AN AGGREGATE BEHAVIORAL APPROACH TO URBAN TRANSPORTATION MODAL SPLIT MODELLING

by

TREVOR ANTHONY TOWNSEND

A Thesis
presented to the University of Manitoba
in partial fulfillment of the
requirements for the degree of
MASTER OF SCIENCE IN CIVIL ENGINEERING
in
THE DEPARTMENT OF CIVIL ENGINEERING

Winnipeg, Manitoba, 1980

AN AGGREGATE BEHAVIORAL APPROACH TO URBAN TRANSPORTATION MODAL SPLIT MODELLING

BY

TREVOR ANTHONY TOWNSEND

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

TREVOR ANTHONY TOWNSEND

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in whole or in part, at the request of other institutions or individuals for the purpose of scholarly research.

TREVOR ANTHONY TOWNSEND

The University of Manitoba requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

ABSTRACT

Utilizing the binomial logit model this thesis develops a model of modal choice for a medium-sized Canadian city.

A zonal aggregation was made considering the zones as behavioral units. Thus the number of commuters in each zone choosing a particular mode was used as the dependent variable instead of the probabilities of a particular modal choice.

Three separate models were developed for a bimodal situation. The models can be extended to multinomial form provided that the relevant data are available.

The data used in the development and testing of the model were from the City of Winnipeg, 1976. The modelling was done with peak-a.m.-hour work trips to the central business district. Testing was done on a university zone situated on the outskirts of the city, and on a zone adjacent to the central business district.

It was found that the models accurately predicted modal choice in the latter case but not in the former.

Discussions on the implications and potential uses of the models, as well as areas for possible research are presented.

ACKNOWLEDGEMENTS

The author wishes to thank all those who made the completion of this thesis possible.

Dr.A.H.Soliman, his Supervisor, for his guidance and counsel: The Information Services of the Winnipeg Transit System for their information about bus fares and schedules; The students of the 1980-81 University of Manitoba third year Highway Engineering course and the members of the West Indian Students' Association of the University of Manitoba for their travel time data: D. Cox and A. Warren for their advice on the computer programming necessary for this thesis; Mr. Bruce Ball of the Energy Council for his data on fuel prices; Professors A. Clayton and O. Hawaleshka for their valuable advice; and, his wife Pamela, for her support, encouragement and faith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

1.1 GENERAL INTRODUCTION

In attempting to model modal choice behavior, transportation planners have endeavoured to quantify the effects of both the transportation system characteristics and the socioeconomic characteristics of the system users.

The variables expected to influence modal choice have usually included, but not been limited to, in-vehicle travel times, out of vehicle travel times, out of vehicle travel times by competing modes and the characteristics of the trip-maker. These have usually been compared as ratios or differences. More recent thinking has resulted in the transportation choice decision of the individual being scrutinized more closely. Attempts have been made to model this decision-making process using choice theory instead of purely empirical data. These models reflect the attitudes of the commuter to his, and the transportation system characteristics.

It would be expected that the 1973 oil embargo and the resulting energy crisis would affect the attitudes of the commuter. Models developed before this time would not be able to effectively simulate the effects of changes in transportation system or user characteristics after the oil embargo.

To date no behavioral model of the modal choice process of commuters in Winnipeg for the "post-energy-crisis" era has been developed. However, it cannot be assumed that commuters would place the same value on time, comfort, or convenience savings as they did prior to 1973. What is needed then, is a behavioral model of modal choice that would be developed on data collected after the expected changes in commuter attitudes.

1.2 OBJECTIVE OF THE STUDY

The objectives of the study were as follows:

1. To develop a behavioral model that would simulate modal choice.

2. To identify the transportation system variables and user characteristics that would play a significant role in the choice behavior of the commuter.

3. To study the implications of changes in these variables and characteristics on modal choice behavior.

1.3 SCOPE OF THE STUDY

Aggregate models of modal choice based upon the binomial logit formulation were obtained utilizing peak-a.m.-hour work trips into the central business district of the City of Winnipeg in 1976.

The models were tested on two other zones in the study area.

1.4 ORGANIZATION OF THE THESIS

The following is an outline of the organization of the report:

Chapter 2: reviews modal split modelling

Chapter 3: discusses the theory behind the models developed and the mathematical formulation used

Chapter 4: describes the transportation system variables and user characteristics used in the modelling

Chapter 5: deals with the estimation of the model coefficients and goes into the statistical and sensitivity analyses

Chapter 6: shows the results of testing of the models on two other zones in the study area

Chapter 7: discusses the assumptions and limitations of the modelling; the implications and possible uses of the models and areas for future research.

Chapter II

REVIEW OF EXISTING MODAL SPLIT MODELS

Modal split is the division of the flow of people or goods between two points into the alternative travel modes available. Considering the simplest of cases in urban transportation - for example a small town - the choice may be between one form of public transportation and the private automobile. Many studies[19] of urban transportation mode patronage have suggested that trip-makers must be thought of as belonging to two separate groups:

1. Captive transit riders.

These are trip-makers who do not have access to a car for a particular trip being considered.

2. Choice transit riders.

These are trip-makers who have access to a car and may choose freely between public and private transportation for a particular journey.

There may also be considered a third grouping; those captive automobile riders who, because of the characteristics of their activities at their destinations, need to use their cars.

Modal split may be carried out either between the trip generation and the trip distribution phases, or after the trip distribution phase of transportation modelling. The former types of models are sometimes referred to as trip-end modal split models. The latter are termed trip-interchange modal split models.

Hutchinson et al.[20] categorized modal split models into four types.

1. Traditional models.

2. Two-stage models.

3. Behavioral models

4. Econometric models

Stopher and Reichman[34] give a good description of the historical development of urban transportation modal choice models. The models are broadly classified into three principal groups. The first group, of which the Chicago Area Transportation Study is an example, attempted to predict mode choice at an aggregate level by using only the socioeconomic characteristics of the aggregated areas. These models did not consider the characteristics of the transportation systems. The second group of models were again aggregate models but they incorporated some characteristics of the transportation systems. The third group that was considered comprised the disaggregate stochastic models.

Others[24] prefer to group the models into the following categories:

1. Aggregate deterministic models.

2. Disaggregate deterministic models.

3. Aggregate stochastic models.

4. Disaggregate stochastic models.

2.1 AGGREGATE DETERMINISTIC MODELS

According to Richards[35]

> An aggregate model is a model estimated with a dependent variable which represents a group of observations,whilst a disaggregate model is a model estimated with a dependent variable which represents an observation of a single occurrence.

2.1.1 TRADITIONAL MODELS

Aggregate deterministic models have been the most used models up to the present time. All the "traditional" models are of this type. They deal with groups of travellers who are assumed to have the same characteristics. Some, like the model used in the Metropolitan Toronto and Region Transportation Study[19], use diversion curves relating per cent of transit usage from a zone to variables such as relative travel time, travel cost and travel service (excess travel time) of competing modes, as well as socioeconomic variables describing the trip-makers. The Southeastern Wisconsin model[19] relates the zonal percentage public transit usage to the average number of cars per household (commuter characteristic) and the ratio of highway to transit accessibility (destination and transportation system characteristic). A diversion curve relationship is also used.

In some cases a mathematical relationship is used instead of diversion curves. For example Zupan[49] used U.S.census journey to work data to develop distribution and modal split models for work trips in the New York region. After distribution by gravity model, the walk to work trips were modelled separately. Then the data were stratified by income and relationships developed for percentage transit usage for each stratum, as well as for the unstratified data, by regression analysis. The unstratified modal split equation developed was:

$$\%\ \text{Transit} = 7.756(\ln ED)+2.723\sqrt{RD}\ +17.844SF +20.474(TA/TT) +0.112(L+P)-14.50$$

where:

ED = employees(thousands)per developed square mile at destination zone

RD = resident workers(thousands) per net residential square mile at origin zone.

SF = rail service factor (dummy variable with a value of 1 for O-D pair with rail transit service that allowed travel without transferring, and 0 for the other O-D pairs).

TA = zone to zone time by auto, in minutes

TT = zone to zone time by transit, in minutes

L = zone to zone toll cost, in cents

P = parking cost at destination, in cents.

2.1.2 TWO-STAGE MODELS

Two-stage models are usually aggregate deterministic models with the data stratified into captive transit and choice transit riders. Each stratum is modelled separately. Ferreri and Cherwony[14] developed separate modal split equations for choice and captive riders using data from Miami, Florida, for a variety of trip types. The captive rider models directly generated trips from zonal and population data while the choice rider model used trip-interchange modal split techniques with diversion curves.

The models so far discussed do not only have the drawback of being aggregate, with a resultant loss of variability within the aggregated units, but also, they do not consider the commuter's decision-making processes.

2.1.3 ECONOMETRIC MODELS

Domencich, Kraft and Valette[10] attempted to combine trip generation and modal choice decisions in a single model. Approaching the problem from an economist's viewpoint, it was assumed that the consumer makes decisions about transportation as he would about any other commodity, and simultaneously decides trip purpose as well as when and how the trip would be made. Using logarithmic, linear, and mixed log-linear models they tested data from California. The dependent variable was the number of round trips using a particular mode, stratified by trip purpose. The independent variables included socioeconomic and land use data pertaining to the trip purpose, and interzonal travel times and costs by all modes and for all zonal pairs. It was felt that this formulation would remove the implicit assumptions in previous models that the number of trips generated is independent of transportation system performance. These models, however, have not proven very effective and require complex mathematical formulations.

2.2 DISAGGREGATE DETERMINISTIC MODELS

These are models that predict the behavior of the individual rather than of a group of individuals. No random element of behavior is assumed. The development of these models has been reported by Manheim[24] to be purely theoretical and they are usually used for expository purposes.

2.3 DISAGGREGATE STOCHASTIC MODELS

Stochastic models are those which include an explicit random element. The stochastic modal split models combine both behavioral and econometric theories in their derivation. Pratt[31] and others attempt to explain choice behavior in terms of utility (or disutility) of various alternatives. In formulating disaggregate stochastic models, it is assumed that the consumer will endeavour to maximize the utility, or, more correctly, minimize the disutility involved in a travel action. Usually, the utility of an alternative is a function of the variables describing the characteristics of the alternative and variables describing the characteristics of the traveler. The randomness of these models can be attributed to either pure probabilistic behavior by the consumer or to the fact that the consumer does not have perfect knowledge about the utilities of his alternatives. Because of this randomness, probabilities of modal choice, rather than actual choices, are considered. Pratt has concluded that:

> Choice of travel mode can be treated as an economic response to the transportation system characteristics. As such, it can be predicted using quantitative disutility comparisons and probability mathematics.

Disaggregate stochastic models have been studied quite extensively over the past few years by Domencich and McFadden(11) and others. Different assumptions about either the form of the utility functions or the nature of the randomness lead to different forms of these models. The three most popular models are discriminant, probit, and logit models. DeDonnea(7) has done a comparison of these models.

Discriminant analysis attempts to classify travellers into two groups according to modal choice. This is a bimodal model. The aim is to find a linear combination of explanatory variables that minimizes misclassification. These variables are usually those describing modal and commuter characteristics.

Probit analysis is based on the premise that if the elements of a population are subjected to a stimulus that can range over an infinite scale, the frequency of responses to the stimulus will be normally distributed. It assumes that the response is a 0,1 response, (i.e., it either occurs or it does not occur). The stimuli in a modal choice situation are the relative utilities of the alternatives.

Logit analysis assumes that the random variable is distributed with a Weibull distribution.

Discriminant analysis techniques have been used by Quarmby[32] to develop a modal choice model for estimating car-bus and car-train modal split for work trips to the central business district of Leeds.

The model obtained from the car-bus study was:

$$p(c/z) = \frac{2.26\, e^{1.04(z - 0.431)}}{1 + 2.26\, e^{1.04(z - 0.431)}}$$

where p(c/z) = the probability of choosing the car mode given that the relative travel disutility is z.

and z = disutility function based upon differences in overall and excess travel time and cost between car and bus and socioeconomic data.

Bock[4] and McGillivray[26] also used discriminant analysis in developing models. McGillivray studied data collected by the San Francisco Bay Area Transportation Study Commision in an attempt to see the effects of several policies on work trip modal split. Shunk and Bouchard(39) used discriminant analysis with marginal disutility as the independent variable on 1958 data from St.Paul and Minneapolis. The equation developed for the marginal utility of auto over transit for non-CBD trips was:

$$U=2.5(Ta+Tw-At)+(Tr-Ar)+(F-0.5P-4.0D)/C$$

where U = marginal utility
Ta = walk time to/from transit
Tw = wait time for transit
Tr = transit running time
F = transit fare
At = auto terminal time
P = parking cost
D = highway distance
C = cost of time

Warner, Lisco and Lave used probit analysis to develop modal choice models. Lisco studied work trips from Skokie, a Chicago suburb, to the "Loop", the central business district. Although the model is developed in terms of routes rather than modal choice, the empirical work estimates a modal choice model. The two choices were a fast toll road and a slow free road.

Lave(21) developed a modal split model on 1956 Chicago data. The model was:

$$Y = -2.08+0.000759KW\Delta T+0.0186\Delta C-0.0254IDC+0.0255A$$

where Y = a binary index number determining the choice decision of an individual. Positive Y means that an individual chooses transit. Negative Y means that the individual chooses auto.
$KW\Delta T$ = The product of the relative time difference between the two modes, the commuter's hourly wage and a factor whose size depends on the individual's marginal preference for work time (income) vs. leisure time.
ΔC = Difference in costs between transit and auto.

IDC = Product of income, distance and a binary variable defining comfort.
A = Age of the trip-maker.

Much work has been done using logit models. Stopher[42], in a study of a group of London commuters, developed a bimodal logit modal split function. Stopher's model was of the form:

$$p = a(C2-C1) + b(T2-T1) + d$$

where p = probability of using a car.
d = probability of using a car when costs and time by public transport and car are the same.
C = cost of (1)automobile and (2) public transport
T = travel time of (1) automobile and (2) public transport.

However, it was found that this formulation yielded values of p in excess of 1 and smaller than 0 for very large positive or negative values of the cost and time differences. Because of this, Stopher postulated a logistic relationship:

$$\ln \frac{p}{1-p} = y = a'(C2-C1)+b'(T2-T1)+d'$$

Recker and Stevens[33] developed attitudinal multinomial logit models for four nonwork activities. These activities were major grocery shopping, shopping for odds and ends, shopping for personal goods and visiting friends and acquaintances. Explanatory variables were the individual's beliefs about the attributes of four modal alternatives, bus, car, taxi and walking. The estimated coefficients were obtained using maximum likelihood techniques.

Golob and Recker[16] put forward a procedure for using data on travellers' attitudes to predict changes in choice of transportation modes accompanying changes in the characteristics of the modes. These characteristics typically included such qualitative attributes as "opportunity to relax" and "security from the undesirable acts of others" as well as quantitative variables like travel time. The study was based on data collected for Buffalo, N.Y.

Hartgen and Tanner[18] have also carried out investigations into the effects of traveller attitudes on modal choice behavior.

Richards and Ben-Akiva[36] developed a work choice model and a simultaneous shopping destination-mode choice model for an urban area in the Netherlands. The latter was very similar in specification to Wilson's entropy maximization models[48]. The modal choices considered for the study were car, public transport, bicycle, moped and train.

Ben-Akiva and Atherton[2] used a multinomial logit model to predict the effects of carpool strategies on modal choice. Atherton in an earlier study[1] combined three separate models into a single model system. These models were:

1. A joint automobile ownership/work-trip mode choice model (for the head of the household only);

2. A work trip mode choice model for all workers; and,

3. A simultaneous frequency, destination and mode choice model for nonwork trips.

Others who investigated simultaneous demand models of the multinomial logit form include Lerman[22], who developed a simultaneous location/housing/automobile ownership/mode-to-work model for skilled single worker households, working and residing in Washington, D.C. in 1968.

Stopher, Talvitie[44] and others have also compared the three most popular forms of disaggregate stochastic models. While opinions differ, the general consensus appears to favour the logit models for ease of formulation and predictive ability.

2.4 AGGREGATE STOCHASTIC MODELS

Aggregating stochastic models remains a trying problem. Domenchich and McFadden, Manheim[24] and others have postulated methods of doing this. They involve making assumptions about the distributions of the independent variables over the population of travellers. To date however, no firmly based aggregate stochastic models have been developed.

The Bibliography lists literature pertaining to modal choice modelling.

Chapter III

GENERAL MODEL FORMULATION

3.1 GENERAL DISCUSSION

This thesis is based on the disaggregate binomial logit model of modal split. The binomial model is a special bimodal case of the more general multinomial model. As previously stated, this model, which attempts to simulate consumer behavior, uses the principle of utility maximization. This principle is embodied in standard consumer theory. There are, however, some differences between the demand approach of economic consumer theory and the transport modelling choice theory. These differences can be summarized as follows:[1]

1. The demand approach of standard consumer theory assumes choices are to be made from among continous, infinitely variable combinations of commodities. Transportation choices, however, are made from among a finite set of discrete, mutually exclusive and exhaustive alternatives. Examples of this would be choices of destinations, modes or routes which are all discrete.

---

[1]This section is after Domencich and McFadden (1975) and Manheim (1978).

2. In standard consumer theory it is assumed that choices are based upon the quantities of the desired commodities in the mix of commodities. Transportation choices are based upon the values of the attributes of these commodities. Thus in making a decision on a choice of destination, the characteristics of each destination are considered; not the "amount" of each destination.

3. Standard consumer theory assumes that the choice made maximizes utility based upon certain resource constraints (usually time or money). There is no need to have a budgetary constraint in modelling transportation choices. The reasons for this are two-fold. Firstly, the money spent on transportation is usually only a modest proportion of income. Secondly, there are other transportation system attributes which play a greater role than monetary cost in determining choice.

Assuming that trip generation rates are constant, then the benefits that may accrue as a result of the trip destination are irrelevant. What are important to the modal choice evaluation are the utilities associated with the different modes. The utility of a trip by a particular mode can be considered to be a common measure converting all trip "costs" into equivalent units.

These "costs" may include, but not be limited to, monetary as well as temporal, spatial and convenience costs. The converted costs can be summed for each mode to give a value of the utility of the mode. Thus a particular consumer will associate a particular value of utility for each modal alternative.

It is assumed that the consumer attempts to make a choice that maximizes the sum total of his personal utility or minimizes the sum total of his personal disutility that accrues from a trip-making decision. It is also assumed that the consumer will, in general, behave rationally and that his behavior will tend to accurately reflect his preference. The consumer, however, does not have and so cannot act with perfect knowledge of the utilities of all alternatives. Thus, a probabilistic behavior mechanism has to be introduced.

3.2 GENERAL EQUATION

The utility function can be divided into two parts - a deterministic part, and an additive random part. Thus the utility of an alternative (a) to a consumer (c) can be defined as:

$$uac = ua(Xa,Sc) + eac \qquad (3.1)$$

where uac = the utility of alternative (a) to consumer (c).

ua(Xa, Sc) = the deterministic part of the utility based upon the characteristics of alternative (a) and consumer (c).

Xa = characteristics of alternative (a).

Sc = socioeconomic characteristics of consumer (c).

eac = random element involved in the perception of the utility of alternative (a) to consumer (c).

ua = a linear function combining (a) and (c).

Furthermore, if (Ac) is defined as the set of all possible alternatives from which consumer (c) must choose one and only one, then the probability of the consumer choosing an alternative (a) is:

$$P(a:Ac)=Prob[uac \geq ubc, \forall b \in Ac] \quad (3.2)$$

Substitution of eq.(3.1) into eq.(3.2) gives:

$$P(a:Ac)=Prob[ua(Xa,Sc)+eac \geq ub(Xb,Sc)+eab, \forall b \in Ac] \quad (3.3)$$

Rearranging the terms of eq.(3.3) gives:

$$P(a:Ac)=Prob[eac-ebc \geq ub(Xb,Sc)-ua(Xa,Sc), \forall b \in Ac] \quad (3.4)$$

This equation implies that the mathematical form of the choice model depends upon the assumption made about the joint probability distribution of the random elements. If it is assumed that the random elements have a Weibull distribution then the multinomial logit model is formed.

$$P(a:Ac)= \frac{e^{Uac}}{\Sigma e^{Ubc}} \quad (3.5)$$

for all $b \in Ac$

Where $Uac = ua(Xa,Sc)$ (3.6)

The logit model lends itself readily to mathematical manipulation and in many applications corresponds to a plausible stochastic specification.

The Weibull distribution is stable under maximization. Thus the maximum of two independent Weibull distributions is itself a Weibull distribution. This is very important when one considers that the problem of modal choice is one of utiity maximization. The second important property of this distribution is that the difference of two independent Weibull distributed random variables has a binary logit distribution. The logistic distribution is an ogive which maps the range of utilities onto a zero-one interval.

The definition of the Weibull distribution and the mathematical derivation of the multinomial logit function are given in Appendix A.

If it is assumed that Uac is linear in its parameters, it is convenient, for any alternative (a) open to consumer (c), to transform the variables (Xa) and (Sc) into variables (Xac) which can be considered as linear combinations of both transportation system variables and user characteristics.

The utility function can then be expressed as:

$$Uac = Q1Xac1 + Q2Xac2 + \ldots + QnXacn. \quad (3.7)$$

Where Xack = k^{th} variable affecting the utility of alternative (a) to consumer (c).

