STOCHASTIC MODELS OF TRAVEL-DEMAND BEHAVIOUR: A COMPARISON OF THREE DISAGGREGATE MODEL FORMS USING INCOMPLETE DATA

by

SIAVOCHE KAHKESHAN

Diplôme d'Ingénieur Civil, Ecole Polytechnique Fédérale de Lausanne, 1980

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Civil Engineering

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1982

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Civil Engineering

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date: 26 April 1982

Abstract

This thesis is concerned with calibrating a behavioural travel demand model by adjustment of an incomplete data base and applying it to three statistical methods: Logit, Probit and Discriminant Analysis, comparing the forecasting ability of each of the models and analysing the responsiveness of decision variables to potential changes in the urban transportation system. In this research, the current state of the art is indicated and the mathematical structure of the models are discussed. Attempts are also made to determine the value of time for the Vancouver Population as a means of estimating the validity of the model calibration, and also to discuss the results in transportation policy terms.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am sincerely indebted to my supervisor, Dr. G.R. Brown, for his constructive criticism and guidance during the preparation of this thesis, and for his patience and diligence in going over the first draft.

I would like to express my appreciation to Dr. F.P.D. Navin and Dr. W.F. Caselton for their contributions as readers.

I am grateful to Malcolm Greig, senior analyst at the Computing Center, University of British Columbia , for his valuable advices on several computer programming problems; and to Victoria Lyon-Lamb, consultant at the Computer Center, University of British Columbia , who patiently solved several text processing problems for me.

I would also like to thank the National Science and Engineering Research Council of Canada for their financial support.

In addition, I am indebted to my parents for their continued moral support during the preparation of this thesis and over the course of the past years.

Finally, I would like to mention Jody Butler and Jeff Smyth, two friends whose help greatly improved the style of this presentation.

Having received the kind help and support of so many people, I must nonetheless take total responsability for any errors and omissions which might be discovered in this research.

I. INTRODUCTION

Transportation Engineering decisions require substantial insight into the prediction of travel demand, in general and in the specific region understudy. Methodology and travel data need to be brought together in a meaningful way to calibrate a mathematical model which can then be used for predicting travel demand.

This research is to examine a local data base (The Vancouver Area Travel Study, VATS) and to apply it to three commonly used demand model structures (the Logit, Probit and Discriminant models) for the purpose of:

i. determining adjustments necessary, and feasible for an incomplete data base,

ii. calibrating, and comparing each of the model structures for application in predicting travel demand, and

iii. to test the sensitivity of the models on the relevant variables selected.

1. THE MODE CHOICE PREDICTION PROBLEM

In order to help transportation authorities decide on different investment schemes, transportation planners should be able to diagnose the impact on traffic flows of changes in transportation policies. Such changes may include the transit operation policy, modifying the pricing structure of the transportation system and/or adding a new facility.

For this purpose, forecasting models which accurately assess the consequences of alternative policies on the travel behaviour of the population becomes necessary. The derivation of the models should be based on the theory of human behaviour and should include the necessary variables to formulate traveller's socio-economic characteristics as well as the attributes of the transportation system.

Substantial efforts in the art of modal choice modelling have been expended in the past 20 years. The earlier modal choice models, trip-end and trip-interchange modal split models, used zonally aggregated data containing both captive and choice riders. For this reason they are not adequately policy sensitive. Two examples of the earlier model are the Southeastern Wisconsin and Toronto modal split model(Hutchinson, 1974).

The version of modal split model structure discussed in this thesis uses disaggregated data which will overcome some of the problems of aggregation by providing estimates of trip disutility. These models attempt to associate a probability to a traveller's decision to use one mode over another and hence the name stochastic, disaggregate modal choice model has been applied. However, these models need intensive data which renders their application less practical with currently available data bases. Therefore, this thesis is an attempt to use an existing data base by estimating statistically the information needed to calibrate three forms of disaggregate stochastic mode choice models.

II. TRANSPORTATION MARKET AND DETERMINATION OF PRINCIPAL VARIABLES

The calibration of the econometric modal split models for Vancouver is based on a survey conducted in the Vancouver Metropolitan Area in the early spring 1972. This survey, called Vancouver Activity Travel Study (VATS), was collected over approximately 3600 households representing 1% of the households in the Region.

1. SURVEY DATA SAMPLE

Relevant data for this study are available in four VATS data files. These files contain household and personal information, trip records and modal choice information and are linked together by means of identification numbers assigned to each record .

A total of 26,652 trips were reported. According to Figure 1, 12.18% of the trips were made for work purposes. This relatively low proportion of work trips might be due to the time of the day that the survey was conducted. The VATS survey was carried out during the day and the trips collected are those for the day before. Therefore, most of the workers were not personally interviewed but the work trip information was collected from at home members of the household.

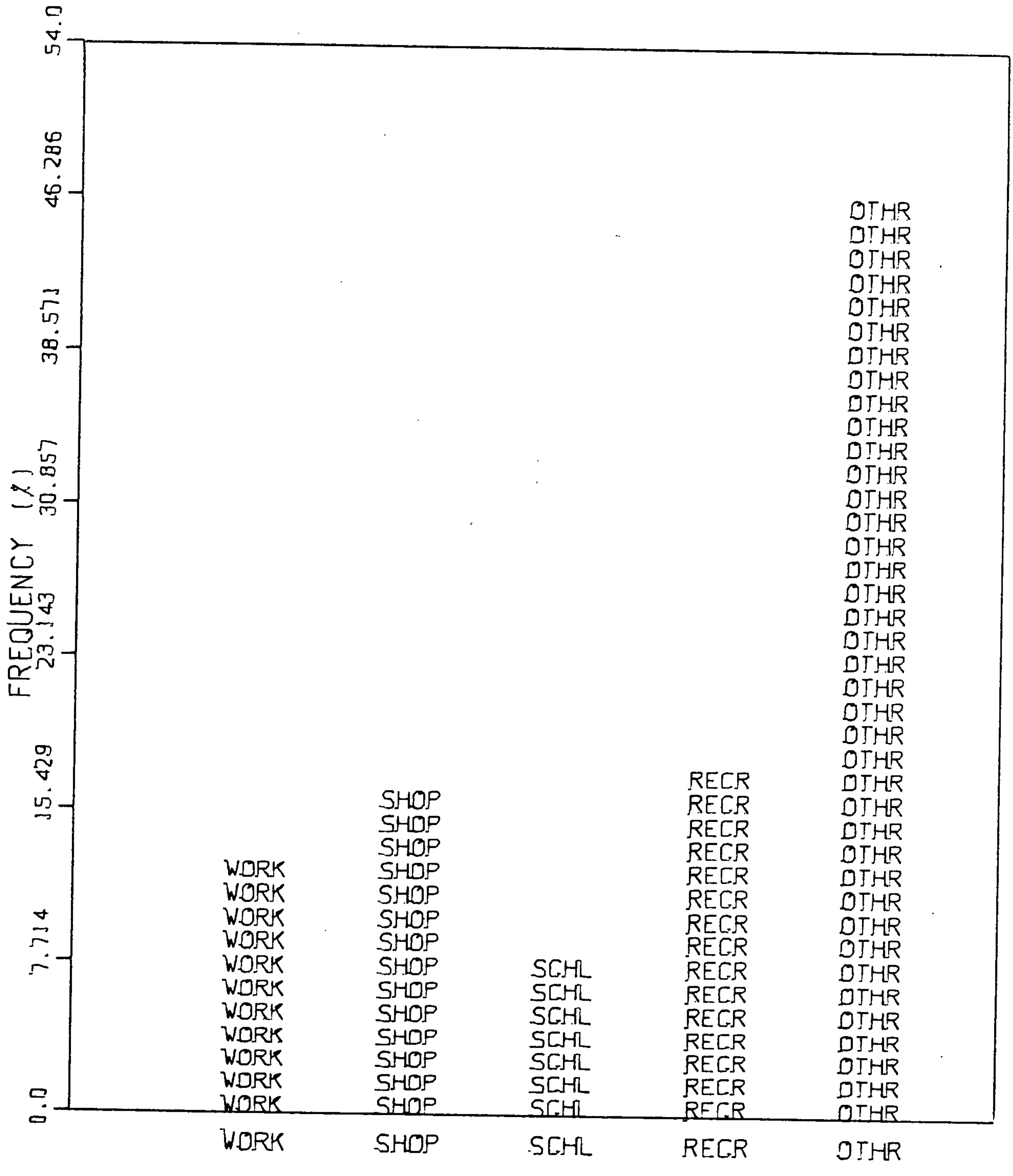

Figure 1 - Trip Purpose Distribution

The largest category is that of "OTHER"(46.26%). This is mostly due to the aggregation of purposes other than those illustrated in Figure 1 for this category. The "OTHER" category embodies purposes such as: eating, personal business, riding along, strolling etc..

Trip length distribution for different purposes is illustrated in Figure 2. All trips have a distribution common to other surveys. Average work trip length is between 20 and 30 minutes. Compared to the others, the work trip distribution has the gentlest slope, which might be due to people having less flexibility to choose their work location.

The modal split distribution of the sample is shown in Figure 3. Travelling by automobile presents the largest proportion of trips (46.64%). The second most popular travel mode is walking which is 23.71% of the 26419 trips. However, for work trips, the most important means of travel are auto and bus (See Table1).

2. WORK-TRIP SAMPLE

The sample of journey to work forms the basic sample for this study. It consists of 3172 cases. A case is considered incomplete if only one piece of information about one variable (i.e. Travel Time, Age, Income, ...) is missing.

From 3172 work-trip records collected from the first three VATS data files (household information, personal information

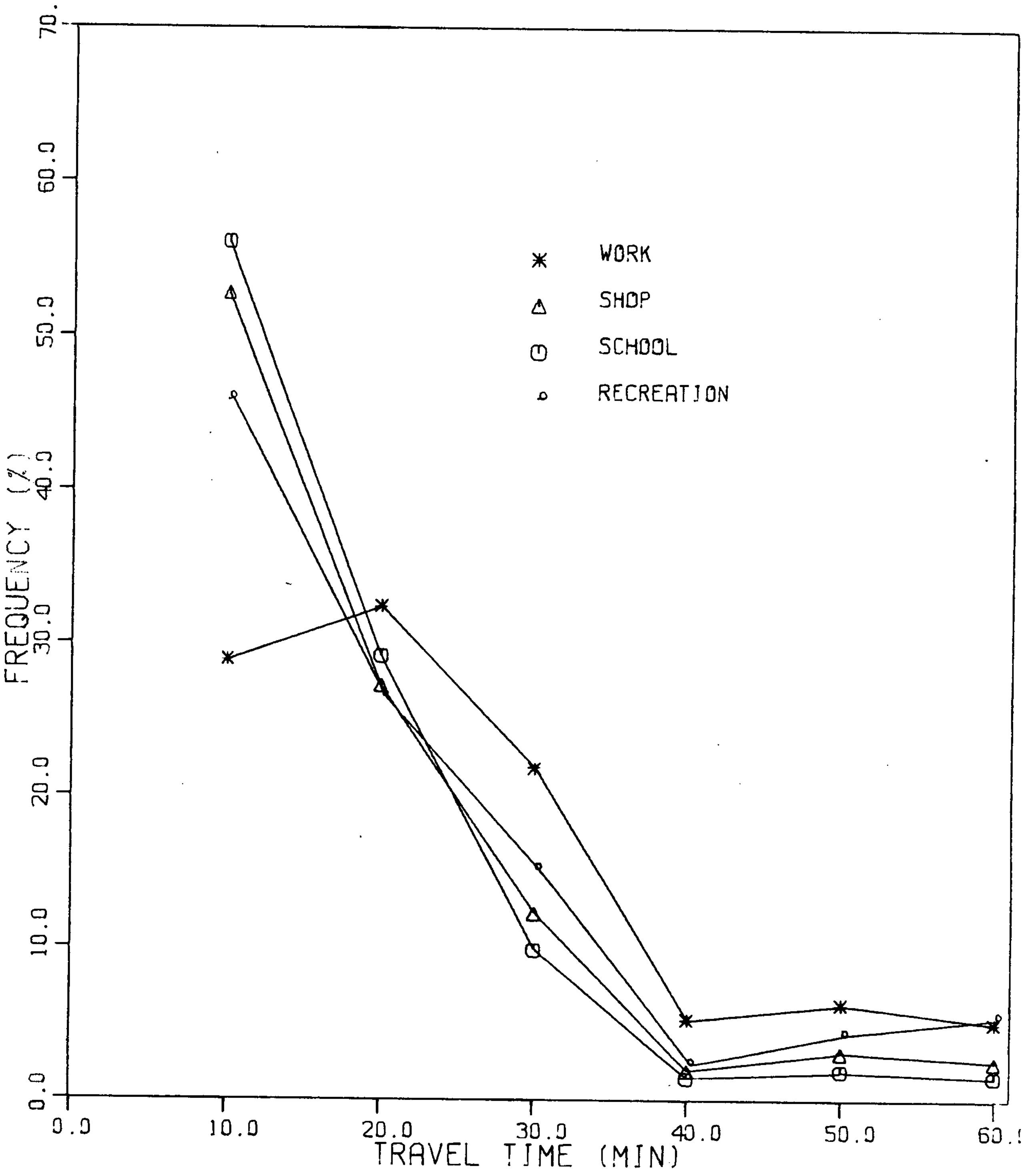

Figure 2 - Trip Length Distribution for Different Purposes

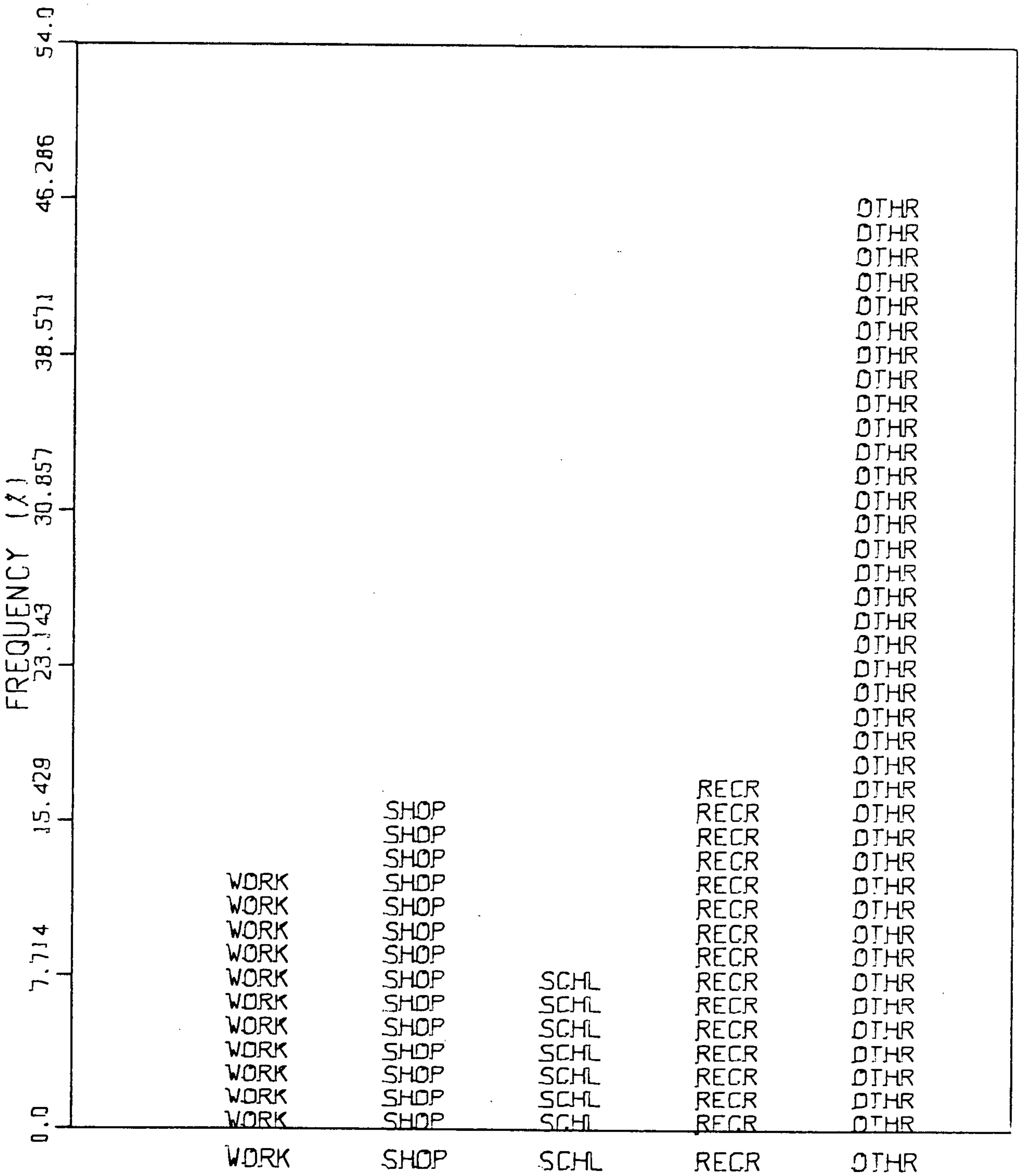

Figure 3 - Travel Mode Distribution

TRAVEL MODE	NUMBER	%
Auto	538	79.23
Auto Pass.	58	8.54
Bus	83	12.23
Walk	0	0.0
Σ	679	100

Table 1 - Composition of Work Trip Data

and trip records), only 841 cases have complete information about home-based work-trips made between the two morning peak hours(7-9 AM).

Out of these 841 responders, 130 persons did not have a drivers licence and 32 persons belonged to a household without any cars. Therefore, this sample has 162 captive transit passengers (19.26%) who were excluded from this set as changes in service policy related do not influence their travel behaviour. This reduces the sample size to 679 cases. The modal choice composition of the underlying sample is given in Table 1. Since this study is concerned about two travel modes: private car and bus, the cases corresponding to auto passengers are also deleted.

The remaining data set of 621 records is further considerably reduced in size when an attempt is made to link the modal choice information, since the file containing this information is the most incomplete file. This is due to the fact that people can not easily estimate the service characteristics of their travel mode. To solve this problem, an attempt should be made to complete the data set. This is discussed in detail in the next chapter.

3. DETERMINATION OF PRINCIPLE VARIABLES

Before completing the data base, the selection of the relevant variables and their appropriate form to be included in the modal choice analysis should be determined. In order to define the set of the explanatory variables, we shall use the results of previous studies as well as engineering judgement.

We should also try to find a compromise between the number of explicatory variables, the simplicity of the variable form and the goodness of fit obtained, since models with a large number of variables are expensive to manipulate.

Variables influencing modal choice decisions can be classified in two categories:

i. System variables, including variables related to the functioning of the transportation system, and

ii. Users' variables, including variables describing the socio-economic characteristics of the users as well as their taste.

3.1 System Variables

3.1.1 Time Variables

There is ample evidence that the components of overall travel time (walking,waiting,in-vehicle and transfer time) are perceived differently from travellers since the degree of inconvenience associated with these components are not equal.

Merlin and Barbier(1965) found that Parisians considered walking time to be 1.75 times more inconvenient than in-vehicle time. Waiting and transfer time was valued three times and twice the value of in-vehicle time respectively. Another study carried out in Manchester showed that walking and waiting time was valued 2.60 and 3.60 times the value of in-vehicle time respectively (Rogers, Townsend and Metcalf, 1970).

Despite this evidence, in several studies the overall travel time was introduced into the model since detailed information regarding the components of the travel time was unavailable. Lisco(1967), Lave(1969), and De Donnea(1971) considered the overall travel time, whereas Brown(1972), Navin(1974), O' Farrel and Markham(1975) and Segal(1978) investigated measuring the effect of different time activities on the travel disutility.

Travel time(either walking, waiting or in-vehicle) can be expressed in different ways. Warner(1962) and McGillivray(1970) have used the logarithm of travel time ratio, whereas Lisco, Quarmby(1967), Lave and De Donnea have considered the difference between travel times, arguing that commuters perceive relative times in terms of difference than in terms of ratio. In his study, Brown has used both difference and ratio of travel times, and Watson(1974) formulated the ratio between travel times difference and the trip length as being the average of actual and alternative mode travel time ((T1-T2)/[(T1+T2)/2]). He argued that five minutes saved on a ten-minutes journey may be more important than on a four-hour journey. Ben Akiva and Atherton(1977) used in-vehicle travel times difference and the ratio between excess travel time and traveled distance to express the effect of time variables on mode choice probability.

In the present study, we will test two formulations of each time activities. The first one is to introduce into the model the differences and the second is to adopt the difference-ratio formulation, where for each tripmaker the ratio of the difference between the actual and the alternative travel time to the actual overall travel time is computed. For instance, in the case of an individual travelling by car, the relative in-vehicle time is equal to: [(Tb-Ta)/TTa] and similarly, in the case of a person travelling by bus we have: [(Tb-Ta)/TTb].

3.1.2 Cost Variables

Cost variables reflect all out-of-pocket expenses. In the case of bus riders the individual cost of a trip is straightforward. It is the bus fare. Whereas for car drivers, the one-way trip cost per capita can not be determined easily, since it is more convenient to gather data regarding the monthly operating cost and monthly parking charge than the daily cost. To compute the total cost of a trip made by car, we shall use the following hypothesis: we suppose that two trips are made during business days and there are 20 business days in a month. Therefore, the total cost per trip will be equal to:

(Monthly Operating Cost+Monthly Parking Charges) / 20x2 and we assume that the above figure represents the one-way trip cost per capita since the daily vehicle occupancy rate is not available from the survey.

The formulation of cost variable should follow the principle of the travel time formulation since both reflect the cost occured to a commuter due to travelling. Therefore, we will use costs difference(Cb-Ca) wherever time differences are adopted and use costs difference-ratio[(Cb-Ca)/C] wherever travel time components are formulated in this manner.

3.2 Users' Variables

This family of variables represent socio-economic qualities of tripmakers. They are normally introduced into modal choice models to express the change in the probability of a mode use resulting from a change in the socio-economic characteristics of individuals. In this category, one may classify variables such as: Household Income, Marital Status, Age, Sex, number of Cars Owned by the Household, Type of Household, Household Composition, Education, Occupation,...etc.

De Donnea(1971) has studied the effect of Income, Age, Sex and Marital Status, number of People in the Household and the Relation of the Tripmaker to the Head of Household. O'Farrel and Markham(1975) considered Income, Age, Sex and Marital Status, Car-Demand ratio, Importance of car at work and Household Composition as users' variables. Lave(1969) included variables such as Auto Ownership, Family Size and Composition, Income, Sex and Age of commuters in his study. Quarmby(1967) has measured the impact of Income, Car-Demand ratio, Ownership of car by firm and Use of car for work on the probability of taking transit mode.

In short, a variety of socio-economic variables affects the decision of an individual facing several alternatives. However, two major reasons prevent one from using some of these variables: their availability and their ease of forecasting. For instance, variables such as Marital Status, Household Composition and Size, Car-Demand ratio, Importance of car at work, while they might increase the model refinement, are very difficult to predict and consequently their inclusion limits drastically the prediction boundary of the model.

The present study will use Income, Age, Sex, Occupation and Car Ownership to formulate the personal characteristic of transportation demand. Their significance in improving the capability of the model will be statistically tested.

