GO TO 30
  273 KLAG= 0
      DO 276 K=KX,KXX
      DO 276 L=1,NBV
      IF(SPIN(K,KC)-BV(L)) 276,277,276
  277 KLAG = 1
  276 CONTINUE
      IF(KLAG) 275,275,274
  275 SPO(JX,KA) = YT/(TVAR * YBAR )
      KA= KA +1
      SPO(JX,KA) = (YVAR - (YT*YT)/TVAR)/(YBAR*YBAR)
      GO TO 30
  280 KLAG=0
      KXX=JX
      IF(KB-21)  312,313,312
  313 KXX=KXX-1
  312 KX=KXX-KD
      IF(KX)   274,274,281
```

```
  281 DO 282 K=KX,KXX
      IF(SPIN(K,KC)) 2831,2831,283
 2831 KLAG = 1
      GO TO 282
  283 DO 282 L=1,NBV
      IF(SPIN(K,KC)-BV(L)) 282,284,282
  284 KLAG= 1
  282 CONTINUE
      IF(KLAG) 285,285,274
  285 IF(KXX-KX-2) 274,287,287
  287 TVAR=0.0
      BETA = 0.0
      YBAR = 0.0
      YT = 0.0
      YVAR = 0.0
      DO 286 K= KC,KD
      BETA = BETA + 1.0
      YLOG=ALOG(SPIN(K,KC))
      TVAR = TVAR + BETA * BETA
      YBAR = YBAR + YLOG
      YT =  YT + YLOG * BETA
  286 YVAR = YVAR + YLOG * YLOG
      TBAR = (BETA + 1.0) / 2.0
      YBAR = YBAR/BETA
      YVAR = YVAR/BETA - YBAR*YBAR
      TVAR = TVAR/BETA - TBAR*TBAR
      YT   = YT/BETA  -  YBAR* TBAR
      SPO(JX,KA) = YT/TVAR
      KA=KA + 1
      SPO(JX,KA) = YVAR - YT*YT/TVAR
      GO TO 30
  290 KLAG = 0
      KXX=JX
      IF(KB-22) 314,315,314
  315 KXX=KXX-1
  314 KX=KXX-KD
      IF(KX) 274,274,316
316   DO 1100 L=KX,KXX
      IF(MODE(1))1101,1101,1102
1101  VVV(L)=SPO(L,KC)
      GO TO 1100
1102  VVV(L)=SPIN(L,KC)
1100  CONTINUE
      DO 291 L=1,NBV
      IF(VVV(KX)     -BV(L)) 292,293,292
  292 IF(VVV(KXX)     -BV(L)) 291,293,291
  293 KLAG= 1
  291 CONTINUE
      IF(KLAG) 294,294,274
  294 IF(VVV(KX))       274,274,295
  295 IF(VVV(KXX))       274,274,296
  296 TR=FLOAT(KXX-KX)
      GRL=(ALOG(VVV(KXX)/VVV(KX)))/TR
      SPO(JX,KA)=EXP(GRL)-1.0
       IF(SPO(JX,KA)+.09)370,371,370
371   SPO(JX,KA)=SPO(JX,KA)+.001
370   KA = KA + 1
```

```
      IF(KXX-KX-1) 304,304,305                                          TSAT2650
  305 KE=KX+1                                                           TSAT2660
      YVAR= 0.0                                                         TSAT2670
      KLAG= 0                                                           TSAT2680
      DO 306 K=KE,KXX                                                   TSAT2690
      M= K-1                                                            TSAT2700
      DO 298 L= 1,NBV                                                   TSAT2710
      IF(VVV(K)    -BV(L)) 298,299,298
  299 KLAG= 1                                                           TSAT2730
  298 CONTINUE                                                          TSAT2740
      IF(KLAG) 300,300,301                                              TSAT2750
  300 IF(VVV(K))      301,301,302
302   TBAR=ALOG(VVV(K))-ALOG(VVV(M))-GRL
      YVAR= YVAR + TBAR * TBAR                                          TSAT2780
      GO TO 306                                                         TSAT2790
  301 KLAG= 1                                                           TSAT2800
  306 CONTINUE                                                          TSAT2810
      IF(KLAG) 303,303,304                                              TSAT2820
  304 SPO(JX,KA) = -.09                                                 TSAT2830
      GO TO 30                                                          TSAT2840
  303 SPO(JX,KA) = YVAR/(TR - 1.0)                                      TSAT2850
   30 CONTINUE                                                          TSAT2860
3000   CONTINUE
28    CONTINUE
      IF(LTO)4010,4010,4011
4011  WRITE(3)((SPO(III,JJJ),JJJ=1,J),III=NSTRT,NSTOP)
      GO TO 4012
4010  WRITE(2)((SPO(III,JJJ),JJJ=1,J),III=NSTRT,NSTOP)
4012  IF(IK-NRT)40,40,24
   40 DO 807 K = NSTRT,NSTOP
  807 WRITE (6,42) (SPO(K,I),I =1,J)
   42 FORMAT(/ (9E13.3))
   24 CONTINUE                                                          TSAT2910
      REWIND   2
      IF(NBT-5)787,788,787
787   REWIND 4                                                          TSAT2930
788   N=NSTOP - NSTRT + 1
      IF(LTO) 4061,4061,4060
4060  REWIND 3
4061  CONTINUE
   61 RETURN                                                            TSAT2950
8777  WRITE(6,8778)IK,I
8778  FORMAT(6X,12HEND OF FILE      ,2I6)
      NFIR=IK
      DO 8779 JJJJ=I,NSTOP
      DO 8779 KKK=1,MLIN
8779  SPIN(JJJJ,KKK)=-.09
      GO TO 778
      END
      SUBROUTINE HEAD(NAME,DATE,KAGE,NORUN,NOSUB,PRONAM)                HEAD0020
C     WRITES PAGE HEADING                                               HEAD0030
C     PARAMETERS -  1 - PROGRAM NAME                                    HEAD0040
C                   2 - DATE                                            HEAD0050
C                   3 - PAGE NUMBER                                     HEAD0060
C                   4 - RUN  NUMBER                                     HEAD0070
C                   5 - SUB NO                                          HEAD0080
C                   6 - PROBLEM NAME                                    HEAD0090
```