Qk = coefficient of the k^{th} variable affecting the utility of alternative (a) to consumer (c).

Xack may be either a generic or an alternative specific variable. In the former case, the variable appears in the utility functions of all the alternatives. In the latter, it may only be relevant to a specific alternative.

The aim of the modal split modelling is to get the best estimates for these coefficients. In disaggregate models, the coefficients must be estimated using maximum likelihood techniques. This is because the observed dependent variable, (i.e., the actual choice made) is a 0,1 variable. One and only one choice of mode can be made.

If behavioral units are aggregated according to socioeconomic categories or geographical location, then the observed dependent variable is a share with a value that lies between zero and one. The data available for this thesis were aggregated zonally and it is assumed that the commuters in any one zone are homogeneous in both socioeconomic characteristics and in how they perceive the characteristics of the transportation system.

Meyburg and Stopher[27] and Cherwony and Lutin[6] have used binomial logit models with aggregate data with varying results. Both maximum likelihood and regression techniques may be used to estimate coefficients using grouped data.

In the bimodal case eq.(3.5) can be transformed thusly:

$$P(m=1) = \frac{e^{Um1}}{e^{Um1} + e^{Um2}} \quad (3.8)$$

where M = set of alternative modes m=1,2

and Umk = utility of mode k

From eq. (3.8)

$$P(m=1:M) = \frac{1}{1 + e^{Um2 - Um1}} \quad (3.9)$$

This can be written as :

$$\ln\left(\frac{1}{P(m=1:M)}-1\right) = Um2 - Um1 \quad (3.10)$$

If Vm1 = volume of commuters choosing mode m1.

and Vm2 = Volume of commuters choosing mode m2.

$$\text{then } P(m=1:M) = \frac{Vm1}{Vm1 + Vm2} \quad (3.11)$$

Substitution of eq.(3.11) in eq.(3.10) gives:

$$\ln \frac{Vm2}{Vm1} = Um2 - Um1 \quad (3.12)$$

But Um1 can be written as:

$$Um1 = Qm1 + Q1X11 + Q2X21 + \ldots + QnXn1 \quad (3.13)$$

Where Qm1 = mode-specific constant for mode 1

Qk = coefficient of the k^{th} generic variable.

Xk1 = value of the k^{th} generic variable for mode 1.

And Um2 can be written as:

$$Um2 = Qm2 + Q1X12 + Q2X22 + \ldots + QnXn2. \quad (3.14)$$

Where Qm2 = mode specific constant for mode 2

Qk = coefficient of the k^{th} generic variable

Xk2 = value of the k^{th} generic variable for mode 2

Finally defining the L.H.S of eq.(3.12) as Y

Then it can be written as:

$$Y = (Qm2 - Qm1) + Q1(X12 - X11) + Q2(X22 - X21)$$

$$+...+ Qn(Xn2 - Xn1) \qquad (3.15)$$

In some cases the mode-specific term may be composed of a constant and a mode-specific variable.

Equation(3.15) is of the form:

$$Y = A + B1X1 + B2X2 +...BnXn \qquad (3.16)$$

which is the form with which multiple regression can be used to obtain estimates for A,B1,B2..Bn.

Chapter IV

DATA BASE OF MODEL

4.1 GENERAL INTRODUCTION

The model was developed using data collected for the City of Winnipeg origin-destination study done in 1976[43]. In 1962, the City of Winnipeg had done an areawide 100% survey of households. Data collected included information regarding zonal employment, socioeconomic characteristics of the population and the transportation system characteristics.

In 1976 the Streets and Transportation Division of the City of Winnipeg updated these findings with a 20% sample of homes in the metropolitan area.

The population of Winnipeg in 1976 was estimated at over 560,000 people. There were 220,817 registered passenger vehicles and 25,929 commercial vehicles. For collection of the 1976 data, the city was divided into thirty-six zones. Figure 1 shows the configuration of these zones. Zone 36 is the central business district.

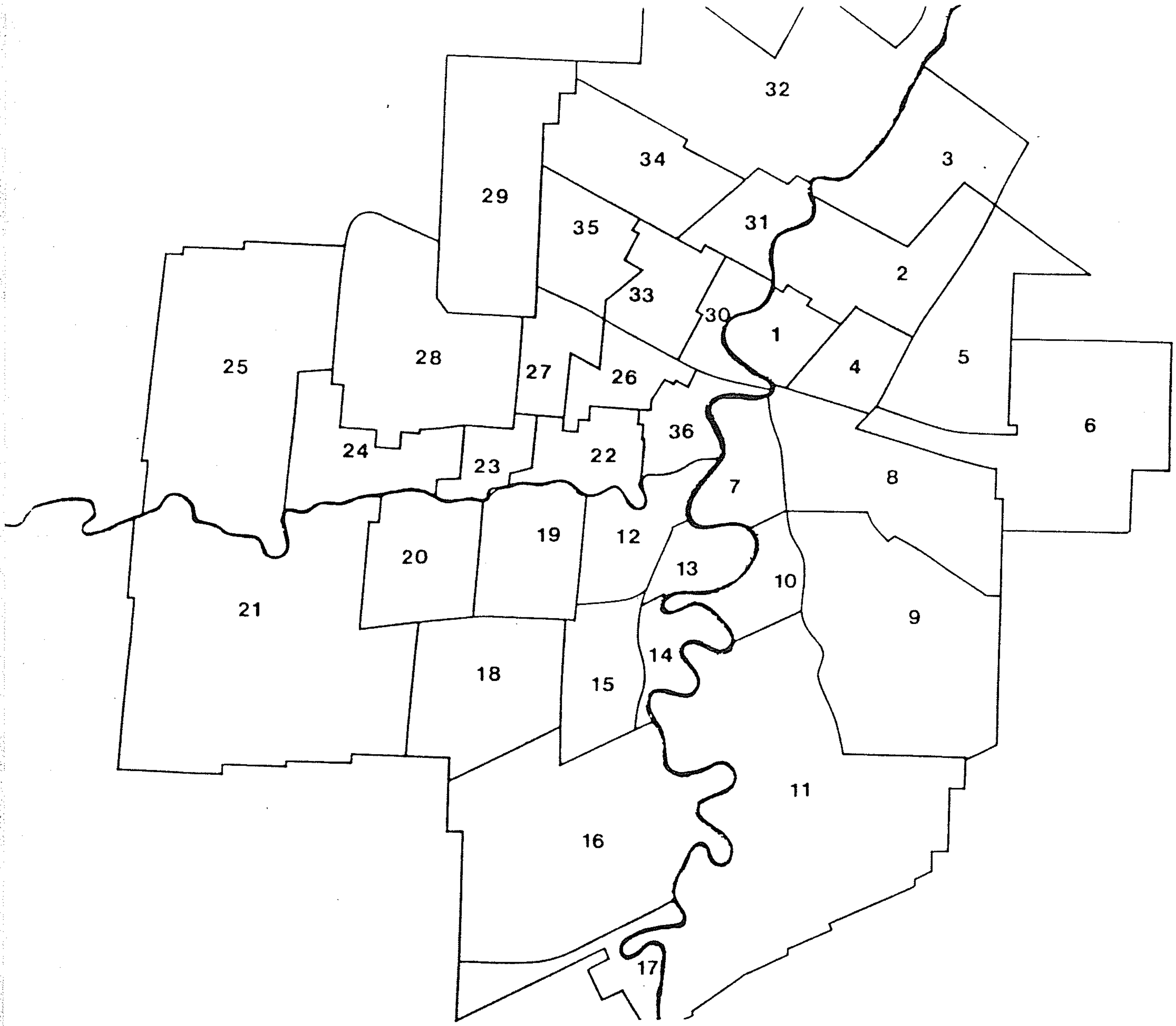

Figure 1: CITY OF WINNIPEG SUPERZONES

The daily and peak-a.m.-hour interzonal trip interchanges were obtained from the 1976 study. These interchanges were given as total, automobile driver and bus passenger trips. It was not possible to accurately separate the shared-ride from the walking passengers, so only auto drivers and bus passengers were used in the development of the model.

The model was developed on these interzonal bus passengers and auto driver trips into the central business district in the peak-a.m.-hour. It should be noted that approximately 21,400 out of 26,500 peak-a.m.-hour trips into this zone were either bus passengers or auto drivers. Thus making an estimate of trips by a third or fourth mode would have introduced unnecessary and unwarranted complexities. Zones from which there were no trips by any mode within the time period were omitted. Those zones from which zero trips emanated by a particular mode, were assigned a value of 0.4 trips for computational purposes. This was done because the calculation of some of the probabilities and statistics required division by the observed number of trips. Division by zero led to suspended functions in the APL computer programs.

4.2 VARIABLES USED IN THE MODELLING

Several variables were considered in the development of the model. These variables were related to both the transportation system as well as the traveller characteristics. It was expected that these variables would accurately represent the penalties involved in making a trip by a particular mode. The variables, which were calculated for each zone, are as follows:

1. In Vehicle Travel Time By Automobile (TA).

This is defined as the time spent travelling in an automobile (in minutes) from the centroid of the origin zone to the centroid of the destination zone. It was obtained from the 1976 O-D study. The zonal centroids were taken to be the centroids of the areas of developed housing in each zone.

2. In Vehicle Travel Time By Bus (TB).

This is defined as the time spent travelling on board a bus (in minutes) from the centroid of an origin zone to the centroid of the destination zone. It was obtained from the 1976 O-D study. The zonal centroids were taken to be the centroids of the areas of developed housing in each zone.

3. Out Of Vehicle Travel Time By Automobile (XA).

This is the total trip time spent not in transit. It is made up of time spent walking from home to the vehicle, time spent in parking the vehicle and time spent in walking from the vehicle to the place of employment.

Rosin[38] estimated an average value of six minutes for XA for work trips to the CBD . However, that estimate was based upon 1966 data. Between 1966 and 1976 the number of vehicles travelling to the CBD in the peak-a.m.-hour decreased from about 12,400 to 10,979. During the same period, the total number of parking spaces available increased by almost one thousand. For this reason, it could be expected that both the parking time and the time spent walking from the parked vehicle to work, would be decreased. Thus an average value of four minutes was used for XA.

4. Out Of Vehicle Travel Time By Bus (XB).

This time(in minutes) is made up of walking time from home to the nearest bus-stop, waiting time, transfer time (if any) and walking time to work from the nearest stop.

The average walking distance to the nearest bus route was taken to be half of the distance between the furthest point of housing development and the route. The walking time was calculated by assuming a rate of twenty minutes per mile. The waiting time was assumed to be half of the headway up to a maximum wait time of six minutes. For headways greater than twelve minutes it was assumed that seasoned travellers would be familiar with bus schedules and would be able to time the buses.

The bus headways were obtained by consulting the relevant schedules. When more than one route to the destination zone ran through an origin zone, the sum of waiting time plus walking tine was averaged over all possible combinations and that value was used. The transfer time was assumed to be one half of the headway of the interconnecting buses up to a maximum headway of twelve minutes. When the headways of interconnecting buses were more than twelve minutes, the tranfer time was obtained in the following manner. The schedules were examined and where, based upon the approximate running times, the schedules were seen to complement each other, (eg.the Dalhousie-St.Norbert feeder and the King's Park Express) a value of six minutes was used. Where there was no evidence of such complementarity, a value of half of the headway of the interconnecting buses was used. Since actual destinations within the destination zone were not known, a value of three minutes was assumed for walking time from the nearest stop to the place of employment.

5. Total Travel Time By Automobile (OTA).

This is the sum of the in-vehicle and the out of vehicle times for the automobile. OTA = TA + XA.

6. Total Travel Time By Bus. (OTB).

This is the sum of the in-vehicle and the out of vehicle times for the bus. OTB = TB + XB.

7. Out Of Vehicle Time By Auto/Distance (XAD).

This is the ratio of the out of vehicle time by automobile and the interzonal distance (in miles). The distance was scaled off a large-scale map of Winnipeg and measured from centroid to centroid along major arterials.

8. Out Of Vehicle Travel Time By Bus/Distance (XBD).

This the ratio of the out of vehicle time by bus to the interzonal centroidal distance. The distance was obtained as before.

9. Out Of Pocket Costs By Automobile (CA).

This is the cost of operating an automobile for a one-way trip from the origin to the destination zone. This cost (in cents) consists of two parts. The first part is dependent upon the distance between the two zones while the other is dependent upon the cost of parking. The first cost, the running cost, can be further divided into cost of fuel and cost of oil for lubrication. Based on data supplied by Mr. Bruce Ball of the Energy Council, the average price of a gallon of gasoline in Manitoba in 1976 was estimated at 77.4 cents per gallon.

No definite data were available for the average fuel consumption of cars driven in Manitoba in 1976. However, it was possible to get a reasonable estimate.

Based on data supplied by R.L.Polk and Co. on the quantities and types of cars sold in Manitoba from 1970 to 1978, and the publication "The Car Mileage Book" published by Energy, Mines and Resources Canada, a fuel economy of eighteen miles per gallon(city driving) was estimated for cars in 1978. Rosin used an estimate of twelve miles per gallon in 1971. A value of fifteen miles per gallon was therefore assumed for 1976.

This value is substantiated by a study done by Orski[29] in which he stated that due to increased vehicle weight, the average fuel economy of cars in the United States had decreased sixteen per cent from 13.9 miles per U.S. gallon in 1962 to 11.6 miles per U.S. gallon in 1973. In terms of Canadian gallons this is a decrease from 16.7 to 14.0 miles per gallon. The cost of gasoline per mile can thus be calculated to be 5.16 cents per mile.

The cost of oil in terms of cents per mile was estimated by assuming ten dollars per oil change and a rate of oil change of one per two thousand miles. This works out to 0.5 cents per mile. The total running costs therefore were estimated at 5.66 cents per mile.

The cost of parking in the CBD for 1976 was obtained from information supplied by operators of parking lots and structures. A value of one dollar(one-way) per day was used.

Thus the total out of pocket cost in cents for an automobile per one way trip is:

CA = 100.00 + 5.66 D

where D = Distance in miles between the origin and destination zones.

10. Out of Pocket Cost By Bus (CB).

The out of pocket cost by bus is the fare that was charged to commuters. This was twenty-five cents per one-way trip in 1976.

11. Out Of Pocket Cost By Auto/Income (CAI).

This is the ratio of the out of pocket cost by automobile to the average zonal income per dwelling unit.

12. Out Of Pocket Cost By Bus/Income (CBI).

This is the ratio of the out of pocket cost by bus, to the average zonal income per dwelling unit.

13. Registered Automobiles Per Dwelling Unit (CO).

14. Registered Automobiles Per Person (CPP).

15. Out Of Pocket Cost Per Passenger By Auto (CAP).

This is the out of pocket cost by auto divided by the number of passengers per auto. The number of passengers per car going into the central business district in the peak-am-hour was assumed to be the same as the number leaving the origin zone in the same time period.

16. Out Of Pocket Cost Per Passenger By Auto/Income (CAPI).

This is the ratio of the out of pocket cost per passenger of auto to the average zonal income per dwelling unit.

The form of Equation(3.16) means that the dependent variable would be regressed against differences between these aforementioned variables. Because of this, it is convenient to define new variables in terms of differences between auto and bus variables where necessary. These variables can be defined symbolically as:

17. TATB = TA - TB

18. XAXB = XA - XB

19. OTAB = OTA - OTB

20. XABD = XAD - XBD

21. CAB = CA - CB

22. CABI = CAI - CBI

23. CABP = CAP - CB

24. CABPI = CAPI - CBI

Variables CO and CPP were expected to be related to the mode specific variable for automobile, i.e., a variable that would help to explain the attractiveness of automobile over bus, all other things being equal.

Chapter V

DEVELOPMENT OF THE MODEL

5.1 GENERAL INTRODUCTION

The model was developed using three functions copied from the APL public libraries. These were "REG", "STREG" and "CM". The first two functions calculated, by multiple and stepwise regression respectively, the best fitting least-squares straight line.

Equation (3.16) can be modified to be of the form:

$$Yi = A + B1X1i + B2X2i + B3X3i + \ldots + BnXni$$

where Yi is the observed dependent variable of the i observation and X1i,X2i...Xni are the observed independent variables of the i observation. Then the least squares method estimates a value of the dependent variable Yi^1 such that $(Yi^1-Yi)^2$ is minimized. The summation is over n values of Yi where n is the number of observations. Table 1 shows the simple correlation coefficients of the variables. Generally, high correlation between the dependent variable and an independent variable implies that a significant amount of variance in the former can be explained by variance in the latter. Thus we see that the dependent variable is highly correlated to CO and OTAB and, to a lesser extent, CPP and XAXB. This gives an idea of what variables would be expected to explain a large amount of the variance in the dependent variable. However, once one variable goes into the regression, the relationships change. Partial correlation coefficients give an indication at each stage of the regression of the correlation of each variable not yet in the regression.

The table of simple correlation can also be used to see the correlation between two independent varibles. The higher the numerical value of the simple correlation coefficient between two independent variables used in the regression, the greater will be the standard error associated with their estimated regression coefficients. In some cases, however, the apparent correlation between two variables may be due to the correlation of each variable with a third variable.

One effect of colinearity is that two variables, say X1 and X2, may be highly positively correlated with a dependent variable and with each other, but the net effect of X2 taking X1 into account may be negative.

Several combinations of variables were used. They were examined by statistical tests and by comparison of certain calculated statistics. In all cases the first run was done considering all zones in the study area. It was felt, however, that certain zones, while not contributing much in terms of number of trips, were nevertheless adversely affecting the accuracy of the regression by acting as

TABLE 1

TABLE OF SIMPLE CORRELATION COEFFICIENTS

	TATB	XAXB	OTAB	XABD	CAB	CABI
TATB	1.0000					
XAXB	0.3862	1.0000				
OTAB	0.8407	0.8242	1.0000			
XABD	-0.5023	0.3225	-0.1191	1.0000		
CAB	-0.8678	-0.4705	-0.8090	0.4825	1.0000	
CABI	-0.0259	-0.0795	-0.0625	-0.1537	0.0199	1.0000
CABP	-0.7755	-0.6062	-0.8320	0.2704	0.8834	0.1446
CABPI	-0.2310	-0.3785	-0.3641	-0.1809	0.2381	0.8560
CO	-0.6783	-0.6812	-0.8164	0.2143	0.6848	-0.3305
CPP	-0.5214	-0.4378	-0.5771	0.3172	0.4980	-0.5256
Y	-0.6238	-0.7116	-0.8008	-0.0613	0.5265	-0.1946

TABLE 1 CONTINUED,

	CABP	CABPI	CO	CPP	Y
CABP	1.0000				
CABPI	0.5252	1.0000			
CO	0.6656	0.0040	1.0000		
CPP	0.4996	-0.1849	-0.3305	1.0000	
Y	0.6731	0.2134	0.8411	0.7076	1.0000

outliers. The zones that were particularly suspect were large zones with low population densities and low trip volumes. The combination of large size and low population densities increased the probability of error in the estimation of travel times and interzonal distances.

Zones 5, 8, 17, 18, 20, 29 and 32 all have relatively low population densities and trip volumes of less than 100 bus passengers in the time period under consideration. It was felt that the error that could be introduced by the inclusion of these zones could not be justified by the number of trips involved. Therefore these zones were examined and those with relatively large values of $(Yi^{1} - Yi)$ were not used in the subsequent regression.

Aggregate models can be extremely sensitive to zoning. The measurement of trip distances and travel times for short distances would carry a high percentage error. This is especially true for intrazonal trips, i.e., trips that begin and end in the same zone. For this reason zone 36, the destination zone, was not used in the regression analysis.

However, all of the zones were used in the analysis of the models developed with the estimated regression coefficients.

5.2 STATISTICAL TESTS ON THE REGRESSION EQUATIONS

The statistical tests used in the analysis of the regression equations were the F test and the T test. The Multiple Correlation Coefficient, a calculated statistic, was also used.

5.2.1 THE F TEST

This is a test of the total regression, i.e., for all k independent variables considered simultaneously. If b1,b2...bn are sample estimators for the population parameters B1, B2...Bn, then the Null Hypothesis for this test is:

$$Ho : B1 = B2 = \ldots = Bk = 0$$

$$vs.\ Ha : Bi \neq 0 \text{ for some } i$$

Thus testing that the variables do not explain a signifant proportion of the variation in Y. F is the ratio of the Mean Square Regression to the Mean Square Residual. This is compared with a tabled value and the null hypothesis is rejected if the calculated F is greater than the tabled F (1-a,k,n-k-1).

Where a = level of significance (usually 0.05)

k = number of variables used in the regression.

n = number of observations used in the regression.

5.2.2 THE T TEST

This is a test on the significance of each regression coefficient considered individually. It tests whether or not the estimated coefficients are significantly different from zero. The Null Hypothesis is:

Ho : Bi = 0

vs Ha : Bi ≠ 0

T is the ratio of the estimated regression coefficient to its standard error. Since the test is done on all k coefficients and since it is a two-tailed test, the calculated T value is compared with a tabled T(1-a/2k,n-k-1).

Where a = overall level of significance (usually 0.05)

k = number of variables considered

n = number of observations used.

We reject the null hypothesis if the calculated T value is greater numerically than the tabled T value.

Another way of looking at it would be to construct a 100(1-a)% interval for Bi as bi ± T(1-a/2k,n-k-1) x Standard Error of bi.