3.2.1 Sex

There is evidence which strongly suggests that sex of a trip maker may influence modal choice probability. According to Morall(1971), the greater proportion of transit riders are female. He used several surveys(1969-1976) and found that in most Canadian cities, the percentage of female transit users was above 60%. This might be due to the historical dominance of male owned and operated cars, and the difference between the threshold level of preference and comfort.

The analysis of our data reveals that only 22.12% of auto-drivers population are female(see Table 2). We will use the indicator variable SEX to formulate the sex influence in the models.

SEX = 1 if the tripmak er is male and

0 if the tripmaker is female

Mode	Number			Percentage		
	Total	Male	Female	of Tot.	Male	Female
Auto	538	419	119	86.63	77.88	22.12
Bus	83	35	48	13.37	42.16	57.84
Total	621	454	167	100	---	---

Table 2 - Influence of Sex on Modal Split

3.2.2 Age

Table 3 displays the effect of age on the probability of transit use. Regarding modal decision, travellers may be divided into two groups:

i. The first group consists of young and elderly people since their behaviour toward transit use may be formulated in a similar fashion. A majority of them use transit mode due to the difficult accessibility to cars: the lack of a driver's licence and a budget constraint in the case of young people and the lower psychological preference for the car and different perception of the degree of comfort in the case of elderly persons. They are less sensitive to travel time, less sensitive to loss of privacy but very affected by the out-of-pocket expenses.

ii. The second group consists of the remaining people.

We will consider the dummy variable AGE to express the influence of age on the transit use probability. It is formulated as below:

AGE = 0 if traveller's age belongs to the semi-closed interval [25,60[

and

AGE = 1 otherwise

AgeBracket	Total	AutoDriv.	TransPass	%Auto	%Trans.
<18	11	9	2	82	18
18-25	129	95	34	74	26
25-30	102	89	13	87	13
30-35	93	86	7	92	8
35-40	87	76	11	87	13
40-45	80	75	5	94	6
45-50	60	54	6	90	10
50-55	40	38	2	95	5
55-60	14	12	2	86	14
>60	5	4	1	80	20
Total	621	538	83		

Table 3 - Influence of Age on Modal Split

3.2.3 Income

Intuitively, income should play a great role in the modal selection process. Table 4 displays different income brackets available in the sample population and their observed modal split figures. It also reveals that the formulation of the Income variable in the model is more likely to be a combination with other variables rather than to be a separate variable, since there is no tendency of the transit ridership to vary with the income variable in a sigmoid(S-shape) fashion.

Income is combined with the components of the generalized travel cost on the basis of time-money trade-off situation. Two classes of individuals may be considered:

i. The class of individuals who are willing to give up time to save money.

ii. The class of individuals who prefers to spend extra money to save time.

There always exists a threshold value beyond which an individual changes groups. For example, a time chooser who remains a time chooser in a situation where he should spend 4 dollars to save 1 minute, but becomes a money chooser when he is faced with the situation where spending 6 dollars is required, has a threshold value between 4 and 6 dollars. This value is called the individual's marginal value of time.

However, in the real world, the determination of this value

Income Bracket	No.	%of Total	in Number		in Percentage	
			Auto	Bus	Auto	Bus
<3000	11	1.77	10	1	90.91	9.09
3000- 5999	34	5.48	30	4	88.24	11.76
6000- 8999	105	6.91	88	17	83.81	16.19
9000-11999	148	23.83	128	20	86.49	13.51
12000-14999	131	21.10	113	18	86.26	13.74
15000-17999	85	13.69	74	11	87.07	12.94
18000-20999	54	8.70	48	6	88.89	11.11
21000-23999	16	2.58	13	3	81.25	18.75
24000-26999	20	3.22	20	--	100.	-----
27000-29999	5	0.81	5	--	100.	-----
>30000	12	1.93	9	3	75.	25.
Total	621	100.	535	86		

Table 4 - Influence of Income on Modal Split

is very difficult. One way of surmounting this difficulty is to relate the value of time to income, assuming that a saving in travel time can be assigned to more production and hence lead to more employee-hours. However, there are many objections to this method which are not discussed in the present study (i.e. see Quarmby and Harrison(1969)).

Lave(1969) and De Donnea(1971) have assumed that travel time savings are proportional to income, whereas Quarmby(1967), Ben Akiva and Atherton(1977) have related the income to the out-of-pocket expenses; the formers have formulated an Income-Time variable ($I.\Delta T$) and the latters have used an Income-Cost variable ($\Delta C/I$).

We will test both approaches for our model and keep the one which leads to a better fit. The value of income used for each bracket is its mid-range value in 1000$.

3.2.4 <u>Occupation</u>

Even though occupation might be highly interrelated with income, we have decided to incorporate it into the model, since income might not express adequately the impact of the individual's social status and prestige on modal choice probability.

Individuals are categorized into 4 occupation-groups according to Statistic Canada: Primary, Professional and Managerial, Clerical-Sales and Labour-Services. Two binary dummy variables, OCC1 and OCC2, were used to formulate this classification as follows:

Primary	OCC1 = 0	OCC2 = 0
Professional and Managerial	OCC1 = 0	OCC2 = 1
Clerical-Sales	OCC1 = 1	OCC2 = 0
Labour-Services	OCC1 = 1	OCC2 = 1

3.2.5 Car Ownership

Number of cars in a household may affect the modal choice. However, similar to the occupation variable, Car Ownership might be considered only if its inclusion ameliorates the fitness of the model and does not lead to large variances of the coefficients.

Categorical variable CO formulates this factor as follows:

CO = 0	if the household possesses one car
CO = 1	if the household possesses more than one car

No.Car/Hous.	in Number	in Percent.
1	249	40.10
2	266	42.83
3	67	10.79
>3	39	6.28
Total	621	100.

Table 5 - Car Ownership Distribution

4. SUMMARY

In summary, we believe that the modal choice model specified should consist of those variables which not only affect the travellers' preferences but are easy to forecast. Refering to the literature available on modal choice modelling, we established the following set of variables and variable forms:

In-Vehicle Travel Time	ΔT	$\Delta T/T$	$I.\Delta T$	$I.\Delta T/T$
Waiting Time	WAIT	WAIT/T	I.WAIT	I.WAIT/T
Walking Time	ΔWALK	ΔWALK/T	I.ΔWALK	I.ΔWALK/T

Out-of-pocket Expenses	ΔC	$\Delta C/C$	$\Delta C/I$	$\Delta C/I.C$
Age	AGE			
Sex	SEX			
Occupation	OCC1 and OCC2			
Car Ownership	CO			

where:

ΔT	=	Tb-Ta
T	is	the actual overall travel time
WAIT	is	the waiting time at bus stops
ΔWALK	=	(walk to from bus stop) - (walk to from car)
ΔC	=	Cb-Ca
C	is	the actual total travel cost
I	is	the household annual income (in 1000 $)

We shall build appropriate models which enable us to deduce those variables (system and users') that appear most promising from the point of view of a tripmaker. This is treated in the next chapters.

III. DATA COMPLETION AND MODEL SIMULATION

This chapter is concerned with the construction of models in order to complete the data base. For this purpose, it was decided first to collect those cases which are complete on the following pair of variables : In-Vehicle Travel Time(IVTT) and Transit Travel Time(TTT)[1], and IVTT and Monthly Operating Cost(MOCOST)[2]; to find by means of statistical tools a relationship between them, and then apply this relationship to those deleted cases which are incomplete on only one-variable in order to estimate the missing value of the other. Note that this approach may be applied logically only to simulate TTT and MOCOST, since one may relate TTT and MOCOST to IVTT but not the others. For instance, as IVTT increases, fuel consumption and hence travel cost increases. Also the riding time in a bus logicaly has some relationship with the in-vehicle travel time. On the other hand, waiting time seems to be independant of TTT or IVTT. Therefore, to complete missing values on waiting time other methods should be investigated, and this is further discussed. For parking costs, the average zonal parking charges are considered.

---

[1] TTT = f(complete IVTT)
[2] MOCOST = f(complete IVTT)

1. LINEAR DATA SIMULATION MODELS

Linear regression models are applied to the pairs of variables indicated above. Therefore, the dependant variables TTT and MOCOST are regressed on IVTT. Tables 6 and 7 show the results of this attempt.

In order to select the best model, the following criteria are used :

i. Relative magnitude and sign of regression coefficients

ii. Squared coefficient of correlation

iii. Variation in sum of square due to regression (SSR) and sum of square due to error term (SSE) from one model to the other.

1.1 Regression of TTT on IVTT

Five models are candidates to express the variation of TTT as a function of IVTT. According to Table 6, the magnitude of the constant term in models 1 and 3 renders the inclusion of the constant term into the models questionable, since the constant term represents the mean of the probability distribution of TTT at IVTT=0, that is an overestimation of transit travel time by 16 and 13 minutes respectively. A further reason which might support the deletion of the constant term from the models is the variation in SSR in relation to SST(the total sum of square).

No	Y	X	Models	R^2	SSR	SSE	SST	Overall Signif.	Std. Error of Coeff.			Signif		
									Cnst.	X	X^2,LogX	Cnst.	X	X^2,LogX
1	TTT	IVTT	TTT=17.30+1.51(IVTT)	.27709	97400	.25 +6	.35 +6	.0000	3.90	.15	---	.0000	.0000	---
2	TTT	IVTT	TTT= 2.1(IVTT)	.27709	.86 +6	.27 +6	.11 +7	0.	---	.071	---	---	0.	---
3	TTT	IVTT, $IVTT^2$	TTT=15.43+1.68(IVTT) -.003$(IVTT)^2$	.27734	97489	.25 +6	.35 +6	.0000	7.20	.55	.009	.0330	.0026	.7570
4	TTT	IVTT, $IVTT^2$	TTT=2.79(IVTT)-.02$(IVTT)^2$	.27079	.86 +6	.26 +6	.11 +7	0.	---	.193	.005	---	.0000	.0001
5	TTT	LogIVTT	TTT=18.04Log(IVTT)	.25688	.85 +6	.28 +6	.11 +7	0.	---	---	.62	---	---	0.

Table 6 - Linear Models to Estimate Missing Values of Transit Travel Time

For instance, by deleting the constant term from Model 1, the ratio SSR/SST increases from 0.278 to 0.782. This increase indicates that the total variability in TTT which is accounted for by Model 2, is greater than that accounted for by the first model, since SSR may be considered as a measure of the variability of TTT associated with the regression line, and the larger SSR is in relation to SST, the greater is the effect of the relationship. Therefore, Models 2, 4, 5 are candidates for further analysis.

1.2 Regression of MOCOST on IVTT

Table 7 shows three models relating the monthly operating cost of an auto driver to his driving time to work.

Since the driving time to work is not the only variable which determines the monthly operating cost, the presence of a constant term in the models may be justified by assuming that it would take account of other relevant variables which are not included in the model due to the lack of data. This constant term can be considered to be the average fixed cost of owning a car. The low squared coefficient of correlation and the large standard error of the regression coefficients are a matter of serious concern. These are mainly due to the nonconstancy of error variance (Heteroskedasticity). In other words: $\sigma^2(\epsilon) = k(\text{IVTT})$

To avoid the problem of heteroskedasticity, the classical transformation :

No	Y	X	Models	R^2	SSR	SSE	SST	Overall Signif.	Model Expression	Coef.Std.Err		Signif.	
										CNST.	X,1/X	CNST.	X,1/X
1	MOCOST	IVTT	MOCOST=35.30+.054IVTT	.00052	77.175	.15 +6	.15 +6	.7661	MOCOST=35.30-.054(IVTT)	4.98	.18	.0000	.7661
2	$\frac{MOCOST}{IVTT}$	$\frac{1}{IVTT}$	$\frac{MOCOST}{IVTT} = .42 + \frac{27.70}{IVTT}$	.27090	199.44	536.79	736.23	.0000	MOCOST=27.70+.41(IVTT)	.24	3.48	.0797	.0000
3	$\frac{MOCOST}{IVTT}$	$\frac{1}{IVTT}$	$\frac{MOCOST}{IVTT} = .03 + \frac{5}{IVTT}$	.72964	6.5517	2.4277	8.9794	.0000	MOCOST=25+.3(IVTT) +.001$(IVTT)^2$	.016	.234	.0310	.0000

Table 7 - Linear Models to Estimate Missing Values of Monthly Operating Cost

$(MOCOST)' = MOCOST/IVTT$, $(IVTT)' = 1/IVTT$

and

$(MOCOST)' = SQRT(MOCOST)/IVTT$, $(IVTT)' = 1/IVTT$

are used.

Note the increase in the squared coefficient of correlation and the reduction in the coefficients' standard error in relation to their magnitude. These changes are specially considerable for Model 3. Therefore, the two following regression models are chosen:

$$MOCOST = 27.90 + .40(IVTT)$$

$$MOCOST = 25 + .30(IVTT) + (IVTT)^2$$

2. POPULATION COMPARISON

In order to select the final models, it is useful to compare for each variable the mean of survey population (Population 1:with missing data) and the mean of population resulted from completing the missing data by means of regression models (Population 2).

To illustrate the problem, assume that each population is normally distributed, even though the distribution of some variables are largely skewed (see Tables 8, 9). Let μ_1 and μ_2 represent the two population means, σ_1 and σ_2 the two population variances and m_1, m_2, s_1^2 and s_2^2 their estimators, respectively.

Population 1				No	Models	Population 2				F	Signif
N	Mean	Var.	Skew.			N	Mean	Var.	Skew.	Stat.	
463	51.61	1461.5	2.76	2	TTT=2.1(IVTT)	679	51.27	916.37	1.056	1.594	.0000
				4	$TTT=2.79(IVTT)-.02(IVTT)^2$	679	52.24	511.24	.107	2.858	.0000
				5	TTT=18.04 Log(IVTT)	679	54.74	131.81	-.476	11.088	0.

Table 8 - Comparison of Two Transit Travel Time Population Variances

Population1				No	Models	Population 2				F	Signif
N	Mean	Var.	Skew.			N	Mean	Var.	Skew.	Stat.	
275	35.36	644.10	3.77	1	MOCOST=27.70+.42(IVTT)	679	37.95	36.42	1.27	17.68	0.
				2	$MOCOST=25+.30(IVTT)+.001(IVTT)^2$	679	33.24	28.40	1.35	22.68	0.

Table 9 - Comparison of Two Operation Cost Population Variances

If the underlying populations have the same variances, then the sampling distribution of (m_1-m_2) will have the t-distribution with (N_1+N_2-2) degrees of freedom. On the other hand, if the variances differ, then the sampling distribution of (m_1-m_2) will have a t-distribution with DF degrees of freedom, where DF is equal to(see Affifi and Azen(1972), Dixon(1969)):

$$DF = \frac{(A_1 + A_2)^2}{A_1^2/(N_1+1) + A_2^2/(N_2+1)} - 2$$

where $Ai = si^2 / Ni$.

2.1 Comparison of Population Variances

For this purpose, we shall use the statistic s_1^2/s_2^2 and test the null hypothesis $H_0 : \sigma_1^2=\sigma_2^2$ vis-a-vis $H_1 : \sigma_1^2\neq\sigma_2^2$. Under the assumption of normality, this statistic has an F-distribution with (N_1-1, N_2-1) degrees of freedom. The analysis of variances of Populations 1 and 2 are shown in Tables 8, 9. These tables clearly indicate that the variance of Population 1 (Survey Population) differs from that of the Simulated Population, since the null hypothesis H_0 can be rejected even at 1% level of significance. In other words, the probability of having $\sigma_1^2\neq\sigma_2^2$ is more than 99%.

This difference in population variances is due to:

i. the size of populations. One expects larger deviation from the mean as the sample size decreases.

ii. the property of regression models. Model parameters estimated by least square method are unbiased and have minimum variance among all unbiased linear estimators. This property implies that Y' (the value of the estimated regression function) is an unbiased estimator of E(Y), with minimum variance in the class of unbiased linear estimators. Therefore, the simulated populations TTT and MOCOST will have the smallest variance.

2.2 Comparison of Population Means

In order to compare the population means, we form the statistic

$$\frac{(m_1-m_2) - (\mu_1 - \mu_2)}{\mathrm{SQRT}[(s_1{}^2/N_1)+(s_2{}^2/N_2)]}$$

and test the null hypothesis $H_0 : \mu_1-\mu_2=0$. Since the variances are not equal, the t-distribution has DF degrees of freedom, where DF is defined in Section(2) .

Table 10 shows the comparison of Transit Travel Time population means for Models 2, 4, and 5. By testing at 5% level of significance, Model 5 can be deleted. Model 4 yields to the smallest 95% confidence interval length. However,

NO	MODELS	T-Statistic	D.F.	Signif	Confidence Interv.(.95)		
					L.Limit	U.Limit	Range
2	TTT = 2.1(IVTT)	0.160	838.71	0.4365	-3.825	4.496	8.321
4	TTT = 2.79(IVTT) - .020(IVTT)^2	-0.319	683.17	0.3749	-4.499	3.251	7.750
5	TTT = 18.08 Log(IVTT)	-1.710	519.48	0.0439	**	---	---

** Not significant at 5%

Table 10 - Comparison of 'TTT' Population Means

No	Models	T-Statistic	D.F.	Signif	Confidence Interv.(.95)		
					L.Limit	U.Limit	Range
1	$TTT = 27.70 + .42(IVTT)$	-1.673	286.72	0.0477	**	---	---
2	$TTT=25+2.79(IVTT)-.020(IVTT)^2$	1.373	283.90	0.0855	-0.903	5.149	6.052

** Not significant at 5%

Table 11 - Comparison of 'MOCOST' Population Means

Dep.Var.	Indep.Var	Model Expression	R-Square
TTT	IVTT	$TTT = 2.1(IVTT)$	.27709
MOCOST	IVTT	$MOCOST = 25 + .30(IVTT) + .001(IVTT)^2$	.72964

Table 12 - Selected Models

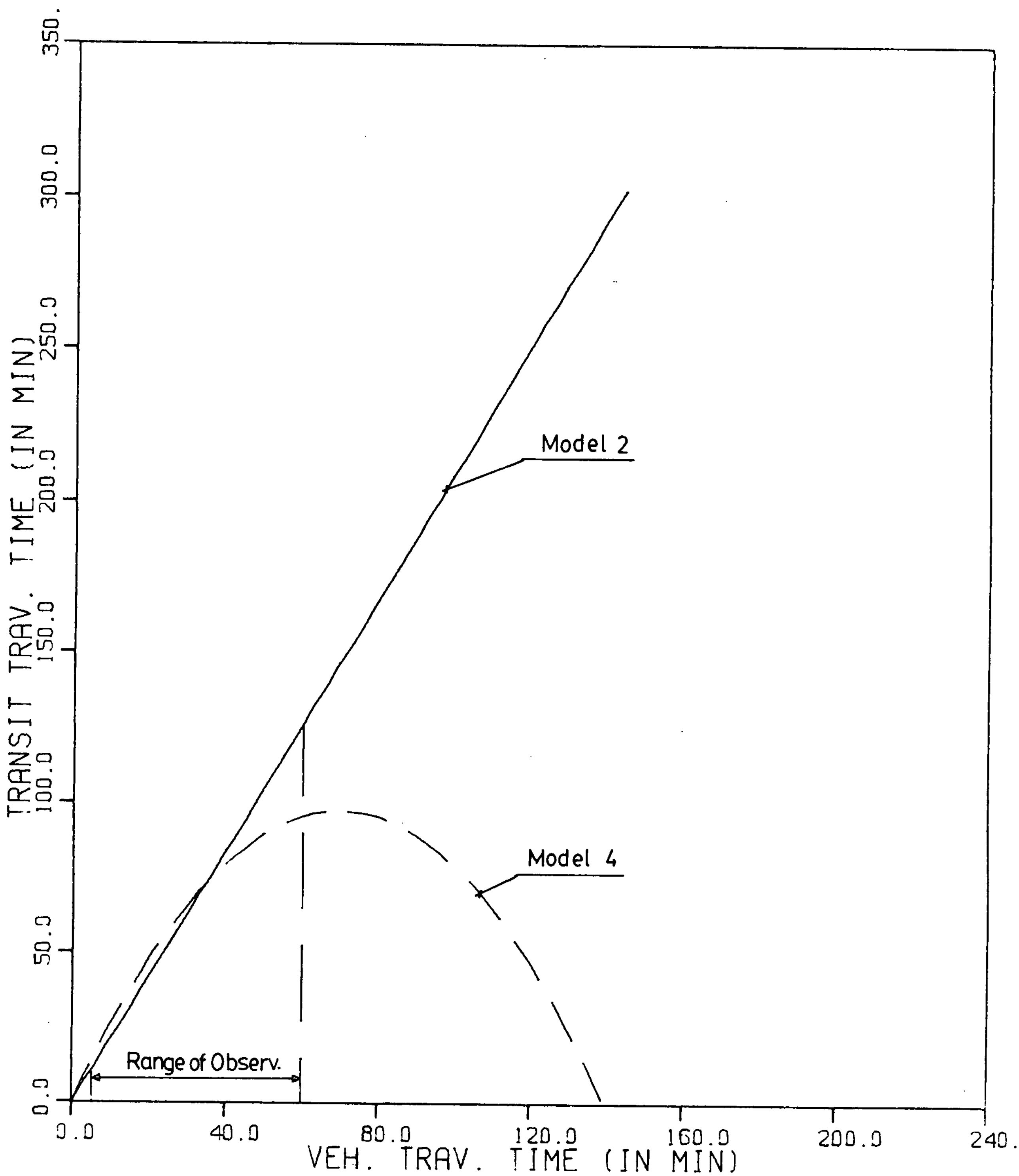

Figure 4 - Transit v.s. Private Car Travel Time

by inspecting Figure 4, the output of Models 2 and 4 differs slightly for in-vehicle travel time less than 40 minutes. Note that the range of observations is limited to about 60 minutes. Model 2, therefore, was chosen as the final model explaining the relationship of TTT and IVTT due to the simplicity of its equational form.

For MOCOST, the inspection of Table 11 indicates that Model 2 can be selected for the final model expressing the variation of MOCOST in terms of IVTT, since the null hypothesis H_0 can be rejected at 5% level of significance.

Selected models are represented in Table 12. The low squared coefficient of correlation is due to the nature of data and is considered acceptable. One should expect more variability in disaggregate data than in zonally aggregated data, since the latter tends to submerge the observed variability at an individual level by giving only one figure representing the zonal resident characteristics. Moreover, we need to make sure that our prediction models are used only for values falling within the range of observations, since beyond these limits the results may not longer apply.

3. SIMULATION OF WAITING TIME VARIABLE

As previously mentioned, waiting time seems likely to be uncorrelated with the in-bus travel time. Therefore, the regression approach was not used to simulate this variable.

The approach taken in the present study is often used in practice, when the a-priori probability distribution of the variable is known to analysts, the so-called random generating model consists of generating random waiting times utilizing their estimated probability distribution. This estimation is made by the method of moments on the sample of the complete values of the waiting time variable.