```
      DIMENSION DATE(2),PRONAM(12)                                       PRSH0110
      KAGE = KAGE +1                                                     HEAD0120
      WRITE(6,1)NAME,(DATE(I),I=1,2),NORUN,NOSUB,KAGE,NORUN,   (PRONAM(I PRSH0130
     1),I=1,12)                                                          PRSH0140
    1 FORMAT( 1H1,5X,8HPROGRAM ,A6,14H (J.G. CRAGG)   ,                  HEAD0150
     1 13X,A6,A2,40X,4HRUN    ,I3,7H    SUB  ,I3,7H   PAGE  ,I3 /        PRSH0160
     2 6X, 4HRUN  ,I3,20X,12A6)                                          PRSH0170
      RETURN                                                             HEAD0180
      END                                                                HEAD0190
      SUBROUTINE SUBHED (NOS   ,SUBNAM,KOPT)                             SUBH0020
      DIMENSION SUBNAM(8)                                                SUBH0030
      WRITE(6,1)NOS,KOPT,(SUBNAM(I),I=1,8)                               SUBH0050
    1 FORMAT (/6X,11HSUBPROBLEM   ,I3,8H   OPTION,I3,6X,8A6)             SUBH0060
      RETURN                                                             SUBH0070
      END                                                                SUBH0080
      SUBROUTINE INVERT(A,DET,N,NMAX)
      LOGICAL COL
      DIMENSION A(NMAX,NMAX),NROW(100),COL(100)
      COMMON /SCRTCH/COL,NROW
      DET=1.0
      DO 5 I=1,N
      NROW(I)=I
    5 COL(I)=.FALSE.
      DO 40 I=1,N
      PIVOT =0
      DO 15 J=1,N
      IF (COL(J)) GO TO 15
      DO 10 K=1,N
      IF(COL(K).OR.(ABS(A(J,K)).LT.ABS(PIVOT)))GO TO 10
      PIVOT=A(J,K)
      JROW=J
      KCOL=K
   10 CONTINUE
   15 CONTINUE
      DET=DET*PIVOT
      IF(JROW.EQ.KCOL)GO TO 25
      DET=-DET
      NEMP=NROW(KCOL)
      NROW(KCOL)=NROW(JROW)
      NROW(JROW)=NEMP
      DO 20 K=1,N
      TEMP=A(JROW,K)
      A(JROW,K)=A(KCOL,K)
   20 A(KCOL,K)=TEMP
   25 COL(KCOL)=.TRUE.
      A(KCOL,KCOL)=1.0
      DO 30 K=1,N
      A(KCOL,K)=A(KCOL,K)/PIVOT
      GO TO 30
 30   CONTINUE
      DO 40 J=1,N
      IF (J.EQ.KCOL) GO TO 40
      TEMP=A(J,KCOL)
      A(J,KCOL)=0.
      DO 35 K=1,N
   35 A(J,K)=A(J,K)-A(KCOL,K)*TEMP
   40 CONTINUE
```

```
      DO 60 J=1,N
      IF(NROW(J).EQ.J) GO TO 60
      DO 45 K=J,N
      I=K
      IF(NROW(K).EQ.J) GO TO 50
   45 CONTINUE
   50 DO 55 K=1,N
      TEMP=A(K,J)
      A(K,J)=A(K,I)
   55 A(K,I)=TEMP
      NROW(I)=NROW(J)
   60 CONTINUE
      RETURN
      END
         END OF RECORD
```

CD TOT   0977

APPENDIX III

The listing below is the STAT3 program developed by the author to assess the multi-mode models which the previous program generated. The program is a more sophisticated version of a binary testing program developed by Dr. Peter Stopher for logit, probit and discriminant models.

This program has one major fault in that it consumes a disproportionate amount of time for the usefullness which is derived. It uses many times as much central memory time as the previous program, and as the size of the models increases, the central memory core that is required also increases greatly. Improved computer techniques could probably improve these deficiencies to a great extent. Also the use of smaller data sets than the 400 used in this case would bring these parameters down in magnitude.

```
      PROGRAM STAT3(INPUT,OUTPUT,TAPE1,TAPE2,TAPE5=INPUT,TAPE6=OUTPUT)   ING00010
C PROGRAM TO ANALYZE THEIL MODELS ON A LOGIT FRAMEWORK   APRIL 9/71
C
C NO=NUMBER OF OBSERVATIONS
C LI=THE CHOICE VARIABLE NUMBER
C VAL(I)=INTEGER VALUES TO FLAG EACH CHOICE IN ORDER
CK=NUMBER OF CHOICES
C L=NUMBER OF VARIABLES
C L=NUMBER OF COEFFICIENTS
C NFMT=NUMBER OF FORMAT CARDS
C  XDAT IS THE TITLE
C XMODES ARE THE MODE TITLES
C N(I) ARE THE NUMBERS OF THE LAST COEFFICIENTS FOR EACH EQUATION
C K4=K-1
CK5=K-2
      DIMENSION Z(400),CH(400),DEP(400),XXXX(400),KRITE(10)
      DIMENSION VARNAM(20),NCO(50),VAR(50),Y(10),COEFF(50)
      DIMENSION N(10),F(10),ETA2(10),NU1(10),NU2(10)
      DIMENSION  Y3(10)
      DIMENSION KH(10)                                                   ING00060
      DIMENSION VAL(10), Y1(10)                                          ING00070
      DIMENSION VARI(40)
      DIMENSION XDAT(60)                                                 ING00080
      DIMENSION Y2(10)
      DIMENSION XMODES(10)                                               ING00090
      DIMENSION FMT(80)
      DIMENSION KCNT(5),KVAR(5),JR(10)                                   ING00110
      COMMON XXX(50000)
      PRINT 91                                                           ING00117
      PRINT 300
  300 FORMAT(///25X,*$$$$$$$$$$     $$$$$$$$$$     $$$$$$$$$$     $$$$$$$$$
     1$     $$$$$$$$$$$$*/25X,*$$$$$$$$$$     $$$$$$$$$$     $$$$$$$$$$
     2$$$$$$$$$$     $$$$$$$$$$$*/25X,*$$      $$        $$        $$
     3  $$        $$        $$        $$*/25X,*$$                  $$
     4  $$      $$        $$                $$*/25X,*$$
     5$$       $$       $$        $$                $$*/25X,*$$
     6      $$        $$       $$        $$               $$*/25X,*$$
     7           $$        $$       $$        $$               $$*/25X,*$$
     8                $$        $$       $$        $$              $$*)
      PRINT 301
  301 FORMAT(25X,*$$$$$$$$$$*,8X,*$$*8X,*$$$$$$$$$$*,8X,*$$*11X,*$$$$$$
     1$$$$*/25X,*$$$$$$$$$$$*8X,*$$*,8X,*$$$$$$$$$$$*,8X,*$$*,       10X,
     2*$$$$$$$$$$$*/33X,*$$*,8X,*$$*8X,*$$*6X,*$$*,8X,*$$*,19X,*$$$*/33X
     3,*$$*,8X,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,20X,*$$*/33X,*$$*,8X,*$$*,8X,*$$*,6X
     4,*$$*,6X,*$$*,8X,*$$*,20X,*$$*/33X,*$$*,8X,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,
     5*$$*,20X,*$$*/33X,*$$*,8X,*$$*,8X,*$$*6X,*$$*,8X,*$$*,20X,*$$*/33X
     6,*$$*,         8X,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,20X,*$$*/33X,*$$*,8X,*$$*,8X
     7,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,20X,*$$*/25X,*$$*,6X,*$$*,8X,*$$*,8X
     9,*$$*,6X,*$$*,8X,*$$*,10X,*$$*,8X,*$$*)
      PRINT 302
  302 FORMAT(25X,*$$$$$$$$$$$*,8X,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,10X,*$$$$$
     2$$$$$$$*/25X,*$$$$$$$$$$$*,8X,*$$*,8X,*$$*,6X,*$$*,8X,*$$*,10X,*$$$
     3$$$$$$$$$*)
   91 FORMAT(18X,* PROGRAM STAT3 DEVELOPED BY PAUL INGLIS IN THE DEPARTMING00118
     1ENT OF CIVIL ENGINEERING AT MCMASTER UNIVERSITY/ MAY , 1971*//14X,ING00118
     2* THIS PROGRAM ANALYSES MULTINOMIAL LOGIT MODELS STATISTICALLY ANDING00118
     3 HAS BEEN USED ON A CDC 6400*)                                     ING00118
      READ 41,KSUBS,NEWS,NBV,NSBV
```