5.2.3 THE MULTIPLE CORRELATION COEFFICIENT

The Multiple Correlation Coefficient R^2, measures the total variability in the dependent variable which can be explained by variability in the independent variables. Thus it is the ratio of explained variance to total variance. R^2 lies between zero and one, with higher values indicating a higher percentage of variance explained.

5.3 STATISTICAL TESTS ON THE MODELS

The models obtained by using the regression coefficients were tested for their ability to accurately estimate the volume of bus travellers. The statistics used for this were The Standard Deviation and the Paired T test.

5.3.1 THE STANDARD DEVIATION

The Standard Deviation, SD,can be defined mathematically as:

$$SD = \sqrt{\frac{\Sigma Di^2}{n}}$$

Where Di = deviation between estimated and actual volume of bus travellers from zone i.

n = the number of zones considered.

The smaller the SD value, the more accurately the model predicts modal choice. Assuming normality of distribution of the actual volumes, then approximately 66.7% of the deviations between the actual and the estimated volumes would be less than or equal to one SD. In other words two-thirds of the time, the estimated volumes would be within one standard deviation of the actual volume of bus passengers from a zone.

5.3.2 THE PAIRED T TEST

The Paired T Test compares corresponding observed and estimated volumes. The Null Hypothesis is that the volumes are samples of two populations whose mean difference is zero, i.e., they belong to the same population. If ud is the mean deviation between the populations then:

Ho : ud = 0

vs Ha : ud ≠ 0

The statistic MD , which is the mean of the differences between the estimated and actual volumes, is an estimator for ud. Thus the calculated T value is:

$$T = \frac{MD}{SD \times nr}$$

Where nr is the root of n, the number of zones compared. This value is compared with a tabled T(1- a/2,n-1) at the a=0.05 level of significance. The null hypothesis is rejected if the calculated T is greater than the tabled T value.

5.4 ANALYSIS OF REGRESSION EQUATIONS

Five equations were selected from the several combinations tried. The others were rejected either due to too small R^2 values, insignificant T values, or illogical signs for the regression coefficients (e.g., the coefficient of travel time should have a negative value).

5.4.1 ANALYSIS OF REGRESSION EQUATION 1

This equation combines the variables TATB, XAXB, CAB and CPP. Table 2 gives the regression coefficients with their standard errors and computed T values.

TABLE 2

ANALYSIS OF EQUATION 1

VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	COMPUTED T-VALUE
TATB	-0.08986	0.01674	-5.369
XAXB	-0.11958	0.01367	-8.746
CAB	-0.02551	0.00721	-3.537
CPP	4.15668	0.87707	4.739
CONSTANT TERM = -1.07741			

The zones that were omitted for the reasons previously mentioned were zones 5, 23, 29, 32 and 36. The number of trips associated with these zone were 9, 102, 57, 13 and 190 respectively. Of these only zone 23 and zone 36 have more than 100 trips. Thus the omission of these zones would not greatly detract from the predictive power of the model. Thus the regression was done on thirty zones.

The computed F value for the regression was 85.423. The tabled F value at the a = 0.05 level F(.95,4,25) = 2.78. Thus the null hypothesis of section 5.2.1 is rejected and we can say that the regression does explain a significant proportion of the variation in Y.

The tabled T value for k = 4 and n = 30 at the a = 0.05 level is 2.787. All the computed T values for the regression coefficients are numerically larger than this value. We can, therefore, reject the null hypothesis of section 5.2.2 in all cases and state that all the regression coefficients are significantly different from zero. Alternatively we can construct 95% confidence intervals for the regression coefficients:

1. 95% confidence interval for regression coefficient of TATB.

Interval = $-0.08986 \pm 2.787 \times 0.01674$
= -0.13651 to -0.04321

2. 95% confidence interval for regression coefficient of XAXB.

Interval = $-0.11958 \pm 2.787 \times 0.01367$
= -0.15768 to -0.08148

3. 95% confidence interval for the regression coefficient of CAB.

Interval = $-0.02551 \pm 2.787 \times 0.00721$

= -0.04560 to -0.00542

4. 95% confidence interval for the regression coefficient of CPP.

Interval = $4.15668 \pm 2.787 \times 0.87707$

= 6.60107 to 1.71229

Since none of these intervals cover zero we reject the null hypothesis of section 5.2.2 .

The Multiple Correlation Coefficient for the regression is found to be .932 . That means that 93.2% of the variation in Y is explained by variation in the independent variables. This is a very acceptable R^2 value and indicates a good fit of data to the model.

5.4.2 ANALYSIS OF REGRESSION EQUATION 2

This equation combines the variables TATB, XAXB and CABI. Table 3 gives the regression coefficients with their standard errors and computed T values.

The zones that were omitted for the reasons outlined in section 5.1 were zones 5, 8, 17, 18, 20, 32 and 36. Zone 13 was also found to adversely affect the accuracy of the regression. There were 220 trips from zone 13. However, it

TABLE 3

ANALYSIS OF EQUATION 2

VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	COMPUTED T-VALUE
TATB	-0.05346	0.01093	-4.890
XAXB	-0.08267	0.01927	-4.291
CABI	-327.35436	78.6661	-4.161
CONSTANT TERM = 0.86064			

was found that omission of zone 13, while improving the regression and the overall predictive power of the subsequent model, did not cause any great change in the estimated trips for zone 13. It was therefore not used in the regression. The regression was thus done on twenty-seven zones.

The computed F value for the regression was 38.317. The tabled F value at the a= 0.05 level, F(.95,3,23) = 3.01. The computed F value is very much greater than the tabled F value so the null hypothesis of section 5.2.1 is rejected. That is, the regression does explain a significant proportion of the variation in Y.

The tabled T value for k=3 and n = 27 at the a = 0.05 level is 2.500. The computed T values of all the regression coefficients are numerically greater than this tabled T value. Therefore, we can reject the null hypothesis of section 5.2.2 and say that all the regression coefficients are significantly different from zero. Alternatively, we can construct 95% confidence intervals for the regression coefficients:

1. 95% Confidence Interval for the Regression Coefficient of TATB.

Interval = $-0.05346 \pm 2.500 \times 0.01093$

= -0.08079 to -0.02614

2. 95% Confidence Interval for the Regression Coefficient of XAXB.

Interval = $-0.08267 \pm 2.500 \times 0.01927$

= -0.13085 to -0.03450

3. 95% Confidence Interval for the Regression Coefficient of CABI.

Interval = $-327.35436 \pm 2.500 \times 78.66611$

= -524.01964 to -130.68909

Since none of these intervals cover zero we can reject the null hypothesis of section 5.2.1 in all cases.

The Multiple Correlation Coefficient for the regression was .833. This means that 83.3% of the variation in Y is explained by variation in the independent variables. This is an acceptable amount of explained variance and indicates a good fit of data to the equation.

5.4.3 ANALYSIS OF REGRESSION EQUATION 3

This equation uses the dependent variables TATB, XAXB, CAB and CO. Table 4 gives the regression coefficients with their standard errors and computed T values.

TABLE 4

ANALYSIS OF EQUATION 3

VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	COMPUTED T-VALUE
TATB	-0.07347	0.02019	-3.639
XAXB	-0.09323	0.01937	-4.814
CAB	-0.02816	0.00858	-3.280
CO	1.29004	0.28456	4.533
CONSTANT TERM = -0.32036			

The zones that were omitted from the regression for reasons outlined in section 5.1 were zones 5,8, and 36. The regression was therefore done on thirty-two zones.

The computed F value for the regression was 62.695. The tabled F value at the a = 0.05 level, F(.95,4,27) is 2.69. This the computed F value is greater than the tabled F value and the null hypothesis of section 5.2.1 is rejected. The regression does explain a significant proportion of the variation in Y.

The tabled T value for k = 4 and n = 32 at the a = 0.05 level is 2.771. The computed T values of all the regression coefficients are numerically greater than this value. Therefore we can reject the null hypothesis of section 5.2.2 and say that all the regression coefficients are significantly different from zero. Alternatively we can construct 95% confidence intervals for the regression coefficients:

1. 95% Confidence Interval for the Regression Coefficient of TATB.

$$\text{Interval} = -0.07347 \pm 2.771 \times 0.02019$$
$$= -0.12942 \text{ to } -0.01752$$

2. 95% Confidence Interval for the Regression Coefficient of XAXB.

$$\text{Interval} = -0.09323 \pm 2.771 \times 0.01937$$
$$= -0.14690 \text{ to } -0.03956$$

3. 95% Confidence Interval for the Regression Coefficient of CAB.

Interval = $-0.02816 \pm 2.771 \times 0.00858$

= -0.05194 to -0.00438

4. 95% Confidence Interval for the Regression Coefficient of CO.

Interval = $1.29003 \pm 2.771 \times 0.28456$

= 2.07855 to 0.50151

Since none of these intervals cover zero, we can reject the null hypothesis of section 5.2.2 in all cases.

The Multiple Correlation Coefficient of the regression is .903. This means that 90.3% of the variation in Y is explained by variation in the independent variables. This is a very acceptable amount of explained variance and indicates a good fit of data to the equation.

5.4.4 ANALYSIS OF REGRESSION EQUATION 4

This equation uses the variables TATB, XAXB and CAB as dependent variables. Table 5 gives the regression coefficients along with their standard errors and computed T values.

The zones that were omitted for reasons discussed in section 5.1 were zones 5, 17, 28, 29, 32 and 36. The regression was done on twenty-nine zones.

TABLE 5

ANALYSIS OF EQUATION 4

VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	COMPUTED T-VALUE
TATB	-0.11980	0.02572	-4.657
XAXB	-0.17784	0.02756	-6.452
CAB	-0.03526	0.01262	-2.794
CONSTANT TERM = 0.77545			

The computed F value for the regression was 30.780. The tabled F value at the a = 0.05 level, F(.95,3,25) is 3.01. Thus the tabled F value is much less than the computed F value and we can reject the null hypothesis of section 5.2.1. In other words, the regression does explain a significant amount of the variation in Y.

The tabled T value for k = 3 and n = 29 at the a = 0.05 level is 2.485. The computed T values of all the regression coefficients are numerically greater than this tabled T value. We can therefore reject the null hypothesis of section 5.2.2 and say that all of the regression coefficients are significantly different from zero. Alternatively we can construct 95% confidence intervals for the regression coefficients:

1. 95% Confidence Interval for the Regression Coefficient of TATB.

Interval = $-0.11980 \pm 2.485 \times 0.02572$

$= -0.18371$ to -0.05589

2. 95% Confidence Interval for the Regression Coefficient of XAXB.

Interval = $-0.17784 \pm 2.485 \times 0.02756$

$= -0.24633$ to -0.10935

3. 95% Confidence Interval for the Regression Coefficient of CAB.

Interval = $-0.03526 \pm 2.485 \times 0.01261$

$= -0.06660$ to -0.00392

Since none of the confidence intervals cover zero we can reject the null hypothesis of section 5.2.2 in all cases.

The Multiple Correlation Coefficient for the regression is .787. This means that 78.7% of the variation in Y is explained by the variation in the independent variables. This is a fairly acceptable amount of explained variance.

5.4.5 ANALYSIS OF REGRESSION EQUATION 5

This equation uses dependent variables TATB, XAXB and CABPI. Table 6 gives the regression coefficients along with their standard errors and computed T values.

TABLE 6

ANALYSIS OF EOUATION 5

VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	COMPUTED T-VALUE
TATB	-0.06276	0.01158	-5.421
XAXB	-0.12055	0.02026	-5.950
CABPI	-227.61770	63.66251	-3.575
CONSTANT TERM = -0.51035			

The zones that were omitted for reasons outlined in section 5.1 were zones 5, 18, 32 and 36. It was also found that zones 13 and 23 which contributed 220 and 102 trips respectively were acting as outliers and impairing the accuracy of the regression. Omission of these zones improved both the regression equation and the estimating power of the subsequent model without adversely affecting, to any great extent, the estimates for the omitted zones. The regression was, therefore, done on twenty-nine zones.

The computed F value for the regression was 40.936. The tabled F value at the a = 0.05 level, F(.95,3,25) is 3.01. We can therefore reject the null hypothesis of section 5.2.1 and state that the regression does explain a significant amount of the variation in Y.

The tabled T value for k = 3 and n = 29 at the a = 0.05 level is 2.485. The computed T values of all the regression coefficients are numerically greater than the tabled T value. Therefore we reject the null hypothesis of section 5.2.2 in all cases and state that the regression coefficients are all significantly different from zero. Alternatively we can construct 95% confidence intervals for the regression coefficients:

1. 95% Confidence Interval for the Regression Coefficient of TATE.

$$\text{Interval} = -0.06276 \pm 2.485 \times 0.01158$$
$$= -0.09154 \text{ to } -0.03398$$

2. 95% Confidence Interval for the Regression Coefficient of XAXB.

$$\text{Interval} = -0.12055 \pm 2.485 \times 0.02026$$
$$= -0.17090 \text{ to } -0.07020$$

3. 95% Confidence Interval for the Regression Coefficient of CABPI.

$$\text{Interval} = -227.61770 \pm 2.485 \times 63.66251$$
$$= -385.81904 \text{ to } -69.41636$$

Since none of these intervals cover zero, we can reject the null hypothesis of section 5.2.2.in all cases.

The Multiple Correlation Coefficient for the regression was .831. This means that 83.1% of the variance in Y can be explained by variation in the dependent variables. This is an acceptable amount of explained variance and implies a good fit of data to the equation.

5.4.6 SUMMARY OF REGRESSION EQUATION ANALYSES

So far, the statistical strengths of the regression equations have been discussed. In terms of computed F and T values they are all significant explainers of the variation in the dependent variable. However, on comparing the R^2 values, equations 1 and 2 appear to be superior to equation 3 and 5 which appear to be superior to equation 4.

In fact equations 1 and 2 can be thought of as more advanced versions of equation 4, i.e., versions in which some more variance in Y is explained by the introduction of another important independent variable.

Equations 3 and 5, while not having as high R^2 values as equations 1 and 2, nevertheless have an acceptable proportion of explained variance. A point of interest is the signs of the constant terms of these two equations. The only difference between the equations is that in equation 3, the cost of an automobile trip is assumed to be borne by the driver alone, while in equation 5 it is assumed to be equally shared among all passengers. This difference, however, has led to a difference in sign of the constant term. The significance of this will be discussed later.

5.5 ANALYSIS OF THE ESTIMATION BY THE MODELS

The "acid test" of any model must be its ability to simulate behavior; in this instance, its ability to estimate trip volumes that are close or equal to the observed trip volumes. It must be remembered that the dependent variable of the regression equations is a transformation of the percentage of bus passengers. This transformation is not linear. What is of importance is how accurately the estimated coefficients calculate the utilities of the modes and the subsequent modal shares.

Henceforth, the utility of the bus and of the automobile will be designated as Ub and Ua respectively.

5.5.1 ANALYSIS OF MODEL 1

Using the estimated coefficients from the regression equation for this model, the utility functions can be constructed:

$$Ub = -0.08986TB - 0.11958XB - 0.02551CB$$

$$Ua = -1.07741 - 0.08986TA - 0.11958XA - 0.02551CA + 4.15668CPP$$

Assuming that the mode-specific constant for the bus is zero, then the mode-specific term for automobile is -1.07741 + 4.15668CPP. The coefficient for excess time is numerically larger than that of in vehicle travel time. This means that one unit of excess time contributes more to disutility than one unit of in-vehicle travel time. This is in agreement with intuitive thinking that travellers would prefer to spend their time in a moving vehicle rather than walking or waiting. Table 7 shows the mean values of the main variables used in the study. Substitution of these values in the above equations gives:

Ub = -2.123 -1.517 -0.637

Ua = -1.077 - 0.988 - 0.478 - 3.189 + 1.650

TABLE 7

MEAN VALUES OF THE VARIABLES

VARIABLE	MEAN
TA	11.00000
XA	4.00000
TB	23.62857
XB	12.68800
OTA	15.00000
OTB	36.31657
XAD	1.22503
XBD	3.45812
CA	125.02043
CB	25.00000
CAI	0.00829
CBI	0.00168
CO	1.26542
CPP	0.39694
CAP	104.19248
CAPI	0.00690
D	4.42057
I	15542.40000

It should be noted that the major contributors to disutility of the bus and the automobile are in-vehicle travel time and cost respectively. The automobile mode specific term has a value of 1.650 - 1.077 = .573. This is positive, implying that, all other corresponding variables being equal, the automobile would be more attractive to travellers than the bus. This is to be expected since the automobile offers greater privacy, mobility and comfort in addition to the perceived status of owning and operating an automobile. Table 8 displays the observed and estimated bus travellers to the destination zone as well as the deviations.

The standard deviation, SD, was found to be 38.59. Thus, approximately two-thirds of all estimated zonal volumes would lie within 38.59 trips of the observed volumes.

The tabled T value for the Paired T test with n = 35 at the a = 0.05 level is 2.030. The computed T value for this model is -0.4096. Since this computed T value is numerically less than the tabled T value, we fail to reject the null hypothesis of section 5.3.2. There is therefore no significant difference between the observed and the estimated zonal volumes. In fact even at the a = 0.5 level, we would fail to reject the null hypothesis since the tabled T value at that level is 0.682 and $0.682 > |-0.4096|$.

TABLE 8

RESULTS OF MODEL 1

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	477	406.0	71.0
2	560	639.5	-79.5
3	528	428.5	99.5
4	377	417.4	-40.4
5	9	2.8	6.2
6	370	395.8	-25.8
7	490	445.2	44.8
8	23	22.7	0.3
9	309	370.2	-61.2
10	263	244.9	18.1
11	601	541.3	59.7
12	1005	997.9	7.1
13	220	255.6	-35.6
14	102	100.1	1.9
15	109	105.4	3.6
16	310	277.4	32.6
17	42	42.4	-0.4
18	0.4	0.4	0.0
19	286	344.2	-58.2
20	15	18.1	-3.1
21	134	129.6	4.4
22	829	837.1	-8.1
23	102	133.8	-31.8
24	542	568.6	-26.6
25	528	534.5	-6.5
26	222	252.7	-30.7
27	181	166.5	14.5
29	57	49.6	7.4
30	321	317.8	3.2
31	259	279.3	-20.3
32	13	27.7	-14.7
33	593	538.7	54.3
34	290	325.7	-35.7
35	154	138.6	15.4
36	190	248.8	-58.8

5.5.2 ANALYSIS OF MODEL 2

Using the estimated regression coefficients for this model the utility functions can be defined thusly:

Ub = -0.05346TB -0.08267XB -327.35436CBI

Ua = .86064 -0.05346TA -0.08267XA -327.35436CAI

Assuming that the mode-specific constant for the bus is zero, then the mode-specific constant for the automobile is .86064. This is positive as should be expected for the reasons outlined earlier. It should be noted that in this model, as in the previous one, the coefficient of out of vehicle time is numerically greater than that of in-vehicle time.

Substitution of the average values of the variables into the utility functions gives:

Ub = - 1.263 - 1.049 - 0.549

Ua = 0.861 - 0.588 - 0.330 - 2.714

As in the previous model, the major contributors to disutility of the bus and the automobile are in-vehicle travel time and cost (in this case cost/income) respectively. Table 9 shows the observed and estimated zonal bus passengers and the corresponding deviations.

The standard deviation, SD, was found to be 36.56. Thus, approximately two-thirds of all estimated volumes would lie between plus or minus 36.56 trips of the actual volume of bus passengers.

TABLE 9

RESULTS OF MODEL 2

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	477	429.7	47.3
2	560	592.0	-32.0
3	528	402.1	125.9
4	377	405.9	-28.9
5	9	4.9	4.1
6	370	324.9	45.1
7	490	426.9	63.1
8	23	26.9	-3.9
9	309	336.3	-27.3
10	263	272.8	-9.8
11	601	597.1	3.9
12	1005	1004.3	0.7
13	220	269.5	-49.5
14	102	103.8	-1.8
15	109	122.7	-13.7
16	310	278.4	31.6
17	42	55.5	-13.5
18	0.4	1.8	-1.4
19	286	344.5	-58.5
20	15	17.5	-2.5
21	134	149.0	-15.0
22	829	831.2	-2.2
23	102	118.3	-16.3
24	542	505.5	36.5
25	528	544.5	-16.5
26	222	256.1	-34.1
27	181	193.7	-12.7
29	57	55.7	1.3
30	321	311.1	9.9
31	259	306.7	-47.7
32	13	38.9	-25.9
33	593	536.0	57.0
34	290	295.7	-5.7
35	154	137.9	16.1
36	190	226.4	-36.4

The computed T value for the Paired T test is -0.0601. The tabled T value for n = 35 at the a = 0.05 level is 2.030. Thus the null hypothesis of section 5.3.2 is not rejected. In fact we fail to reject the null hypothesis even at the a = 0.5 level since the tabled T value at this level is 0.682 which is numerically greater than -0.0601.

5.5.3 ANALYSIS OF MODEL 3

Using the estimated regression coefficients the utility functions can be constructed thusly:

Ub = -0.07347TB -0.09323XB -0.02816CB

Ua = -0.32031 -0.07237TA -0.09323XA -0.02816CA + 1.29004CO

Assuming that the mode-specific constant for the bus is zero, then the mode-specific term for the automobile is -0.32031 + 1.29004CO . As in the previous two models, the coefficient of out of vehicle time is numerically greater than that of in-vehicle travel time. Substitution of the mean values of the variables into the above functions gives:

Ub = - 1.736 - 1.183 - 0.704

Ua = - 0.320 - 0.808 - 0.373 - 3.520 + 1.632

The results are similar to those of Model 1 . The major contributors to the disutilities are the same variables. Also, the resultant automobile mode-specific term is -0.320 + 1.632 = 1.212 . This is again positive as is expected.