The method of moments assumes that the estimator u of μ should be the sample mean and the estimator s^2 of the variance σ^2 is the sample variance. Since the observed distribution is skewed to the left, two types of distribution are considered :

i. GAMMA distribution[1] :

$$f(x) = \frac{\lambda^{k}.x^{k-1}.\exp(-\lambda x)}{\Gamma(k)}$$

where $x \geq 0$ and λ and k are the model parameters

$\mu = k/\lambda$; $\sigma = SQRT(k)/\lambda$

The use of the method of moments implies that :

$u = \hat{k}/\hat{\lambda} = 5.59$ and $s = SQRT(\hat{k})/\hat{\lambda} = 4.18$

Therefore : $\hat{k} = 1.79$ and $\hat{\lambda} = 0.32$

1- According to the comment made by Dr.Navin, 2-step density functions can also be applied(i.e. linear+negative exponential).

ii. LOGNORMAL distribution :

$$f(x) = 1/[\mathrm{SQRT}(2\Pi)\sigma'.x]\ \exp\{-1/2[(\ln x-\mu')/\sigma']^2\}$$

where μ' and σ' are the mean and standard deviation of the natural logarithm of x.

u' = 1.48 and s' = 0.70

In order to make a final selection between these distributions , the X^2 closeness-of-fit statistic with (k-r-1) degrees of freedom is used. The letter k represents the number of categories considered and r is the number of parameters estimated from the data.

The results of this investigation are shown in Fig. 5 and 6, and Tables 13 and 14. Due to the lower magnitude of the total normalized squared difference and the better fit obtained in the range of 4-7 minutes, one might conclude that the use of a lognormal distribution is more appropriate to the present data set.

4. COMPLETION OF PARKING COST

The Aggregation method is used to complete missing information on parking cost. This method was applied as follows:

Land use characteristics of the work trip destinations are available from the survey. For each type of destination land use, an average parking cost is calculated from

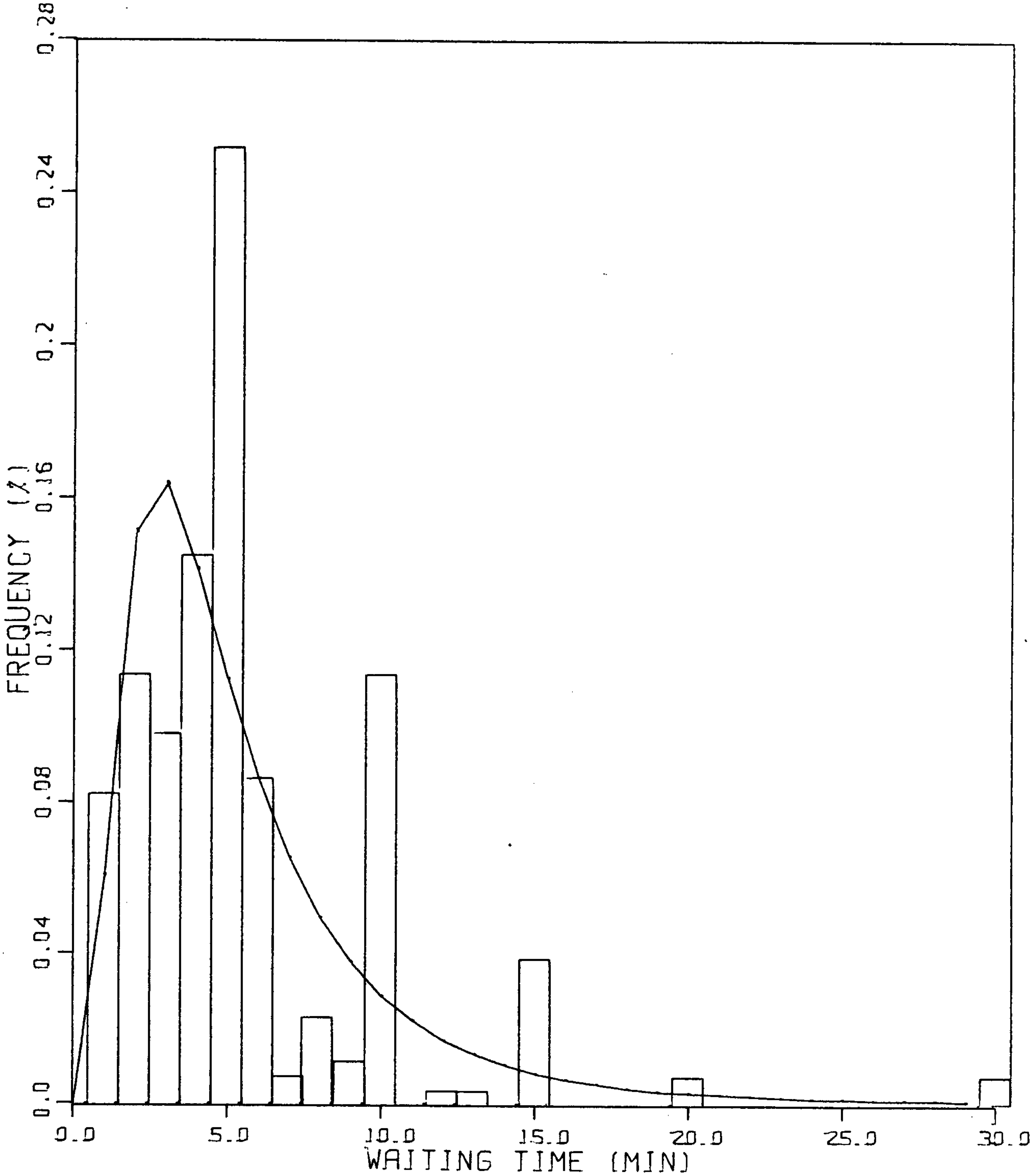

Figure 5 - Gamma Generation Model

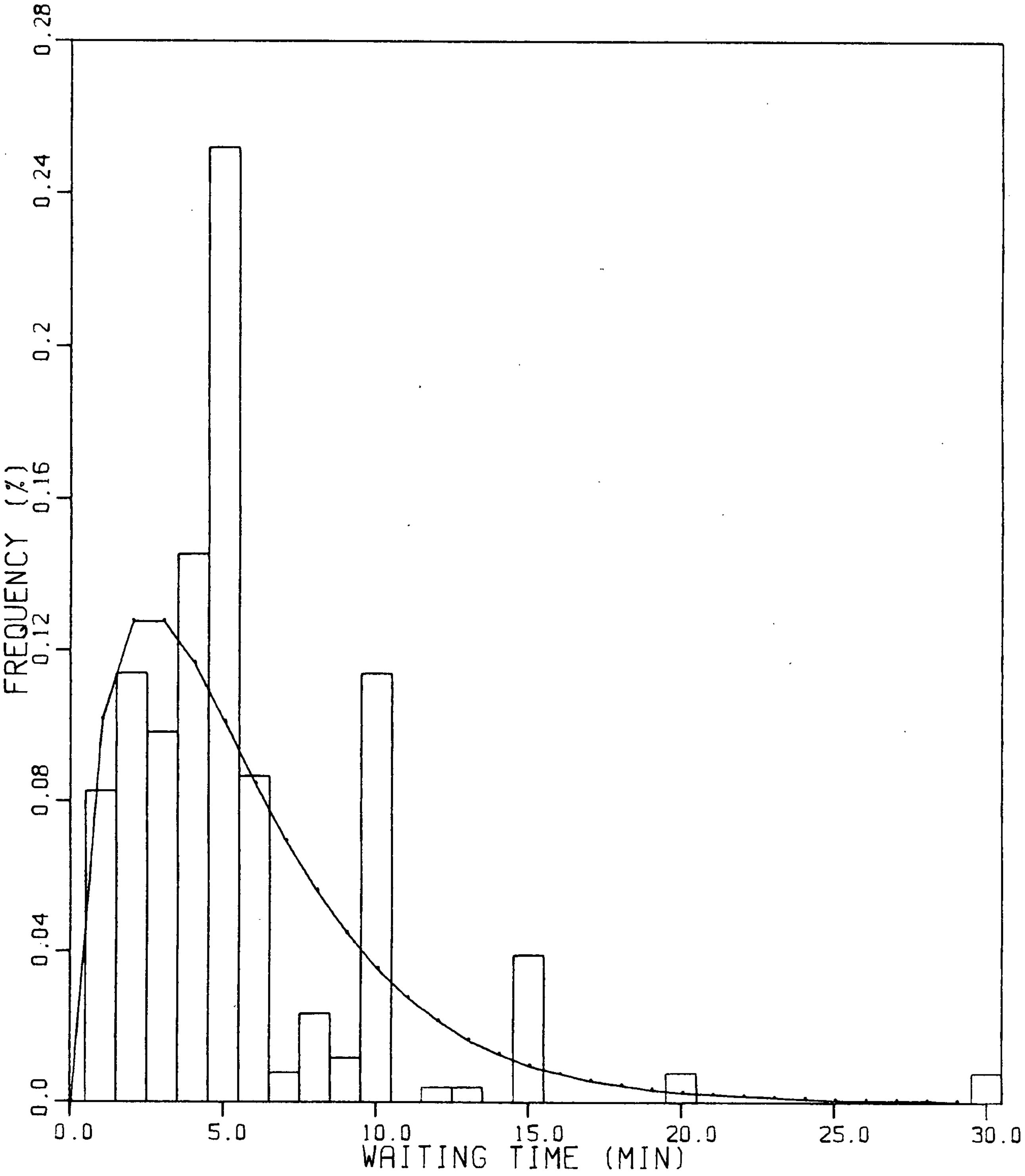

Figure 6 - Lognormal Generation Model

Xi	Ni		f(Xi)	n.f(Xi)		$(Ni-n.fi)^2/n.fi$	
1	21		.1017	25.83		.90	
2	22		.1277	32.43		.36	
3	25		.1278	32.44		1.71	
4	37		.1164	29.57		1.86	
5	64	125[1]	.1008	25.61	94.28	57.52	10.0
6	22		.0846	21.48		.013	
7	2		.0694	17.62		13.85	
8	6		.0560	14.22		4.75	
9	3		.0446	11.33		6.13	
10	29	39	.0352	8.94	46.93	45.00	1.34
11	0		.0276	7.00			
12	1		.0214	5.44		3.63	
13-20	15		.0868	22.05		2.25	
	254			254		16.56	

1- Aggregated Value

X^2-Stat.(.01,4) = 13.28

X^2-Stat.(.005,4)= 14.86

Table 13 - GAMMA Model Goodness-of-Fit

Xi	Ni		f(Xi)	n.f(Xi)		$(Ni-n.fi)^2/n.fi$	
1	21		.0612	15.48		1.96	
2	22		.1515	38.48		2.33	
3	25		.1638	41.59		6.62	
4	37		.1412	35.86		.035	
5	64	125[1]	.1121	28.46	102.74	44.37	4.82
6	22		.0861	21.85		.001	
7	2		.0652	16.57		12.81	
8	6		.0494	12.54		3.41	
9	3		.0375	9.51		4.46	
10	29	39	.0286	7.25	39.17	65.14	0.001
11	0		.0219	5.57			
12	1		.0169	4.30		2.53	
13-20	15		.0646	16.45		.13	
	254			254		15.85	

1- Aggregated Value

X^2-Stat.(.01,4) = 13.28

X^2-Stat.(.005,4)= 14.86

Table 14 - LOGNORMAL Model Goodness-of-Fit

the complete information. These zonal average figures are then used to complete the missing values of parking charges according to their corresponding land use information.

5. SUMMARY

Since the calibration of a logit or probit model requires a relatively large sample size, the completion of the missing data was a necessary pre-requisit for calibration. For this purpose, three data simulation methods were used: Regression Analysis to simulate TTT and MOCOST; Random Generation Model to generate waiting time and Aggregation method to estimate the missing value of parking cost. Statistical inferences were made to select the more appropriate model. However, the use of this simulated data may have some implications in the calibration of the modal choice models. For instance, the true waiting time distribution might not be lognormal which will bias the impact of this variable on the choice probability. Therefore, the comparison of the calibrated model and results found in previous studies may be a relevant way of evaluating this attempt.

IV. BEHAVIOURAL TRAVEL DEMAND THEORY

In this section, we are interested in estimating modal choice probability models relating the probability of using a travel mode to a series of exogenous variables. Similar to other demand models, multivariate regression models can be used. But, since the dependant variable has a binary form (the user will or will not choose the travel mode i), the estimation of the model parameters becomes complicated. Sources of difficulties are as follows :

i. The error variance is not constant

if we let $E[Y=1|X] = p$ then

$Var[Y] = E[Y^2]-\{E[Y]\}^2 = p-p^2 = p(1-p)$

therefore, the variance around Y is function of the estimated value.

ii. The regression function is constrained

$0 \leq E[Y|X] \leq 1$

This latter is the most troublesome, since for some value of the independant variables, the estimated value of the dependant variable exceeds 1 or or is less than 0. Due to the fact that we can not make a transformation of the dependant variable to linearize its expected value, we are forced to fit a model which is non-linear in the parameters. Two models are commonly used for this purpose : the logit and probit model.

1. MATHEMATICAL THEORY OF BEHAVIOURAL MODELS

According to the approach taken by economists to diagnose consumer preferences, a behavioural model should investigate the manner that an individual uses to decide among several characteristics. Applying this approach to transportation demand, it is hypothesized that the traveller reaches a decision by considering the perceived alternative travel modes by valuing the attributes of each mode. His behaviour follows the relative values he associates to different attributes and may be described by his indifference curves since these latters represent all combination of choices among which the traveller is indifferent. The family of indifference curves is called the utility function and can be expressed by the following equation:

$$Umi = V(\underline{X}m, \underline{S}i) \qquad 4.1$$

where $\underline{X}m$ is the set of service attributes of alternative m and $\underline{S}i$ the set of the attributes of individual i.

However, this formulation implicitly assumes that the consumer is aware of all possible combination of attributes and makes a decision with perfect information. In the real world, this assumption is rarely satisfied and therefore a stochastic term should be introduced into (4.1) to express the probabilistic error made each time the utility of a given mode is evaluated. The functional form of U will then be as :

$$U_{mi} = V(\underline{X}_m, \underline{S}_i) + \epsilon_{mi} \qquad 4.2$$

where V is the non-stochastic term and represents a common element shared by a subset of population predefined according to their socio-economic characteristics, and ϵ_{mi} is the i<u>th</u> traveller's taste not shared by others, and since it can not be assessed , the assignment of a probability distribution to any individual taste becomes necessary. Therefore ϵ_{mi} forms the stochastic component of the utility function. Note that in transportation demand, since the service attributes (travel time, travel cost,...) are negatively valued, that is, the utility of a mode increases as travel cost or time decreases, one should use the term disutility for the function U. However, in the present study, we will use the term utility and assign appropriate signs to the parameters of the service attributes.

With this in mind, a consumer is assumed to be a utility maximizer and prefers mode m to remaining modes, only if :

$$U_{mi} > U_{ki} \qquad \text{for } k=1,\dots,M \; ; \; k \neq m \qquad 4.3$$

This is the equivalent of stating that a consumer will choose that mode for which the greatest utility (disutility) is perceived.

In order to analyse the behaviour of commuter i when he is faced with M alternatives, we should assign a probability to his decision, that is :

$$Pmi = P(Umi > Uki) \qquad \text{for } k=1,\dots,M \;;\; k \neq m \qquad 4.4$$

where Pmi is the probability that individual i takes mode m to work.

$$Pmi = P(V(\underline{X}m,\underline{S}i)+\epsilon mi > V(\underline{X}k,\underline{S}i)+\epsilon ki) \text{ for } k=1,\dots,M \;;\; k \neq m$$

or to facilitate the notation

$$Pmi = P(Vmi+\epsilon mi > Vki+\epsilon ki) \text{ for } k=1,\dots,M \;;\; k \neq m \qquad 4.5$$

In the present study, in order to determine a structural solution to Equation 4.5, we will consider the special case of binary choice (M=2). Therefore, 4.5 becomes :

$$Pmi = P(V1i+\epsilon_1 i > V2i+\epsilon_2 i) \qquad 4.6$$

Clearly, to develop a model, it is necessary to assume a specific distribution for the probabilistic components $\epsilon_1 i$ and $\epsilon_2 i$.

1.1 Multivariate Logit Model

Under the assumption of independancy of error terms $\epsilon_1 i$ and $\epsilon_2 i$, their joint probability density in the Cartesian space is given by :

$$f(x,y) = f_1(x) \cdot f_2(y) \qquad 4.7$$

where x and y are the values taken on by random variables $\epsilon_1 i$ and $\epsilon_2 i$ respectively. Therefore, probabilities in the ($\epsilon_1 i, \epsilon_2 i$) plane are assigned in accordance with Equation 4.8 :

$$P(E) = \iint_E f_1(x) \cdot f_2(y) \; dx \, dy \qquad 4.8$$

By arranging (4.6) we have for mode 1:

$$P1i = (\epsilon_2 i - \epsilon_1 i < V1i - V2i) \qquad 4.9$$

Consider a new random variable ϵ such that $\epsilon = \epsilon_1 i - \epsilon_2 i$. Thus the event $E = \{\epsilon < V1i - V2i\}$ is defined by the region of $(\epsilon_1 i, \epsilon_2 i)$ plane such that :

$$y - x < V1i - V2i$$

$$y < V1i - V2i + x \qquad 4.10$$

Therefore,

$$P1i = P(\epsilon_2 i - \epsilon_2 i < V1i - V2i)$$

$$= \iint_{y<V_{1i}-V_{2i}+x} f_1(x).f_2(y)\ dx\ dy$$

$$= \int_{-\infty}^{+\infty} f_1(x) \int_{-\infty}^{V_{1i}-V_{2i}+x} f_2(y)\ dx\ dy \qquad 4.11$$

Let $F_2(y)$ denote the probability distribution of ϵ_2; therefore:

$$P1i = \int_{-\infty}^{+\infty} f_1(x)\ .\ F_2(V1i-V2i+x)\ dx \qquad 4.12$$

Equation 4.12 is called the convolution of two density functions f_1 and f_2.

The logit model results if ϵ has a Weibull distribution with 0 mean value. One of the most important characteristics of the Weibull distribution is that the difference of Weibull distributed variables has a logistic distribution which leads to the desired sigmoid shape. The Weibull distribution is given by:

$$W(x) = Prob(\epsilon<x) = \exp[-\exp(-x)] \qquad 4.13$$

Therefore, the associated frequency distribution is :

$$w(x) = dW(x)/dx = \exp(-x)\ .\ \exp[-\exp(-x)] \qquad 4.14$$

By substituting Eqs. 4.13 and 4.14 into 4.12, we obtain :

$$P1i = \int_{-\infty}^{+\infty} \{\exp(-x)\exp[-\exp(-x)]\}.\{\exp[-\exp(-(V1i-V2i+x))]\}\ dx$$

$$= \int_{-\infty}^{+\infty} \exp(-x).\exp[-\exp(-x).(1+\exp(V2i-V1i))]\ dx$$

Using the transformation t=exp(-x) results in :

$$P1i = \int_{0}^{+\infty} \exp[-t(1+\exp(V2i-V1i))]\ dt$$

$$= -\{\exp[-t(1+\exp(V2i-V1i))] \;/\; (1+\exp(V2i-V1i))\}\Big|_{0}^{+\infty}$$

And hence :

$$P1i = 1 \;/\; (1+\exp(V2i-V1i))$$

$$= \exp(V1i-V2i) \;/\; 1+\exp(V1i-V2i) \qquad 4.15$$

Equation 4.15 states that the probability of taking mode 1 is a function of the difference between the non-stochastic terms V1i and V2i of the utility functions. Note that the non-stochastic term of utility functions has the form:

$$Vmi = \beta_0+\beta_1 X^{\lambda}{}_1+\ldots+\beta_n X^{\lambda}_{mn}+\ldots+\beta_k X^{\lambda}_{\Delta k} = \underline{\beta}.\underline{X}mi$$

and the only restriction imposed on it, is that this relation must be linear in parameters (βi), whereas, $\underline{X}$ may be a complex transformation of the raw data.

It should also be noticed that the parameters of the utility functions (βi) consist of two components : deterministic and random.

$$\beta = \tilde{\beta} + \upsilon im$$

where υim represents the unobserved random taste of the individual i for mode m. However, to enable the formulation to be used for consumer behaviour, it is usually assumed that $\upsilon im=0$.

By letting

$$V1i - V2i = V(\underline{X}i, \underline{\beta})$$

the probability of choosing mode 2 will be equivalent to :

$$P2i = 1-P1i = 1-[\ \exp(V(\underline{X}i,\underline{\beta}))\ /\ 1+\exp(V(\underline{X}i,\underline{\beta}))\ ]$$

$$P2i = 1\ /\ [\ 1+\exp(V(\underline{X}i,\underline{\beta}))\ ] \qquad 4.16$$

By taking the logarithm of the probabilities ratio :

$$L = \log(P1i\ /\ 1-P1i) = \log(P1i)-\log(1-P1i)$$

and substituting P1i by its expression, we form the logit L :

$$L = V(\underline{X}i,\underline{\beta}) \qquad 4.17$$

The major properties of this transformation are as follows :

i. The logit is a linear function of the difference between the non-stochastic components of utility functions (see Eq. 4.17), whereas the probabilities themselves are not (see Eqs. 4.15, 4.16)

ii. While the probabilities are bounded, the logit is unbounded with respect to the value of $V(\underline{X}i, \underline{\beta})$. This is in accordance with the constraint imposed on the regression function (see p.45).

1.2 Multivariate Probit Model

The specification of the probit model is straightforward and follows the definition of the normal probability function. However, several assumptions are needed.

It is assumed that each person possesses a different critical value or a threshold Ui which determines whether the underlying person will take car or bus. This critical value is unobserved and hence unavailable to analysts.

In order to make a decision, individual i compares his own critical value to choice index $G(\underline{X}mi, \underline{S}i, \underline{\beta})$. This index is produced by the linear combination of the exogenous variables or "stimulus". Exogenous variables, themselves, can be a complex transformation of the raw data.

$$G(\underline{X}mi, \underline{S}i, \underline{\beta}) = G(\underline{X}i, \underline{\beta}) = \beta_0 + \beta_1 \dot{x}^{\lambda}_{m1} + \ldots + \beta_n \dot{x}^{\lambda}_{mn} + \ldots + \beta_k \dot{x}^{\lambda}_{\Delta k} \quad 4.18$$

where $\underline{X}mi$ and $\underline{S}i$ are the sets of the service attributes of the mode m and socio-economic characteristics of commuter i, respectively.