```
      KPAUL=KSUBS
 93   CONTINUE
      REWIND 1
      READ 107,NTG,LA
 107  FORMAT(4I4)
      READ(5,3)NO,K,L,NFMT,(XDAT(IZ),IZ=1,10)                           ING00120
      PRINT 92,XDAT(1)
 92   FORMAT(1H1,10X,* THE DATE IS    *,A6,*    FOR THIS RUN *//)       ING00130
      IF(NEWS.LT.10.)NEWS=10                                            ING00140
 41   FORMAT(5I4)                                                       ING00150
      WRITE(6,22)(XDAT(IZ),IZ=2,10)                                     ING00160
      PRINT 42,NO,K,L,NFMT,KSUBS,NEWS                                   ING00170
 42   FORMAT(/10X,* NUMBER OF OBSERVATIONS =*,I4/10X,* NUMBER OF ALTERNAING00175
     1TIVES =*,I4/10X,* NUMBER OF COEFFICIENTS =*,I4/10X,* NUMBER OF FORING00180
     2MAT CARDS =*,I4/10X,* NUMBER OF SUBPROBLEMS  =*,I4/10X,* NUMBER TOING00190
     3 BE PRINTED    =*,I4)                                             ING00200
 3    FORMAT(X,4I4,10A6)                                                ING00210
 22   FORMAT(/17X,10A6)                                                 ING00220
      IF(KPAUL.NE.KSUBS)GO TO 2001
      NI=NO
 2000 KFMT=NFMT*10
      READ(5,4)(FMT(I),I=1,KFMT)                                        ING00240
      WRITE(6,66)(FMT(I),I=1,KFMT)                                      ING00250
 66   FORMAT(/10X,*  THE FORMAT FOR DATA IS *,10A8)                     ING00260
  4   FORMAT(10A8)                                                      ING00270
      READ(5,25)(XMODES(IT),IT=1,10)                                    ING00300
 25   FORMAT(10A8)                                                      ING00310
 2001  CONTINUE
      NO=NI
      READ(5,35)LI,(VAL(I),I=1,K)                                       ING00320
 35   FORMAT(I4,19F4.0)                                                 ING00335
      LZ=L+1                                                            ING00340
      READ 47,(NCO(I),I=1,LA)
      PRINT 403,(NCO(I),I=1,LA)
 403  FORMAT(//* THE VARIABLES IN ORDER ARE */2X,20I4//)
      NOO=0
      LN=LA+1
 47   FORMAT(20I4)                                                      ING00360
      IF(KPAUL.NE.KSUBS)GO TO 2002
 2003 DO 40 IV=1,NO
      READ FMT,(VAR(JJ),JJ=1,LZ)                                        ING00380
      DEP(IV)=VAR(LI)
      DO 406 IVV=1,K
      IF(VAL(IVV).EQ.DEP(IV))GO TO 407
 406  CONTINUE
      NOO=NOO + 1
      GO TO 40
 407  CONTINUE
      WRITE(1)(VAR(K),K=1,LZ)
 40   CONTINUE                                                          ING00490
      NO=NO-NOO
      REWIND  1
C THIS CARD ORDERS THE VARIABLES TO THE COEFFICIENTS                    ING00440
      IF(NBV.EQ.0.AND.NSBV.EQ.0)  GO TO 81
      CALL BAVALU(NBV,NSBV,LZ,NO,VAR)
 81   CONTINUE
```

```
      LZ=LZ+NTG
      IAB=NO*LZ + 1
      IAC=IAB
      IAD=IAC + LZ
      IAE=IAD + NO
      IAF=IAE + LZ
      IAG=IAF + NO
      IAH=IAG+LZ*LZ
      IF(NTG)80,80,82
 82   CALL TRANS(NTG,NO,LZ,VAR,XXX(1),XXX(IAC))
 80   CONTINUE
      WRITE(6,45)NEWS                                                   ING00280
 45   FORMAT(//*  THE FIRST *,I4,* OBSERVATIONS*//)                     ING00290
      REWIND 1
      DO 2005 IN=1,NO
      READ(1)(VAR(K),K=1,LZ)
      WRITE(2)(VAR(K),K=1,LZ)
 2005 CONTINUE
      REWIND 1
      DO 95 IV=1,NO
      READ(1)(VAR(K),K=1,LZ)
      IF(NEWS-IV)49,43,43                                               ING00450
 43   PRINT 44,(VAR(J),J=1,LZ)
 44   FORMAT(/6E20.5/6E20.5/6E20.5/6E20.5/6E20.5/6E20.5/6E20.5)
 95    CONTINUE
 49   CONTINUE
 2002 CONTINUE
      REWIND 2
      REWIND 1
      DO 400 IV=1,NO
      READ(2)(VARI(JJ),JJ=1,LZ)
      DEP(IV)=VARI(LI)
      DO 48 J=1,LA
      NCP=NCO(J)                                                        ING00410
      VAR(J)=VARI(NCP)
      IF(NCP.EQ.0.)VAR(J)=1.
 48    CONTINUE                                                         ING00430
      VAR(LN)=DEP(IV)
      WRITE(1)(VAR(JJ),JJ=1,LN)
 400  CONTINUE
      REWIND 2
      REWIND 1
      CALL STAT(LN,NO,XXX(IAD),XXX(IAE),XXX(IAF),XXX(IAG),XXX(IAH),DEP) ING00509
      READ 51,(VARNAM(J),J=1,LA)                                        ING00510
 51   FORMAT(10A8)                                                      ING00520
      READ 2,(COEFF(II),II=1,LA)
      K4=K-1                                                            ING00550
      K5=K-2                                                            ING00560
      READ(5,5)(N(III),III=1,K4)                                        ING00570
 5    FORMAT(6X,10I4)                                                   ING00580
 73   FORMAT(/2X,* MODE *,I4,*  DEVELOPS TO COEFFICIENT *,I4/)
      PRINT 73,(I,N(I),I=1,K4)                                          ING00531
      DO 69 II=1,LA
      PRINT 52,VARNAM(II),COEFF(II)                                     ING00534
 52   FORMAT(12X,*THE COEFFICIENT OF *A8,*   IS   *,E20.5)              ING00536
 69   CONTINUE                                                          ING00535
```