Table 10 gives the observed and estimated volume of bus travellers as well as the corresponding deviations.

The standard deviation was found to be 36.45. Thus approximately two-thirds of estimated volumes would be less than 36.45 trips away from the actual volume of bus travellers.

The computed T value for this model is -0.3382. This is numerically less than the tabled T value for n = 35 at the a = 0.05 level which is 2.030. Therefore, we fail to reject the null hypothesis of section 5.3.2. In fact this value is numerically less than the tabled T value at the a = 0.5 level. We can state that there is no significant difference between the estimated and the observed volumes.

5.5.4 ANALYSIS OF MODEL 4

Using the estimated regression coefficients the utility functions can be constructed thusly:

Ub = -0.11980TB -0.17784XB -0.03526CB

Ua = 0.77545 -0.11980TA -0.17784XA -0.03526CA

Assuming that the mode-specific constant for the bus is zero, then the mode-specific constant for automobile is 0.77545. This is positive as should be expected. The coefficient for out of vehicle travel time is numerically greater than that for in-vehicle travel time. This is

TABLE 10

RESULTS OF MODEL 3

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	477	416.2	60.8
2	560	602.9	-42.9
3	528	454.7	73.3
4	377	380.8	-3.8
5	9	2.9	6.1
6	370	365.7	4.3
7	490	433.2	56.8
8	23	21.1	1.9
9	309	321.3	-12.3
10	263	261.5	1.5
11	601	570.4	30.6
12	1005	1068.2	-63.2
13	220	239.9	-19.9
14	102	109.7	-7.7
15	109	104.8	4.2
16	310	320.5	-10.5
17	42	40.9	1.1
18	0.4	0.4	0.0
19	286	366.8	-80.8
20	15	11.7	3.3
21	134	127.5	6.5
22	829	879.8	-50.8
23	102	140.2	-38.2
24	542	596.5	-54.5
25	528	498.6	29.4
26	222	235.2	-13.2
27	181	171.5	9.5
29	57	42.8	14.2
30	321	303.0	18.0
31	259	285.5	-26.5
32	13	21.1	-8.1
33	593	524.4	68.6
34	290	276.3	13.7
35	154	134.7	19.3
36	190	253.4	-63.4

similar to the results for the previous models. Substitution of the mean values of the variables into the utility functions gives:

Ub = - 2.830 - 2.257 - 0.882

Ua = 0.775 - 1.318 - 0.711 - 4.408

In this case, as in all the previous cases, the major contributors to disutility of bus and automobile are in-vehicle travel time and cost respectively. Table 11 gives the observed and estimated zonal bus passengers along with their corresponding deviations.

The standard deviation, SD, was found to be 44.63. Thus, approximately two-thirds of the time, an estimated volume should be within 44.63 trips of the actual volume of bus passengers.

The computed T value is -0.2207. This is numerically less than the tabled T value at the a = 0.05 level, which is 2.030. Thus we fail to reject the null hypothesis of section 5.3.2 In fact, the computed T value is numerically less than the tabled T value at the a = 0.5 level so we would still fail to reject the null hypothesis. We can therefore state that the estimated and observed volumes of bus passengers are not significantly different.

TABLE 11

RESULTS OF MODEL 4

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	477	404.5	72.5
2	560	616.9	-56.9
3	528	443.0	85.0
4	377	388.8	-11.8
5	9	1.1	7.9
6	370	382.8	-12.8
7	490	441.8	48.2
8	23	21.9	1.1
9	309	358.2	-49.2
10	263	258.5	4.5
11	601	561.6	39.4
12	1005	1011.5	-6.5
13	220	264.9	-44.9
14	102	104.5	-2.5
15	109	102.2	6.8
16	310	314.0	-4.0
17	42	38.0	4.0
18	0.4	0.2	0.2
19	286	390.9	-104.9
20	15	21.8	-6.8
21	134	90.8	43.2
22	829	906.6	-77.6
23	102	131.8	-29.8
24	542	617.3	-75.3
25	528	538.4	-10.4
26	222	222.7	-0.7
27	181	188.3	-7.3
29	57	34.8	22.1
30	321	282.2	38.8
31	259	281.9	-22.9
32	13	22.2	-9.3
33	593	479.5	113.5
34	290	305.4	-15.4
35	154	122.2	31.8
36	190	218.4	-28.4

5.5.5 ANALYSIS OF MODEL 5

Using the estimated regression coefficients the utility functions for this model can be constructed thusly:

Ub = -0.06276TB -0.12055XB -227.61770CB

Ua = -0.51034 -0.06276TA -0.12055XA -277.61770CAP

Assuming that the mode-specific constant for the bus is zero, then the mode specific constant for the automobile is -0.51034. This is negative and implies that, if all corresponding variables were equal, then the bus would be more attractive than the automobile. This is just the opposite to what all the previous models imply. For reasons expressed in section 5.5.1, this term should be positive. There is no likely explanation for the mode-specific variable for the bus being greater than that of the automobile. The model is intuitively incorrect and will, therefore, be excluded from further analysis.

5.5.6 SUMMARY OF MODEL ANALYSES

Examination of the first four models shows that Models 1, 2 and 3 are clearly superior to Model 4 in terms of the value of SD. Furthermore, while Models 1 and 3 include the income of travellers indirectly in terms of automobile ownership, and Model 2 includes it directly, Model 4 does not include it at all. It is expected that any model that attempts to simulate consumer behavior must consider the income of the consumer. The total exclusion of this important variable may account for the comparatively poor predictive ability of Model 4.

There is nothing to be gained by examining Model 4 any further since the other models not only fit the data as well as or better than it (in terms of R^2 values), but they also give better estimates of the zonal volumes of bus travellers.

We are now left with three models. Of these Model 3 seems to be best in terms of R^2 values and in terms of its predictive ability (low SD value). Model 1 has a very slightly lower R^2 value but only a slighter higher SD value. While the R^2 value of Model 2 is somewhat lower than that of the other two models, the SD value is in fact lower than that of Model 1 and just a shade higher than that of Model 3. Part of the difference in R^2 value may be due to the fact that Model 1 regresses on only three independent variables. Model 1 also has the advantage of including the income of the commuter directly in the utility functions.

All three models, therefore, have their strengths and their weaknesses and it is difficult to choose which one is definitely the best. In fact, making such a choice is unnecessary.They can all be used to estimate modal split depending on the type of data available. The models can also be used to check the validity of data collected for modal split modelling.

5.6 SENSITIVITY ANALYSIS

It is important to know the sensitivity of the models, i.e., how much the estimated volumes change with changes in the independent variables. This is done in two ways. Firstly, the elasticity of the bus volumes with respect to the independent variables is calculated for all zones. Elasticity of V with respect to X is the percentage change in V for a one per cent change in X, all other variables being held constant. Secondly, graphs of %Bus Modal Share vs. changes in the independent variables are drawn.

The relative magnitude of elasticity is often important. When the elasticity of V is numerically less than one, then V is said to be inelastic with respect to X.

5.6.1 SENSITIVITY ANALYSIS OF MODEL 1

Table 12 gives the percentage change in the zonal volume of bus passengers for a one per cent increase of the stated variables.

The total volume is inelastic with respect to out of vehicle travel time,car ownership per person and just barely so for in vehicle travel time. It is elastic with respect to out of pocket cost for automobile but this is mainly due to the parking cost. In fact the volume of bus passengers is less elastic with respect to the price of oil and gas than to any other factor considered.

TABLE 12

ELASTICITY OF MODEL 1

ZONE OF ORIGIN	TB	XB	CA	COST OF PARK	COST OF GAS	CPP
1	-0.60	-0.67	1.32	1.13	0.18	-0.73
2	-0.94	-0.95	1.60	1.27	0.33	-0.84
3	-1.46	-1.11	1.99	1.48	0.50	-1.07
4	-0.81	-0.67	1.40	1.15	0.25	-0.69
5	-2.01	-2.97	3.01	2.23	0.76	-1.36
6	-1.25	-0.50	1.50	1.04	0.46	-0.60
7	-0.37	-0.46	1.08	0.96	0.12	-0.58
8	-0.84	-0.53	1.39	1.14	0.25	-0.69
9	-1.32	-0.81	1.87	1.45	0.42	-0.96
10	-1.01	-0.79	1.72	1.44	0.28	-1.02
11	-1.43	-0.75	1.90	1.46	0.44	-1.01
12	-0.58	-0.36	1.08	0.96	0.13	-0.58
13	-0.73	-0.58	1.38	1.22	0.17	-0.81
14	-1.51	-1.02	2.13	1.72	0.40	-1.28
15	-1.46	-0.92	1.99	1.61	0.38	-1.11
16	-2.12	-0.90	2.36	1.64	0.71	-1.24
17	-2.29	-0.90	2.33	1.52	0.81	-0.97
18	-2.95	-3.13	3.31	2.45	0.84	-1.95
19	-1.51	-1.02	2.13	1.73	0.40	-1.36
20	-2.25	-1.82	2.86	2.17	0.68	-1.97
21	-2.76	-1.77	3.02	1.99	1.01	-1.34
22	-0.34	-0.33	0.96	0.88	0.08	-0.58
23	-0.62	-0.42	1.16	0.98	0.19	-0.57
24	-1.28	-0.57	1.71	1.30	0.41	-0.90
25	-2.16	-0.91	2.33	1.63	0.70	-1.15
26	-0.25	-0.32	0.76	0.71	0.05	-0.32
27	-0.70	-0.51	1.37	1.17	0.21	-0.84
29	-1.35	-1.03	1.80	1.54	0.26	-0.85
30	-0.31	-0.25	0.74	0.67	0.08	-0.29
31	-1.01	-0.56	1.51	1.24	0.26	-0.81
32	-1.69	-1.55	2.41	1.87	0.53	-1.29
33	-0.58	-0.34	1.06	0.91	0.15	-0.45
34	-1.31	-0.88	1.87	1.43	0.44	-0.93
35	-0.99	-0.56	1.46	1.23	0.23	-0.70
36	-0.13	-0.26	0.61	0.59	0.03	-0.23
TOTAL	-0.99	-0.64	1.51	1.20	0.30	-0.79