Therefore, traveller i will choose :

i. mode 1 if $G(\underline{X}i,\underline{\beta}) \geq Ui$

ii. mode 2 if $G(\underline{X}i,\underline{\beta}) < Ui$ 4.19

Since all individuals will not possess the same threshold level, a probability distribution is assigned to critical values. If we assume that this distribution is normal N(0,1), the probability that commuter i possesses a critical value greater than the choice index, and consequently chooses mode 1, is given by :

$$P1i = 1/SQRT(2\Pi) \int_{-\infty}^{G(\underline{X}_i,\underline{\beta})} \exp(-t^2/2) \; dt \qquad 4.20$$

Since we are limiting the present study to the case of binary choice, the probability of choosing mode 2 will be equivalent to:

$$P2i = 1\text{-}P1i = 1-[\; 1/SQRT(2\Pi) \int_{-\infty}^{G(\underline{X}_i,\underline{\beta})} \exp(-t^2/2) \; dt \;]$$

$$P2i = 1/SQRT(2\Pi) \int_{G(\underline{X}_i,\underline{\beta})}^{+\infty} (\exp(-t^2/2)) \; dt \qquad 4.21$$

$G(\underline{X}i,\underline{\beta})$ is defined as the probit of Pmi(m=2)

One of the above assumptions which needs justification is the normality of the critical values. These values are a complex combination of a large number of psychological, physiological, social and cultural factors which can be assumed

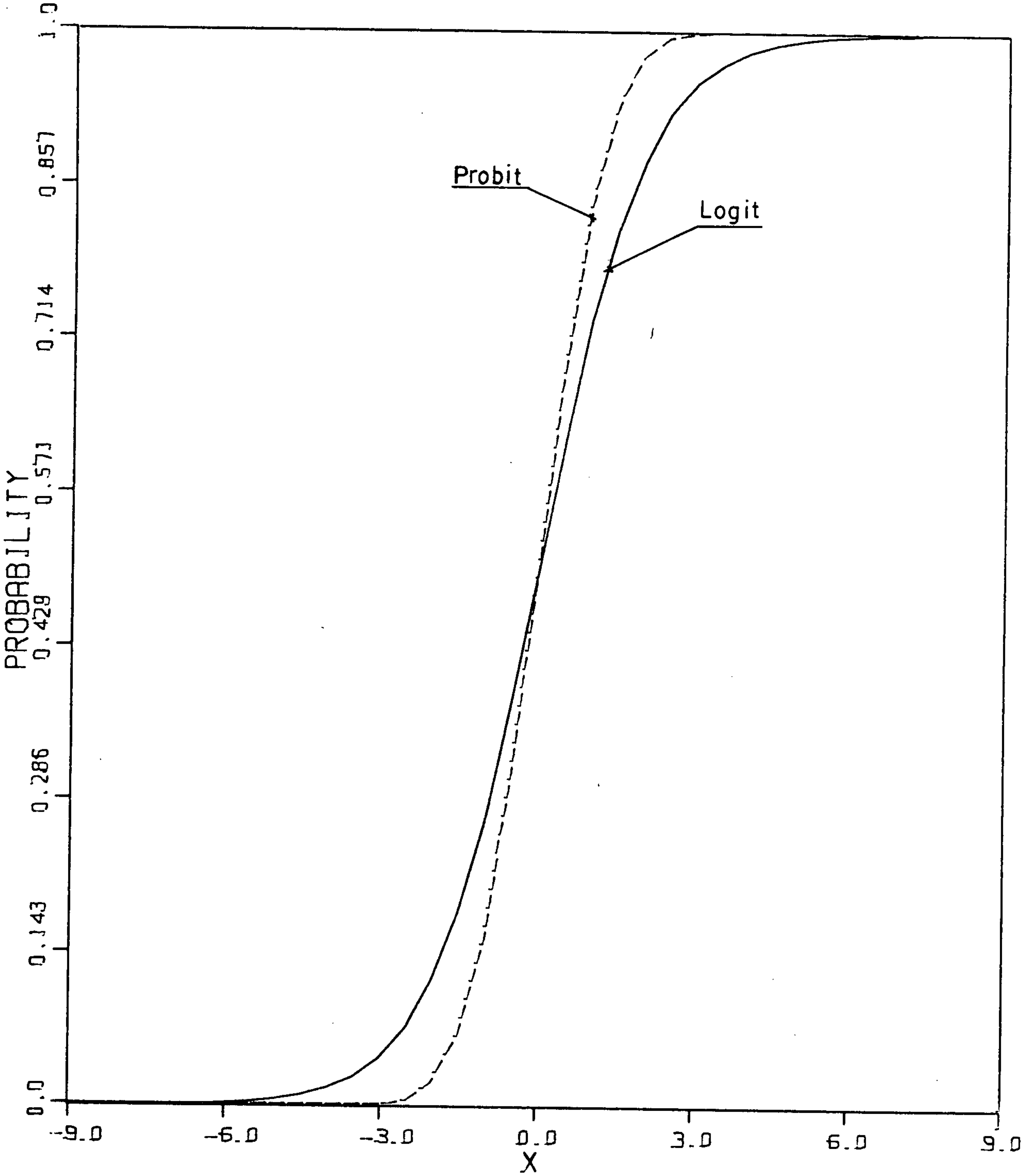

Figure 7 - Comparison of Logistic and Probit Forms

independant.Therefore, according to the Central Limit theorem which states that sums of independant random variables are asymptotically normally distributed, we can accept the fact that critical values are normal.

Similar to the logit transformation, probit transformation of the choice index produces a sigmoid curve. It is worthwhile to remark that the difference between a logistic distribution (obtained from logit transformation), and a cumulative normal distribution (obtained from probit transformation) is not very large(see Fig.7).

2. ESTIMATING MODEL PARAMETERS

The model parameters can be estimated by means of weighted least square or maximum likelihood method.

2.1 Maximum Likelihood Estimation

The method of maximum likelihood chooses as estimators of the parameters βi, a set of statistics which maximize the likelihood function for the given values.

The likelihood function is defined as being the probability of the occurence of random sample (W1,W2,...,Wn) as a function of the unknown parameters ($\beta_0, \beta_1, \ldots, \beta_n$) with the condition that the unknown parameters ($\underline{\beta}$) belongs to the parameter space Ω which has to be specified for the case study

$$g(\underline{\beta}) = \prod_{i=1}^{N} f(\underline{X}i, \underline{\beta}) \qquad \underline{\beta} \in \Omega$$

where N is the number of observations.

Since our problem consists of determining whether or not a person selected randomly chooses a particular travel mode, random variables Wi are defined as Wi=1 if commuter i selects mode 1 and Wi=2 if he does not:

$$P[Wi=1] = P1 \text{ and}$$

$$P[Wi=2] = P2 = 1-P1 \qquad 4.22$$

Assuming Wi's are mutually independant, the likelihood function of the parameters is given by:

$$g(\underline{\beta})=P(W1,W2,\ldots,WN1,\ldots, WN)=\prod_{i=1}^{N_1}(P[Wi=1]).\prod_{i=N_1+1}^{N}(P[Wi=2]) \qquad 4.23$$

where N1 is the number of observations that select mode 1. Substituing 4.22 into 4.23 yields to :

$$g(\underline{\beta}) = \prod_{i=1}^{N_1}(P1i).\prod_{i=N_1+1}^{N}(1-P1i) \qquad 4.24$$

and the parameter space will be defined as :

$$\Omega : \{ R^k \}$$

But, since P1i is a function of service attributes and socio-economic characteristics of the trip makers, 4.23 can be written as :

$$g(\underline{\beta}) = \lambda(\underline{X},\underline{\beta}) \tag{4.25}$$

The aim of the maximum likelihood estimation is to choose the vector of unknown coefficients $\underline{\beta}$ in a manner which make the probability defined in 4.23 as large as possible. For this, the set of $\underline{\beta}$ maximizing $\lambda(\underline{X},\underline{\beta})$ and therefore satisfying the following equation has to be determined.

$$\delta(\lambda(\underline{X},\underline{\beta})) \;/\; \delta(\beta j) \;=\; 0 \text{ for } j=1,\ldots,k \tag{4.26}$$

To facilitate the differentiation task, $\log[\lambda(\underline{X},\underline{\beta})]$ is considered rather than $\lambda(\underline{X},\underline{\beta})$

In general, maximum likelihood estimators for large sample sizes possess some desirable properties. They are as follows :

i. Maximum likelihood estimators are consistent. That is, $\lim_{N\to\infty} \text{Prob}(|b_j-\beta_j|<\epsilon) = 1$ where ϵ is an arbitrary positive value.

ii. Maximum likelihood estimators are unbiased minimum variance. That is, b_j has among all unbiased estimators($E[b_j]=\beta_j$), the smallest variance.

iii. Maximum likelihood estimators are approximately normally distributed.

Since the data available for this study consists of 621 observations, these properties apply.

2.2 Estimating Logit Model Parameters

According to 4.24 and 4.25, the likelihood function for the logit model is :

$$\lambda(\underline{X},\underline{\beta}) = \lambda = \prod_{i=1}^{N_1}(P1i) \;.\; \prod_{i=N_1+1}^{N}(1-P1i)$$

and the logarithm of this function will be :

$$\Lambda = \log(\lambda) = \sum_{i=1}^{N_1} [\log(P1i)] + \sum_{i=N_1+1}^{N} [\log(1-P1i)] \qquad 4.27$$

Substitution of 4.16 and 4.17 into 4.27 yields to :

$$\Lambda = \sum_{i=1}^{N_1} \{\log[\exp(V(\underline{X}i,\underline{\beta})) \;/\; 1+\exp(V(\underline{X},\underline{\beta}))]\}$$
$$+\sum_{i=N_1+1}^{N} \{\log[1 \;/\; 1+\exp(V(\underline{X},\ \underline{\beta}))]\}$$

$$= \sum_{i=1}^{N_1} \{\log[\exp(V(\underline{X}i,\underline{\beta}))]\} - \Sigma\ \{\log[1+\exp(V(\underline{X}i,\underline{\beta}))]\}$$
$$-\sum_{i=N_1+1}^{N} \{\log[1+\exp(V(\underline{X}i,\underline{\beta}))]\}$$

Therefore:

$$\Lambda = \sum_{i=1}^{N_1} \{V(\underline{X}i,\underline{\beta})\} - \sum_{i=1}^{N} \{\log[1+\exp(V(\underline{X},\underline{\beta}))]\} \qquad 4.28$$

In order to find the vector of the estimated value of $\underline{\beta}$ which maximizes Λ, we differentiate partially Λ with respect to βj :

$$\delta\Lambda/\delta\beta j = \sum_{i=1}^{N_1} (xij) - \sum_{i=1}^{N} [xij/(1+\exp(V(\underline{X},\underline{\beta})))] \text{ for } j=1,\ldots,K \qquad 4.29$$

By equating 4.29 to zero, we obtain a system of K non-linear equations that can be solved iteratively (K=number of exogenous variables).

2.3 Estimating Probit Model Parameters

According to 4.27, the log likelihood function for the probit model is :

$$\Lambda = \sum_{i=1}^{N_1} \{\log[P1(\underline{X}i,\underline{\beta})]\} + \sum_{i=N_1+1}^{N} \{\log[P2(\underline{X}i,\underline{\beta})]\} \qquad 4.30$$

where

$$P1(\underline{X},\underline{\beta}) = 1/\mathrm{SQRT}(2\Pi) \int_{-\infty}^{G(\underline{X}_i,\underline{\beta})} \exp(-t^2/2)\, dt$$

and

$$P2(\underline{X},\underline{\beta}) = 1/\mathrm{SQRT}(2\Pi) \int_{G(\underline{X}_i,\underline{\beta})}^{+\infty} \exp(-t^2/2)\, dt$$

And by differentiating 4.30 with respect to βj we have :

$$\begin{aligned} \delta\Lambda/\delta\beta j \;=\; & \sum_{i=1}^{N_1} \{[\delta(P1(\underline{X}i,\underline{\beta}))/\delta\beta j]/P1(\underline{X}i,\underline{\beta})\} \\ & + \sum_{i=N_1+1}^{N} \{[\delta(P2(\underline{X}i,\underline{\beta}))/\delta\beta j]/P2(\underline{X}i,\underline{\beta})\} \end{aligned} \qquad 4.31$$

In order to calculate $\delta(P1(\underline{X}i,\underline{\beta}))/\delta\beta j$, we apply Leibnitz' rule and therefore:

$$\delta(P1(\underline{X}i,\underline{\beta}))/\delta\beta j = 1/\mathrm{SQRT}(2\Pi)\ \exp[-1/2(G(\underline{X}i,\underline{\beta}))^2]xij \qquad 4.32$$

and

$$\delta(P2(\underline{X}i,\underline{\beta}))/\delta\beta j = 1/SQRT(2\Pi)\ \exp[-1/2(G(\underline{X}i,\underline{\beta}))^2]xij$$

And finally, by substituting 4.32 and 4.33 into 4.31 we will have a system of K equations which are non linear in β and which must be solved by an iterative process.

2.4 <u>Goodness of Fit</u>

Since the parameters of logit and probit models are estimated by the maximum likelihood method, hypothesis about the overall significance of the relationship may be tested by the likelihood ratio method. This method tests the null hypothesis that the probability of selecting a particular mode is independant of the values of the explanatory variables of the model. This can be formulated as testing

H_0 : all β=0

versus

H_1 : Not all β=0

Let assume that b_0 is the estimated value of β_0 for which $\Lambda(b_0,0,0,\ldots,0)$ is maximum and $(b_0,b_1,\ldots,b_n)$ are those which maximize $\Lambda(b_0,b_1,\ldots,b_n)$. The likelihood ratio will then be :

$$\Lambda = \Lambda(b_0,0,0,\ldots,0)\ /\ \Lambda(b_0,b_1,\ldots,b_n) \qquad 4.35$$

with the following critical region :

Reject H_0 if $0 \le \lambda \le c$

Accept H_0 if $c < \lambda \le 1$

where c is chosen in a fashion that the critical region has the desired size.

It is found that $-2\log\lambda$ is a more convenient statistic, since for large samples, this statistic is distributed as a chi-square random variable, with a degree of freedom equal to the difference between the number of parameters in the model.

An alternative statistic involving the likelihood ratio λ is the Pseudo R-Square coefficient. This coefficient, also called likelihood ratio index, is denoted by ρ^2 and is analogous with the well-known R^2 index of least-square regression analysis.

$$\rho^2 = \Lambda(b_0,0,0,\ldots,0) \;/\; \Lambda(b_0,b_1,\ldots,b_n) = 1-\lambda$$

However, the ρ^2 index values are considerably lower than those of R^2 index. For instance, if values of 0.8 to 0.9 for R^2 represent an excellent fit, for ρ^2 index these values correspond to 0.2 to 0.4 (McFadden1976).

3. DISCRIMINANT ANALYSIS

The third statistical tool selected to calibrate the relationship between a binary choice variable and a set of exogenous variables is Discriminant Analysis. In this approach, two separate populations are considered : private car and transit users. The aim of this method is to assign probability to an observation as coming from one of these populations such that the cost of misclassification is minimum. For this purpose, explanatory variables are linearly combined to form a scalar called the discriminant score. The linear discriminant function is of the type :

$$ymi = \sum_{k=1}^{K} (\beta_k \, . x^{i}_{mk}) \qquad m=1,\ 2 \qquad 4.36$$

where k=1,...,K=number of exogenous variable

i=1,...,N=number of trip makers(=observations)

m=1,2

The matricial presentation of 4.36 is :

$$\underline{Y} = \underline{\beta}' \, .\underline{X} = \underline{\beta} \, . \, \underline{X}' \qquad 4.37$$

where $\underline{\beta}'$ and $\underline{X}'$ are the transposed matrices of $\underline{\beta}$ and $\underline{X}$ respectively.

Let $f_1(\underline{X})$ and $f_2(\underline{X})$ represent the density functions of these two populations, q and 1-q denote the a-priori probability that an observation comes from population1 and 2 respectively, and c_1 and c_2 the respective misclassification costs. The conditional probability that an observed set of exogenous variables ($\underline{X}$), comes from population 1 is :

$$P1(\underline{X}) = (q.f_1(\underline{X}))/(q.f_1(\underline{X})+(1-q)f_2(\underline{X})) \qquad 4.38$$

and its expected cost of misclassification is:

$$E(C_1) = c_1(q.f_1(\underline{X}))/(q.f_1(\underline{X})+(1-q)f_2(\underline{X})) \qquad 4.39$$

Similarily we have:

$$P2(\underline{X}) = (q.f_2(\underline{X}))/(q.f_1(\underline{X})+(1-q)f_2(\underline{X})) \qquad 4.40$$

and

$$E(C_2) = c_2(q.f_2(\underline{X}))/(q.f_1(\underline{X})+(1-q)f_2(\underline{X})) \qquad 4.41$$

To relate $P(\underline{X})$ to the the set of explanatory variables, we form the log ratio of probabilities:

$$\log(P1(\underline{X})/1-P1(\underline{X})) = \log(q.f_1(\underline{X})/(1-q)f_2(\underline{X}))$$

$$= \log(f_1(\underline{X})/f_2(\underline{X})) + \log(q/1-q) \qquad 4.42$$

Now in order to find a mathematical expression for the conditional probabilities $Pi(\underline{X})$, some assumptions about the characteristics of populations should be made. The most 'convenient' and well known hypothesis is that the explanatory variables of each population are joint normal with mean vectors $\underline{\mu}_1$, $\underline{\mu}_2$ and a common covariance matrix Σ.

Therefore, according to the expression of a multivariate normal density function we have:

$$f_1(\underline{X}) = (1/SQRT(2\Pi))^{K}.\ \Sigma^{1/2}.\ \exp(-1/2A'_1\ \Sigma^{-1}\ A_1)$$

and

$$f_2(\underline{X}) = (1/SQRT(2\Pi))^{K}.\ \Sigma^{1/2}.\ \exp(-1/2A'_2\ \Sigma^{-1}\ A_2)$$

where

$A_1 = \underline{X}-\underline{\mu}_1$, $A_2 = \underline{X}-\underline{\mu}_2$ and k = number of factors.

Hence:

$$f_1(\underline{X})/f_2(\underline{X}) = \exp[-1/2\ (A_1'.\Sigma^{-1}.A_1 - A_2'.\Sigma^{-1}.A_2)] \qquad 4.44$$

and consequently:

$$\log(f_1(\underline{X})/f_2(\underline{X})) = -1/2[(A_1'.\Sigma^{-1}.A_1 - A_2'.\Sigma^{-1}.A_2) = Y(\underline{X}) \qquad 4.45$$

Note that $y(\underline{X})$ is linear in parameters and has the form of Eq. 4.37 since one can write:

$$Y(\underline{X})=-1/2[(\underline{X}-\underline{\mu}_1)'\Sigma^{-1}(\underline{X}-\underline{\mu}_1) - (\underline{X}-\underline{\mu}_2)'\Sigma^{-1}(\underline{X}-\underline{\mu}_2)]$$

$$=-1/2[\underline{X}'\Sigma^{-1}\underline{X} - \underline{X}'\Sigma^{-1}\underline{\mu}_1 - \underline{\mu}'_1\Sigma^{-1}\underline{X} + \underline{\mu}'_1\Sigma^{-1}\underline{\mu}_1$$
$$- \underline{X}'\Sigma^{-1}\underline{X} + \underline{X}'\Sigma^{-1}\underline{\mu}_2 + \underline{\mu}'_2\Sigma^{-1}\underline{X} - \underline{\mu}'_2\Sigma^{-1}\underline{\mu}_2$$

Using the property that Σ^{-1} is symmetric and thus $\underline{X}'\Sigma^{-1}\underline{\mu}_1=\underline{\mu}'_1\Sigma^{-1}\underline{X}$ we have:

$$Y(\underline{X})= \underline{X}'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2) - 1/2(\underline{\mu}_1'\Sigma^{-1}\underline{\mu}'_1 - \underline{\mu}_2\Sigma^{-1}\underline{\mu}_2)$$
$$= \underline{X}'\underline{\beta} - K \qquad 4.46$$

where $\underline{X}'$ is the row vector variables and $\underline{\beta}$ is the column vector of parameters. By substituting 4.45 into 4.46 we obtain:

$$\log(P1(\underline{X})/1-P1(\underline{X})) = Y(\underline{X})+\log(q/1-q) = G(\underline{X}) \qquad 4.47$$

which is equivalent to:

$$P1(\underline{X})/1-P1(\underline{X}) = \exp(G(\underline{X}))$$

and therefore

$$P1(\underline{X}) = \exp(G(\underline{X})) / [1+\exp(G(\underline{X}))] \qquad 4.48$$

Equation 4.48 has the familiar structural form of the modal split models already developed in the case of logit model (Eq. 4.15)

3.1 Classification of an Observation X

Recall from 4.46 and 4.47 that:

$$G(\underline{X}) = \underline{X}'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2) - K + \log(q/1-q) \qquad 4.49$$

The first term of 4.49 represents the linear classification function which will be denoted by $Z(\underline{X})$. Z is normally distributed with parameters:

$$E(Z)i = (\underline{\mu}_1-\underline{\mu}_2).\Sigma^{-1}.\underline{\mu}i \qquad i=1,2$$

and

$$Var(Z) = (\underline{\mu}_1-\underline{\mu}_2)'.\Sigma^{-1}.(\underline{\mu}_1-\underline{\mu}_2) = \sigma^2 \qquad 4.50$$

The variance σ^2 is known as the Mahalanobis distance D^2 of the two multivariate normal populations.

Now, assume that the vector observation $\underline{x}$ is taken on $\underline{X}$. According to Bayes, we will minimize the cost of misclassification by classifying $\underline{x}$ as coming from population 1 if:

$$E(C_1) > E(C_2)$$

That is

$$c_1.q\ f_1(\underline{X}) > c_2(1-q)f_2(\underline{X})$$

$$f_1(\underline{X})/f_2(\underline{X}) \quad > \quad c_2(1-q) \ / \ c_1 q \qquad 4.51$$

According to Eqs. 4.45, 4.46 and 4.49, 4.51 becomes:

$$Y(\underline{X}) \quad > \quad \log(c_2(1-q) \ / \ c_1.q)$$

or

$$Z(\underline{X}) \quad > \quad \log(c_2(1-q) \ / \ c_1.q) + K$$

and if we assume that c_1 and c_2 are equal, the classification rule will be:

The vector observation $\underline{x}$ is drawn from:

population 1 if $Z(\underline{X})>\log((1-q)/c_1 q)+K$

population 2 otherwise

3.2 Goodness of Fit

Tests concerning parameters β can be set up using the t-distribution. The general linear test which tests the overall significance of the function can be set up using the F-distribution.

An alternative overall test of significance is the Mahalanobis statistic D^2. This statistic, defined by Equation 4.50, represents the standardised distance between groups. It is assumed that the larger D^2 is, the better is the fit, since the region of misclassification is smaller.

V. ANALYSIS

The present chapter investigates the selection of relevant variables and variable forms which appear to affect the individual travel preferences. For this purpose, we first find that combination of system variables which best express the variation in individual utility function, and then attempt to obtain the best set of socio-economic variables which seem likely to succeed in explaining the behaviour of travellers.

Before entering into the analysis, it may be noteworthy to outline two problems usually encountered in these sort of studies:

a. The first one is the problem of multicollinearity. This occurs when the independant variables are highly or even perfectly correlated among themselves.

b. The second type of problem is the use of zonally aggregated data. Horowitz (1981) has shown that the use of zonally average variables in maximum likelihood estimation normally will produce inconsistent estimates of disaggregate choice probabilities unless the zonally averaged explanatory variables have the same joint distribution function in each zone and are not correlated with any disaggregate variables .

Talvitie(1976) suggests that zonal average variables are adequate when policy questions affect zones in a relatively homogeneous way, such as parking cost in the case of this study(see Ch.III.4).