```
 2      FORMAT(8F10.4)                                              ING00540
        N(0)=0                                                      ING00590
        DO 1001 I2=1,K                                              ING00600
        Y1(I2)=0.0                                                  ING00610
        KRITE(I2)=0.                                                ING00620
        ZMEAN=0.
        VAR1=0.
        JR(I2)=0                                                    ING00630
        Y(I2)=0.0                                                   ING00640
        KVAR(I2)=0                                                  ING00650
        KCNT(I2)=0                                                  ING00660
 1001   CONTINUE                                                    ING00670
        DO 100 JA=1,NO                                              ING00680
        READ(1)(VAR(J),J=1,LN)
        DEP(JA)=VAR(LN)
        DO 6 K3=1,K4                                                ING00690
        IJ=K3                                                       ING00730
        IR=IJ+1                                                     ING00740
        IF (IJ.EQ.1)GO TO 200                                       ING00740
        IJ1=IJ-1                                                    ING00750
        K6=N(IJ1) + 1                                               ING00760
        IF(K6.NE.1)GO TO 201                                        ING00770
 200    CONTINUE                                                    ING00780
        K6=1                                                        ING00790
 201      CONTINUE                                                  ING00800
        K7=N(IJ)                                                    ING00810
        Y(IR)=0.                                                    ING00815
        DO 6 IA=K6,K7                                               ING00820
        Y(IR)=COEFF(IA)*VAR(IA)+Y(IR)                               ING00830
 6      CONTINUE                                                    ING00860
        IF(DEP(JA).EQ.VAL(1))VAR1=VAR1 + 1.
        DEN=1.                                                      ING00870
        DO 11 K1=2,K                                                ING00880
        DEN=DEN + EXP(Y(K1))                                        ING00910
 11     CONTINUE                                                    ING00920
        Y(1)=1.                                                     ING00930
        DO 12 K2=2,K                                                ING00940
        Y(K2)=EXP(Y(K2))/DEN                                        ING00950
        Y1(K2)=Y1(K2) + Y(K2)                                       ING00960
        Y(1)=Y(1)-Y(K2)                                             ING00970
        Y3(K2)=Y1(K2)/NO
 12     CONTINUE                                                    ING00980
        Y1(1)=Y1(1)+Y(1)                                            ING00990
        Y3(1)=Y1(1)/NO
        FMAX=0.                                                     ING01000
        DO 33 K9=1,K                                                ING01010
        IF(Y(K9).LE.FMAX)GO TO 33                                   ING01020
        FMAX=Y(K9)                                                  ING01030
        KCH=K9                                                      ING01040
 33     CONTINUE                                                    ING01050
        KCNT(KCH)=KCNT(KCH) + 1                                     ING01060
 50     IF(VAL(KCH).EQ.DEP(JA))KRITE(KCH)=KRITE(KCH) + 1            ING01080
        KVAR=IFIX(DEP(JA))                                          ING01080
        DO 37 IM=1,K                                                ING01090
        IF(VAL(IM).EQ.DEP(JA))GO TO 36                              ING01100
        GO TO 37                                                    ING01110
```

```
 36    JR(IM)=JR(IM) + 1                                                    ING01120
       Y2(IM)=FLOAT(JR(IM))/NO
 37    CONTINUE                                                             ING01130
       XXXX(JA)=Y(1)
       Z(JA)=DEP(JA)
       ZMEAN=ZMEAN + Z(JA)
100    CONTINUE                                                             ING01140
C
C      COMPUTE MULTIPLE F AND MULTIPLE CORRELATION RATIOS
C
       ZMEAN=ZMEAN/NO
       VARY=VAR1/NO
       VARY=VARY*(1.-VARY)
       CALL CORRAT(NO,XXXX,Z,VARY,ZMEAN,ETA2(I),F(I),NU1(I),NU2(I))
       PRINT 150,ETA2(I),F(I),NU1(I),NU2(I)
150    FORMAT(//10X,*MULTIPLE CORRELATION RATIO =*,F9.6,5X,* MULTIPLE F-VALUE = *
      1ALUE = *,F12.6,5X,* WITH *,I6,2X, * AND*,I6,2X,* DEGREES OF FREEDOM *//)
      2M *//)
       DO 67 J=1,K                                                          ING01150
       WRITE(6,20)KRITE(J),J,XMODES(J)                                      ING01160
 20    FORMAT(/*  THE NUMBER CORRECTLY PREDICTED IS *I4,*  FOR J  *I4,* IING01170
      1N MODE *,A8)                                                         ING01175
 67    CONTINUE                                                             ING01180
       PRINT 74                                                             ING01185
 74    FORMAT(1H1////9X,*THE COUNTS ACCORDING TO THE MODEL ARE*)            ING01186
       WRITE(6,65)(XMODES(IT),IT=1,10)                                      ING01190
       WRITE(6,21)(KCNT(I5),I5=1,K)                                         ING01200
 21    FORMAT(/4X,8(4X,I6)/)                                                ING01210
       WRITE(6,24)                                                          ING01220
 24    FORMAT(//9X,28HTHE TRUE COUNTS ARE GIVEN AS/)                        ING01230
       WRITE(6,65)(XMODES(IT),IT=1,10)                                      ING01240
       WRITE(6,21)(JR(KVA),KVA=1,K)                                         ING01250
       WRITE(6,34)                                                          ING01260
 34    FORMAT(//9X,*THE COUNTS BY SUMMING PROBABILITIES ARE*)               ING01270
 65    FORMAT(/10X,10A8)                                                    ING01275
       WRITE(6,65)(XMODES(IT),IT=1,10)                                      ING01280
       WRITE(6,71)(Y1(I),I=1,K)                                             ING01290
       PRINT 84
       PRINT 85,NO,(XMODES(IT),IT=1,10)
       PRINT 83,(Y2(I),I=1,K)
 85    FORMAT(/X,*OBS=*,I4,10A8//)
 84    FORMAT(//10X,* THE MEAN TRUE PROBABILITIES ARE*//)
       PRINT 151
       PRINT 85,NO,(XMODES(IT),IT=1,10)
       PRINT 83,(Y3(I),I=1,K)
151    FORMAT(//10X,* THE MODEL MEAN PROBABILITIES ARE*//)
 83    FORMAT(10X,9F8.6)
 71    FORMAT(10X,10F8.1)                                                   ING01295
       NI=NO
       KSUBS=KSUBS - 1
       IF(KSUBS)94,94,93
 94    CONTINUE
       STOP                                                                 ING01300
       END                                                                  ING01240
       SUBROUTINE BAVALU(NBV,NSBV,LZ,NO,VAR)
       DIMENSION VAR(LZ),BAVAL(10),KVAR(10),VAL(10)
```