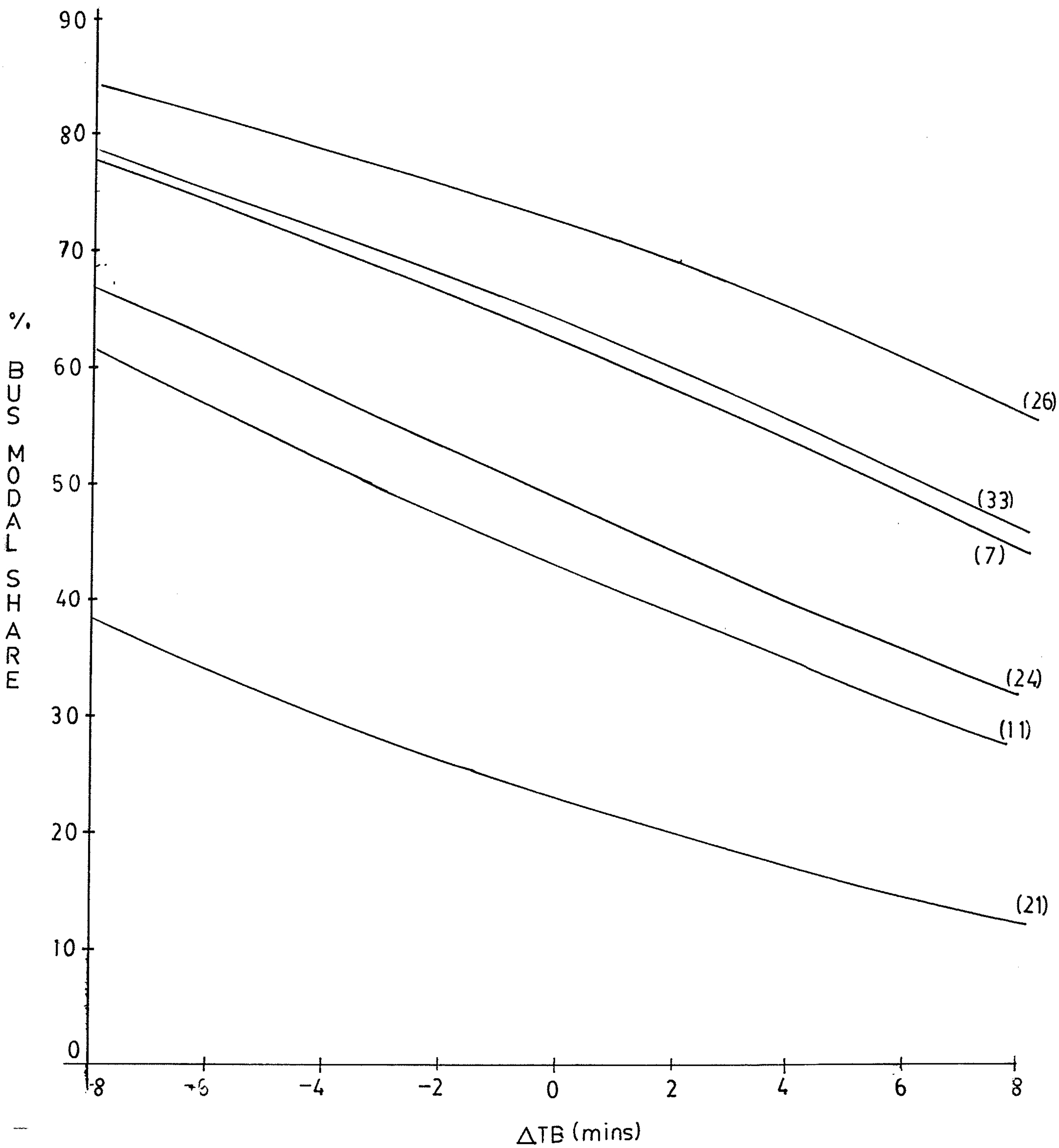

Figure 2: GRAPH OF %BUS MODAL SHARE VS.ΔTB

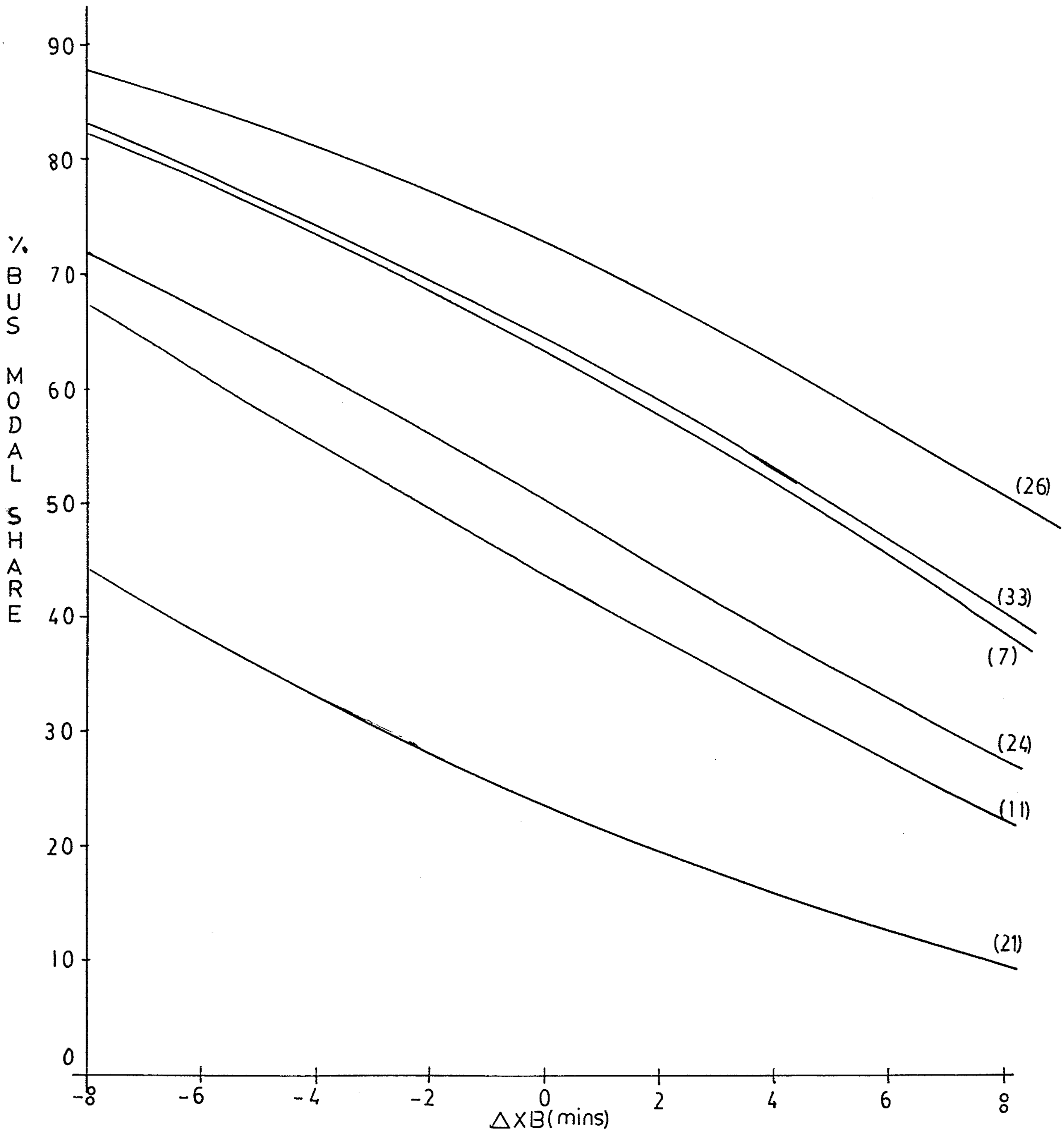

Figure 3: GRAPH OF %BUS MODAL SHARE VS ΔXB

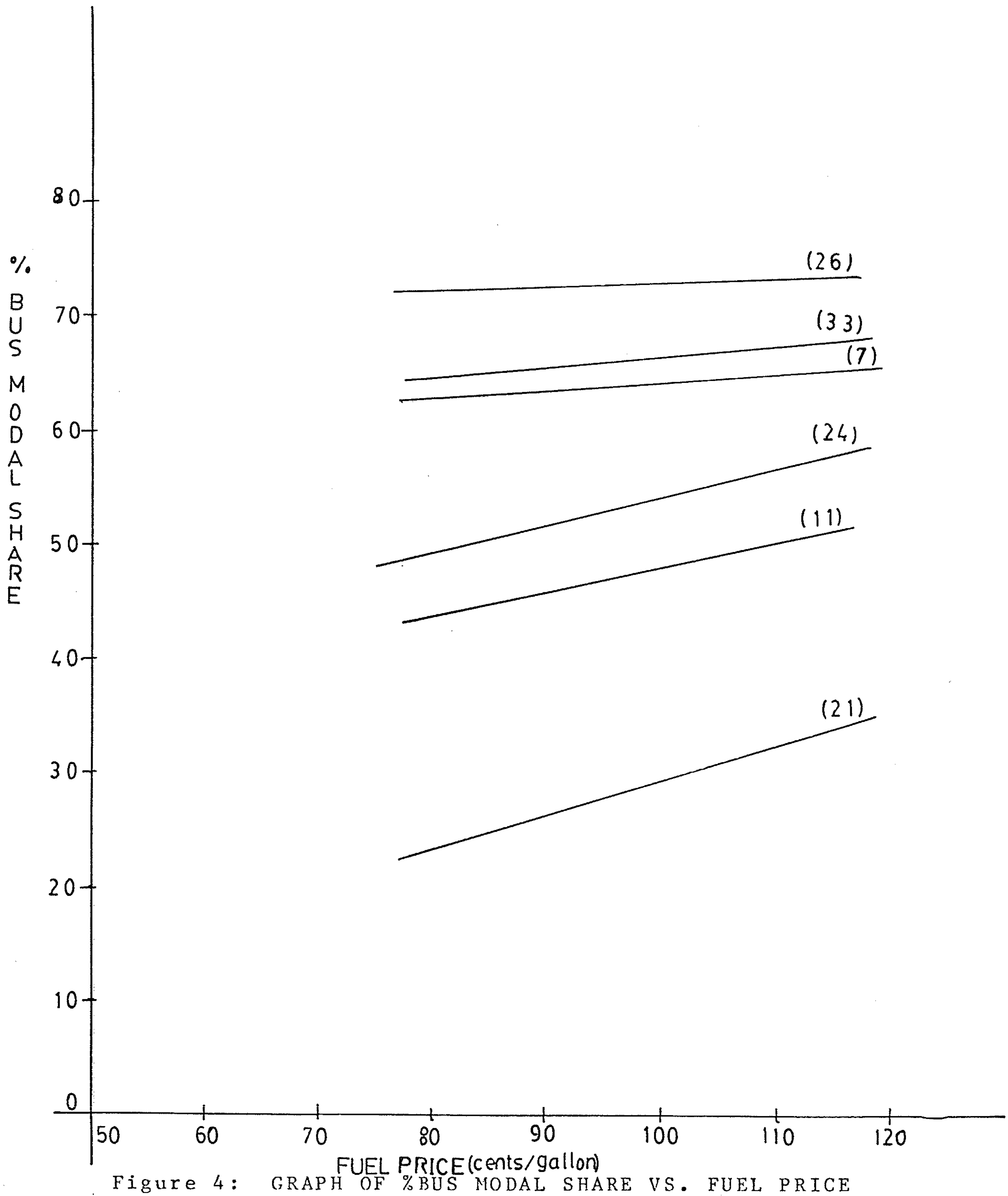

Figure 4: GRAPH OF %BUS MODAL SHARE VS. FUEL PRICE

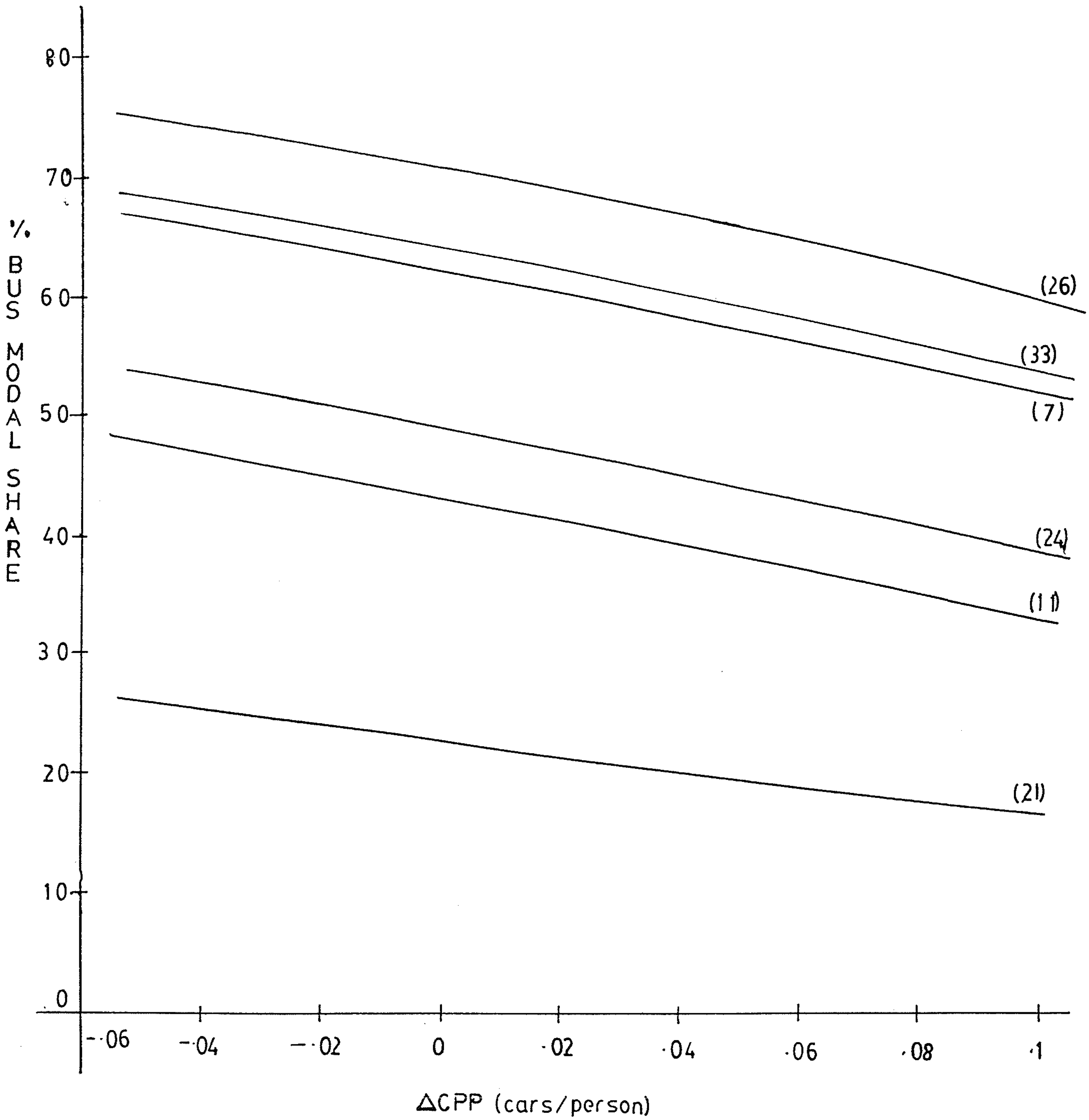

Figure 5: GRAPH OF %BUS MODAL SHARE VS.ΔCPP

The graphs in Figures 2,3,4 and 5 show the relationship of %bus modal share to varying values of changes in TB, XB, price of fuel and CPP. For clarity only six zones are used in the graphs. The zones chosen are zones 7, 11, 21, 24, 26 and 33. These zones vary in percentage of bus travellers as well as in geographical location within the city and in their distances from the CBD. In-vehicle travel times for bus were varied from -8 to +8 minutes of their original values. This was also done for the out of vehicle travel times. Fuel prices were varied from 77.4 cents per gallon to 117.4 cents per gallon. The number of cars per person were varied from their original values up to these values plus .1. Table 13 shows the changes in the %bus modal share for these changes in the independent variables.

TABLE 13

MODEL 1 CHANGES IN % BUS MODAL SHARE

ZONE OF ORIGIN	CHANGE IN TB +8	CHANGE IN TB -8	CHANGE IN XB +8	CHANGE IN XB -8	CHANGE IN FUEL PRICE 40¢/gal	CHANGE IN CPP .1
7	-17.7	14.9	-23.5	18.8	3.4	-10.2
11	-16.1	17.8	-20.5	23.3	9.2	-9.7
21	-10.1	14.8	-12.5	20.5	12.3	-6.5
24	-17.1	17.3	-22.1	22.4	9.4	-10.2
26	-16.4	12.1	-22.3	15.0	1.6	-9.1
33	-17.6	14.4	-23.5	18.2	4.2	-10.7
AVG	-15.8	15.2	-20.7	19.7	6.7	-9.4

The smallest decrease in the % bus modal share for the increase in in-vehicle travel time occurred with zone 21. This zone is the furthest away from the CBD and has the largest in-vehicle travel time by bus. The largest decrease occurred with zone 7, which had the second lowest in-vehicle travel time by bus. The smallest increase in % bus modal share for the decrease in the in-vehicle travel time occurred with zone 26. The largest increase occured with zone 11.

The smallest decrease in % bus modal share for the increase in out of vehicle travel time occurred with zone 21 while the largest decrease occurred with zones 7 and 33. Zone 21 had the largest out out of vehicle travel time by bus and zone 33 the smallest. When out of vehicle travel time was decreased, zone 26 experienced the smallest increase in % bus modal share. Zone 11 experienced the largest. In general the relative changes are similar to but larger than those for changes in in-vehicle travel time.

The greatest increase in % bus modal share for a 40 ¢/gallon increase in fuel price occurred with zone 21. This zone is the furthest away from the CBD and also has the lowest % bus modal share of the six zones considered. The smallest increase occurred with zone 26 the zone closest to the CBD. The greatest increase in % bus modal share for increasing car ownership occured with zone 33. The smallest increase occurred with zone 21.

5.6.2 SENSITIVITY ANALYSIS OF MODEL 2

Table 14 gives the percentage change in the zonal volume of bus passengers for a one per cent change in the stated variables.

Again, as in Model 1, volume is only elastic with respect to out of pocket cost for automobile. This is again mainly due to the parking cost. Volume is just inelastic with respect to income and is most inelastic with respect to fuel cost.

TABLE 14

ELASTICITY OF MODEL 2

ZONE OF ORIGIN	TB	XB	CA	COST OF PARK	COST OF GAS	INCOME
1	-0.33	-0.43	1.17	1.01	0.16	-0.92
2	-0.60	-0.71	1.36	1.08	0.28	-1.08
3	-0.91	-0.80	1.46	1.09	0.37	-1.17
4	-0.56	-0.48	1.27	1.05	0.23	-1.01
5	-1.06	-1.82	2.31	1.72	0.59	-1.84
6	-0.93	-0.43	1.38	0.96	0.43	-1.13
7	-0.24	-0.34	1.08	0.96	0.12	-0.83
8	-0.39	-0.28	1.14	0.93	0.21	-0.90
9	-0.84	-0.60	1.43	1.10	0.32	-1.13
10	-0.55	-0.50	1.33	1.11	0.21	-1.04
11	-0.79	-0.48	1.43	1.09	0.33	-1.14
12	-0.34	-0.25	1.05	0.93	0.12	-0.81
13	-0.41	-0.38	1.18	1.04	0.14	-0.92
14	-0.88	-0.69	1.43	1.16	0.27	-1.12
15	-0.79	-0.58	1.48	1.20	0.28	-1.17
16	-1.26	-0.62	1.55	1.08	0.47	-1.26
17	-1.08	-0.50	1.65	1.08	0.57	-1.37
18	-1.42	-1.76	2.65	1.96	0.68	-2.10
19	-0.90	-0.71	1.36	1.10	0.25	-1.06
20	-1.35	-1.27	1.22	0.93	0.29	-0.97
21	-1.58	-1.17	2.00	1.33	0.67	-1.63
22	-0.21	-0.23	0.94	0.86	0.08	-0.72
23	-0.44	-0.35	1.15	0.96	0.19	-0.90
24	-0.84	-0.43	1.32	1.00	0.32	-1.06
25	-1.28	-0.62	1.62	1.13	0.49	-1.31
26	-0.14	-0.21	0.90	0.84	0.06	-0.69
27	-0.34	-0.29	1.08	0.92	0.17	-0.85
29	-0.74	-0.66	1.55	1.33	0.22	-1.20
30	-0.19	-0.18	0.91	0.82	0.09	-0.71
31	-0.54	-0.35	1.25	1.03	0.21	-0.98
32	-0.86	-0.91	1.75	1.36	0.39	-1.39
33	-0.35	-0.24	1.10	0.95	0.15	-0.86
34	-0.83	-0.65	1.46	1.11	0.34	-1.16
35	-0.59	-0.39	1.32	1.11	0.21	-1.03
36	-0.10	-0.23	0.91	0.87	0.04	-0.69
TOTAL	-0.60	-0.45	1.26	1.02	0.24	-0.99