1. DETERMINATION OF SYSTEM VARIABLES

1.1 Selection Criteria

Below, we will list criteria and set limits of significance which are used to reject hypothesis:

i. Sign Test

We expect that the partial model coefficients have the correct sign. Variables involving in-vehicle time difference(ΔT), walking time difference and waiting time should have a negative sign since an increase in riding time, walking time to bus stop and waiting time results in a reduction of the probability of transit use. On the other hand, the cost difference variable (ΔC) should have a positive sign.

ii. Likelihood Ratio Test

We should at least be able to reject at five percent level of significance the null hypothesis that the selected set of variables does not explain the variation in the utility function. As mentioned before, the use of the likelihood ratio test is a way to measure the existence of this relationship. This ratio is distributed as a chi-square with K degrees of freedom, where K represents the number of independant variables included in the model.

iii. T Test

This test is used to verify whether coefficient βk is significantly different from zero. As indicated in Chapter IV, we can make this inference about βk only in the case where we hold a large data set, since for small sample βk is not assumed to be normally distributed(see IV.2.1.iii)

iv. Correlation between Variables

Since one of the major sources of error is due to the correlation that might exists between variables, we must check the correlation coefficient matrix for each model. In the presence of the correlation, we should take the appropriate remedial measure.

1.2 Model Development

The following models, involving different forms of the four system variables: ΔT, $\Delta WALK$, WAIT and ΔC, are considered and treated with Logit, Probit and Discriminant Analysis:

M1 : $L = \beta_0+\beta_1\Delta T+\beta_2\Delta WALK+\beta_3 IWAIT+\beta_4(\Delta C/I)$

M2 : $L = \beta_0+\beta_1 I.\Delta T+\beta_2 I.\Delta WALK+\beta_3 I.WAIT+\beta_4\Delta C$

M3 : $L = \beta_0+\beta_1(\Delta T/T)+\beta_2(\Delta WALK/T)+\beta_3(WAIT/T)+\beta_4(\Delta C/I.C)$

M4 : $L = \beta_0+\beta_1(I\Delta T/T)+\beta_2(I\Delta WALK/T)+\beta_3 39I.WAIT/T)$
$+\beta_4(\Delta C/C)$

and

$$M1.1 : L = \beta_0+\beta_1\Delta T+\beta_2\Delta EXC+\beta_3(\Delta C/I)$$

$$M2.1 : L = \beta_0+\beta_1(I.\Delta T)+\beta_2(I.\Delta EXC)+\beta_3(\Delta C)$$

$$M3.1 : L = \beta_0+\beta_1(\Delta T/TT)+\beta_2(\Delta EXC/TT)+\beta_3(\Delta C/IC)$$

$$M4.1 : L = \beta_0+\beta_1(I.\Delta T/TT)+\beta_2(I.\Delta EXC/TT)+\beta_3(\Delta C/C)$$

where $\Delta EXC = \Delta WALK + WAIT$

1.2.1 Logit Treatment

Results of this treatment are shown in Table 15. By applying the sign test, we eliminate those models which have a positive ΔT, $\Delta WALK$, WAIT coefficient; or a negative ΔC coefficient.

Models 2 and 2.1 satisfy this criterion:

$$M2 : L = 1.93 - .0015(I\Delta T) - .00175(I\Delta WALK) - .00199(I.WAIT) + .2245\Delta C$$

and

$$M2.1 : L = 1.94 - .0015(I\Delta T) - .00196(I.\Delta EXC) + .2289.\Delta C$$

where I is expressed in 1000$

They lead almost to the same likelihood ratio, root mean square error(RMSE) and sum of absolute error(SAE). However, in the case of M2, the chi-square distribution of the likelihood ratio has 4 degree of freedom whereas, that of M2.1 has 3.

1.3 Probit Treatment

Table 16 displays the estimated coefficients for models M1,...,M4.1. Following the same reasoning made in the case

Models	Constant	In-Vehicle Time	Walking Time	Waiting Time	Out-of-Pocket Expenses	Excess Time	Likelihood Ratio	RMSE[1]	SAE[2]
M1	1.697	-0.0173	-0.0586	-0.0093	-0.1944	----	6.153	.3392	142.65
M1.1	1.573	-0.018	----	----	-0.1845	-0.0158	5.176	.3394	142.85
M2	1.931	-0.0015	-0.00175	-0.0020	0.2245	----	8.136	.3380	141.90
M2.1	1.942	-0.0015	----	----	0.2289	-0.0020	8.130	.3381	141.92
M3	8.661	7.847	1.1955	0.1517	0.1823	----	106.1	.3012	110.78
M3.1	8.731	7.908	----	----	0.1447	0.4203	105.1	.3003	110.54
M4	2.107	0.0617	0.7042	-0.0206	-0.3252	----	13.69	.3362	140.72
M4.1	2.569	0.0628	----	----	-0.0454	-0.0017	11.63	.3372	140.97

1- Root Mean Square of Error

2- Sum of Absolute Error

Table 15 - Parameter Estimation of Logit Models

Models	Constant	In-Vehicle Time	Walking Time	Waiting Time	Out-of-Pocket Expenses	Excess Time	Likelihood Ratio	RMSE[1]	SAE[2]
M1	0.972	-0.01054	-0.0296	-0.0044	-0.0829	----	6.63	.3393	142.59
M1.1	0.938	-0.01012	-----	----	-0.0797	-0.0089	5.54	.3394	142.75
M2	1.125	-0.00084	-0.0011	-0.0010	0.1126	----	8.28	.3382	141.98
M2.1	1.122	-0.00084	-----	----	0.1109	-0.0011	8.28	.3382	141.97
M3	4.415	3.780	0.4847	0.0542	0.0758	----	103.28	.3027	111.47
M3.1	4.462	3.825	-----	----	0.0581	0.1613	102.58	.3020	111.37
M4	1.231	0.03690	0.0431	-0.0103	-0.1943	----	14.22	.3363	139.91
M4.1	1.495	0.03636	-----	----	-0.0317	-0.00094	11.45	.3371	140.86

1- Root Mean Square of Error

2- Sum of Absolute Error

Table 16 - Parameter Estimation of Probit Models

of the logit models, models M2 and M2.1 are selected.

$$\text{M2} \; : \; L = 1.13 - .00084(I.\Delta T) - .00113(I.\Delta WALK) - .00104(I.WAIT) + .1126\Delta C$$

and

$$\text{M2.1} \; : \; L = 1.22 - .00084(I.\Delta T) - .00105(I.\Delta EXC) + .1109\Delta C$$

Their likelihood ratio, obtained level of significance, RMSE and SAE are given in Table 16.

Tables 17, 18, 19 and 20 show the result of testing the null hypothesis that βk is equal to zero. This hypothesis can be rejected at 5% level of significance for all parameters except β_2 and β_4 for M2 and β_3 for M2.1 .

The correlation matrix is shown in Table 21. It indicates that variables included in models M2 and M2.1 have an acceptable level of correlation. The maximum correlation(0.32) is found to be between (I.ΔT) and (ΔC).

Parameters	T-Ratio	Signif.
β_0	3.008	0.001
β_1	-2.3957	0.008
β_2	-0.5904	0.227
β_3	-1.9771	0.024
β_4	0.4148	0.339

(d.f.=616)

Table 17 - Significance of Logit Model M2 Parameters

Parameters	T-Ratio	Signif.
β_0	3.0893	0.001
β_1	-2.4351	0.007
β_2	-2.1136	0.017
β_3	0.4254	0.335

(d.f.=617)

Table 18 - Significance of Logit Model M2.1 Parameters

Parameters	T-Ratio	Signif.
β_0	3.2751	0.000
β_1	-2.4757	0.006
β_2	-0.7484	0.227
β_3	-1.8147	0.035
β_4	0.3869	0.349

(d.f.=616)

Table 19 - Significance of Probit Model M2 Parameters

Parameters	T-Ratio	Signif.
β_0	3.3194	0.000
β_1	-2.5128	0.006
β_2	-2.0079	0.002
β_3	0.3834	0.350

(d.f.=617)

Table 20 - Significance of Probit Model M2.1 Parameters

Models	Variables	1	2	3	4
1	1= ΔT	1.00			
	2=ΔWALK	.08	1.00		
	3= WAIT	.04	-.08	1.00	
	4= ΔC/I	.01	-.01	.02	1.00
2	1=IΔT	1.00			
	2=IΔWALK	-.27	1.00		
	3=IΔWAIT	-.31	.19	1.00	
	4=ΔC	.32	.03	.001	1.00
3	1= ΔT/T	1.00			
	2=ΔWALK/T	-.22	1.00		
	3= WAIT/T	-.28	.54	1.00	
	4= ΔC/IC	-.02	-.02	-.03	1.00
4	1=IΔT/T	1.00			
	2=IΔWALK/T	-.37	1.00		
	3=IΔWAIT/T	-.49	.48	1.00	
	4=ΔC/C	-.09	.21	.14	1.00

Table 21 - Correlation Matrices

1.4 Discriminant Analysis

Recall from Chapter IV, that the expression of modal choice probability obtained by means of discriminant analysis is similar to that resulted from the logit treatment(see Eqs. 4.15 and 4.48). Therefore, time variables should have a negative sign, whereas the cost variable should contribute negatively to the classification function(Eq. 4.49). Table 22 displays the results of this analysis[1]. Only model 4 leads to correct signs. It also leads to the second largest D^2 (Mahalanobis distance). The overall F-statistic found for this model is 4.27 with an attained significance of 0.002.

Therefore, according to 2.49, function G will be:

$$G = -.45 - .07(I\Delta T/T) - .10(I\Delta WALK/T) - .004(I.WAIT/T) + .31(\Delta C/C) - K + \log(83/538)$$

where 83 and 538 are the transit and auto sample size respectively. K is calculated internally by the computer programme according to Eq. 4.46.

---

[1] Since in the case of logit and probit analysis, no significant changes has been detected between the magnitude of coefficients of models 1, 2, 3 and 4 and that of models 1.1, 1.2,1.3 and 1.4 coefficients, these latters were not pursued further

Models	Variables	F-Stat	Signif	Discrim.Function		Z-Func.
				Auto	Transit	
1	Constant ΔT ΔWALK WAIT ΔC/I	 3.5916 1.4566 .2143 .0279	 .0585 .2279 .6436 .8674	-4.933 -.1391 .6644 .2963 -7.786	-4.802 -.1237 .7179 .3060 -7.608	.1310 .0154 .0553 .0097 .1786
D^2 = 0.0807 F-STAT. = 1.4436 SIGNIF. = .2180						
2	Constant IΔT IΔWALK I.WAIT ΔC	 6.0254 0.2646 3.7322 .2037	 .0144 .6072 .0538 .6519	-13.499 -.0036 .0304 .0104 -22.064	-13.670 -.0049 .0318 .0127 -22.305	-.1710 .0013 .0014 .0023 -.2410
D^2 = 0.1099 F-STAT. = 1.9653 SIGNIF. = .0982						
3	Constant ΔT/T ΔWALK/T WAIT/T ΔC/I.C	 175.72 2.044 1.140 .0190	 .0000 .1533 .2883 .8909	-19.907 -47.45 .3830 -2.450 -10.468	-30.882 -59.605 -1.617 -2.864 -10.617	-10.975 -12.152 -1.234 -.415 -.149
D^2 = 2.5012 F-STAT. = 44.745 SIGNIF. = .0000						
4	Constant I.ΔT/T IΔWALK/T IWAIT/T ΔC/C	 8.903 3.465 2.252 .0527	 .0030 .0631 .1340 .8185	-129.25 -.3663 1.2737 .0617 -181.39	-129.70 -.4410 1.172 -.0574 -181.08	-.45 -.07 -.10 -.004 .31
D^2 = 0.2388 F-STAT. = 4.2721 SIGNIF. = .0020						

Table 22 - Parameter Estimation of Discriminant Models

2. DETERMINATION OF SOCIO-ECONOMIC VARIABLES

The set of socio-economic variables which appear to have a potential role in explaining individual behaviour towards travelling is listed Chapter II. They are as follows: Sex, Age, Occupation and Car Ownership.
Note that the effect of income has already been considered by the combination with system variables.

2.1 Selection Criterions

i. Partial Coefficient Sign

According to the sex variable formulation, Female=0 and Male=1, and, since the probability that a male take the car is higher, we should expect a positive sign for Sex coefficient.

Age variable must have a positive sign since it takes on the value 1 when trip maker is younger than 25 or older than 60 years old, and we know that people belonging to these age brackets are mostly transit users.

As for Car Ownership variable, a positive sign is expected. CO is equal to 1 when the household possesses more than one car and therefore, the probability of taking car to work for people belonging to this category of household is higher.

Occupation variables needs more reflexion since two indicator variables OCC1 and OCC2 explain simultaneously its effect

– Primary category is formulated by OCC1=0 and OCC2=0. Therefore, the corresponding utility function is:

$$Lp = \beta_0+\beta_1(I\Delta T)+\beta_2(I\Delta WALK)+\beta_3(I.WAIT)+\beta_4.\Delta C$$
$$+\beta_5.SEX+\beta_6 AGE+\beta_7.0+\beta_8.0+\beta_9.CO$$

– Individuals belonging to Managerial and Professional category have an utility function as:

$$Lm = \beta_0+\beta_1(I\Delta T)+\beta_2(I\Delta WALK)+\beta_3(I.WAIT)+\beta_4.\Delta C$$
$$+\beta_5.SEX+\beta_6 AGE+\beta_7.0+\beta_8.1+\beta_9.CO$$

– Clerical and Salemen's utility are expressed by:

$$Lc = \beta_0+\beta_1(I\Delta T)+\beta_2(I\Delta WALK)+\beta_3(I.WAIT)+\beta_4.\Delta C$$
$$+\beta_5.SEX+\beta_6 AGE+\beta_7.1+\beta_8.0+\beta_9.CO$$

– And finally, we can formulate the utility of Labours by:

$$Ll = \beta_0+\beta_1(I\Delta T)+\beta_2(I\Delta WALK)+\beta_3(I.WAIT)+\beta_4.\Delta C$$
$$+\beta_5.SEX+\beta_6 AGE+\beta_7.1+\beta_8.1+\beta_9.CO$$

In order to find the appropriate sign for β_7, the marginal effect of OCC1, we should compare Lm and Ll. The difference between these two utility functions are given as:

$$L1 - Lm = \beta_7$$

and since we expect that individuals belonging to managerial and professional category are mostly private car users, in other words they attach more disutility to the travel activity($Lm>L1$), therefore β_7 should be negative.

Similarily, when we compare Lm and Lp, we find that the correct sign for β_8 is a positive sign since:

$$Lm - Lp = \beta_8 \quad \text{and} \quad Lm > Lp$$

ii. Pseudo R-square Criterion

As mentioned in Chapter IV, ρ^2 index is a measure of goodness of fit. One should be forewarned that a good fit is expressed by value of .2 to .4, and this criterion is used to select the 'best' set of variables. Since as one adds variables to the model ρ^2 index increases regardless of the explanatory power of the variable[1], the intention will then be not to maximize ρ^2 index but to find whether the inclusion of the new variable worth the increase in ρ^2 index.

iii. Mean-Square of Error Criterion

This is also used to measure the goodness of fit of the model. It is expressed as: MSE = SSE / n-k where k is the number of parameters in the model. The

---

[1] Since by increasing the number of variables, the model approaches the saturated model for which the fit is perfect.

advantage of MSE criterion is that it takes account of the number of parameters present in the model. We should seek to minimize MSE since Min(MSE) can increase as k increases if the reduction in SSE becomes so small that a loss of an additional degree of freedom can not compensate it.

2.1.1 Logit and Probit Models

Figures 7 and 8, and Tables 23 and 24 show the effect of each new users' variable on increasing or reducing the ρ^2 and MSE. The ρ^2 index and MSE value for each entering variable are plotted in these figures. Points are connected by straight lines to express the effect of adding additional independant variables. Both figures clearly display the important effect of SEX and CO, shown by the slope of connecting lines, in improving the model capability. On the other hand, small gains are achieved when age and occupation variables are added to the model as illustrated in Fig. 7. Note also that the T-ratio of OCC1 and OCC2 are low(-.73 and 1.68), whereas that of age variable varies between -3.66 and -3.37. Therefore, only the occupation variables OCC1 and OCC2 were excluded. The final modal choice resulted by means of logit and probit approach are respectively:

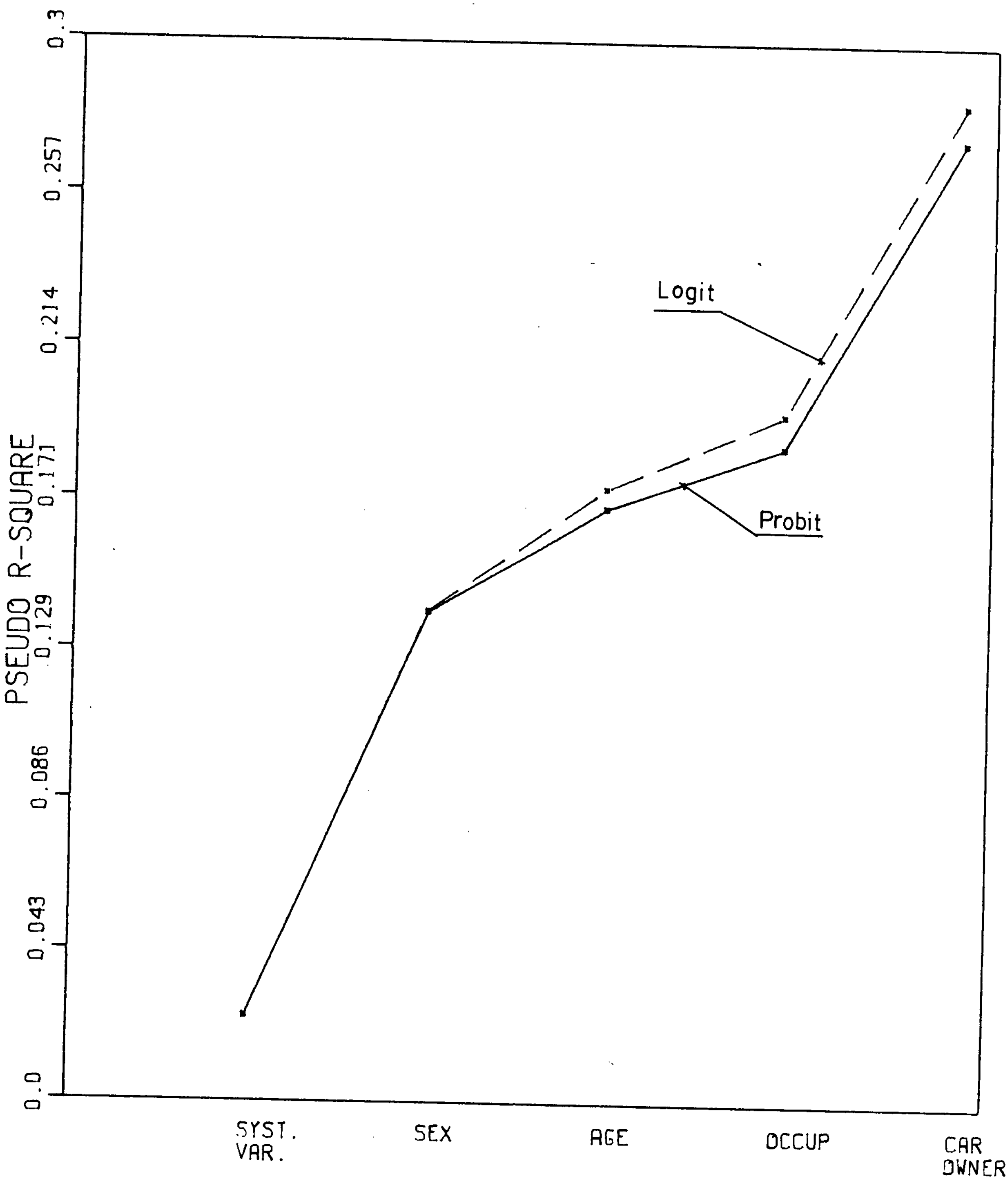

Figure 8 - Pseudo R-Square Variation

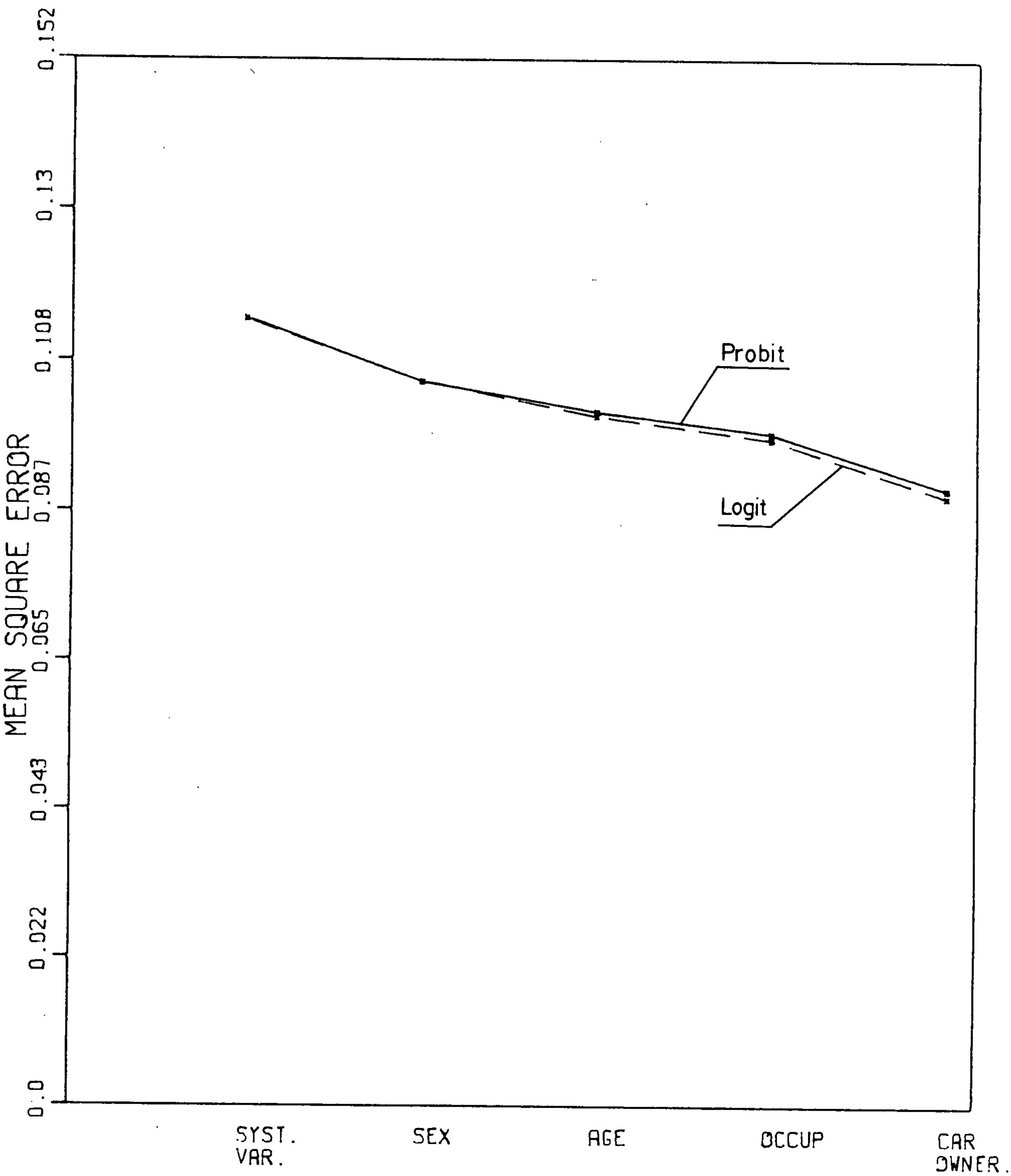

Figure 9 - Mean Square of Error Variation

Entering Variable	Pseudo R^2	MSE	Estimates	T-Ratio	Degree of Freedom	Significance
System Variables	.0239	.1143				
Sex	.1400	.1053	1.591	6.83	615	.0000
Age	.1748	.1003	-.956	-3.66	614	.0001
OCC1			-.259	-0.73		.2330
	.1959	.1974			612	
OCC2			.575	1.66		.0480
CO	.2831	.0889	1.054	5.52	611	.0000

Table 23 - Socio-Economic Variables for Logit Model M2

Entering Variable	Pseudo R^2	MSE	Estimates	T-Ratio	Degree of Freedom	Significance
System Variables	.0243	.1144				
Sex	.1396	.1054	.869	6.37	615	.0000
Age	.1695	.1009	-.494	-3.37	614	.0001
OCC1	.1868	.0980	-.096	-.54	612	.2940
OCC2			.319	1.68		.0460
CO	.2730	.0899	.841	5.60	611	.0000

Table 24 - Socio-Economic Variables for Probit Model M2

$$L = 1.10 - .0012(I\Delta T) - .0053(I\Delta WALK) - .0026(I.WAIT) + .490\Delta C + 1.70SEX - 1.03AGE + 1.56CO$$

Likelihood Ratio	:	95.74
Attained significance level	:	.0000
Pseudo R-Square	:	.2623
Mean Square of Error	:	.0916

$$L = 0.63 - .0006(I\Delta T) - .0025(I\Delta WALK) - .0013(I.WAIT) + .256\Delta C + 0.90SEX - 0.54AGE + 0.82CO$$

Likelihood Ratio	:	92.52
Attained significance level	:	.0000
Pseudo R-Square	:	.2542
Mean Square of Error	:	.0924

Tables 25 and 26 illustrate the standard deviations and the t-ratio statistics obtained for the parameters of the above models.