```
      KCOUNT=KCONT=0
      REWIND 1
      IF(NBV.EQ.0)GO TO 2
      READ 1,(BAVAL(I),I=1,NBV)
1     FORMAT(10F8.4)
      DO 2 II=1,NO
      READ(1)(VAR(I),I=1,LZ)
      DO 3 I=1,LZ
      DO 3 J=1,NBV
      IF(VAR(I).EQ.BAVAL(J))GO TO 31
3     CONTINUE
      WRITE(2)(VAR(I),I=1,LZ)
      GO TO 2
31    KCOUNT=KCOUNT + 1
2     CONTINUE
      REWIND 2
      REWIND 1
      NO = NO-KCOUNT
      IF(NSBV.EQ.0)GO TO 4
      DO 4 K=1,NO
      IF(NBV.EQ.0)GO TO 10
      READ(2)(VAR(I),I=1,LZ)
      GO TO 11
10    CONTINUE
      READ(1)(VAR(I),I=1,LZ)
11    CONTINUE
      DO 40 J=1,NSBV
      READ 5,KVAR(J),VAL(J)
5     FORMAT(I4,F10.5)
40    CONTINUE
      DO 6 J=1,NSBV
      KK=KVAR(J)
      IF(VAR(KK).EQ.VAL(J))GO TO 7
6     CONTINUE
      IF(NBV.EQ.0)GO TO 12
      WRITE(1)(VAR(I),I=1,LZ)
      GO TO 13
12    WRITE(2)(VAR(I),I=1,LZ)
13    CONTINUE
      GO TO 4
7     CONTINUE
      KCONT=KCONT+1
4     CONTINUE
      NO=NO-KCONT
      IF(NBV.EQ.0.OR.NSBV.EQ.0) GO TO 14
      GO TO 15
14    CONTINUE
      REWIND 1
      REWIND 2
      DO 15 KK=1,NO
      READ(2)(VAR(I),I=1,LZ)
      WRITE(1)(VAR(I),I=1,LZ)
15    CONTINUE
      NRONG=KCONT + KCOUNT
      PRINT 9,NRONG
9     FORMAT(/2X,*          THE NUMBER DELETED BY BAD VALUES IS*,I4)
```

```
      RETURN
      END
      SUBROUTINE TRANS(NTG,NO,LZ,VAR,SPO,VVV)
C     TRANSGENERATION CODES                                              TSAT0220
C      1 =  A+K                  7 = LOGF(A)          14 = A/B           TSAT0230
C      2 =  A-K                  8 = 1/A              15 =A**B           TSAT0240
C      3 =  A*K                  9 = EXPF(A)          16=X=1 IF A GREATER TSAT0250
C      4 =  A/K                 10 = ABSF(A)             OR EQUAL B      TSAT0260
C      5 =  A**K                11 = A+B                 X=0 IF A LESS B TSAT0270
C      6 = X=1 IF A ABOVE K     12 = A-B              17=ARITH REG GR    TSAT0280
C          X=0 IF A BELOW K     13 = A*B              18=LOG  REG  GR    TSAT0290
C          X=1 IF A =K                                19=RATIO GR        TSAT0300
C     20  -  LAG ARITHMETIC GROWTH RATE
C     21  -  SQUARE ROOT OF A
      DIMENSION VVV(LZ),VAR(LZ)
      DIMENSION SPO(NO,LZ),ILT(10,3),TLF(10)
      DIMENSION MODE(10)
      LZ=LZ-NTG
    7 READ (5   ,8)((ILT(I,JK),JK=1,3),TLF(I),MODE(I),I=1,NTG)           TSAT0620
    8 FORMAT(3I4,F10.0,3X,I1)                                           TSAT0630
  615 WRITE (6,15)                                                      TSAT0760
   15 FORMAT ( /6X,17HTRANSGENERATIONS ,12X,5HPLACE,6X,3HOP.,5X,6HA VAR.TSAT0770
     1,9X,6HB VAR. )
      WRITE(6,16)((ILT(I,JK),JK=1,3),TLF(I),MODE(I),I=1,NTG)
16    FORMAT(36X,I3,7X,I3,7X,I3,7X,F11.3,4X,I2)
C
C     TRANSFORMATION LOOP
C
      LZ1=0.
      REWIND 1
      DO 3000 JX=1,NO
      READ(1)(SPO(JX,K),K=1,LZ)
   29 DO 30 I= 1,NTG                                                    TSAT1070
      KA= ILT(I,1)                                                      TSAT1080
      IF(LZ.LT.KA.AND.LZ1.LT.KA)LZ1=KA
      KB= ILT(I,2)                                                      TSAT1090
      KC= ILT(I,3)                                                      TSAT1100
801   C =  SPO (JX,KC)
803   IF (KB -10) 32,32,31
   31 KD=IFIX(TLF(I))                                                   TSAT1130
      IF(KB.GE.17) GO TO 32
805   D = SPO (JX,KD)
806   IF (  D     +.09) 32,34,32                                        TSAT1140
34     SPO(JX,KA)=-.09
      GO TO 30                                                          TSAT1160
32    IF(KB.GT.16.AND.KB.LT.27)GO TO 36
      IF(       C  +.09) 36,34,36
   36 GO TO(110,120,130,140,150,155,160,170,180,190,210,220,230,        TSAT1180
     1240,250,260,270,280,290,270,422)KB                                TSAT1190
422   IF(C)34,426,426
426   SPO(JX,KA)=SQRT(C)
      GO TO 30
  110 SPO(JX,KA) = C         + TLF(I)                                   TSAT1280
      GO TO 30                                                          TSAT1290
  120 SPO(JX,KA) = C -TLF(I)                                            TSAT1300
      GO TO 30                                                          TSAT1310
  130 SPO(JX,KA) = C* TLF(I)                                            TSAT1320
```