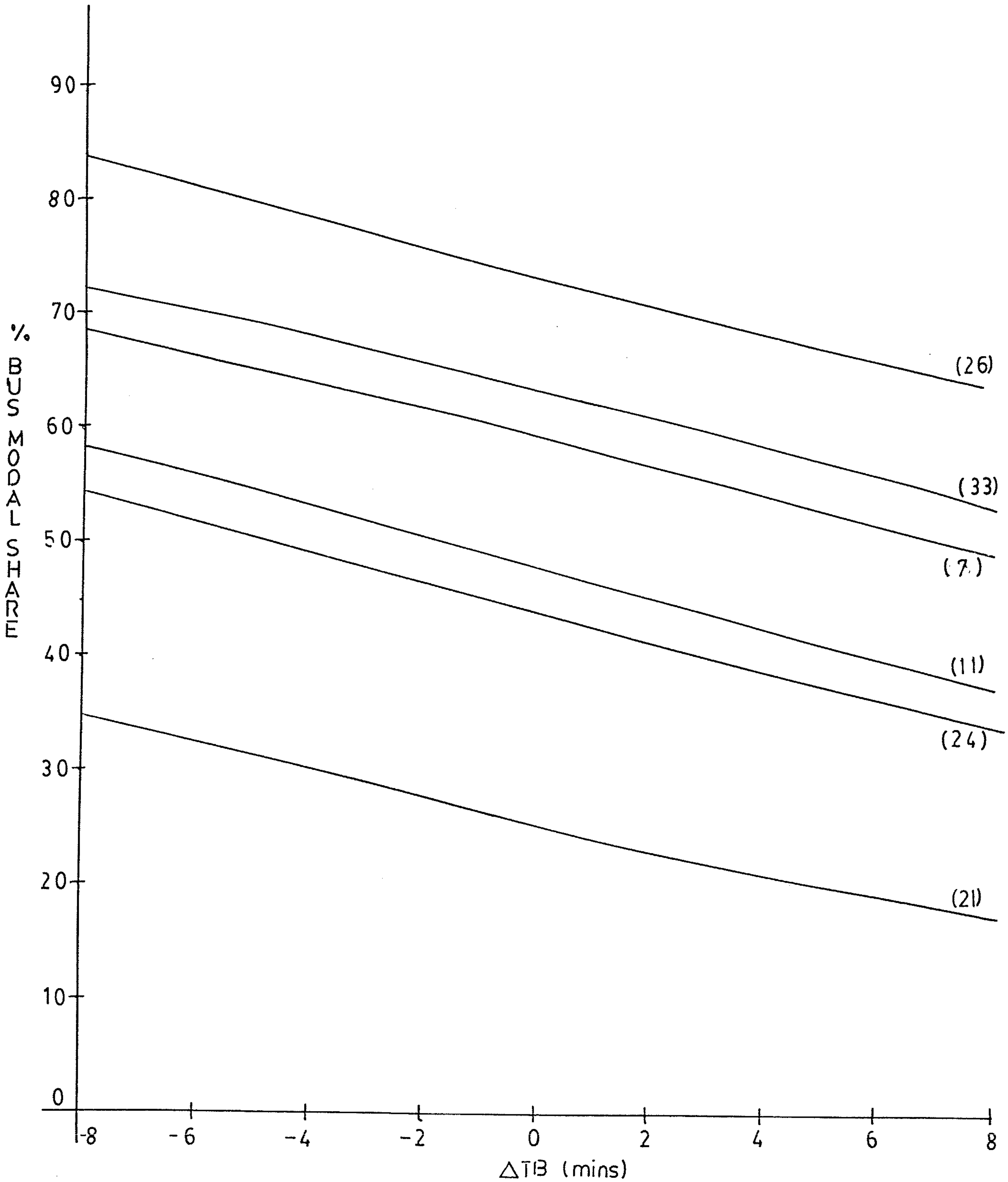

Figure 6: GRAPH OF %BUS MODAL SHARE VS.ΔTB

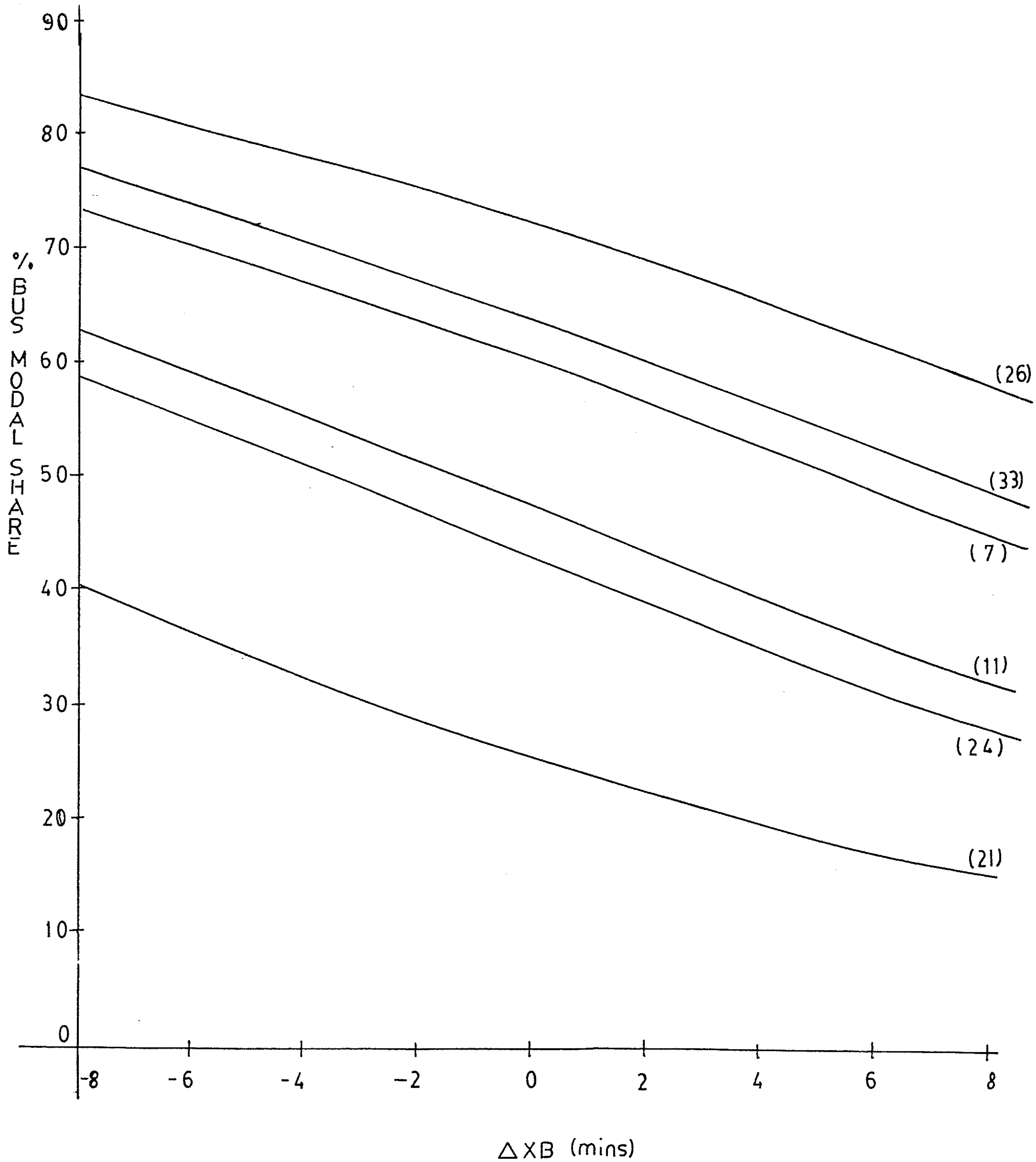

Figure 7: GRAPH OF %BUS MODAL SHARE VS.△XB

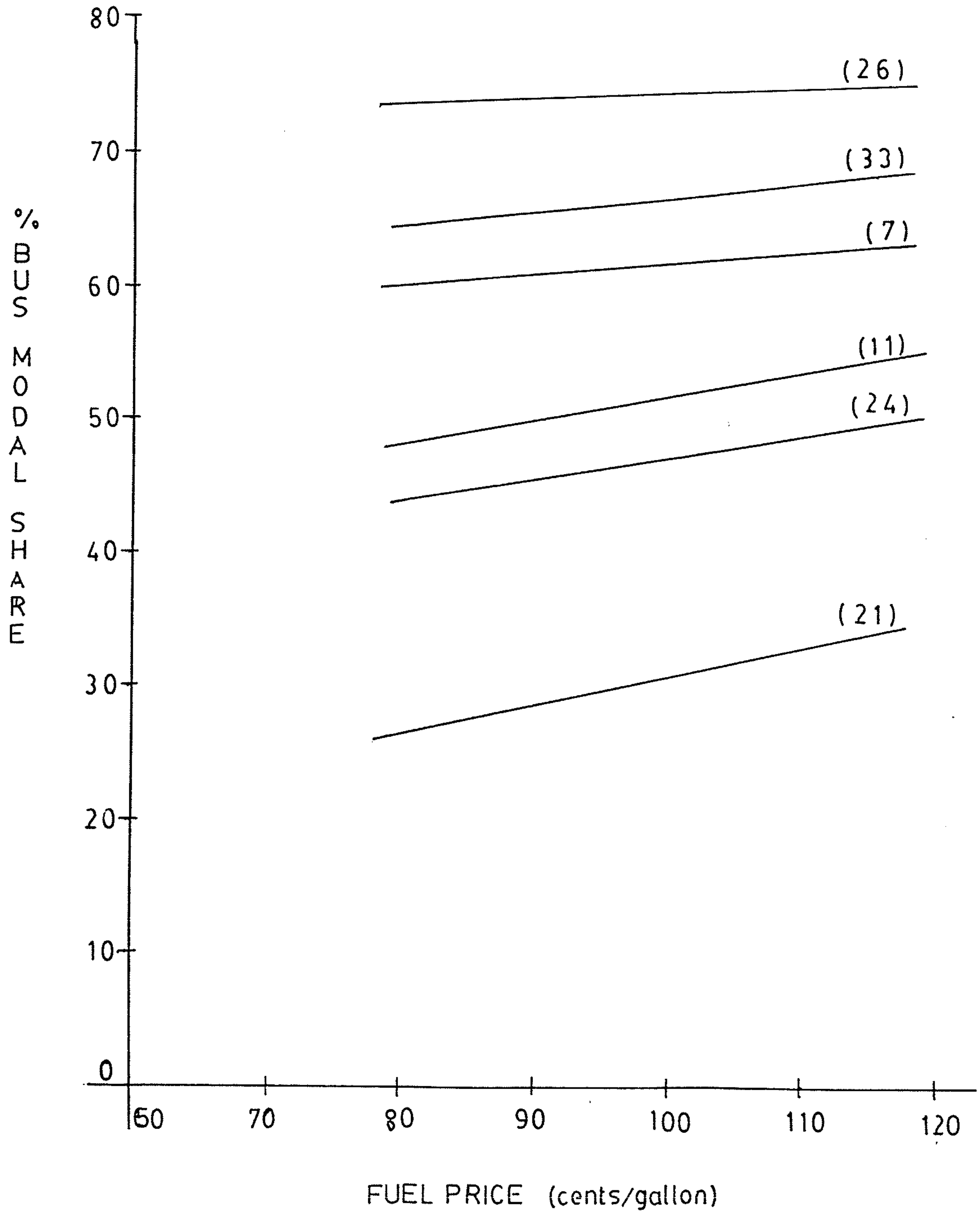

Figure 8: GRAPH OF %BUS MODAL SHARE VS. FUEL PRICE

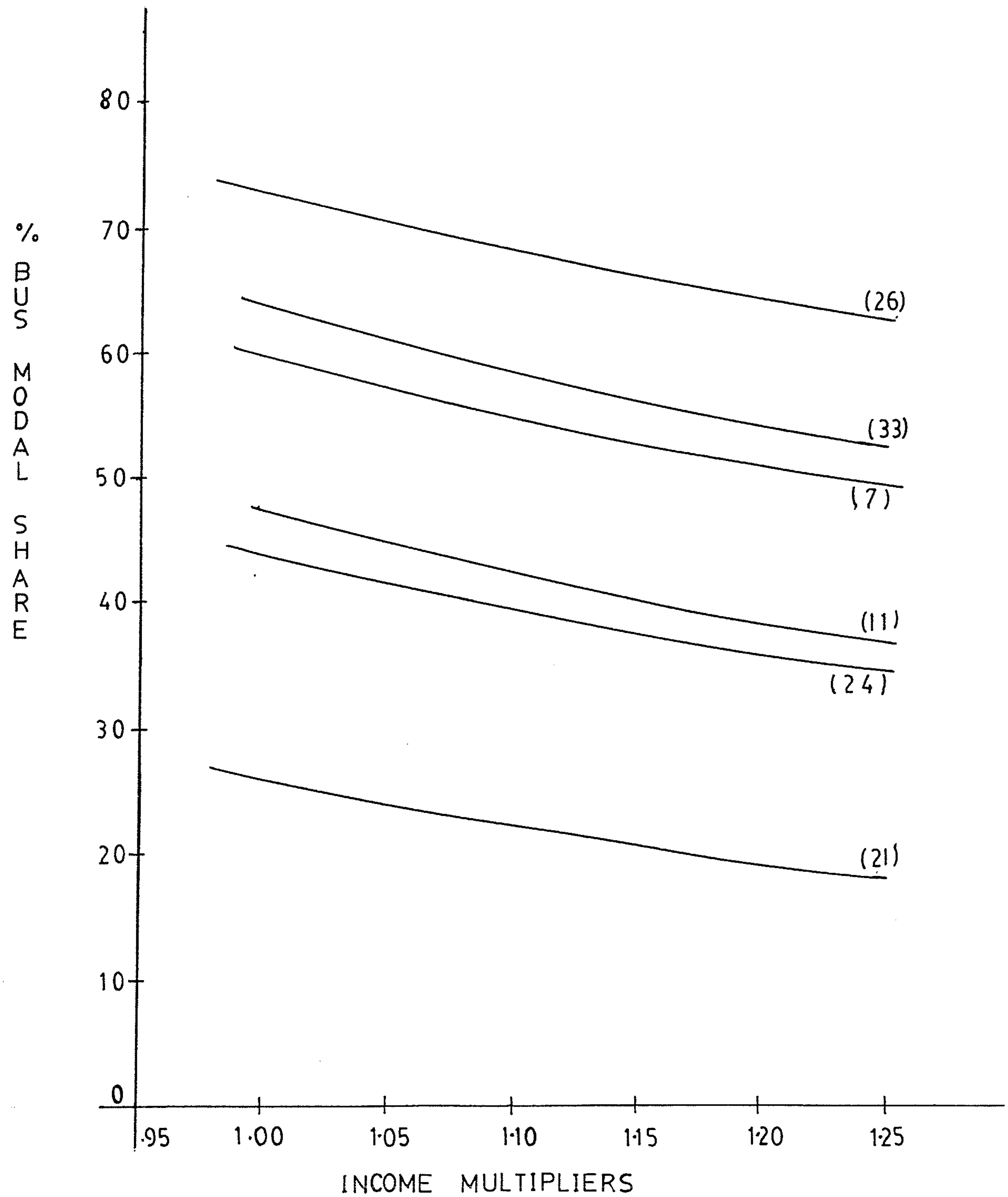

Figure 9: GRAPH OF %BUS MODAL SHARE VS. INCOME

The graphs of Figures 6, 7, 8 and 9 show the relationship of %bus modal share to various values of changes in TB, XB, fuel price and income. The zones chosen are the same as for Model 1. The in-vehicle time, out of vehicles travel time and fuel prices are varied over the same intervals as before. The income is varied up to one hundred and twenty-five per cent of its original value. Table 15 shows the changes in % bus modal share for these changes in the independent variables.

TABLE 15

MODEL 2 CHANGES IN % BUS MODAL SHARE

ZONE OF ORIGIN	CHANGE IN TB +8	CHANGE IN TB -8	CHANGE IN XB +8	CHANGE IN XB -8	CHANGE IN FUEL PRICE 40¢/gal	CHANGE IN INCOME 25%
7	-10.6	9.7	-16.4	14.5	3.3	10.3
11	-10.4	10.6	-15.7	16.1	7.4	10.7
21	-7.3	9.0	-10.6	14.5	9.0	7.6
24	-10.0	10.7	-15.0	16.4	6.6	8.9
26	-9.2	7.5	-14.7	11.0	1.9	11.2
33	-10.3	9.2	-16.1	13.6	4.5	11.6
AVG	-9.6	9.5	-14.8	14.4	5.5	10.1

The smallest decrease in % bus modal share for the increase in in-vehicle travel time occurred with zone 21. This zone is the furthest away from the CBD, has the largest in-vehicle travel time by bus and the lowest % bus modal share of all six zones. The largest decrease occurred with zone 7. The smallest increase in % bus modal share for the decrease in in-vehicle travel time occurred with zone 26. Zone 24 experienced the largest increase.

When out of vehicle travel times were increased, zones 21 and 7 experienced the smallest and largest decreases in % bus modal share respectively. For the decrease in out of vehicle travel time, zones 26 and 24 experienced the smallest and largest increase in % bus modal share respectively. In general the changes are larger than but similar to those of in-vehicle travel time.

An increase of 40¢/gallon in the price of fuel caused the largest increase in % bus modal share in zone 21. Zone 26 experienced the smallest increase. These results are similar to those of Model 1.

The greatest increase in % bus modal share for a 25% increase in average zonal income per dwelling unit occurred with zone 33 while the smallest occurred with zone 21.

5.6.3 SENSITIVITY ANALYSIS OF MODEL 3

Table 16 gives the percentage change in zonal volumes of bus passengers for a one per cent increase in the stated variables.

TABLE 16

ELASTICITY OF MODEL 3

ZONE OF ORIGIN	TB	XB	CA	COST OF PARK	COST OF GAS	CO
1	-0.48	-0.50	1.41	1.21	0.20	-0.60
2	-0.81	-0.79	1.87	1.49	0.40	-0.90
3	-1.14	-0.83	2.10	1.56	0.53	-0.97
4	-0.73	-0.58	1.71	1.40	0.31	-0.79
5	-1.64	-2.31	3.31	2.46	0.84	-1.68
6	-1.13	-0.43	1.84	1.27	0.57	-0.75
7	-0.32	-0.37	1.24	1.11	0.14	-0.52
8	-0.75	-0.45	1.66	1.36	0.30	-0.75
9	-1.19	-0.70	2.28	1.76	0.51	-1.19
10	-0.78	-0.59	1.79	1.50	0.29	-0.84
11	-1.12	-0.56	2.01	1.54	0.47	-0.90
12	-0.42	-0.25	1.05	0.93	0.12	-0.36
13	-0.64	-0.48	1.63	1.43	0.19	-0.82
14	-1.18	-0.76	2.24	1.81	0.42	-1.12
15	-1.20	-0.72	2.20	1.78	0.42	-1.10
16	-1.59	-0.64	2.37	1.65	0.72	-1.06
17	-1.92	-0.72	2.64	1.72	0.91	-1.14
18	-2.39	-2.43	3.63	2.68	0.92	-2.21
19	-1.20	-0.77	2.28	1.85	0.43	-1.23
20	-1.96	-1.52	3.38	2.56	0.80	-2.67
21	-2.27	-1.39	3.35	2.21	1.12	-1.62
22	-0.25	-0.23	0.95	0.87	0.08	-0.36
23	-0.47	-0.30	1.18	0.99	0.19	-0.41
24	-1.00	-0.42	1.80	1.37	0.43	-0.78
25	-1.84	-0.74	2.68	1.87	0.80	-1.33
26	-0.24	-0.29	0.99	0.92	0.06	-0.35
27	-0.55	-0.39	1.46	1.24	0.22	-0.68
29	-1.20	-0.88	2.17	1.85	0.31	-1.06
30	-0.29	-0.22	0.93	0.84	0.10	-0.29
31	-0.81	-0.42	1.62	1.34	0.28	-0.70
32	-1.51	-1.31	2.90	2.25	0.64	-1.78
33	-0.50	-0.28	1.22	1.06	0.17	-0.42
34	-1.19	-0.77	2.32	1.77	0.55	-1.21
35	-0.84	-0.45	1.66	1.39	0.27	-0.68
36	-0.10	-0.19	0.64	0.61	0.03	-0.12
TOTAL	-0.81	-0.50	1.66	1.32	0.33	-0.74

As was the case for the previous models, the volume of bus passengers is elastic only with respect to out of pocket cost for automobile. The component of this cost that leads to the elasticity is again the parking cost. In fact volume is more elastic with respect to in-vehicle travel time, car ownership per dwelling unit and out of vehicle travel time in that order, than to the price of fuel.