2.1.2 Discriminant Analysis

In the present stepwise routine, we first examine the sign of the new variables according to the sign test criterion, then choose those variables which lead to the largest increase in D^2 and F statistics. In addition, we should examine that the significance level attained by these F-values falls below our predetermined limit of significance(5%).

Param.	Estimates	Std.Dev.	T-Ratio
β_0	1.10	.72	1.52
β_1	-.0012	.0007	-1.75
β_2	-.0053	.0032	-1.67
β_3	-.0026	.0012	-2.26
β_4	.490	.6007	.82
β_5	1.70	.2681	6.35
β_6	-1.30	.2745	-3.74
β_7	1.56	.2835	5.15

Table 25 - Point Estimation of Logit Model Parameters

Param.	Estimates	Std.Dev.	T-Ratio
β_0	.63	.3901	1.61
β_1	-.0006	.0004	-1.71
β_2	-.0025	.0017	-1.52
β_3	-.0013	.0006	-2.01
β_4	.256	.3208	.80
β_5	.90	.1460	6.20
β_6	-.54	.1528	-3.56
β_7	.82	.1483	5.53

Table 26 - Point Estimation of Probit Model Parameters

Table 27 shows the results of the analysis. It is noticed that SEX and CO variables contribute to a significant increase in F and D^2 statistics, whereas OCC1 and OCC2, expressing the effect of occupation, reduce the F-value and slightly augment the D^2.

Table 28 displays the estimated coefficients when System, Sex, Age and Car Ownership variables are present in the model. According to these results, the following discriminant modal choice model is formulated:

$$Z = -6.91 - .124(I\Delta T/T) - .066(I\Delta WALK/T) - .0065(I.WAIT/T) - 1.25(\Delta C/C) + 2.08SEX - 1.30AGE + 1.94CO$$

Mahalanobis Distance : 1.955

F - Statistic : 19.89

Attained Significance Level : .0000

Note that by including socio-economic variables, the sign of the cost variable becomes negative and hence violates sign criteria. Therefore, we decided to take the reverse 'cheminement' which is: having the present set of socio-economic variables (SEX, AGE and CO), look for the 'best' model(M1, M2 or M3) which respects the sign criterion and leads to an acceptable F-value. Table 29 compares these three models and indicates that only M3 satisfies the sign criterion.

Entering Variable	Sign	F-Stat.	Signif.	D-Square	ΔD²	Overall F-Stat.	ΔF	Signif.
System Variables				.2388		4.2721		.0020
Sex	+	47.653	.0000	.9230	.6842	13.189	8.9169	.0000
Age	-	15.893	.0001	1.1697	.2467	13.905	0.7160	.0000
OCC1	-	1.789	.1818					
				1.3483	.1768	11.982	0.1923	.0000
OCC2	+	3.343	.0682					
CO	+	49.837	.0000	2.1604	.8121	17.038	5.0561	.0000

Table 27 - Discriminant Analysis:Selection of Socio-Economic Variables

Variables	Discrim. Function		Z-Function[1]
	Auto	Transit	
Constant	-154.24	-161.75	-6.910
IΔT/T	-0.6087	-0.7332	-0.124
IΔWALK/T	1.4999	1.4340	-0.066
IWAIT/T	0.1054	0.0402	-0.065
ΔC/C	-185.24	-186.49	-1.250
SEX	10.679	-12.759	2.08
AGE	8.110	6.810	-1.30
CO	10.368	12.312	1.944

1- See Eq. 4.49

Table 28 - Coefficient Estimation of Discriminant and Z Functions

Variables	Z-Function		
	Model 1	Model 2	Model 3
Constatnt	-3.167	-4.277	-13.042
InVehTime	0.011	0.001	-11.36
WalkTime	0.089	0.005	-1.272
WaitTime	0.014	0.003	-0.354
O.P.E.[1]	-2.187	-0.586	1.584
SEX	2.047	2.083	1.908
AGE	-1.278	-1.236	-1.069
CO	1.682	1.647	1.430

1- Out-of-Pocket Expenses

Table 29 - Discriminant Models 1, 2 and 3

Therefore, the final model is formulated as below:

$$Z = -13.04 - 11.36(\Delta T/T) - 1.27(\Delta WALK/T) - .35(WAIT/T)$$
$$+ 1.58(\Delta C/I.C) + 1.91SEX - 1.07AGE + 1.43CO$$

Mahalanobis Distance : 3.69

F - Statistic : 37.58

Attained Significance Level : .0000

3. COMPARISON OF THE METHOD OF ANALYSIS

Although the strategy employed to develop the modal choice models was the same for the three approaches, the form of System variables obtained by means of Discriminant analysis differs from that obtained by Logit and Probit methods. However, the use of all three methods suggests that Occupation variable does not significantly affect the individual utility function.

It is very difficult to base the comparison and the assessment of the three approaches, on the one hand, to the quality of estimates since they all produce significant overall statistic, and on the other hand, to the magnitude of estimates since their definitions are not the same(see ChapterIV). Therefore, we decided to evaluate the efficiency of each method by its prediction capability. For this purpose, we apply the models obtained from these three methods to the present data and measure their ability to reproduce the actual situation.

Observation-prediction[1] Tables 30, 31 and 32 show the results of this comparison . Note that entries tij represent the mode i observation followed by mode j prediction. Therefore, all off-diagonal cases represent a prediction error. The overall frequency of correct prediction is then calculated by $\Sigma(tij) / t_{..}$. In some cases this score might be misleading. For instance, the overall frequency score obtained by means of Discriminant analysis is 78% which represents the highest among the two others (73% for Logit and 69% for Probit). But the percent of correct prediction for transit mode obtained through Discriminant approach is only 2% whereas, Logit and Probit analysis lead to better scores(16.9% and 15%)). In order to take account of this fact, we shall use the success index which is the normalized prediction success proportion:

$$(tij/t_{.1}) \ / \ (t_{1.}/t_{..}) \qquad \text{for } i=j$$

According to the above three tables, one may choose the Logit approach since it leads to a greater success index. Note also the closeness of Probit success index to that of Logit model.

---

[1] See for example Theil(1966) and McFadden(1976)

P R E D I C T I O N

O B S E R V A T I O N

	Auto	Transit	Total	Observed %
Auto	435	103	538	87
Transit	62	21	83	13
Total	497	124	621	
Predicted %	80	20	100	
Correct %	87.5	16.9	73	
Success Index	1.01	1.30		

Table 30 - Logit Observation-Prediction Table

P R E D I C T I O N

O B S E R V A T I O N

	Auto	Transit	Total	Observed %
Auto	407	131	538	87
Transit	59	24	83	13
Total	466	155	621	
Predicted %	75	25	100	
Correct %	87.3	15.5	69	
Success Index	1.003	1.19		

Table 31 - Probit Observation-Prediction Table

PREDICTION

OBSERVATION

	Auto	Transit	Total	Observed %
Auto	486	52	538	87
Transit	82	1	83	13
Total	568	53	621	
Predicted %	91.5	8.5	100	
Correct %	85.5	1.9	78	
Success Index	.98	.15		

Table 32 - Discriminant Observation-Prediction Table

4. SUMMARY

In this chapter, we applied three classical approaches: Logit, Probit and Discriminant analysis to our data set, in order to determine the set of service attributes and travellers' socio-economic characteristics which 'best' affect the transportation consumers' behaviour. This analysis attempted to obtain those uncorrelated variables with the appropriate form which lead to a significant overall statistic and to significant coefficient estimates with the correct sign. We found that, while the models resulted from Logit and Probit treatment include the same set of variables with similar forms, the Discriminant approach leads to different variable forms. Although not very different from the forecasting ability of Probit model, the Logit model has the greatest prediction ability.

VI. SENSITIVITY ANALYSIS

This chapter investigates assessing the impact of a change in service attributes and users' socio-economic characteristics on the mode choice probability. The logit model since, as it was shown, the discriminant analysis performs badly on prediction, and as it will be seen, the variation of probit model variables follows the same trend as those of logit model. First, the interpretation of coefficients is given, then a sensitivity analysis is carried out and finally, the value of time is derived.

1. COEFFICIENT INTERPRETATION

It should be recalled that in the logit and probit models, the dependant variable is not the index L, the linear combination of independant variables, but the probability of choosing a mode given a set of independant variables which is found by means of logit or probit transformation of index L. Due to this fact, the interpretation of coefficients needs to be explained.

1.1 Probit

Recall from Eq. 4.20:

$$P(X) = 1/SQRT(2\Pi) \int_{-\infty}^{L(X)} \exp(-t/2)\, dt$$

where $L(X) = \underline{\beta}'.\underline{X}$

A one unit change in Xk will lead to a change of β in the index L which changes the area under the standard normal curve. In other words, the marginal effect of Xk is equivalent to the β standard deviation units. Since this ordinate is larger near the center of the distribution, therefore, the largest variation in the probability is obtained in this area. Moreover, one can easily derive the definition of the constant term β_0; it is the probability corresponding to β_0 standard deviation of N(0,1) of taking the bus when all independant variables are zero. In the present case, it represents the probability that a female traveller belonging to age group 25-60 and to a household owing one car, takes bus when all the system characteristics(travel time, travel cost,...) of alternatives are similar; it is equal to 74%.

1.2 Logit

According to the logit transformation of the index L, the probability of selecting the transit mode is:

$$P(X) = \exp(L) / 1+\exp(L) = 1 / (1+\exp(-L))$$

The marginal effect of one of the variables can be seen by taking the partial derivative of P with respect to the underlying variable, for example Xk:

$$\delta P/\delta Xk = \delta[1/(1+\exp(-\underline{\beta}'.\underline{X}))]/\delta Xk$$
$$= \beta_k \exp(-\underline{\beta}'.\underline{X})\ 1/(1+\exp(-\underline{\beta}'.\underline{X}))^2$$

and by substituting the expression of P(X) we have:

$$\delta P/\delta Xk = \beta_k.P.(1-P)$$

As one can notice, the marginal effect of variable Xk depends on where the probability P is evaluated. Similar to the probit transformation, the highest effect occurs at the mid-point of the distribution(P=50%).

The straightforwardness and the simplicity of the logistic function parameters render the use of this approach more convenient. The interpretation of β_0 follows the principle of the logit transformation; it corresponds to the logarithm of the ratio of bus choice and car choice probabilities when all Xk are zero. In the actual case, the transit use probability associated to β_0 for the type of traveller mentioned before is 75%.

2. SENSITIVITY ANALYSIS

Before evaluating the effect of each variable, it is interesting to note a point deduced from the fact that the highest marginal effect occured at the center of the distribution. In the transportation context it means that disaggregate modal choice models are primarily useful only in cities where the whole transportation system is well developed and public transit is highly competitive with the private mode in a manner that makes the probability of taking a bus or a car to work equal. Below, the effect of each variable on the transit use probability is studied.

2.1 Effect of Income

It is expected that an increase in income reduces the probability of transit use. This is due mostly to the fact that the disutility associated with travel activity varies directly with the level of income: higher income people attach a greater value to their privacy and level of comfort which makes the use of transit less probable.

Figure 10 compares the probability of transit use obtained by logit and probit model for three annual income levels: 10,000$, 20,000$ and 30,000$. It can be concluded that higher income persons are more sensitive to changes occured in travel time difference. For instance, a difference of ±15 minutes between car and bus travel times only marginally affects mode choice probability of low-income people(72%<Pt<75%) whereas, it has a large effect on the high-income people's mode choice probability(64%<Pt<84%).

Note also that probabilities obtained by probit model are slightly lower and favor the private mode relative to the transit mode up to a certain point varying with the income level. For instance, for a 30,000$ level of income this point

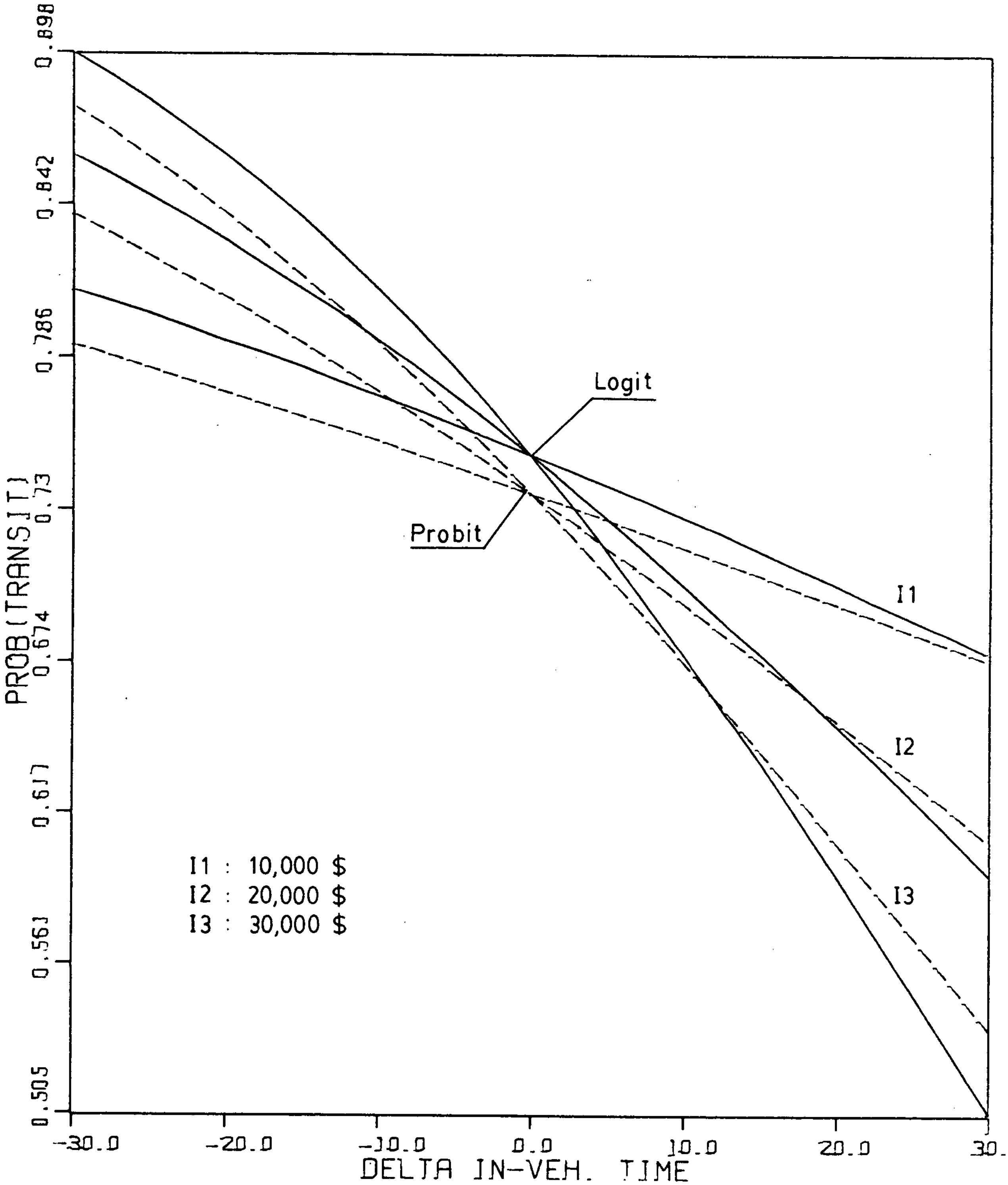

Figure 10 - Effect of Income on Transit Use Probability with Respect to ΔT

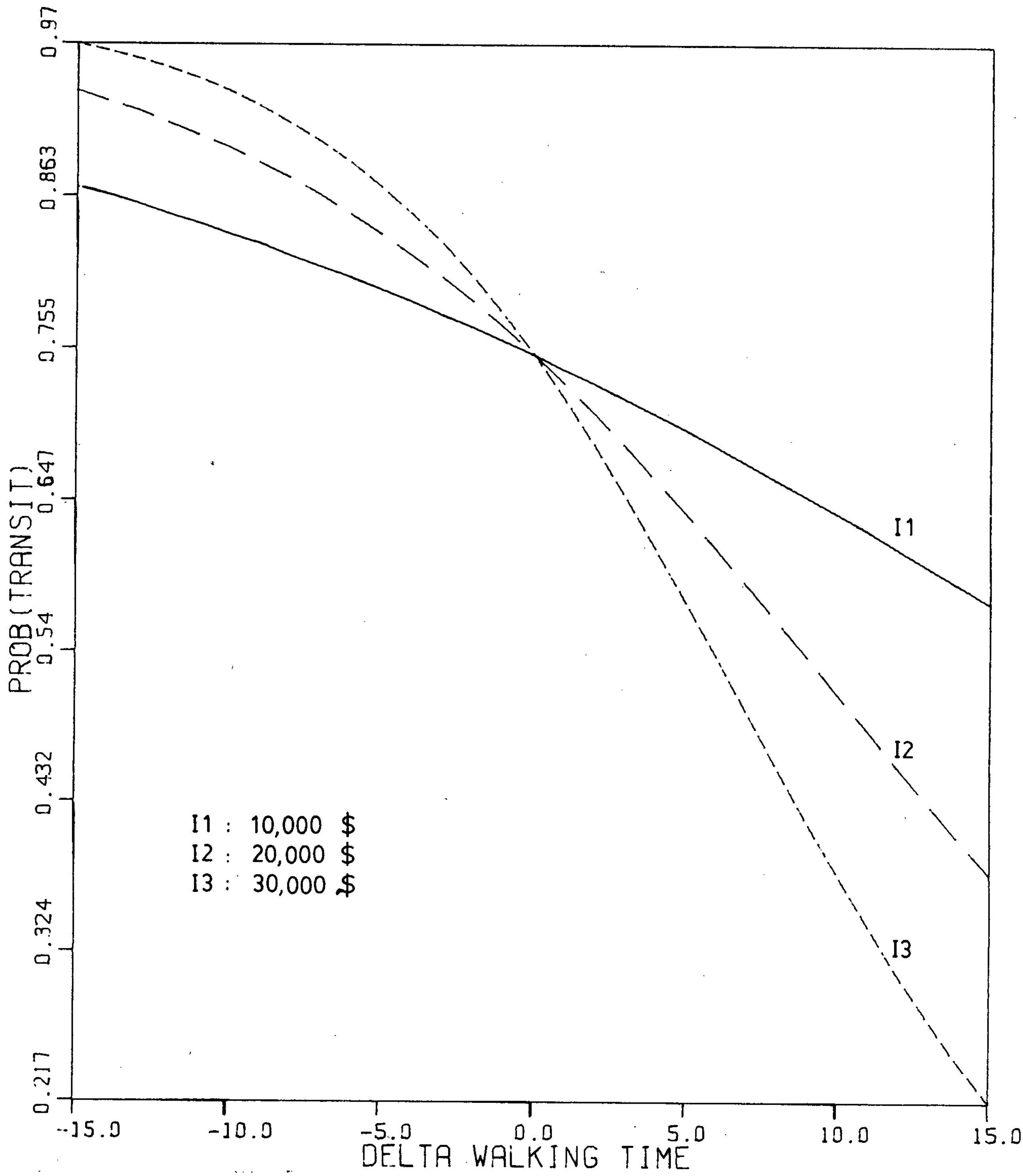

Figure 11 - Effect of Income on Transit Use Probability with Respect to Walking Time

is about $\Delta T=10$minutes.

From this comparison, it can be seen that the variations of the logit and probit models variables follows the same trend, and for this reason, only the logit model will be considered further. Figure 11 displays the effect of income on the variation of transit use probability in terms of the relative walking time. It clearly illustrates the fact that the disutility of walking is higher than that of in-vehicle time. A change of ± 1.5minutes in walking time to or from the bus stop modifies the underlying probability from 22% to 97% for high-income people and from 55% to 83% for low-income people. These results are used in the next chapter for transportation policy proposals.

2.2 Income-Sex Interaction

Figure 12 displays the interaction of income level and the sex of the tripmaker on the bus choice probability. It is interesting to notice that income seems to have a small influence on mode choice probability of women whereas, for men, it seems to more drastically affect their decisions. In addition, it might be concluded that women are less sensitive and tolerate more variation of in-vehicle time than men. Therefore, women may be less affected by transportation policies modifying the in-vehicle travel time. However, the effect of an increase in walking time to/from bus stop on the transit use probability of high income women is important. This is shown in Figure 13.