```
      GO TO 30                                                          TSAT1330
  140 IF(TLF(I) ) 141, 34, 141                                          TSAT1340
  141 SPO(JX,KA) = C/ TLF(I)                                            TSAT1350
      GO TO 30                                                          TSAT1360
 150  IF(C)1,151,151                                                    TSAT1370
  151 SPO(JX,KA) = C     ** TLF(I)                                      TSAT1380
      GO TO 30                                                          TSAT1390
  160 IF      (C) 34,34   ,   161                                       TSAT1400
161   SPO(JX,KA)=ALOG(C)
      GO TO 30                                                          TSAT1420
  170 IF      (C)  171,34  , 171                                        TSAT1430
  171 SPO(JX,KA) = 1.0/C                                                TSAT1440
      GO TO 30                                                          TSAT1450
  180 SPO(JX,KA) = EXP(C)                                               TSAT1460
      GO TO 30                                                          TSAT1470
  155 IF(C -TLF(I)) 156,157,157
  156 SPO(JX,KA) = 0.0                                                  TSAT1490
      GO TO 30                                                          TSAT1500
  157 SPO(JX,KA) = 1.0                                                  TSAT1510
      GO TO 30                                                          TSAT1520
  190 SPO(JX,KA) = ABS(C)                                               TSAT1530
      GO TO 30                                                          TSAT1540
  210 SPO(JX,KA) = C+D                                                  TSAT1550
      IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
      GO TO 30                                                          TSAT1560
  220 SPO(JX,KA) = C-D                                                  TSAT1570
      IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
      GO TO 30                                                          TSAT1580
  230 SPO(JX,KA) = C*D                                                  TSAT1590
      IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
      GO TO 30                                                          TSAT1600
  240 IF(D)241,34,241                                                   TSAT1610
  241 SPO(JX,KA) = C / D                                                TSAT1620
      IF(SPO(JX,KA).EQ.(-.09)) SPO(JX,KA)=SPO(JX,KA)+.000001
      GO TO 30                                                          TSAT1630
 1    SPO(JX,KA)=ABS(C)
      SPO(JX,KA)=SPO(JX,KA)**TLF(I)
      SPO(JX,KA)=1/SPO(JX,KA)
      GO TO 30
 250  IF(C)1,251,251                                                    TSAT1640
  251 SPO(JX,KA) = C**D                                                 TSAT1650
      GO TO 30                                                          TSAT1660
  260 IF(C-D)261,262,262                                                TSAT1670
  261 SPO(JX,KA) = 0.0                                                  TSAT1680
      GO TO 30                                                          TSAT1690
  262 SPO(JX,KA) = 1.0                                                  TSAT1700
      GO TO 30                                                          TSAT1710
  270 KXX=JX                                                            TSAT1720
      IF(KB-20) 310,311,310                                             TSAT1730
  311 KXX=KXX-1                                                         TSAT1740
  310 KX=KXX-KD                                                         TSAT1750
      IF(KX) 274,274,271                                                TSAT1760
  271 TVAR= 0.0                                                         TSAT1770
      BETA= 0.0                                                         TSAT1780
      YBAR= 0.0                                                         TSAT1790
      YT   = 0.0                                                        TSAT1800
      YVAR= 0.0                                                         TSAT1810
```

```
      DO 272 K=KX,KXX                                                   TSAT1820
      BETA= BETA +1.0                                                   TSAT1830
      TVAR= TVAR + BETA* BETA                                           TSAT1840
      YBAR=YBAR + SPO(K,KC)
      YT=YT + SPO(K,KC)*BETA
  272 YVAR=YVAR+SPO(K,KC)*SPO(K,KC)
      TBAR= (BETA+1.0)/2.0                                              TSAT1880
      YBAR= YBAR/BETA                                                   TSAT1890
      YVAR= YVAR/BETA - YBAR * YBAR                                     TSAT1900
      TVAR= TVAR/BETA - TBAR * TBAR                                     TSAT1910
      YT  = YT/BETA -YBAR*TBAR                                          TSAT1920
      IF(YBAR)274,274,273                                               TSAT1930
  274 SPO(JX,KA) = -.09                                                 TSAT1940
      KA=KA+1                                                           TSAT1950
      SPO(JX,KA) = -.09                                                 TSAT1960
      GO TO 30                                                          TSAT1970
  273 KLAG= 0                                                           TSAT1980
      IF(KLAG) 275,275,274                                              TSAT2040
  275 SPO(JX,KA) = YT/(TVAR * YBAR )                                    TSAT2050
      KA= KA +1                                                         TSAT2060
      SPO(JX,KA) = (YVAR - (YT*YT)/TVAR)/(YBAR*YBAR)                    TSAT2070
      GO TO 30                                                          TSAT2080
  280 KLAG=0                                                            TSAT2090
      KXX=JX                                                            TSAT2100
      IF(KB-21)  312,313,312                                            TSAT2110
  313 KXX=KXX-1                                                         TSAT2120
  312 KX=KXX-KD                                                         TSAT2130
      IF(KX)   274,274,281                                              TSAT2140
  281 DO 282 K=KX,KXX                                                   TSAT2150
      IF(SPO(K,KC))2831,2831,283
 2831 KLAG = 1                                                          TSAT2170
  283 CONTINUE
  282 CONTINUE                                                          TSAT2220
      IF(KLAG) 285,285,274                                              TSAT2230
  285 IF(KXX-KX-2) 274,287,287                                          TSAT2240
  287 TVAR=0.0                                                          TSAT2250
      BETA = 0.0                                                        TSAT2260
      YBAR = 0.0                                                        TSAT2270
      YT = 0.0                                                          TSAT2280
      YVAR = 0.0                                                        TSAT2290
      DO 286 K= KC,KD                                                   TSAT2300
      BETA = BETA + 1.0                                                 TSAT2310
      YLOG=ALOG(SPO(K,KC))
      TVAR = TVAR + BETA * BETA                                         TSAT2330
      YBAR = YBAR + YLOG                                                TSAT2340
      YT =  YT + YLOG * BETA                                            TSAT2350
  286 YVAR = YVAR + YLOG * YLOG                                         TSAT2360
      TBAR = (BETA + 1.0) / 2.0                                         TSAT2370
      YBAR = YBAR/BETA                                                  TSAT2380
      YVAR = YVAR/BETA - YBAR*YBAR                                      TSAT2390
      TVAR = TVAR/BETA - TBAR*TBAR                                      TSAT2400
      YT   = YT/BETA  -  YBAR* TBAR                                     TSAT2410
      SPO(JX,KA) = YT/TVAR                                              TSAT2420
      KA=KA + 1                                                         TSAT2430
      SPO(JX,KA) = YVAR - YT*YT/TVAR                                    TSAT2440
      GO TO 30                                                          TSAT2450
  290 KLAG = 0                                                          TSAT2460
```

```
      KXX=JX
      IF(KB-22) 314,315,314
  315 KXX=KXX-1
  314 KX=KXX-KD
      IF(KX) 274,274,316
 316  DO 1100 L=KX,KXX
      IF(MODE(I))1101,1101,1102
 1101 VVV(L)=SPO(L,KC)
      GO TO 1100
 1102  VVV(L)=SPO(L,KC)
 1100 CONTINUE
  291 CONTINUE
      IF(KLAG) 294,294,274
  294 IF(VVV(KX))      274,274,295
  295 IF(VVV(KXX))      274,274,296
  296 TR=FLOAT(KXX-KX)
      GRL=(ALOG(VVV(KXX)/VVV(KX)))/TR
      SPO(JX,KA)=EXP(GRL)-1.0
       IF(SPO(JX,KA)+.09)370,371,370
 371  SPO(JX,KA)=SPO(JX,KA)+.001
 370  KA = KA + 1
      IF(KXX-KX-1) 304,304,305
  305 KE=KX+1
      YVAR= 0.0
      KLAG= 0
      DO 306 K=KE,KXX
      M= K-1
  298 CONTINUE
      IF(KLAG) 300,300,301
  300 IF(VVV(K))      301,301,302
 302  TBAR=ALOG(VVV(K))-ALOG(VVV(M))-GRL
      YVAR= YVAR + TBAR * TBAR
      GO TO 306
  301 KLAG= 1
  306 CONTINUE
      IF(KLAG) 303,303,304
  304 SPO(JX,KA) = -.09
      GO TO 30
  303 SPO(JX,KA) = YVAR/(TR - 1.0)
   30 CONTINUE
 3000  CONTINUE
 28   CONTINUE
      REWIND 1
      IF(LZ1.GT.LZ)LZ=LZ1
      DO 4001 I=1,NO
      DO 4000 J=1,LZ
      VAR(J)=SPO(I,J)
 4000 CONTINUE
      WRITE(1)(VAR(J),J=1,LZ)
 4001 CONTINUE
      REWIND 1
      RETURN
      END
      SUBROUTINE STAT(MVAR,NOBS,Z,PARAM,X,SETA2,SF,DEP)
      DIMENSION Z(NOBS),PARAM(MVAR),X(NOBS),SETA2(MVAR,MVAR)
      DIMENSION DEP(NOBS)
      DIMENSION SF(MVAR,MVAR)
```