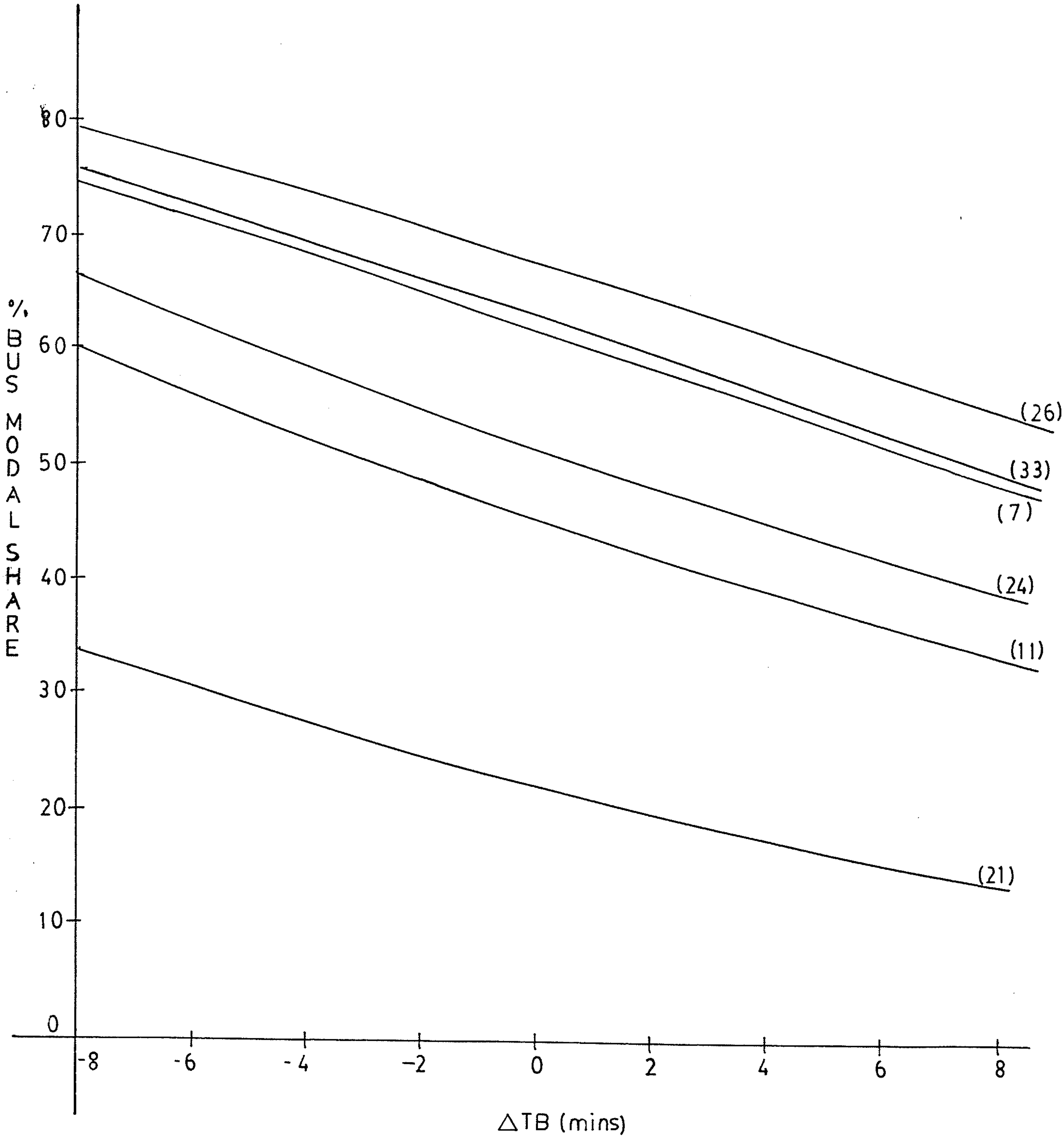

Figure 10: GRAPH OF %BUS MODAL SHARE VS △TB

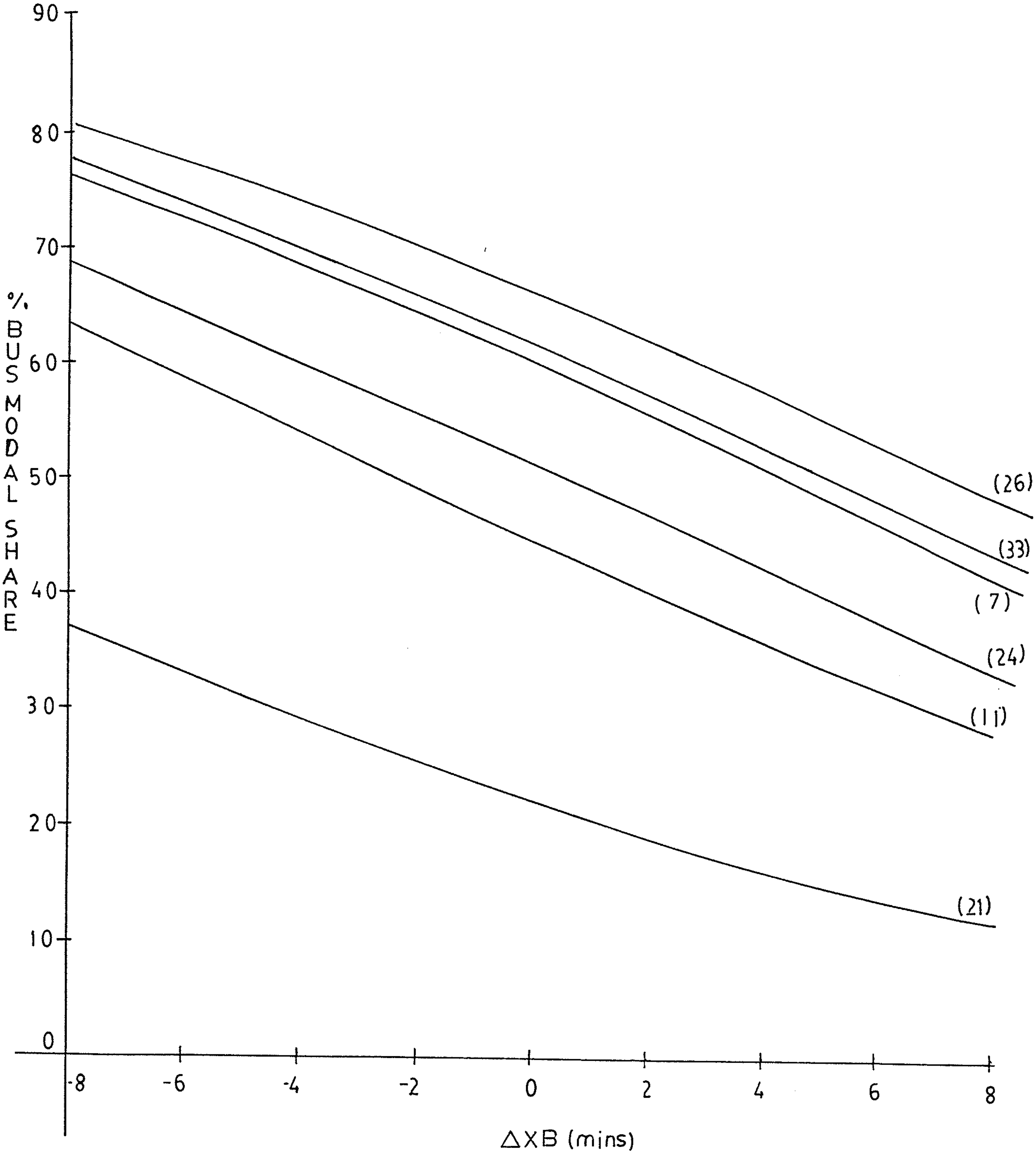

Figure 11: GRAPH OF %BUS MODAL SHARE VS ΔXB

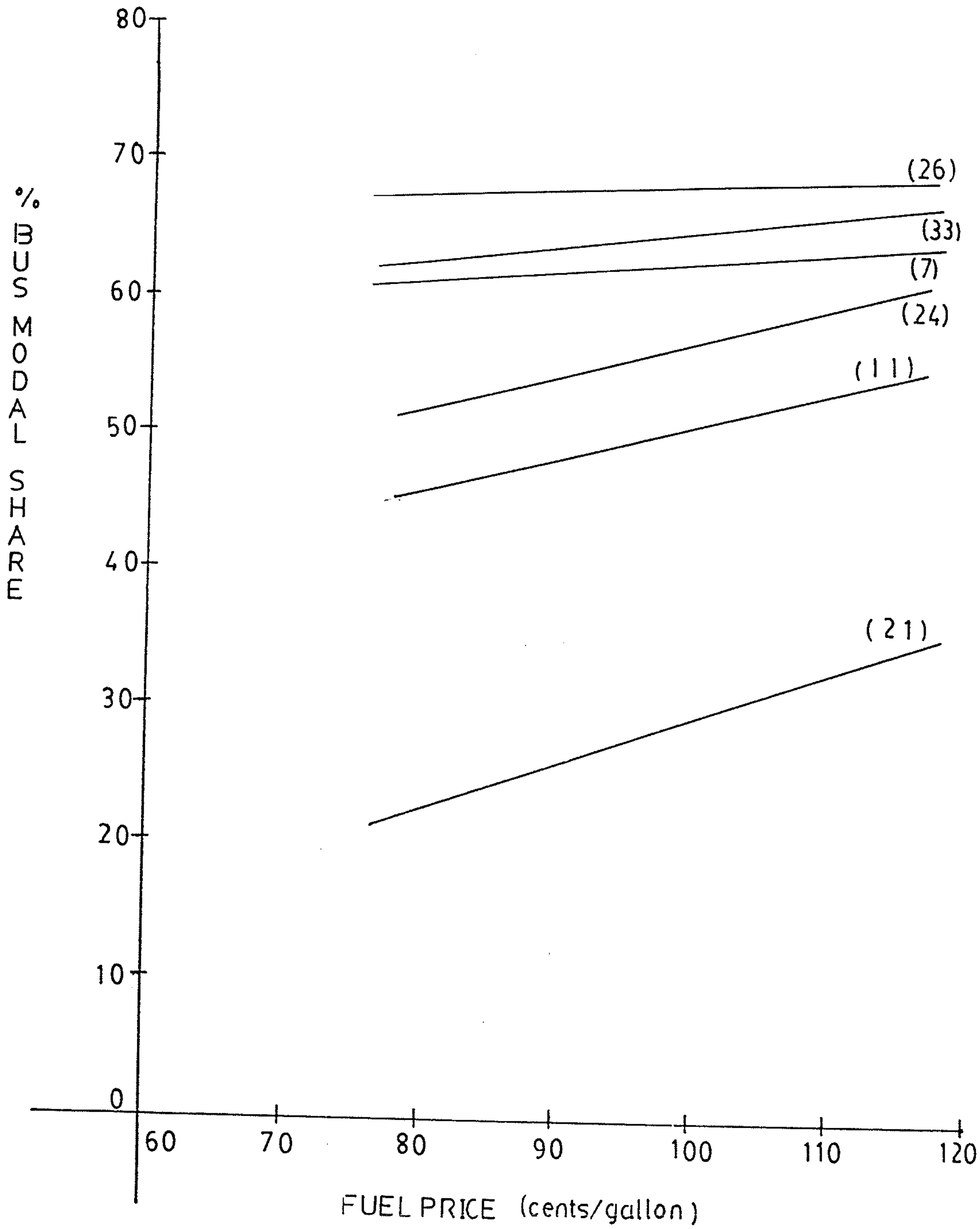

Figure 12: GRAPH OF %BUS MODAL SHARE VS. FUEL PRICE

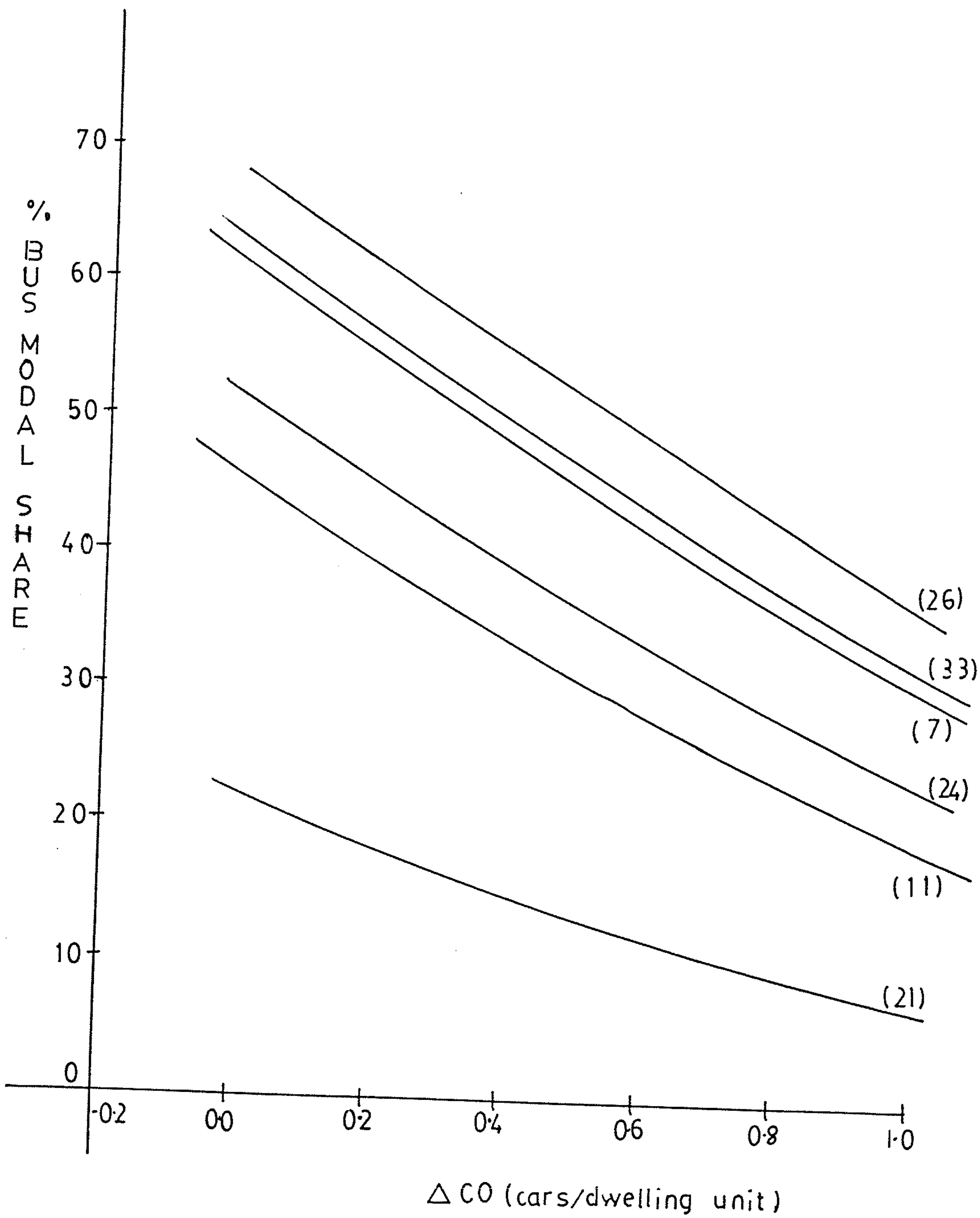

Figure 13: GRAPH OF %BUS MODAL SHARE VS.△CO

The graphs in Figures 10, 11, 12 and 13 show the relationship of %bus modal share to various values of changes in TB, XB, fuel price and CO. The in-vehicle travel time, out of vehicle travel time and fuel price are varied as before. CO values are increased up to one plus the original values. Table 17 shows the changes in % bus modal share for these changes in the independent variables.

TABLE 17

MODEL 3 CHANGES IN % BUS MODAL SHARE

ZONE OF ORIGIN	CHANGE IN TB +8	CHANGE IN TB -8	CHANGE IN XB +8	CHANGE IN XB -8	CHANGE IN FUEL PRICE 40¢/gal	CHANGE IN CO 1
7	-14.5	12.9	-18.4	15.8	3.8	-30.9
11	-13.7	14.6	-17.1	18.3	10.2	-26.7
21	-8.5	11.7	-10.2	15.3	13.7	-14.9
24	-14.4	14.1	-18.0	17.6	10.3	-28.9
26	-14.0	11.5	-17.9	14.1	2.0	-31.1
33	-14.5	12.5	-18.4	15.3	4.0	-31.1
AVG	-13.3	12.9	-16.7	16.1	7.3	-27.3

The smallest decrease in % bus modal share for the increase in in-vehicle travel time occurred with zone 21;the largest decrease occurred with zones 7 and 33. When the in-vehicle times were decreased, the smallest increase in % bus modal share occurred with zone 26 while the largest increase occurred with zone 11.

Changes in out of vehicle travel time caused similar relative changes in the % bus modal share of the zones.

When the cost of fuel was increased by 40 ¢/gallon, zones 21 and 26 experienced the largest and smallest increases in % bus modal share respectively. These results are similar to those of Models 1 and 2.

An increase in car ownership per dwelling unit caused the greatest decrease in % bus modal share in zones 26 and 33. These zones originally had the lowest car ownership per dwelling unit. Conversely, the smallest decrease occurred with zone 21 which originally had the largest car ownership level of all six zones.

5.6.4 SUMMARY OF SENSITIVITY ANALYSES

The sensitivity and elasticity relationships are similar for all three models. The changes in volume of bus passengers or % bus modal share brought about by changes in the independent variables are generally greatest for Model 1, followed by Model 3 and then Model 2.

In descending order, the volume of bus passengers is more elastic with respect to parking cost, income, in-vehicle travel time, car ownership and out of vehicle travel time, than fuel prices. These elasticities are considered in terms of percentages changes in the independent variables. This means that a one per cent change in parking cost has more effect than a one per cent change in any other variable.

It is not always practical to consider percentage change. Absolute changes may be more important. The effects of absolute changes may be different than the effects of percentage changes. Thus, while the volume of bus travellers is more elastic with respect to changes in in-vehicle travel time than to changes in out of vehicle time, an 8-minute change in the latter has a greater effect on the % bus modal share than an equivalent change in the latter. This is because of the average values of the variables.

An increase of 40¢/gallon in the price of fuel generally caused a 6 - 7% increase in the % bus modal share. A reduction of 8 minutes in in-vehicle travel time by bus has an effect two times greater than that. On the other hand, if car ownership levels were to increase by 1 car for every 10 persons, there would be a reduction of 9 percentage points in the % bus modal share. As is expected, an increase in the average income per dwelling unit has an adverse effect on the % bus modal share; a 25% increase in income causing a reduction of ten percentage points.

Chapter VI

TESTING OF THE MODELS

All three models were tested for their estimation ability on two zones. The zones were zone 16 and zone 22.

6.1 TESTING OF MODELS ON ZONE 16

Zone 16 is located about seven miles south of the central business district. The major attraction of this zone is the University of Manitoba, to which about twenty thousand trips are made daily.

The coefficients of Model 1 were used along with the independent variables for zone 16 to obtain utilities from which the modal shares were estimated. This was also done with the coefficients of Models 2 and 3.

The estimated and observed volumes of bus passengers were compared. None of the models accurately simulated the peak-a.m.-hour trips into zone 16.

6.2 TESTING OF MODELS ON ZONE 22

Zone 22 is a zone directly adjacent to the central business district. Unlike zone 16, this zone has a wide mix of employment. There are also quite a few major arterials running through this zone. Zone 22 attracts just over ten thousand work trips per day.

6.2.1 TESTING OF MODEL 1

The coefficients of Model 1 were used with the independent variables of zone 22 to obtain utilities and thus the estimated zonal volumes of bus passengers. The observed and estimated volumes were compared using the standard deviation and, more importantly, the Paired T test of section 5.3.2. Table 18 gives the observed and estimated zonal bus passenger volumes and their corresponding deviations.

The standard deviation was found to be 14.4. This is less than the average zonal bus passenger volume of 25.9 although it is high.

The computed T value for the Paired T test was found to be -0.193. The tabled T value for n =33 at the a = 0.05 level is 2.040. Thus the computed T value is numerically less than the tabled T value and we fail to reject the null hypothesis of section 5.3.2.In fact even at the a = 0.5 level the null hypothesis is not rejected since the tabled T value at the a = 0.5 level is 0.683, which is numerically greater than the computed T value. We can therefore state that the estimated zonal bus passenger volumes are not significantly different from the observed zonal bus passenger volumes. The model works.

TABLE 18

RESULTS OF MODEL 1 ZONE 22

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	29	20.5	8.5
2	73	26.2	46.8
3	12	16.7	-4.7
4	5	9.6	-4.6
5	0.4	0.3	0.1
6	5	9.4	-4.4
7	28	18.7	9.3
8	16	2.7	13.3
9	11	18.4	-7.4
10	5	7.9	-2.9
11	36	28.1	7.8
12	76	44.0	32.0
13	6	17.6	-11.6
14	0.4	1.3	-0.9
15	6	3.2	2.7
16	6	4.0	2.0
17	0.4	1.3	-0.9
19	8	9.0	-1.0
20	0.4	1.1	-0.7
21	6	16.9	-10.9
22	59	82.3	-23.3
23	10	37.0	-27.0
24	52	61.5	-9.5
25	64	84.1	-20.1
26	45	62.4	-17.4
27	0.4	9.2	-8.8
29	0.4	0.9	-0.5
30	6	8.0	-2.0
31	10	13.6	-3.6
33	37	26.9	10.1
34	5	6.5	-1.5
35	10	14.4	-4.4
36	71	52.8	18.2

6.2.2 TESTING OF MODEL 2

The coefficients of Model 2 were used with the independent variables for zone 22 and the zonal volume of bus passengers were estimated. As before a comparison between observed and estimated volumes was done using the SD and the Paired T test. Table 19 gives the observed and estimated zonal bus passenger volumes with their deviations.

The standard deviation was found to be 15.2. This is high, although it is less than the mean zonal bus passenger volume of 25.9.

The computed T value for the Paired T test was found to be -1.673. The tabled T value for n = 33 at the a = 0.05 level is 2.040. This is numerically greater than the computed T value so we fail to reject the null hypothesis of section 5.3.2. In fact, even at the a = 0.1 level we would fail to reject the null hypothesis since the tabled T for a = 0.1 is 1.697 which is numerically greater than the computed T. Thus, we state that there is no significant difference between the estimated and the observed zonal bus passenger volumes. The model works.

6.2.3 TESTING OF MODEL 3

The coefficients for Model 3 were used with the independent variables for zone 22 and the zonal bus passenger volumes were estimated. These were compared with the observed

TABLE 19

RESULTS OF MODEL 2 ZONE 22

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	29	26.7	2.3
2	73	32.1	40.9
3	12	23.1	-11.1
4	5	12.2	-7.2
5	0.4	0.9	-0.5
6	5	10.3	-5.3
7	28	23.6	4.4
8	16	4.5	11.5
9	11	27.8	-16.8
10	5	14.1	-9.1
11	36	42.3	-6.3
12	76	61.0	15.0
13	6	24.6	-18.6
14	0.4	2.9	-2.5
15	6	7.2	-1.2
16	6	9.9	-3.9
17	0.4	3.1	-2.7
19	8	18.8	-10.8
20	0.4	1.8	-1.4
21	6	24.0	-18.0
22	59	82.5	-23.5
23	10	30.0	-20.0
24	52	59.9	-7.9
25	64	101.5	-37.5
26	45	52.7	-7.7
27	0.4	15.0	-14.6
29	0.4	1.5	-1.1
30	6	8.0	-2.0
31	10	20.8	-10.8
33	37	30.0	7.0
34	5	10.2	-5.2
35	10	17.2	-7.2
36	71	39.4	31.6

volumes using the same criteria as before. Table 20 gives the observed and estimated volumes of bus passengers with their deviations.

The standard deviation was found to be 13.0. This is a fairly high value, although it is lower than the mean zonal bus passenger volume of 25.9.

The computed T value for the Paired T test was found to be -0.413. This is numerically less than the tabled T values at the a = 0.05 level(2.040), and at the a = 0.5 level(0.683). We therefore fail to reject the null hypothesis and state that there is no significant difference between the estimated and observed zonal bus passenger volumes. The model works.

In conclusion, the models did not accurately estimate modal shares for peak-am-hour work trips to zone 16, but they did estimate peak-a.m.-hour work trips to zone 22. This will be discussed in the following chapter.

TABLE 20

RESULTS OF MODEL 3 ZONE 22

ZONE OF ORIGIN	OBSERVED VOLUME OF BUS PASSENGERS	ESTIMATED VOLUME OF BUS PASSENGERS	DEVIATION
1	29	20.9	8.1
2	73	23.3	49.7
3	12	17.8	-5.8
4	5	8.3	-3.3
5	0.4	0.3	0.1
6	5	8.2	-3.2
7	28	18.6	9.4
8	16	2.5	13.5
9	11	15.6	-4.6
10	5	9.2	-4.2
11	36	29.6	6.4
12	76	53.6	22.4
13	6	15.6	-9.6
14	0.4	1.6	-1.2
15	6	3.5	2.5
16	6	5.6	0.4
17	0.4	1.2	-0.8
19	8	10.0	-2.0
20	0.4	0.5	-0.1
21	6	13.4	-7.4
22	59	82.9	-23.9
23	10	33.6	-23.6
24	52	54.5	-2.5
25	64	62.6	1.4
26	45	49.4	-4.4
27	0.4	9.9	-9.5
29	0.4	0.8	-0.4
30	6	7.2	-1.2
31	10	14.4	-4.4
33	37	24.9	12.1
34	5	5.3	0.3
35	10	13.0	-3.0
36	71	50.6	20.4

Chapter VII

DISCUSSION AND CONCLUSIONS

7.1 REVIEW OF ASSUMPTIONS

The first assumption on which the models are based is that transportation can be thought of as a commodity. The commuter tries to maximize the utility of this commodity in making transportation choices. This is done by determining the characteristics of his choices and their effects on him.

The models are developed by explicitly considering all the transportation system variables and consumer characteristics which influence the commuter's choice. The commuter, however, does not have perfect knowledge about the characteristics of his alternatives and therefore a randomness in his decisions can be expected.

The randomness element is assumed to be additive and has a Weibull distribution. This leads to the multinomial logit formulation.

Finally, The City of Winnipeg zones are considered to be behavioral units. Each unit is assumed to be composed of homogeneous travellers who have the same characteristics and who perceive the transportation system in the same way.

The last assumption is perhaps the most important one of this thesis. It enables a hitherto disaggregate model to be used with aggregate data. This assumption was necessary because the data were not readily available in a disaggregated form.

Aggregation, unfortunately, causes many problems. Firstly, the aggregation procedure, especially zoning, is almost always done arbitrarily and with no objective criteria for defining the areas and shapes of the zones. Secondly, aggregate models assume that the modal split is only a function of the mean values of the explanatory variables and not the distribution of these variables. Thirdly, there is a definite probability that the variance among the units in a zone is, in fact, greater than the variance between the zones. Even if this were not the case, aggregation causes loss of variability within zonal units. The number of observations available for analysis is considerably reduced. In the case of this thesis, for example, the number of observation are reduced to thirty-five.

The ideal model should not only be disaggregate, but should also be stratified socioeconomically since different socioeconomic groups may place differing emphases on the transportation system characteristics. Most disaggregate models, however, usually use some aggregated data, especially socioeconomic data. The models developed in the thesis, although not ideal, are nevertheless useful to transportation planners.

7.2 DISCUSSION OF VARIABLES USED IN THE MODELS

The independent variables used for development of the models were in-vehicle travel time, out of vehicle travel time, out of pocket cost, the ratio of out of pocket cost to income, car ownership per person and car ownership per dwelling unit. The first two variables were common to all models. Model 1 also included the third and fifth variables. Model 2 included the fourth and Model 3 the third and sixth variables.

As is expected, the out of vehicle travel time is weighted more heavily than the in-vehicle travel time. However, the sensitivity analysis shows that the bus modal share is more elastic with respect to in-vehicle travel time than to out of vehicle travel time.

It is interesting to compare the relationships between the coefficients in these models to results from other studies. Morrison[(28)], in Winnipeg, found that the ratios of the coefficients of excess time to in-vehicle travel time were 1.68 for bus and 1.85 for the automobile. Domencich and McFadden[(11)] felt that this ratio should be between 2 and 3 for both modes. The ratio was 1.33, 1.55 and 1.29 for Models 1, 2 and 3 for both modes.

The difference between the ratios obtained with the models developed in this thesis and the ratios obtained by Morrison may be due to the time at which data were collected. The difference between the ratios and those of Domencich and McFadden may be due to the geographical location of Winnipeg.

The value of travel time, as obtained from the models, is also worth considering. The average minimum wage for Winnipeg in 1976 was $2.69 per hour. Out of vehicle travel time has been valued by the three models at an average of $2.38 per hour. The in-vehicle travel time has been valued at $1.73 per hour. The value put on the out of vehicle travel time is reasonably close to the minimum wage. As is expected the in-vehicle travel time is considered to be less costly. In-vehicle travel time can usually be used to peform other activities (e.g. preparing for the day's work or getting rid of domestic worries) and is therefore not completely wasted.

One variable that was not considered in the analysis was the capital cost of owning an automobile. This is because it does not affect the immediate decision of either taking a bus or driving to work. Once the car is available, then the only cost that would affect the choice is the out of pocket cost for the particular trip.

Of the two car-ownership variables, car-ownership per dwelling unit should more accurately model the availability of the automobile for the peak-a.m.-hour work trip.

7.3 DISCUSSION OF THE TESTING

The models were tested on two quite different zones. With zone 16, the university zone, the volume of bus passengers in the peak-a.m.-hour was consistently underestimated.

There may be a number of reasons why the models underestimated bus passenger volumes to zone 16. Firstly, there is no distinct peak-a.m.-hour for trips into this zone as there is for trips into the downtown area. Secondly, there are only two groups of commuters into the zone. These are employees with higher than average incomes (e.g. professors) and students with low or no income. Thus the income distribution of the trip-makers into zone 16 would be very different from that of trip-makers into the CBD. Thirdly, the car availability distribution of commuters into zone 16 would be very different to that of commuters into the CBD.

7.4 IMPLICATIONS OF THE MODELS

The models fit well into widely used urban transportation planning methods. Moreover, because they are based upon utilities, they enable planners to gauge what tradeoffs travellers are willing to make between conflicting system characteristics without shifting mode. Thus a planner can deduce what extra fares bus passengers would tolerate to obtain lower in-vehicle or out of vehicle travel times. Planners could also predict the impact of certain policies, e.g.,parking cost increases, on the modal split of peak-a.m.-hour CBD work trips. The models could also be useful in predicting what future facilities would be required to keep public transport patronage in the light of increases in income and car ownership.

The effects of policy options can be seen by considering an increase of $2.30 per gallon in the price of gasoline over a five year period. Suppose this is done through a fifty cents per gallon increase each year for the first three years, and a forty cents per gallon increase for each of the last two years.