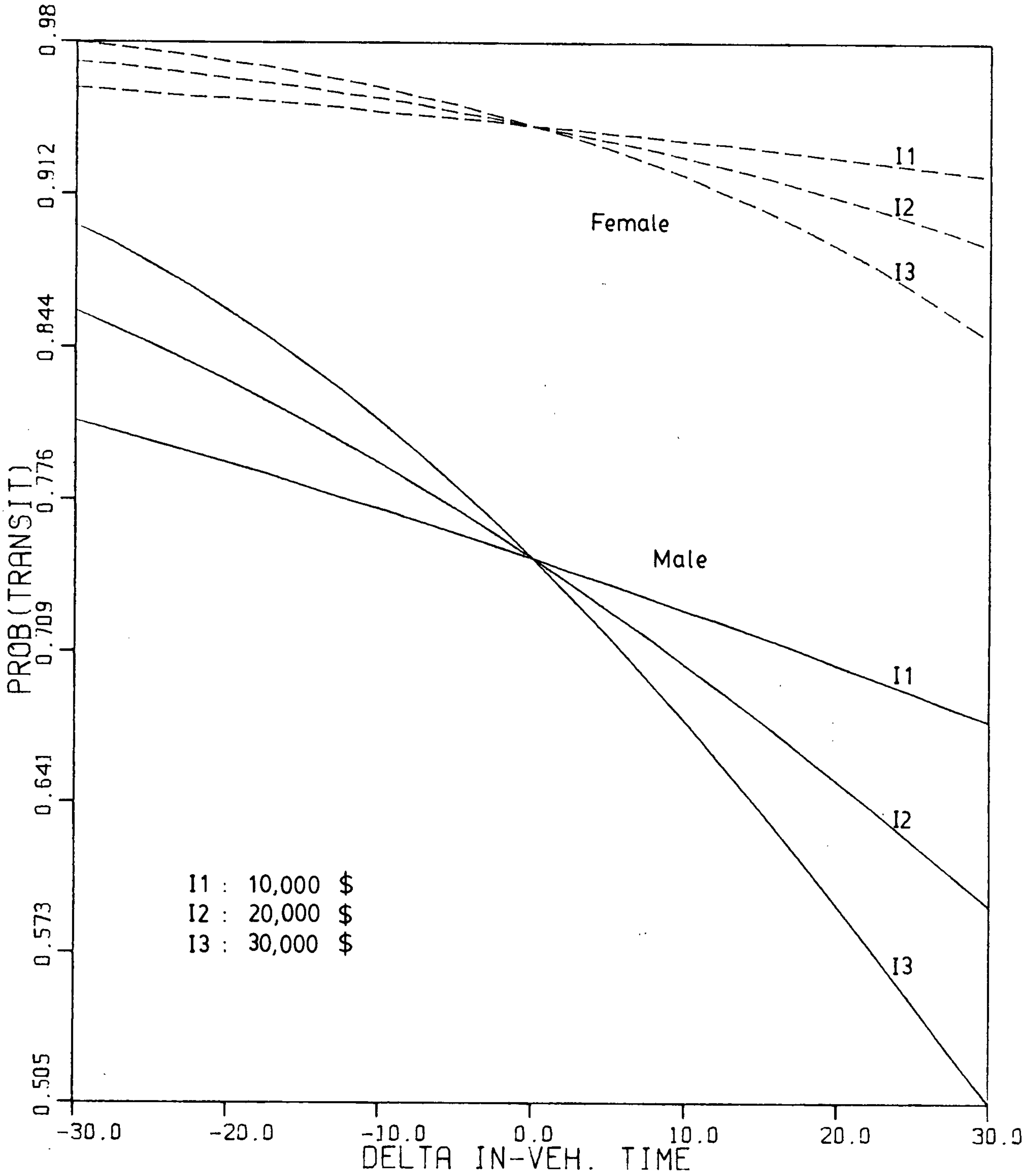

Figure 12 - Interaction Effect of Income and Sex on Transit Use Probability with Respect to ΔT

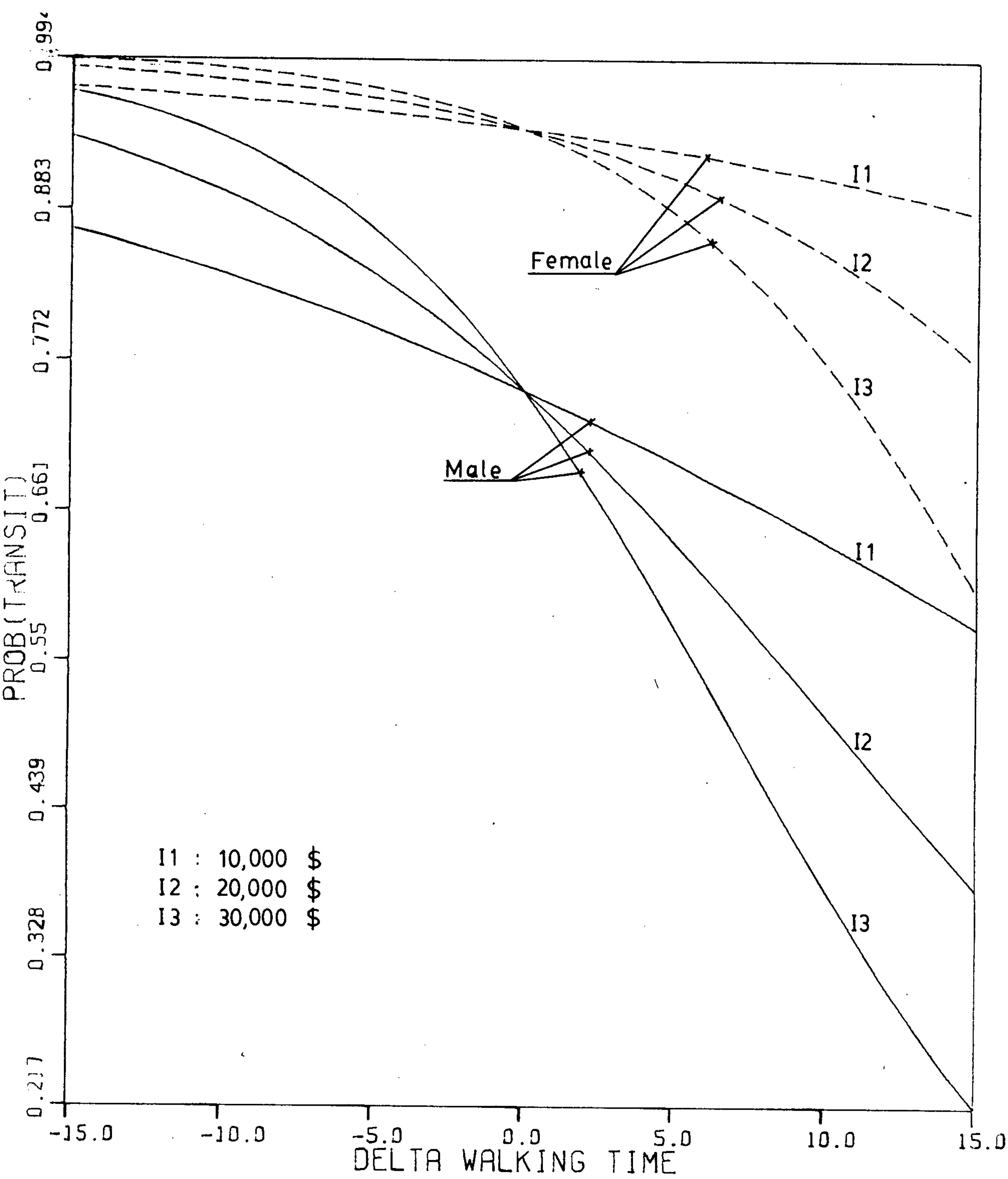

Figure 13 - Interaction Effect of Income and Sex on Transit Use Probability with Respect to Walking Time

2.3 Income-Age Interaction

The analysis of Figure 14 follows the same line as for the above case. People belonging to age group 25-60 are more sensitive to variation in travel time. It should also be noted that the level of income plays a similar role in mode choice decision making for both age groups. If the probability curves of two age groups are completely separated when the in-vehicle time is considered, they become closer when the probabilities are calculated in terms of walking time (see Figure 15). This implies that the disutility occured to people under 25 or above 60 years old towards walking is close to that of the people belonging to the other age group. Note also that for low-income people of 25-60 age group the probability of transit is practically zero when the walking time to/from bus stop increases by over 15 minutes.

2.4 Effect of Travel Cost

Recall that in our model the formulation of travel cost is:

ΔC = Fare - ((MOCOST+PKCOST)/20x2)

where MOCOST is the monthly operating cost

PKCOST is the monthly parking cost

and 20 working days per month is assumed

The effect of travel cost can be assessed by changing on one hand the transit fare and on the other hand the cost of operating a car. For the latter the intervention of policy makers can be either on MOCOST component, by increasing the price of fuel or creating a toll system, on PKCOST component

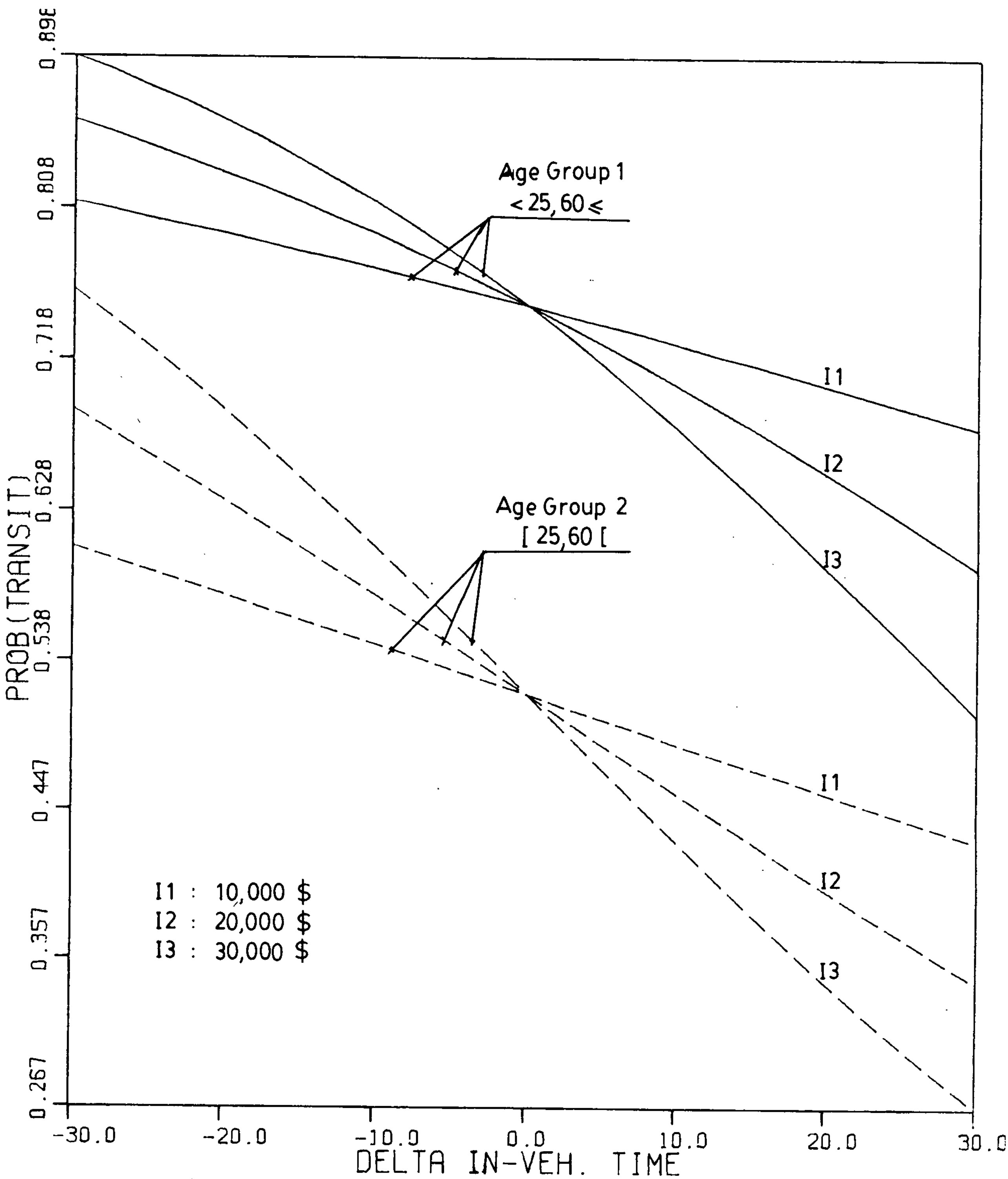

Figure 14 - Interaction Effect of Income and Age on Transit Use Probability with Respect to ΔT

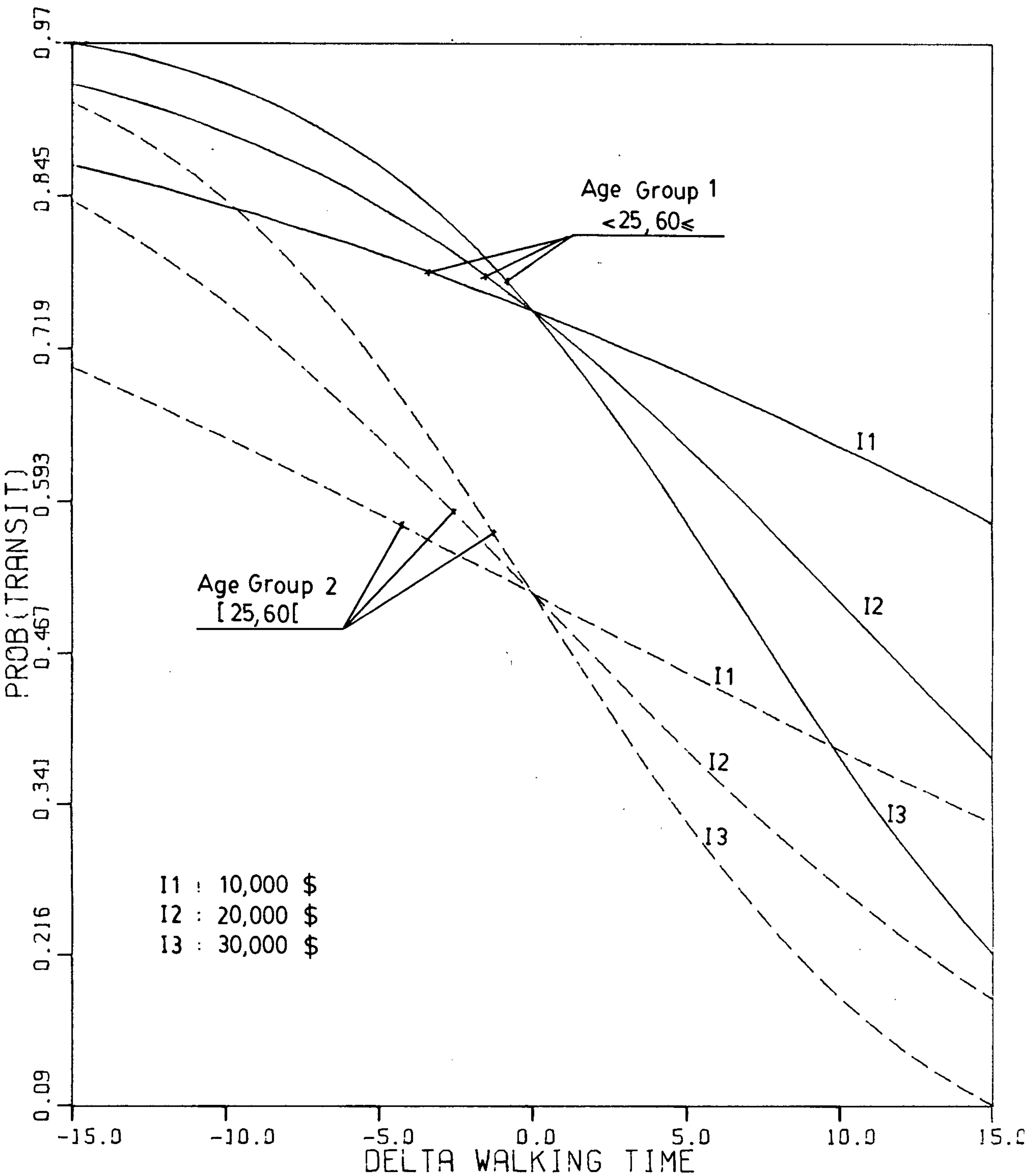

Figure 15 - Interaction Effect of Income and Age on Transit Use Probability with Respect to Walking Time

or on both. In this study, the effect of changes in fare and parking costs are only considered, since a variation in MOCOST will have a similar effect on the choice probability as that of PKCOST.

2.4.1 Fare

Figure 16 illustrates the expected results that the probability of transit selection decreases as the transit fare augments. Figure 17 depicts the interaction effect of fare increase and income. For high level income class, the downward shift of the curve in the result of a fare increase is the smallest, since they are less concerned about the cost of travel. This shift becomes more important as the income level decreases. The above curves correspond to the three income levels mentioned before and for fares of .50 and .75 dollars.

Figure 18 displays the effect of a fare change in terms of relative walking time. Since the range of the variation of the transit use probability is almost the same as when the in-vehicle time was taken into account, the consideration of relative walking time is not further pursued.

2.4.2 Parking Cost

Figure 19 presents the interaction effect of parking charge increase and income. Three levels of income 10,000 $, 20,000 $ and 30,000 $; and two monthly parking costs 10$ and 40$ were used. They produce similar curves as in the above case. In these figures the effect of income level on the curves shift

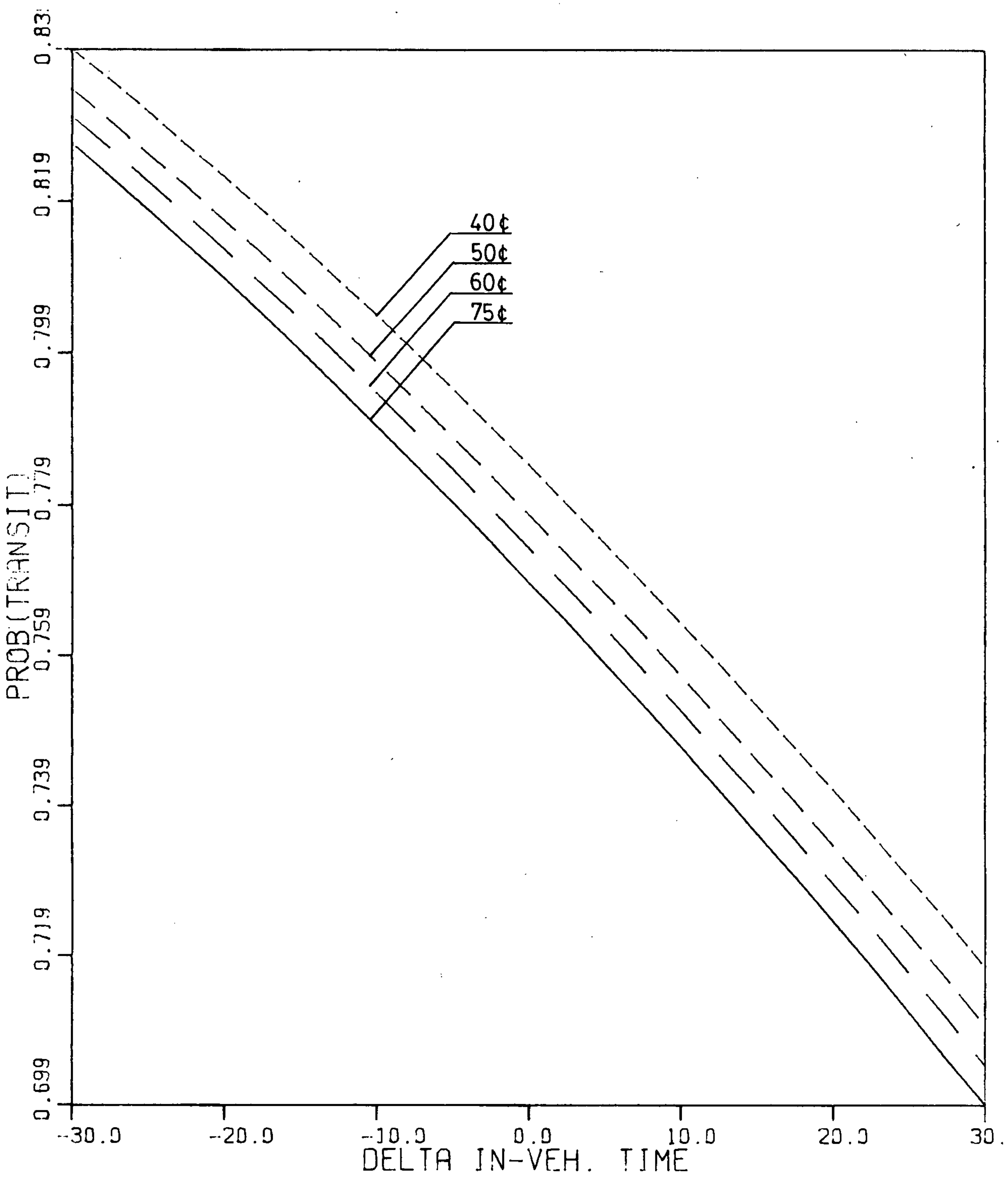

Figure 16 - Effect of Fare Variation on Transit Use Probability with Respect to ΔT

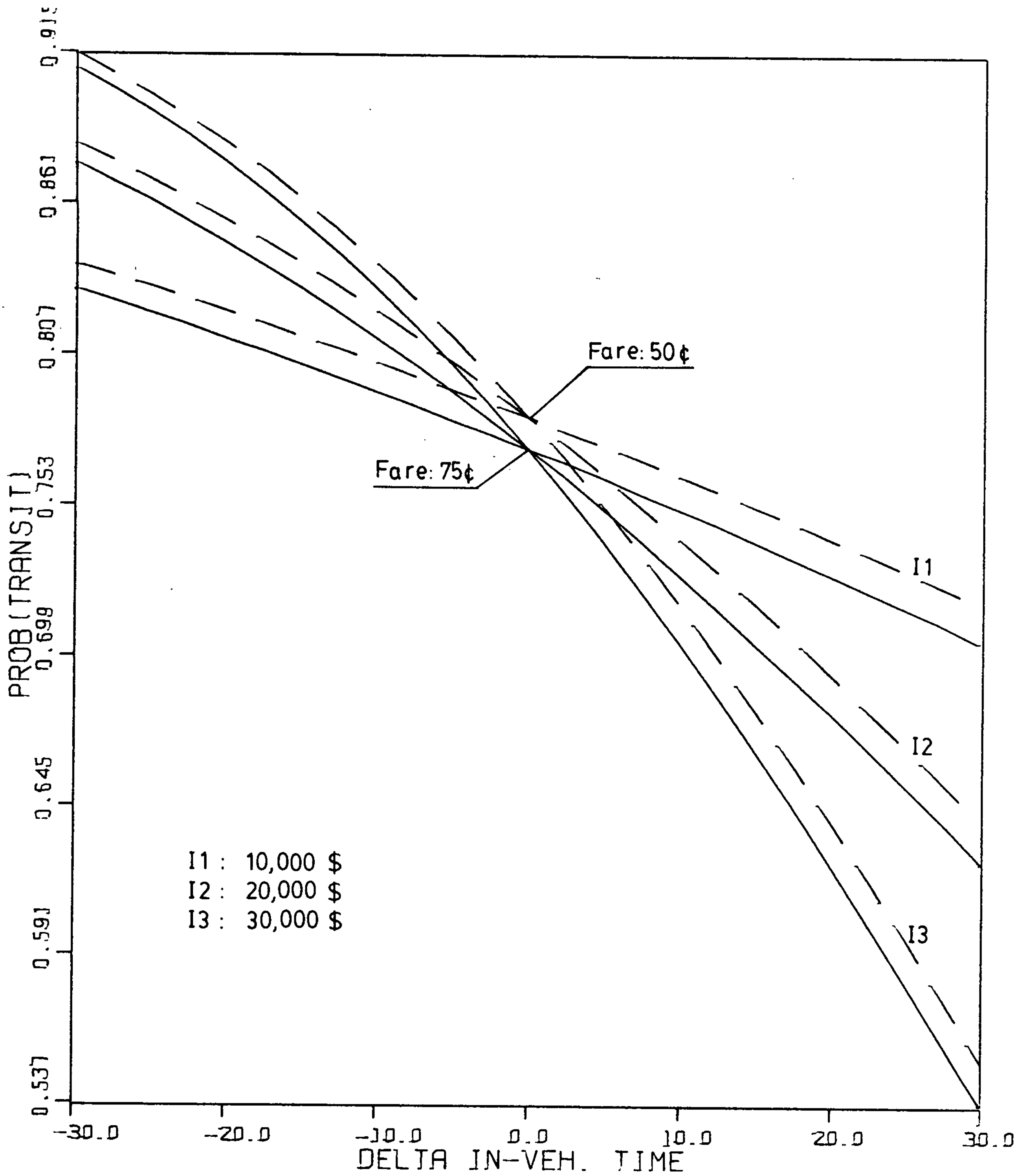

Figure 17 - Interaction Effect of Income and Fare Variation on Transit Use Probability with Respect to ΔT