```
C
C COMPUTE CORRELATION RATIO MATRIX                                     STAT3010
C
      NVAR=MVAR-1
      REWIND 1                                                          STAT3011
      DO 1001 I=1,MVAR                                                  STAT3030
      IF(I-MVAR)701,702,702
 701  DO 1024 LL=1,NOBS
      READ(1)(PARAM(K),K=1,NVAR)
 1024 X(LL)=PARAM(I)                                                    STAT3023
      REWIND 1                                                          STAT3024
      GO TO 703
 702  DO 704 LL=1,NOBS
 704  X(LL)=DEP(LL)
 703  DO 1001 J=1,MVAR
      ZMEAN =0.0           $  VARZ=0.0                                  STAT3020
      IF(I-J)1002,1003,1002                                             STAT3050
 1003 SETA2(I,J)=1.00                                                   STAT3060
      SF(I,J)=-0.0                                                      STAT3071
      GO TO 1001                                                        STAT3070
 1002 IF(J-MVAR)705,706,706
 706  DO 707 LL=1,NOBS
      Z(LL)=DEP(LL)
      ZMEAN=ZMEAN + Z(LL)
 707  VARZ=VARZ + Z(LL)**2
      GO TO 708
  705 DO 1004 LL=1,NOBS
      READ(1)(PARAM(K),K=1,NVAR)
      Z(LL)=PARAM(J)                                                    STAT3100
      ZMEAN=ZMEAN+Z(LL)                                                 STAT3110
      VARZ=VARZ+Z(LL)**2                                                STAT3120
 1004 CONTINUE                                                          STAT3121
      REWIND 1
 708  ZMEAN=ZMEAN/NOBS
      VARZ=VARZ-NOBS*ZMEAN**2                                           STAT3140
      IF(VARZ)1050,1050,1051                                            STAT3141
 1051 VARZ=VARZ/(NOBS-1.)                                               STAT3150
      CALL CORRAT(NOBS,X,Z,VARZ,ZMEAN,SETA2(I,J),SF(I,J),NU1,NU2)       STAT3160
      GO TO 1001                                                        STAT3161
 1050 SETA2(I,J)=0.0                                                    STAT3162
      SF(I,J)=0.0                                                       STAT3163
 1001 CONTINUE                                                          STAT3170
C
C PRINT OUT SIMPLE CORRELATION RATIOS AND F-SCORES                      STAT3180
C
      IF(MVAR-10)1010,1010,1011                                         STAT3181
 1010 PRINT 1005,(J,J=1,MVAR)                                           STAT3190
 1005 FORMAT(1H1//10X,*CORRELATION RATIO MATRIX*//10X,10(I6,4X))        STAT3200
      DO 1006 I=1,MVAR                                                  STAT3210
 1006 PRINT 1007,I,(SETA2(I,J),J=1,MVAR)                                STAT3220
 1007 FORMAT(/2X,I6,2X,10(F7.4,3X))                                     STAT3230
      PRINT 1008,(J,J=1,MVAR)                                           STAT3240
 1008 FORMAT(1H1//10X,*SIMPLE F-SCORES MATRIX*//10X,10(I6,4X))          STAT3250
      DO 1009 I=1,MVAR                                                  STAT3260
 1009 PRINT 1012,I,(SF(I,J),J=1,MVAR)                                   STAT3270
 1012 FORMAT(/2X,I6,2X,10(F10.4))                                       STAT3280
```

```
      GO TO 1023                                                        STAT3290
 1011 PRINT 1005,(J,J=1,10)                                             STAT3300
      DO 1013 I=1,MVAR                                                  STAT3310
 1013 PRINT 1007,I,(SETA2(I,J),J=1,10)                                  STAT3320
      PRINT 1014,(J,J=11,MVAR)                                          STAT3330
 1014 FORMAT(1H1//10X,10(I6,4X))                                        STAT3340
      DO 1015 I=1,MVAR                                                  STAT3350
 1015 PRINT 1007,I,(SETA2(I,J),J=11,MVAR)                               STAT3360
      PRINT 1008,(J,J=1,10)                                             STAT3370
      DO 1016 I=1,MVAR                                                  STAT3380
 1016 PRINT 1007,I,(SF(I,J),J=1,10)                                     STAT3390
      PRINT 1014,(J,J=11,MVAR)                                          STAT3400
      DO 1017 I=1,MVAR                                                  STAT3410
 1017 PRINT 1007,I,(SF(I,J),J=11,MVAR)                                  STAT3420
C
 1023 CONTINUE
      RETURN
      END
      SUBROUTINE CORRAT(NOBS,X,Z,VARZ,ZMEAN,ETA2,F,NU1,NU2)             CORR0010
      DIMENSION CMEAN(10),NUM(10),FLIM(10)                              CORR0020
      DIMENSION X(NOBS),Z(NOBS)
C
C SUBROUTINE TO COMPUTE CORRELATIONS RATIOS AND F SCORES                CORR0030
C
      ETA2=0.                                                           CORR0040
      DO 21 II=1,10                                                     CORR0050
      NUM(II)=0                                                         CORR0060
   21 CMEAN(II)=0.0                                                     CORR0070
      FLIM(1)=0.1                                                       CORR0080
      DO 1 KOUNT=2,10                                                   CORR0090
      LL=KOUNT-1                                                        CORR0100
    1 FLIM(KOUNT)=FLIM(LL)+0.1                                          CORR0110
      XMAX=X(1)                                                         CORR0120
      XMIN=X(1)                                                         CORR0130
      DO 2 JOUNT=2,NOBS                                                 CORR0140
      IF(X(JOUNT).GT.XMAX)101,103                                       CORR0150
  103 IF(X(JOUNT).LT.XMIN)102,2                                         CORR0160
  101 XMAX=X(JOUNT)                                                     CORR0170
      GO TO 2                                                           CORR0180
  102 XMIN=X(JOUNT)                                                     CORR0190
    2 CONTINUE                                                          CORR0200
      DIFF=XMAX-XMIN                                                    CORR0210
      DO 3 JJ=1,NOBS                                                    CORR0230
    3 X(JJ)=X(JJ)-XMIN                                                  CORR0240
      IDIFF=IFIX(DIFF+0.99)                                             CORR0250
      DO 4 K=1,10                                                       CORR0260
    4 FLIM(K)=FLIM(K)*IDIFF                                             CORR0270
C
      DO 5 J=1,NOBS                                                     CORR0280
      DO 6 KOUNT=1,10                                                   CORR0290
      IF(X(J)-FLIM(KOUNT))106,106,6                                     CORR0300
  106 CMEAN(KOUNT)=CMEAN(KOUNT)+Z(J)                                    CORR0310
      NUM(KOUNT)=NUM(KOUNT)+1                                           CORR0320
      GO TO 5                                                           CORR0330
    6 CONTINUE                                                          CORR0340
    5 CONTINUE                                                          CORR0350
```