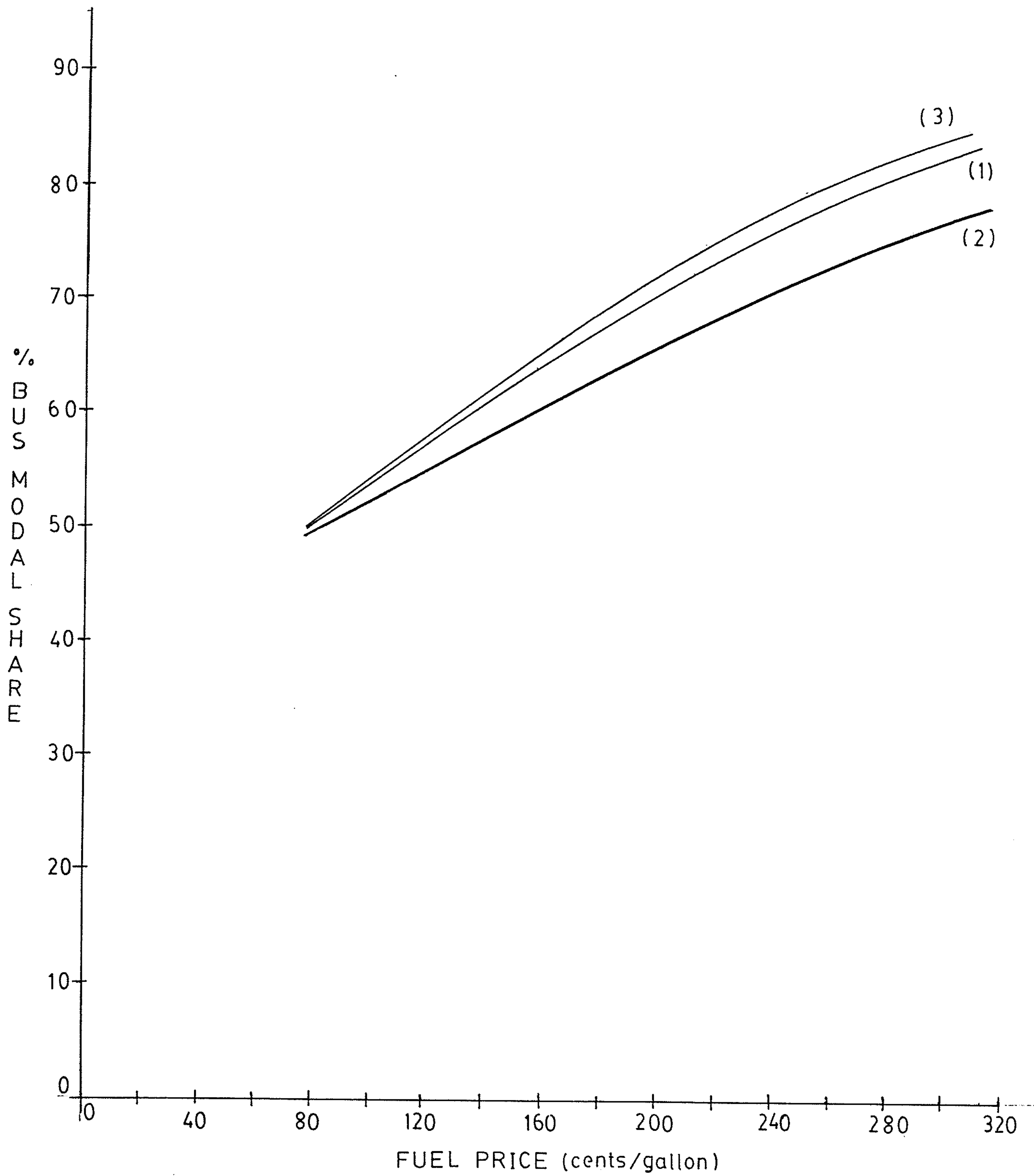

Figure 14: GRAPH OF CBD % BUS MODAL SHARE VS. FUEL PRICE

Figure 14 shows the effect of this policy on the % bus modal share for peak-a.m.-hour work trips into the CBD using all three models. If this is averaged, then there would be increases in the % bus modal share of 8.6, 8.3 and 7.2 percentage points for the first three years, and 4.9 and 3.8 percentage points for the last two years respectively. Thus, over a period of five years, bus patronage could be increased by almost thirty-three percentage points, i.e., from 49% to 82%.

Even if incomes increase by 8% during this time period, this would only reduce the expected bus patronage by about 3%. Furthermore, increased patronage would bring increased revenues and this leads to better service. Better service would encourage more patronage.

Although the highway gas consumption of automobiles will improve, it is not expected that the peak-a.m.-hour consumption would improve significantly. Thus increased gasoline prices could bring about important changes in peak-a.m.-hour bus modal shares.

The models may be transfered to medium-sized urban areas with similar geographical and transportation system characteristics as Winnipeg. Because of their aggregate nature, however, the transferability is limited. The assumption that would have to be made in transferring the models to any city is that commuters in that city have the same attitudes toward travel as Winnipeg commuters.

7.5 CONCLUSIONS AND RECOMMENDATIONS

An aggregate behavioral approach to modal split modelling has been used with 1976 Winnipeg data. Three models have been developed for peak-a.m.-hour work trips to the central business district.

Based on the models' estimation of peak-a.m.-hour work trips to the CBD and to an adjacent zone, it can be concluded that the models are good estimators of modal choice in this particular situation.

There are a number of areas for further research. Modal split models should be developed for various other trip purposes. An example of this would be a joint shopping destination-modal choice model. The models should not only be disaggregate, but, if possible, should be stratified according to the socioeconomic characteristics of the travellers. Once a set of these models have been developed, they can then be used simultaneously to predict the effects of policy options, (e.g., taxes on gasoline), on overall modal choice. In this model, for instance, it was found that raising the price of fuel by forty cents per gallon caused an increase of 6 - 7% in the bus modal share.

These models should be developed for at least three modes, since car-pooling is expected to play an ever-increasing role as a transportation option.

Finally, the models presented in this thesis should be investigated to attempt to improve upon their estimating accuracy by the inclusion of previously unconsidered variables, improved data collection, stratification and/or disaggregation.

BIBLIOGRAPHY

(1) Atherton, T.J.Suhrbier, J.H., and Jessiman, W.A. Use of Disaggregate Travel Demand Models to Analyze Car Pooling Policy Incentives. Transportation Research Record No. 599, 1976, pp.35-40

(2) Ben-Akiva, M., and Atherton, T.J., Choice-Model Predictions of Car-Pool Demand: Methods and Results. Transportation Research Record No.637, 1971, pp 13 - 17.

(3) Ben-Akiva, M., and Lerman, S.R. Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work. Transportation Vol.3 No.4 December 1974, p 357.

(4) Bock, F.C. Factors Influencign Modal Trip Assingment. National Cooperative Highway Research Program Report No.57 . Washington, D.C.: Highway Research Board, 1968.

(5) Brown, G.R. Policy Sensitive Model for Mode Split Planning. ASCE Transportation Engineering Journal, Vol.99, No.TE4, November 1973, pp.711 -723.

(6) Cherwony, W., and Lutin, J. Two-Dimensional Logit Modal Split Model. ASCE. Transportation Engineering Journal, Vol.102, No.TE3, August 1976, pp.525 - 530.

(7) DeDonnea, F.X. The Determinants of Transport Mode Choice in Dutch Cities, Some Disaggregate Stochastic Models. Belgium: Rotterdam University Press, 1971.

(8) Demetsky, M.J., and Hoel, L.A. Modal Demand:A User Perception Model. Transportation ResearchVol.6, No.4, 1972, pp.293 - 308.

(9) Dobson, R. Toward the Analysis of Attitudinal and Behavioral Response to Transportation System Characteristics. Transportation Vol.4, No.3, September 1975, p.267.

(10) Domencich, T., Kraft, D., and Valette, J.P. Estimation of Urban Passenger Travel Behavior : An Economic Demand Model. Highway Research Record , No.238, 1968.

(11) Domencich, T.A., and McFadden, D. Urban Travel Demand : A Behavioral Analysis. Amsterdam: North-Holland Publishing Co.: New York: Elsevier Publishing Co.Inc. 1975.

(12) Dunn, O.J. and Clark, V.A. Applied Statistics : Analysis of Variance and Regression. Toronto: John Wiley and Sons, Inc.1974.

(13) Energy, Mines and Resources Canada. The Car Mileage Book. How to Buy and Maintain Your Car To Save Money and Energy. Ottawa: Energy, Mines and Resources Canada1978.

(14) Ferreri, M.G., and Cherwony, W. Choice and Captive Modal-Split Models. Highway Research Record No.369, 1971, pp.80-90.

(15) Gillen, D.W. Effects of Parking Cost on Urban Transport Modal Choice. Transportation Research Record No.637, 1977, pp.46 - 50

(16) Golob, T.F. and Recker, W.W. Mode Choice Prediction Using Attitudinal Data : A Procedure and Some Results. Transportation Vol.6, No.3, September 1977, pp.265 - 286.

(17) Hartgen, D.T. Attitudinal and Situational Variables Influencing Urban Mode Choice : Some Empirical Findings. Transportation Vol.3, No.4, December 1974, p.377.

(18) Hartgen, D.T.and Tanner, G.H. Investigations of the Effects of Traveller Attitudes in a Model of Mode-Choice Behavior. Highway Research Record No.369, 1971, pp.1 - 14.

(19) Hutchison, B.G. Principles of Urban Transport Systems Planning. Washington, D.C.: Scripta Book Company, 1974.

(20) Hutchinson, B.G.et.al. Report on Urban Modal Modal Split Models. Ottawa: Roads and Transportation Association of Canada.

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

(22) Lerman, S.R. Location, Housing, Automobile Ownership and Mode to Work : A Joint Choice Model. Transportation Research Record No.610, 1976, pp.6 - 11.

(23) Makowski, G.G., Baker, M.M.and Sirha, K.C. Estimating Transportation Mode Utilities. ASCE. Transportation Enginnering Journal Vol.103No.TE4, November 1976, p.807

(24) Manheim, M.L. Fundamentals of Transportation Systems Analysis : Volume 1, Basic Concepts. Cambridge, Massachusetts: M.I.T. Press, 1978.

(25) McGillvray, R.G. Mode Split and Value of Time. Transportation Research Vol.6, No.4, December 1972, pp.309 - 316.

(26) McGillvray, R.G. Demand and Choice Models of Modal Split. Journal of Transport Economics and Policy Vol.4, No.2, 1970, pp.192 - 207.

(27) Meyburg, A.H. and Stopher, P.R. Aggregate and Disaggregate Travel Demand Models. ASCE. Transportation Engineering Journal No.101, Vol.TE2, May 1975, pp.237 - 245.

(28) Morrison, J.B. A Proposed Modal Choice Model For Urban Transportation Systems. Unpublished M.Sc. Thesis University of Manitoba, 1978.

(29) Orski, C.K. The Potential For Fuel Conservation : The Case of The Automobile. Transportation Research Vol.8, August 1974, pp.247 - 257.

(30) Parody, T.E. Analysis of Predictive Qualities of Disaggregate Modal-Choice Models. Transportation Research Record No.637, 1977, pp.51 - 56.

(31) Pratt, R.H. A Utilitarian Theory of Travel Mode Choice. Highway Research Record No.322, 1970, pp.40 - 53.

(32) Quarmby, D.A. Choice of Travel Mode for the Journey to Work. Journal of Transport Economics and Policy Vol.1, 1967, pp.273 - 314.

(33) Recker, W.W. and Stevens, R.T. Attitudinal Models of Mode Choice : The Multinomial Case for Selected Non-Work Trips. Transportation Vol.5, 1976, pp.355 - 375.

(34) Reichman, S. and Stopher, P.R. Disagregate Stochastic Models of Travel-Mode Choice. Highway Research Record No.369, 1971, pp.91 - 103.

(35) Richards, M.G. Disaggregate Simultaneous Urban Travel Demand Models : A Brief Introduction. Transportation Vol.3, No.4, December 1974, pp.335 - 342.

(36) Richards, M.G. and Ben-Akiva, M. A Disaggregate Travel Demand Model. Westmead: Saxon House, Lexington Books, 1975.

(37) Richards, M.G. and Ben-Akiva, M. A Simultaneous Destination and Mode Choice Model For Shopping Trips. Transportation Vol.3, No.4, December 1974, p.343.

(38) Rosin, K.A. An Urban Transportation Demand Model. Unpublished M.Sc. Thesis, University of Manitoba, 1972.

(39) Shunk, G.A. and Bouchard, R.J. An Application of Marginal Utility to Travel Mode Choice. Highway Research Record No.322, 1970, pp.30 - 39.

(40) Statistics Canada, Prices Division. Industry Price Indexes, August 1976. Ottawa : Statistics CanadaDecember 1976.

(41) Statistics Canada, Prices Division. Industry Price Indexes, November 1979. Ottawa : Statistics CanadaJanuary 1980.

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

(43) Streets and Transportation Division, City of Winnipeg. The City of Winnipeg, Travel and Demographic Trends 1962 - 1976. Winnipeg : Streets and Transportation Division, City of Winnipeg 1978.

(44) Talvitie, A. Comparison of Probablistic Modal Choice Models : Estimation Methods and System Inputs. Highway Research Record No.392, 1971, pp.111 - 120.

(45) Traffic Research Corp. Ltd. and Metro. Corp. of Winnipeg. The Winnipeg Area Transportation Study. Winnipeg, 1966-1968.

(46) Underwood, J.R.and Leake, G.R. Comparison of Cost Functions in Modal Split Models. ASCE. Transportation Engineering Journal Vol.103, No.TE6, November 1977, p.763.

(47) Watson, P.L. The Value of Time : Behavioral Models of Mode Choice. Lexington, Mass.: Lexington Books, 1974.

(48) Wilson, A.G. Entropy Maximizing Models in the Theory of Trip Distribution. Journal of Transport Economics and Policy Vol.3, No.1, 1967, pp.108 - 126.

(49) Zupan, J.M. Mode Choice : Implications for Planning. Highway Research Record No.251, 1965, pp.6 - 25.

Appendix A

MULTINOMIAL LOGIT MODEL DERIVATION

A random variable, n_i, has a Weibull(extreme value Gnedenko) distribution[2] if:

$$\text{Prob}\ [n_i < n] = e^{-e^{-(n+a)}}$$

Where a is a parameter. The associated frequency function is

$$\Psi(n) = e^{-(n+a)} \exp\left[-e^{-(n+a)}\right]$$

The Weibull distribution has two important properties.

1. It is stable under maximization, i.e., the maximum of two independent Weibull random variables is itself a Weibull random variable.

2. It is the only distribution that has a logistic distribution in the multinomial case.

If random variables n_i, have independent Weibull distributions with parameters a_i for $i = 1,2...n$, then:

1. $n_i + v$ has a Weibull distribution with parameter $a_i - v$ for any real v.

----

[2]This Appendix is after Domencich and McFadden(1975).

2. Maximum of ni has a Weibull distribution with parameter

$$-\log \Sigma e^{-ai}$$

3.

$$\text{Prob}[v1+n1 \geq v2+n2] = \frac{e^{v1-a1}}{e^{v1-a1} + e^{v2-a2}}$$

4.

$$\text{Prob}[vi+ni \geq vj+nj] = \frac{e^{vi-ai}}{\Sigma e^{vj-aj}}$$

For j = 1,....n.

Appendix B

VARIABLES FOR ZONE 36

ZONE OF ORIGIN	TB	XB	TA	D	CA	INCOME	CO	CPP
1	15	12.5	9	2.87	116.24	13360	1.07	0.392
2	21	16.0	5	4.60	126.04	16300	1.33	0.403
3	28	16.0	18	6.00	133.96	18234	1.36	0.444
4	20	12.5	11	3.86	121.85	14516	1.23	0.366
5	20	29.0	13	6.06	134.30	14648	1.52	0.380
6	34	10.2	17	7.84	144.37	17543	1.29	0.356
7	11	10.2	5	2.17	112.28	13688	1.03	0.366
8	21	10.0	9	3.89	122.02	12012	1.20	0.370
9	26	12.0	11	5.13	129.04	18015	1.48	0.408
10	20	11.8	9	3.42	119.36	15125	1.22	0.437
11	28	11.0	12	5.40	130.56	15758	1.28	0.427
12	17	8.0	8	2.30	113.02	13148	0.83	0.369
13	17	10.2	8	2.40	113.58	14079	1.25	0.408
14	25	12.7	10	4.10	123.21	18748	1.38	0.461
15	26	12.3	11	4.15	123.49	15536	1.36	0.427
16	37	11.8	17	7.69	143.53	19402	1.40	0.469
17	43	12.7	20	9.36	152.98	14334	1.45	0.391
18	35	28.0	14	6.12	134.64	12825	1.85	0.498
19	25	12.7	11	4.07	123.04	20161	1.46	0.485
20	30	18.2	14	5.56	131.47	30000	2.33	0.565
21	40	19.2	18	8.98	150.83	18403	1.62	0.417
22	11	7.9	6	1.57	108.89	13273	0.90	0.400
23	18	9.2	8	3.44	119.47	15426	0.90	0.360
24	28	9.3	12	5.60	131.70	18439	1.24	0.426
25	38	12.0	17	7.63	143.19	18291	1.57	0.435
26	10	9.4	5	1.21	106.85	10435	0.81	0.271
27	17	9.4	8	3.16	117.89	13079	1.19	0.439
29	25	14.4	12	3.00	116.98	13619	1.26	0.341
30	13	7.8	5	2.00	111.32	11092	0.74	0.260
31	23	9.5	10	3.65	120.66	13891	1.13	0.397
32	26	17.9	12	5.04	128.53	14901	1.75	0.429
33	18	8.0	6	2.82	115.96	12424	0.87	0.304
34	26	13.2	12	5.47	130.96	17673	1.50	0.400
35	23	9.8	9	3.36	119.02	14287	1.07	0.348
36	6	9.3	3	0.80	104.53	11319	0.44	0.242

Appendix C

PEAK-AM-HOUR VOLUMES FOR ZONE 36

ZONE OF ORIGIN	BUS PASSENGERS	AUTO DRIVERS
1	477	255
2	560	716
3	528	493
4	377	382
5	9	12
6	370	298
7	490	225
8	23	18
9	309	546
10	263	297
11	601	657
12	1005	597
13	220	269
14	102	204
15	109	175
16	310	463
17	42	63
18	0	7
19	286	778
20	15	101
21	134	443
22	829	450
23	102	115
24	542	619
25	528	941
26	222	129
27	181	126
29	57	67
30	321	111
31	259	288
32	13	89
33	593	248
34	290	450
35	154	113
36	190	134

Appendix D

VARIABLES AND VOLUMES FOR ZONE 16

ZONE OF ORIGIN	TB	XB	TA	D	CA	BUS PASS	AUTO DRIVE
1	45.0	14.4	26.9	9.53	68.66	42	45
2	51.0	17.9	32.0	11.39	79.19	81	152
3	58.0	17.9	37.0	13.21	89.49	22	67
4	50.0	13.0	28.9	10.19	72.40	15	63
5	56.0	30.9	34.9	11.97	82.47	0	4
6	64.0	12.1	34.0	13.57	91.53	23	101
7	41.0	12.1	17.0	7.00	54.34	118	113
8	51.0	14.6	21.0	8.89	65.04	8	17
9	56.0	16.6	20.7	7.79	58.81	49	220
10	42.0	11.1	15.5	6.17	49.64	58	104
11	46.0	14.4	21.0	6.82	53.32	40	286
12	24.0	8.4	12.5	4.64	40.98	627	349
13	23.7	9.1	13.5	4.97	42.85	118	86
14	12.0	11.2	8.0	2.97	31.53	156	100
15	13.0	11.2	8.5	3.32	33.51	150	102
16	14.0	10.3	5.0	1.75	24.63	341	359
17	20.0	14.0	6.0	3.97	37.19	38	87
19	26.1	10.9	14.8	7.16	55.25	317	337
20	25.8	18.5	17.8	8.22	61.25	50	114
21	34.8	13.3	22.0	11.48	79.70	66	178
22	35.7	8.2	23.0	6.99	54.28	320	111
23	38.1	14.1	20.0	8.91	65.15	27	62
24	48.1	14.3	24.0	11.12	77.66	40	185
25	58.1	17.0	29.0	13.09	88.81	56	194
26	40.0	11.3	27.5	7.24	55.70	65	79
27	47.0	11.3	30.5	8.59	63.34	25	28
29	59.8	16.3	27.8	9.69	69.57	3	41
30	43.0	9.8	30.0	8.60	63.40	52	73
31	53.0	9.9	34.0	10.26	72.79	39	52
32	56.0	19.9	36.0	11.68	80.83	3	20
33	48.0	8.4	32.0	9.46	68.26	93	90
34	56.0	19.9	36.0	11.40	79.24	33	145
35	53.0	10.2	32.0	9.32	67.47	37	57
36	30.0	9.5	20.0	6.56	51.85	158	20

$CA = 14.17 + 5.66 \times D.$

Appendix E

VARIABLES AND VOLUMES FOR ZONE 22

ZONE OF ORIGIN	TB	XB	TA	D	CA	BUS PASS	AUTO DRIVE
1	26.0	12.5	15.0	4.44	75.13	29	65
2	32.0	16.0	21.0	6.15	84.41	73	70
3	39.0	16.0	24.0	7.57	92.85	12	108
4	31.0	14.3	17.0	5.43	80.73	5	48
5	37.0	30.8	19.0	7.63	93.19	0	12
6	45.0	12.0	23.0	9.04	101.17	5	42
7	22.0	12.0	8.6	3.73	71.11	28	69
8	32.0	11.8	13.6	5.46	80.90	16	0
9	37.0	13.8	14.6	6.70	87.92	11	173
10	31.0	13.6	13.1	4.99	78.24	5	70
11	33.8	12.8	16.1	6.97	89.45	36	146
12	25.4	9.8	8.5	2.43	63.75	76	204
13	22.8	12.0	10.3	2.95	66.70	6	102
14	36.0	14.5	11.2	4.59	75.98	0	28
15	37.0	14.1	12.2	4.88	77.62	6	50
16	48.0	15.7	18.2	7.00	89.62	6	106
17	54.0	14.5	21.2	9.16	101.85	0	20
19	30.8	14.5	10.0	3.64	70.60	8	164
20	20.0	17.5	10.0	3.76	71.28	0	15
21	25.0	23.3	14.0	7.28	91.20	6	173
22	5.0	9.3	2.0	0.80	54.53	59	195
23	7.0	9.1	5.0	1.99	61.26	10	76
24	17.0	9.3	8.0	4.49	75.41	52	153
25	27.0	12.0	12.0	6.55	87.07	64	379
26	10.0	8.3	5.5	1.34	57.58	45	89
27	16.0	15.4	7.3	3.16	67.89	0	64
29	25.0	20.4	11.3	3.40	69.24	0	11
30	24.0	9.6	11.0	3.56	70.15	6	18
31	34.0	11.3	16.0	5.21	79.49	10	76
33	22.0	9.6	6.1	2.88	66.30	37	82
34	25.0	20.6	10.7	4.88	77.62	5	82
35	20.0	13.7	7.6	2.82	65.96	10	76
36	11.0	7.8	6.0	1.57	58.89	71	31

CA = 50 + 5.66 x D.