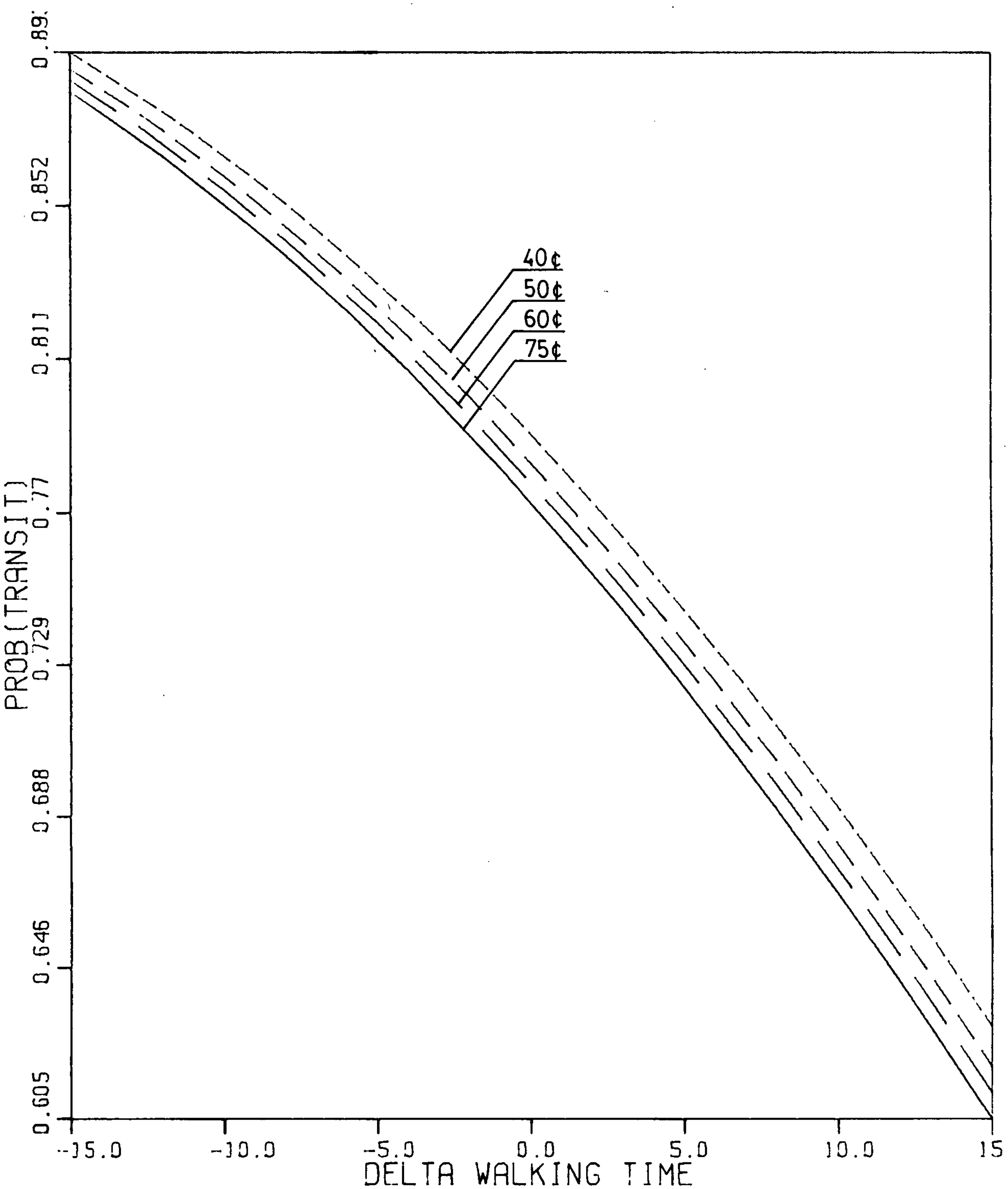

Figure 18 - Effect of Fare Variation on Transit Use Probability with Respect to Walking Time

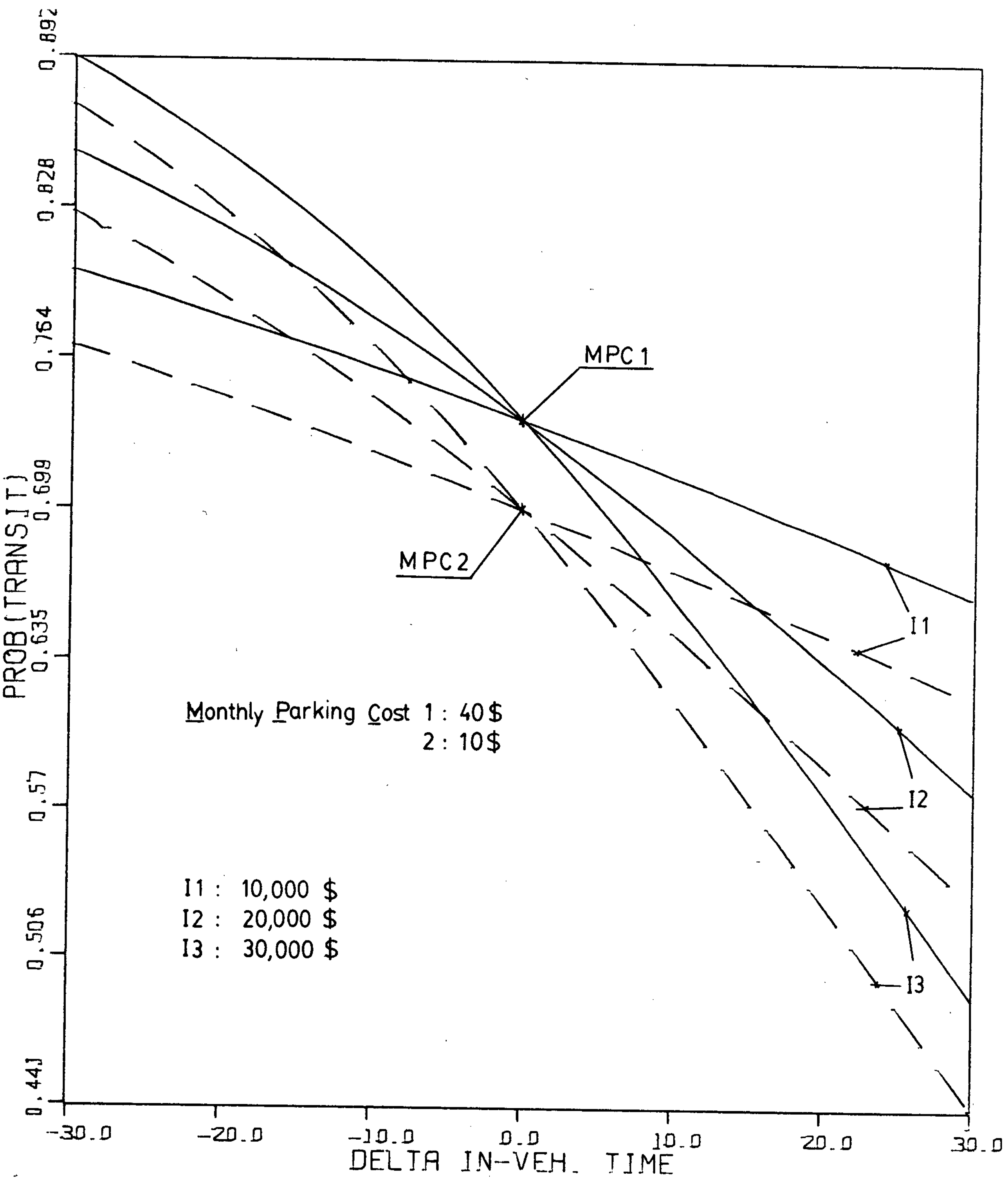

Figure 19 - Interaction Effect of Income and Parking Cost Variation on Transit Use Probability with Respect to ΔT

is more accentuated. Everything being equal in terms of system attributes, an increase of 30$ in monthly parking cost rises the probability of transit use from 70% to 74%. However, one might keep in mind that this study considers only work trips which are inelastic towards parking cost variation. The effect of parking cost on the transit use probability of other purposes is important.

3. VALUE OF TIME

The value of time is defined as to be the amount of money that a traveller is ready to pay in order to save 1 minute of his/her travel time.

For this purpose, first, the linear function L is differentiated with respect to in-vehicle travel time($\delta L/\delta(\Delta T)$) and then with respect to travel cost($\delta L/\delta(\Delta C)$); the value of in-vehicle travel is derived by the $\delta L/\delta(\Delta T)$ to $\delta L/\delta(\Delta C)$ ratio when $\delta(\Delta T)$ is equal to unity.

According to the calibrated logit model, the function L is equal to:

$$L = 1.10-.0012(I\Delta T)+...+.490\Delta C$$

By differentiating the above function with respect to ΔT and ΔC, we have:

$$\delta L/\delta(\Delta T)=-.0012I \quad , \quad \delta L/\delta(\Delta C)=.49$$

Thus:

$$[\delta L/\delta(\Delta T)]/[\delta L/\delta(\Delta C)] = \delta(\Delta C)/\delta(\Delta T) = -(.0012/.49)\ I$$

And by equating $\delta(\Delta T)$ to 1 minute, the value of in-vehicle time is found:

$$Vt = |\delta(\Delta C)| = 2.45\ 10^{-3}\ I\ [\$/min] = 0.147\ I\ [\$/hr]$$

where I is the household annual income in 1000 $.
Usually, the value of time is expressed in terms of the hourly wage rate(w). If 2000 hours per year is assumed, the value of the in-vehicle time will be:

$$Vt = (.147 \times 2000/1000)w = .29\ w\ [\$/min]$$

The determination of walking and waiting time follows the same line of reasoning:

$$Vwalk = (.0053/.490)I = 1.08\ 10^{-2}\ I\ [\$/min] = 0.65\ I\ [\$/hr]$$
$$= 1.30\ w\ [\$/min]$$

and

$$Vwait = (.0026/.490)I = 5.31\ 10^{-2}\ I\ [\$/min] = 0.32\ I\ [\$/hr]$$
$$= 0.64\ w\ [\$/min]$$

The use of the probit model leads to the following value of time:

$$Vt = 0.28\ w\ [\$/hr]$$
$$Vwalk = 1.18\ w\ [\$/hr]$$
$$Vwait = 0.60\ w\ [\$/hr]$$

One might take the average and determine the value of different components of travel time in $/hour for commuters in the year of the survey(1972):

$$V_t = 0.285\ w$$
$$V_{walk} = 1.240\ w$$
$$V_{wait} = 0.620\ w$$

These values in terms of in-vehicle travel time value(Vt) are:

$$V_t = 1$$
$$V_{walk} = 3.50$$
$$V_{wait} = 2.17$$

4. COMPARISON WITH OTHER STUDIES

From the above section, it can be concluded that commuters consider walking and waiting time 3.50 and 2.17 times more onerous than in-vehicle time. This difference of walking time being valued 1.6 times as highly as waiting time is not in agreement with the results obtained by previous studies (see Chap.II). Two possible explanations for this discrepency are:

i. The waiting time generation model did not simulate correctly the survey situation. This needs more detailed analysis and can be the subject of further research.

ii. The second explanation assumes that the waiting time was correctly simulated and commuters indeed consider walking time twice as inconvenient as waiting time. This may be true when environmental factors such as weather are considered; for example on rainy days, the disutility of walking is greater than that of waiting under a shelter. And, since the data set used for this analysis were obtained from a survey carried out during spring, the obtained result is reasonable.

The value of in-vehicle time is estimated to be about 28.5% of the hourly wage rate. The travel time value found in the literature have a high variance. It varies between 25% and 67% of the hourly earning rate. However, a value of 30% of the hourly wage rate is usually considered (Foster and Beesley, 1963).

VII. GENERAL CONCLUSION

This thesis determines the components of the journey to work demand in the Vancouver Metropolitan Area.

1. SIMULATION

The major problem encountered in estimating the coefficients of the underlying demand function was the amount of missing information on the responders' alternative travel mode attributes. It was decided to simulate the missing data rather to attempt obtain their obsrved value since it is believed that the important element in the behavioural model development is the travellers' perception of the system attributes, and to supplement the existing data would be too costly.

Three types of simulation model were used: regression model to estimate the missing values of transit travel time, the random generation model to simulate the missing values on waiting time from an estimated a-priori frequency distribution and an aggregation method to complete parking cost data.

The aptness of the above models was evaluated by comparing the obtained value of different components of travel time with those available from previous studies (see Section3: Variable Sensitivity).

2. MODEL STRUCTURE

From the three statistical approaches used to calibrate the demand model, only logit and probit estimation methods succeeded in reproducing the actual situation satisfactorily. Models calibrated with these two methods embody exactly the same variables and variables form, whereas discriminant analysis led to a different form of the variables.

The determinant of transportation modal choice which were statiscally significant for the available data set are as follows:
In-vehicle travel time, walking time to and from the modal interface, waiting time at the modal interface, the travel out-of-pocket expenses, the sex and age of the tripmaker, the household income and the number of cars available to the traveller's household.

3. VARIABLE SENSITIVITY

The value of in-vehicle time obtained in this study (28.5% of w, the hourly earning rate) corresponds approximately to the value used in practice (30% of w). But the walking time was valued 1.6 times more onerous than waiting which is not in agreement with other studies. It is, however, believed that environmental factors had highly influenced the attitude of responders and modified the shape of the perceived waiting time distribution.

Since on an intuitive ground, income plays an important role in modal decision, the effect of other variables on the modal split probability was combined with three income levels: 10,000$, 20,000$ and 30,000$.

It is found that a difference of ±15 minutes in in-vehicle travel time changes the probability of low-income people from 72% to 75% whereas, for high-income people this change is between 64% and 84%. However, with respect to walking time, this range of variation is greater(55% to 83% for low-income people and 22% to 97% for wealthy people).

Although female travellers seems to be less sensitive to changes in in-vehicle time, they react strongly to policies related to walking time. This also applies to all travellers below 25 and above 60 years old. These results might be used as proposals in order to increase the transit ridership:

i. To capture the higher social strata, the performance of transit system in terms of in-vehicle time should be highly competitive with that of private transportation. For this, modern technology, offering a better acceleration and deceleration rate should be incorporated into the system, stop spacing should be optimized and boarding and alighting time should be reduced to its minimum.

ii. Since people regardless their age and sex are strongly affected by walking time increases, a restriction on parking places or on parking development in the downtown core will help increase the transit ridership. Park and ride configurations combined with modern transit technology such as Automated Light Rail Transit(ALRT), Rapid Rail Transit(RRT) and so forth, would provide desirable results since the walking time can be reduced to its minimum and the in-vehicle time to a point which can compete with the private car.

Although parking cost can have a great importance in the determination of the magnitude of the transit use, in this study an increase in parking cost had no considerable effect on the probability of transit use. This is possibly due to the fact that in the present research, only work trips were considered and the parking demand of these trips is highly inelastic with respect to parking cost.

BIBLIOGRAPHY

1. Afifi, A.A. and Azen, S.P. (1972); Statistical Analysis: A Computer Oriented Approach ; Academic Press, New York.

2. Anderson, T.W. (1958); An Introduction to Multivariate Statistical Analysis ; Wiley, New York.

3. Ben Akiva, M. and Atherton, T. (1971); "Methodology for Short-Range Travel Demand Prediction: Analysis of Carpooling Incentives", in Journal of Transport Economics and Policy, Vol.9, pp. 224-261.

4. Benjamin, J.R. and Cornell, C.A. (1970); Probability, Statistics and Decision for Civil Engineers ; McGraw-Hill, New York.

5. Brown, G.R. (1972); "An Analysis of Preferences for System Characteristics to Cause a Modal Shift", in Highway Research Record, No.417, pp.25-36.

6. Cooley, W.W. and Lohnes, P.R. (1971); Multivariate Data Analysis ; Wiley, New York.

7. Cox, D.R. (1970); The Analysis of Binary Data ; Methuen and Co. Ltd., London.

8. Cramer, H. (1955); The Elements of Probability Theory ; Wiley, New York.

9. De Donnea, F.X. (1971); The Determinants of Transport Mode Choice in Dutch Cities ; Rotterdam University Press.

10. Dixon, W.J. (1969); Introduction to Statistical Analysis ; McGraw-Hill, New York.

11. Domencich, T.A. and McFadden, D. (1975); Urban Travel Demand, A Behavioral Analysis ; North-Holland, Amsterdam.

12. Feller, W. (1968); An Introduction to Probability Theory and its Application Vol.2; Wiley, New York.

13. Finney, D.J. (1975); Statistical Method in Biological Assay ; Cambridge University Press.

14. Foster, C.D. and Beesley, M.E. (1963); "Estimating the SocialBenefit of Constructing an Underground Railway in London"; Journal of the Royal Statistical Society , Vol.126.

15. Hanushek, E.A. and Jackson J.E. (1977); Statistical Methods for Social Scientists ; Academic Press, New York.

16. Harrison, A.J. and Quarmby D.A. (1969) "The Value of Time in Transport Planning: a Riview", in Theoretical and Practical Research on an Estimation of Time Saving , European Conference of Ministers of Transports, Report of the Sixth Round, Paris.

17. Horowitz, J.L. (1981); "Sources of Error and Uncertainty in Behavioral Travel Demand Models", in New Horizon in Travel Behavior Research , P.R. Stopher, A.H. Meyburg and W. Brog; Lexington Books, Mass.

18. Hutchinson, B.G. (1974); Principles of Urban Transport Systems Planning ; McGraw-Hill, New York.

19. Johnston, J. (1963); Econometrics Methods ; McGraw-Hill, New York.

20. Kendall, M.G. and Stuart A. (1969); The Advanced Theory of Statistics , Vol.2; Hafner, New York.

21. Lave, C.A. (1969); "A Behavioural Approach to Modal Split Forecasting", in Transportation Research , Vol.3, pp. 463-480.

22. Lisco, T. (1967); The Value of Commuters' Travel Time: A Study in Urban Transportation . Department of Economics, University of Chicago, Ph.D. Thesis.

23. Mc Gilliway, K.G. (1970); "Demand and Choice Models of Mode Split", in Journal of Transport Economics and Policy, Vol.4, pp.192-207.

24. McFadden, D. (1974); "Measurment of Urban Travel Demand", in Journal of Public Economics , Vol.3, pp. 303-328.

25. ——————— (1976); The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation , presented at the Seminar on Engineering Transportation Planning Methods, Daytona Beach, Florida.

26. ——————— (1976); "The Mathemetical Theory of Demand Models", in Behavioral Travel Demand Models , P.R. Stopher and A.H. Meyburg; Lexington Books, Mass.

27. Manheim, M.L. (1979); Fundamentals of Transportation System Analysis , Vol.1; MIT Press, Mass.

28. Merlin, P. and Barbier, M. (1965); Study of the Modal Split Between Car and Public Transport in Paris , IAURP.

29. Morall, J. (1971); "Work Trip Distribution and Modal Split in the Metropolitan Torronto Region", in Canadian Transit Handbook , edited by Soberman, R.M. and Hazard H.A.

30. Morrison, D.P. (1967); Multivariate Statistical Methods ; McGraw-Hill, New York.

31. Navin, F.D.P. (1974); "Time Costs in Personal Rapid Transit", Personal Rapid Transit II , University of Minnesota, Minneapolis, Minn. pp.589-602.

32. Neter, J. and Wasserman, W. (1974); Applied Linear Statistical Models ; Irwin Inc., Homewood, Ill.

33. O'Farrel, P.N. and Markham, J. (1975); The Journey to Work: A Behavioural Analysis ; Pergamon Press, Oxford.

34. Quandt, R.E. (1970); The Demand for Travel: Theory and Measurement ; Lexington Books, Mass.

35. Quarmby, D.A. (1967); "Choice of Travel Mode for the Journey to work: Some Findings", in Journal of Transport Economics and Policy , Vol.1, No.3, pp. 273-312.

36. Rao, C.R. (1952); Linear Statistical Inference and its Applications ; Wiley, New York.

37. Reichman, S. and Stopher, P.R. (1971); Towards Disaggregate, Stochastic Models of Travel Mode Choice , paper presented to the 50th annual meeting of the Highway Research Board.

38. Rogers, K.; Townsend, S. and Metcalf, A. (1970); Planning for the Work Journey , Local Government Operational Research Unit, Report No. 667, April.

39. Rosenberg, L (1971); Statistical Reasoning ; Merill, Colombus, Ohio.

40. Segal, D. (1978); A Discrete Multivariate Model of Work-Trip Mode Choice , Department of City and Regional Planning, Harvard University, Camb. Mass., Discussion Paper D78-7.

41. Stopher, P.R. (1969); "A Probability Model of Travel Mode Choice for the Work Journey", Highway Research Record , No.283, pp. 57-65

42. ——————— and Meyburg, A.H. (1975); Behavioural Travel-Demand Models ; Lexington Books, Mass.

43. Talvitie, A.P. (1981); " Inaccurate or Incomplete Data as a Source of Uncetainty in Econometric or Attitudinal Models of Travel Behavior", in New Horizon in Travel Behavior Research , P.R. Stopher, A.H. Meyburg and W. Brog, Lexington Books, Mass.

44. Theil, H. (1966); Applied Economic Forecasting ; North-Holland, Amsterdam.

45. ————— (1969); "A Multinominal Extension of the Linear Logit Model", in International Economic Review , Vol.10, No.3, pp. 251-259.

46. ————— (1971); Principles of Econometrics ; Wiley, New York.

47. Warner, S.L. (1962); Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice ; Northwestern University Press, Evanston.

48. Watson, P.L. (1974); The Value of Time; Behavioral Models of Modal Choice ; Lexington Books, Mass.

49. Wilks, S.S. (1962); Mathematical Statistics ; Wiley, New York.