```
C
      DO 7 JX=1,10                                                      CORR0360
      IF(NUM(JX))7,7,107                                                CORR0370
  107 ETA2=ETA2+CMEAN(JX)**2/NUM(JX)                                    CORR0380
    7 CONTINUE                                                          CORR0390
      IF(ETA2)108,108,109                                               CORR0400
  109 ETA2=ETA2/NOBS-ZMEAN**2                                           CORR0410
      ETA2=ETA2/VARZ                                                    STAT0420
      NU2=NOBS-10                                                       CORR0430
      NU1=9                                                             CORR0440
      F=ETA2*NU2/((1.-ETA2)*9)                                          CORR0450
      GO TO 110                                                         CORR0460
  108 ETA2=0.0                                                          CORR0470
      F=0.0                                                             CORR0480
 110  RETURN
      END                                                               CORR0500
```

APPENDIX IV

BIBLIOGRAPHY

1. Adams, W. T., "Factors Influencing Transit and Automobile Use in Urban Areas", Bureau of Public Roads, Vol. 30, 1959.

2. Betak, J, and Betak, C., "Urban Modal Choice: A Critical Review of the Role of Behavioural Decision Variables"

3. Biciunnas, A. E., "Modal Split of First Work Trip" CATS Research News, Chicago, Illinois, 1964, Vol. 6 # 1

4. Bock, F. C., "Factors Influencing Modal Trip Alignment" HRB NCHRP Report # 57. Washington, D.C. 1968.

5. Buchanan, C., et al, "Traffic in Towns", HMSO, London, 1963.

6. Carroll, J. D., "Relation Between Land Use and Traffic Generation" B.P.R. Research Digest # 1, Feb. 1954.

7. Charles Rivers Associates, Inc., "An Evaluation of Free Transit Service" Federal Clearing House Report PB 179 845, Washington, D.C., August, 1968.

8. Davis, H. E., "Urban Transportation Planning-Introduction to Methodology". A.S.M.E. Transportation Engineering Conf., 1968. "Defining Transportation Requirements". Papers and Discussions, 1969.

9. de Donnea, F., "Probit and Logit Modal Split Models for Rotterdam", Discussion Paper VI, Netherlands Economic Institute, Rotterdam, December, 1970.

10. Deen, T., Mertz, W., Irwin, N., "Application of a Modal Split Modes to a Travel Estimate for the Washington Area" HRB Record # 38, Washington, 1963.

11. Finney, D. H., "Probit Analysis". Cambridge University Press, 1964.

12. Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems", Ann. Eugen. Lond., 1936, Vol. 7.

13. Hamburg, J. R., and Guinn, C. R., "A Modal Split Model: Description of Basic Concepts", New York State Dept. of Public Works, Albany, 1966.

14. Hill, D. M., and Von Cube, H. G., "Development of a Model for Forecasting Travel Mode Choice in Urban Areas", HRB Record # 38, Washington, 1963.

15. Howe, J. J., "Modal Split on CBD Trips" CATS Research News, 1958, Vol. 2, # 12, Chicago.

16. Hadden, J. K., "The Use of Public Transportation in Milwaukee": Traffic Quarterly, 1962, Vol. 18.

17. Lave, C. A., "Modal Choice in Urban Transporation: A Behavioural Approach". Phd., Stanford University, 1968.

18. Lisco, T. E., "The Value of Commuter's Travel Time: A Study in Urban Transportation", PhD., Univ. of Chicago, 1967.

19. Martin, B. V., Memmott, F. W., and Bone, A. J., "Principles and Techniques of Predicting Future Demand for Urban Area Transportation". Research Report No. 38, M.I.T., June, 1961.

20. McGillivray, R. G., "Binary Choice of Transport Mode in the San Francisco Bay Area", PhD., Univ. of California, 1969.

21 McLynn, J. M., and Watkins, R. H., "Multimode Assignment Model", Paper of the National Bureau of Standards, Washington, D.C., 1965.

22. Mongini, A., "Some Aspects of Discriminant Functions and Other Interurban Modal Split", "Modela", Northeast Corridor Transportation Project Report, 1965.

23. McLynn, J. M., Goldman, A. J., Meyers, P. R., Watkins, R. H., "Analysis of a Market-Split Model:. Northeast Corridor Transportation Project Technical Paper # 8, April, 1967.

24. Moskowitz, K., "California Method of Assigning Diverted Traffic to Proposed Freeways", HRB Bulletin # 130, Washington, 1956.

25. Oi, W. Y., and Shuldiner, P. W., "An Analysis of Urban Travel Demands", Northwestern University Press, 1969.

26. Paine, F. T., "Consumer Conceived Attributes of Transportation: An Attitude Study". Univ. of Maryland, June, 1967.

27. Quarmby, D. A., "Choice of Travel Mode for the Journey to Work: Some Findings", Journal of Transport Economics and Policy, 1967, Vol. 1, # 3.

28. Quinby, H. D., "Traffic Distribution Forecasts: Highway and Transit" Traffic Engineering, 1961, Vol. 31, # 5.

29. Rassam, P. R., Ellis, R. H., Bennett, J. C., "The N-Dimensional Logit Model; Development and Application", Washington, D.C., 1970 (unpublished).

30. Reichmann, S., and Stopher, P., "Towards Disaggregate, Stochastic Models of Travel Mode Choice", Paper of the HRB 50TH Annual Meeting, Washington, D.C., January 1971.

31. Sharkey, R. H., "A Comparison of Modal Stratification of Trips by Distance from the CBD", CATS Research News, Chicago, Illinois, 1958, Vol. 2, # 5.

32. Sharkey, R. H., "The Effect of Land Use and Other Variables on Mass Transit Usage in the CATS Area", CATS Research News, Chicago, Illinois, 1959, Vol. 3, # 1.

33. Stopher, P. R., "Report On The Journey to Work Survey", Research Memo, January 1968, London.

34. Stopher, P. R., "A Probability Model of Travel Mode Choice for the Work Journey", HRB Record # 283, Washington, D.C., 1969.

35. Theil, H., "A Multinomial Extension of the Linear Logit Model", International Economic Review, 1969.

36. Thomas, T. C. and Thompson, G. I., "The Value of Time Saved by Purpose", October, 1970, Stanford Research Institute, California.

37. Warner, S. L., "Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice". Northwestern University Press, Chicago, Illinois, 1962.

38. Wynn, F. H., "Land Use Relationships with Travel Patterns", B.P.R. Research Digest, 2 # 2, Nov. 1955.